

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉVALUATION DU RÔLE DU GÈNE SLC35A4 DURANT LE STRESS OXYDATIF

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN BIOCHIMIE

PAR

IMADEDINE TIAIBA

MARS 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

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

## REMERCIEMENTS

Aujourd'hui, je marque la fin de deux longues années de travaux et de recherche, une période qui a été exceptionnelle tant sur le plan scientifique que personnel. Tout d'abord, je tiens à exprimer ma profonde gratitude envers mon directeur de recherche, le Professeur Benoît Vanderpere, pour avoir cru en moi en me donnant l'opportunité de mener ce projet et m'avoir guidé tout le long de ma maîtrise. Sa confiance, sa disponibilité et son écoute ont été le moteur de ma motivation, me permettant d'explorer ce domaine de recherche passionnant. Je suis honoré d'avoir pu bénéficier de son mentorat.

Un grand merci à la Professeure Diana Alison Averill et au Professeur Saïd Kourrich, membres de mon jury, pour leur précieuse contribution lors de l'évaluation de mon travail. J'exprime également ma reconnaissance envers mes collègues de laboratoire, Ikram Ajala, Damien Lipuma et Amitha Muraleedharan, ainsi qu'à notre assistant de recherche Benoit Grondin, qui m'ont apporté leur aide et ont enrichi mon expérience de recherche, créant un environnement collaboratif et motivant.

Je remercie chaleureusement ma famille, en particulier ma mère, qui a joué un rôle important dans cette aventure académique grâce à son soutien inestimable.

## TABLE DES MATIERES

REMERCIEMENTS .....	ii
TABLE DES MATIERES .....	iii
LISTE DES FIGURES.....	vi
LISTE DES TABLEAUX .....	vii
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	viii
LISTE DES SYMBOLES ET DES UNITÉS .....	xiii
RÉSUMÉ.....	xiv
ABSTRACT .....	xv
CHAPITRE 1 INTRODUCTION .....	16
1.1 Le dogme de la biologie moléculaire .....	16
1.2 Les gènes multi codants .....	18
1.3 Les cadres de lecture alternatifs (AltORF).....	19
1.4 Méthodes d'identification des AltORF .....	20
1.4.1 Approche computationnelle .....	21
1.4.2 Le profilage des ribosomes (RiboSeq).....	21
1.4.3 La protéomique basée sur la spectrométrie de masse (SM) .....	22
1.4.4 Édition de gènes par CRISPR-Cas9 .....	22
1.5 La différence entre les sORF et les AltORF et leur base des données .....	23
1.6 L'importance des protéines alternatives .....	23
1.7 Le stress cellulaire .....	25
1.8 La réponse intégrée au stress (RIS) .....	25
1.9 La synthèse des protéines : La traduction.....	26
1.10 La traduction coiffe-dépendante et coiffe-indépendante .....	27
1.11 L'importance du facteur eIF2 $\alpha$ dans l'initiation de la traduction .....	27
1.12 Régulation de la traduction : la phosphorylation d'eIF2 $\alpha$ .....	29
1.13 La régulation de la traduction par les uORF.....	29
1.14 La famille de eIF2 $\alpha$ kinase .....	31
1.14.1 PKR-like ER kinase (PERK).....	31
1.14.2 Protein kinase Double-stranded RNA-dependent (PKR) .....	32
1.14.3 General control non-derepressible 2 (GCN2) .....	32
1.14.4 Heme-regulated eIF2a kinase (HRI) .....	33

1.15 Modifications post-traductionnelles.....	33
1.16 La glycosylation .....	34
1.17 Les transporteurs de solutés SLC .....	34
1.17.1 La famille de transporteurs de solutés SLC35.....	34
1.17.2 L'importance du CDP-ribitol.....	35
1.17.3 Le gène double codant <i>SLC35A4</i> .....	35
1.18 Hypothèse et objectif.....	37
CHAPITRE 2 MATÉRIELS ET MÉTHODES .....	39
2.1 Culture cellulaire et traitement .....	39
2.2 Clonage.....	39
2.2.1 Plasmide pLJM1_SLC35A4-3xHA.....	39
2.2.2 Plasmide pLJM1_AltSLC35A4_SLC35A4-3xHA .....	40
2.2.3 Transformations des bactéries .....	41
2.2.4 Extraction des plasmides par Miniprep .....	41
2.3 Génération de populations cellulaires stables.....	41
2.3.1 Production des lentivirus .....	41
2.3.2 Infections lentivirales.....	42
2.4 Immunobuvardage ( <i>Western blot</i> ).....	42
2.5 Immunofluorescence (IF) .....	43
2.6 Génération des cellules SLC35A4 knock-out par CRISPR Cas9 ( <i>Clustered regularly interspaced short palindromic repeats</i> ) .....	44
2.6.1 Designs et protocole .....	44
2.6.2 Transfection .....	44
2.6.3 Cytométrie en flux (FACS, <i>Fluorescence-Activated Cell Sorting</i> ).....	45
2.6.4 Validation de l'édition du gène.....	45
2.7 Mesure de la prolifération cellulaire.....	46
2.8 Mesure de la viabilité cellulaire : essai CellTiter-Glo® 2.0 .....	46
2.9 Immunoprécipitations (IP) .....	47
CHAPITRE 3 RÉSULTATS.....	48
3.1 La génération des cellules knock-out (KO) SLC35A4 avec la technique CRISPR/Cas9 .....	48
3.2 Évaluation de l'expression de SLC35A4 durant le stress cellulaire .....	49
3.2.1 Résultats négatifs.....	49
3.2.2 Résultats Positifs .....	53
3.3 Évaluation du rôle de SLC35A4 dans la réponse au stress cellulaire .....	58
3.3.1 Analyse comparative des cellules knock-out SLC35A4, des cellules <i>Wild type</i> et des cellules knock-out AltSLC35A4 .....	58
CHAPITRE 4 DISCUSSION .....	62

4.1	Évaluation de l'expression de SLC35A4 durant le stress cellulaire .....	62
4.1.1	Détection de la protéine endogène .....	62
4.1.2	L'apparition de deux bandes additionnelles dans WB lors du stress .....	63
4.1.3	Nos hypothèses sur l'origine des bandes additionnelles.....	64
4.1.4	L'augmentation de l'expression de la protéine SLC35A4 par IF .....	65
4.2	Tester l'importance de SLC35A4 dans le stress cellulaire.....	67
4.2.1	SLC35A4 est invalidée fonctionnellement dans les cellules knock-out .....	67
4.2.2	L'absence d'impact de SLC35A4 sur la prolifération cellulaire sous condition normale .....	67
4.2.3	L'impact de SLC35A4 sur l'aptitude cellulaire durant le stress.....	68
	CONCLUSION .....	71
	ANNEXE A Analyse de l'expression de la protéine endogène SLC35A4 avec l'anticorps SLC35A4 .....	72
	ANNEXE B Test de différentes méthodes d'extraction .....	73
	ANNEXE C Origine potentielle des isoformes.....	75
	ANNEXE D Les cellules SLC35A4 KO produisent une protéine non-fonctionnelle .....	76
	RÉFÉRENCES .....	77

## LISTE DES FIGURES

Figure 1.1. Le dogme central.....	17
Figure 1.2. Les ORF alternatifs, une source de diversité protéique.....	20
Figure 1.3. L'importance du facteur eIF2 $\alpha$ dans la traduction en condition normale.....	28
Figure 1.4. L'importance du facteur eIF2 $\alpha$ dans la traduction en condition de stress.....	30
Figure 1.5. Le gène double codant <i>SLC35A4</i> .....	36
Figure 1.6. Comparaison de l'occupation des ribosomes en réponse à la phosphorylation d'eIF2 $\alpha$ .....	38
Figure 2.1. Édition de la séquence codante de <i>SLC35A4</i> .....	44
Figure 3.1. Génération des lignées 143B <i>SLC35A4</i> KO par CRISPR/Cas9.....	49
Figure 3.2. Analyse de l'expression de la protéine endogène <i>SLC35A4</i> .....	50
Figure 3.3. Absence de détection de la protéine <i>SLC35A4</i> par IP-SM.....	52
Figure 3.4. Évaluation de l'expression de la protéine <i>SLC35A4</i> durant le stress cellulaire par <i>Western blot</i> .....	54
Figure 3.5. Évaluation de l'expression de la protéine <i>SLC35A4</i> durant le stress cellulaire par Immunofluorescence.....	57
Figure 3.6. Évaluation de l'importance de <i>SLC35A4</i> par un test de prolifération.....	58
Figure 3.7. Évaluation de l'importance de <i>SLC35A4</i> par un test de viabilité cellulaire.....	60
Figure 3.8. Diagramme résumant la partie résultats.....	61

## LISTE DES TABLEAUX

Tableau 1-1. Exemples de AltProt fonctionnels dans différentes localisations d'AltORF .....	24
Tableau 2-1. Amorces utilisées.....	40
Tableau 2-2. Anticorps utilisés .....	43
Tableau 2-4. Liste des ARNsg et des amorces utilisées pour l'édition de SLC35A4 par CRISPR/Cas9.....	45



## LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

AA	Acide Aminé
ADN	Acide Désoxyribonucléique.
ADNc	Acide Désoxyribonucléique Complémentaire
AltORF:	Cadre de Lecture Ouvert Alternatif (« Alternative Open Reading Frame »)
AltPro:	Protéines Alternatives (« Alternative Proteins »)
AltSLC35A4:	Famille de transporteurs de solutés 35, membre A4 (« Solute Carrier Family 35 Member A4 »)
ARN:	Acide Ribonucléique
ARNcirc :	Acide Ribonucléique Circulaire
ARNm:	Acide Ribonucléique Messenger
ARNnc :	Acide Ribonucléique Non Codant
ARNsg:	RNA short guide=ARN petit guide
ARNt:	Acide Ribonucléique de Transfert
AS:	Arsénite de Sodium
ATF4:	Facteur de transcription activateur 4 (« Activating Transcription Factor 4 »)
ATF5	Facteur de transcription activateur 5 (« Activating Transcription Factor 5 »)
ATF6	Facteur de transcription dépendant de l'AMP (« AMP-dependent transcription factor »)
CDP-ribitol	Cytidine Diphosphate donneur de groupe phospho-ribitol

CDS	Séquence Codante (« Coding Sequence »)
CITE	Amplificateur de traduction indépendant de la coiffe (« Cap-Independent Translation Enhancer »)
CMP-sialique	Cytidine Monophosphate N-acétylneuraminique
CPI	Complexe de Pré-Initiation
CREP	Constitutive Repressor of eIF2 $\alpha$ Phosphorylation = Répresseur Constitutif de la Phosphorylation de eIF2 $\alpha$
CRISPR-Cas9	Répétitions Palindromiques Courtes Inter-régulières Groupées et Régulièrement Espacées (« Clustered Regularly Interspaced Short Palindromic Repeats »)
DMEM	Dulbecco's Modified Eagle Medium = Milieu de l'Aigle Modifié de Dulbecco
dORF	« Downstream Open Reading Frame »
ECL	Enhanced Chemiluminescence »
ECM	« Extracellular Matrix »
EcoRI	Escherichia coli souche RY13 (« Escherichia coli Strain RY13 »)
EGFP	« Enhanced Green Fluorescent Protein »
eIF2 $\alpha$	« Eukaryotic Initiation Factor 2 alpha »
eIF4F	« Eukaryotic Initiation Factor 4A »
eIF5	« Eukaryotic Initiation Factor 5 »
FACS	Cytométrie en flux (« Fluorescence-Activated Cell Sorting »)

GADD34	Protéine 34 Induite par l'Arrêt de la Croissance et les Dommages de l'ADN (« Growth Arrest and DNA Damage-inducible Protein 34 (« Fluorescence-Activated Cell Sorting »)
GCN2	Kinase 2 Non-Dépressible de Contrôle Général (« General Control Nonderepressible kinase 2 »)
GDP	Guanosine Diphosphate
GEF	« Guanosine Exchange Factor »
GTP	Guanosine Triphosphate
GTP-eIF3	Guanosine Triphosphate- Eukaryotic Initiation Factor 3
HDR	Recombinaison homologue (« Homology-Directed Repair »)
HEK293	« Human Embryonic Kidney cells 293 »
HRI	Heme-Regulated eIF2 $\alpha$ Kinase ou Eukaryotic Translation Initiation factor 2 alpha kinase 1= Kinase eIF2 $\alpha$ régulée par l'hème
HSP90 et HSP70	Protéines de choc thermique (« Heat Shock Proteins »)
IF	Immunofluorescence
IFRED	Contenant un domaine d'édition d'ARN à repliement d'immunoglobuline (« Immunoglobulin Fold RNA Editing Domain Containing »)
IP	Immunoprécipitation
IRES	Site d'entrée ribosomique interne (« Internal ribosome entry site »)
JENH	Jonction d'extrémité non homologue (« Non-Homologous End Joining »)

KO	Knock-out
m7GTP	7-méthyl guanosine triphosphate
MPT	Modifications Post-Traductionnelles
NEAA	« Non Essential Amino Acid »
Nhel	Extrêmophile de Nouvelle-Hampshire I (« New Hampshire Extremophile I »)
NST	Transporteurs de Sucres Nucléotidiques
ORF	Cadre de Lecture Ouvert
PBS	Phosphate Buffered Saline
PFA	Paraformaldéhyde
PKR ou eIF2AK2	RNA-Dependent Protein Kinase
RE	Réticulum Endoplasmique
RefORF	Région Codante de Référence
RHBDL1	Rhomboid Like 1
RiboSeq	Séquençage Ribosomique
RIS	Réponse Intégrée au Stress
RNAInc	ARN long non codant
RPNR/UPR	Réponse aux protéines non repliées (« Unfolded Protein Response »)
RPP14	Ribonuclease Protein subunit p14
S2P ou MBTPS2	Protéase de Site-2
SIT	Site d'Initiation de la Traduction

SLC	Transporteurs de solutés (« Solute Carrier »)
SM	Spectrométrie de Masse
SPP-like 3	« Signal Peptide Peptidase-like 3 »
SVF	Sérum de Veau Fœtal
TBS	« Tris-Buffered Saline »
TBST	« Tris-Buffered Saline with Tween 20 »
uORF	Cadre de lecture ouvert en amont (« Upstream Open Reading Frame »)
UTR	Région Non Traduite (« Untranslated Region »)
UV	Rayons Ultraviolets
WB	Immunobuvardage (« Western Blot »)
βME	Mercaptoéthanol

## LISTE DES SYMBOLES ET DES UNITÉS

$\mu\text{g}$	Microgramme
$\mu\text{l}$	Microlitre
$\mu\text{M}$	Micromolaire
$\mu\text{m}$	Micromètre
5x	Cinq fois
h	Heure
mg	Milligramme
mL	Millilitre
mM	Millimolaire
$\alpha$ -	Alpha
KDa	Kilo Dalton

## RÉSUMÉ

Le gène double codant SLC35A4 code pour deux protéines : la protéine SLC35A4, encodée à partir du CDS (séquence codante), une protéine transmembranaire exprimée majoritairement dans le Golgi, qui transporte le CDP-Ribitol à l'état basal, ainsi que la protéine AltSLC35A4, dont la séquence codante est localisée dans la région 5'UTR. Il s'agit d'une protéine transmembranaire exprimée dans la mitochondrie dont la fonction n'est pas encore déterminée. De plus, des études précédentes basées sur le profilage ribosomique ont révélé que, lors du stress oxydatif et du stress du réticulum endoplasmique, l'ARNm du gène SLC35A4 échappe à l'inhibition globale de la traduction. De surcroît, la région codante de la protéine SLC35A4 subit la plus grande augmentation d'efficacité traductionnelle de tous les ARNm cellulaire. Ceci caractérise les ARNm produisant des protéines essentielles à la réponse intégrée au stress, qu'il s'agisse de protéines favorisant la résistance au stress ou de protéines délétères induisant l'apoptose. Ce mémoire a pour but de valider les résultats du profilage ribosomique en examinant la dynamique d'expression de la protéine SLC35A4 en condition de stress oxydatif, ainsi que de déterminer l'importance jouée par les protéines codées par *SLC35A4* lors du stress. En effet, nous avons montré l'émergence d'isoformes de la protéine SLC35A4 lors du stress oxydatif, avec une augmentation proportionnelle de leur expression en fonction de la durée du stress. De plus, nous avons démontré l'importance de SLC35A4, qui confère une résistance aux cellules face au stress oxydatif, ainsi que l'importance de son uORF (cadre de lecture ouvert en amont) dans la traduction du CDS de SLC35A4 lors du stress. Ces résultats ouvrent des perspectives pour approfondir l'étude de ce gène double codant, présentant un potentiel thérapeutique pour les maladies liées au stress cellulaire, notamment le stress oxydatif.

Mots clés : SLC35A4 et isoformes, protéine alternative, stress oxydatif, CRISPR/Cas9, régulation traductionnelle.

## ABSTRACT

The double-coding gene *SLC35A4* encodes for two proteins: the *SLC35A4* protein, encoded from the CDS (coding sequence), a transmembrane protein expressed mainly in the Golgi, that transports CDP-Ribitol at the basal state, as well as the Alt*SLC35A4* protein, which is encoded from the 5'UTR region, a transmembrane protein expressed in the mitochondria whose function is not yet determined. In addition, previous studies based on ribosome profiling revealed that, under oxidative stress and endoplasmic reticulum stress, the mRNA of the *SLC35A4* gene escapes global inhibition of translation. Moreover, the CDS of the *SLC35A4* protein undergoes the largest increase in translational efficiency of all cellular mRNAs. This characterizes the mRNAs producing proteins essential for the integrated stress response, whether they are proteins promoting stress resistance or deleterious proteins inducing apoptosis. The aim of this work is to validate the results of the ribosome profiling by examining the expression dynamics of the *SLC35A4* protein under oxidative stress conditions, as well as to determine the importance played by *SLC35A4* encoded proteins during stress. Indeed, we showed the emergence of isoforms of *SLC35A4* protein during oxidative stress, with a proportional increase of their translation according to the duration of the stress. In addition, we demonstrated the importance of *SLC35A4*, which confers resistance to cells against oxidative stress, as well as the importance of its uORF (upstream open reading frame) in the translation of the *SLC35A4* CDS during stress. These results open perspectives to further study this dual coding gene, presenting a therapeutic potential for diseases related to cellular stress, especially oxidative stress.

Keywords: *SLC35A4* and isoforms, alternative protein, oxidative stress, CRISPR/Cas9, Translational regulation.



# CHAPITRE 1

## INTRODUCTION

### 1.1 Le dogme de la biologie moléculaire

Le dogme fondamental de la biologie moléculaire décrit comment l'information génétique codée dans l'ADN est convertie en protéines fonctionnelles dans un système biologique (Pukkila, 2001). Le génome d'un organisme est en fait l'ensemble des informations génétiques qui résident dans les noyaux des cellules. La transcription est la première étape du processus de conversion des informations stockées dans le génome en une protéine fonctionnelle. Lors de la transcription, une portion de l'ADN est transcrite et copiée en ARN messager (ARNm), qui transmet le message de l'ADN au cytoplasme où se produit la traduction. La traduction est l'étape suivante de la synthèse des protéines. La molécule d'ARNm est décryptée par le ribosome, un grand complexe de protéines et de molécules d'ARN dans le cytoplasme. Le ribosome traduit le message de l'ARNm en une séquence spécifique d'acides aminés (Figure 1.1), qui se replie en une protéine fonctionnelle (Clancy et Brown, 2008). Bien que la traduction produise une protéine avec une séquence d'acides aminés spécifique, la protéine résultante n'est pas toujours immédiatement fonctionnelle. Dans de nombreux cas, la protéine subit des modifications post-traductionnelles, c'est-à-dire des changements qui interviennent dans la protéine après sa synthèse. Les modifications post-traductionnelles, telles que la glycosylation, influent sur le repliement, la localisation, la stabilité et l'activité des protéines. Elles confèrent une flexibilité importante aux protéines, leur permettant d'exercer différents rôles au sein de la cellule (Mann *et al.*, 2003). Finalement, il est essentiel d'avoir une régulation précise de ces mécanismes afin d'assurer un bon fonctionnement cellulaire et de permettre aux cellules d'exprimer le bon répertoire génique nécessaire pour se diviser, se différencier et s'adapter. Cela est particulièrement important lors de conditions de stress cellulaire.

De plus, pour accomplir une variété de rôles biologiques et de réactions biochimiques essentielles à la survie, une diversité protéique est indispensable. Traditionnellement, l'évaluation de cette diversité se basait sur le nombre estimé de gènes codant pour des protéines dans le génome d'un organisme. Cependant, contrairement à une estimation antérieure de 100 000 gènes, l'humain aurait environ 20 300 gènes qui encodent pour des protéines (Pruitt *et al.*, 2007). Malgré ce nombre relativement faible, le protéome humain est estimé à comprendre plus d'un 1 million de protéoformes (Jensen, 2006; Munoz et Heck, 2014; Ponomarenko *et al.*, 2016; Aebersold *et al.*, 2018). En effet, les protéoformes sont les différentes formes d'une protéine produites à partir du même gène, mais présentant des variations dans

leur séquence, leur épissage ou leurs modifications (Smith *et al.*, 2018). Cette observation indique que la complexité du protéome résulte de mécanismes biologiques autres que l'expansion génique conventionnelle, qui se mesure par le nombre de gènes dans le génome d'un organisme (Schluter et Conte, 2009). Ainsi, il est essentiel d'explorer l'origine de cette diversité. En effet, cette vaste gamme de protéines découle de mécanismes de diversification présents à toutes les étapes du processus de biosynthèse des protéines, débutant au niveau de l'ADN, se poursuivant à travers l'ARN, et s'étendant jusqu'au niveau post-traductionnel. La complexité peut découler des variations alléliques (Lo *et al.*, 2003), de la duplication du gène (Kondrashov, 2012), de l'épissage alternatif des transcrits ARN (Wang *et al.*, 2008) et de nombreuses modifications post-traductionnelles (Mann *et al.*, 2003). Ces exemples de mécanismes contribuent à accroître la diversité fonctionnelle des protéines.

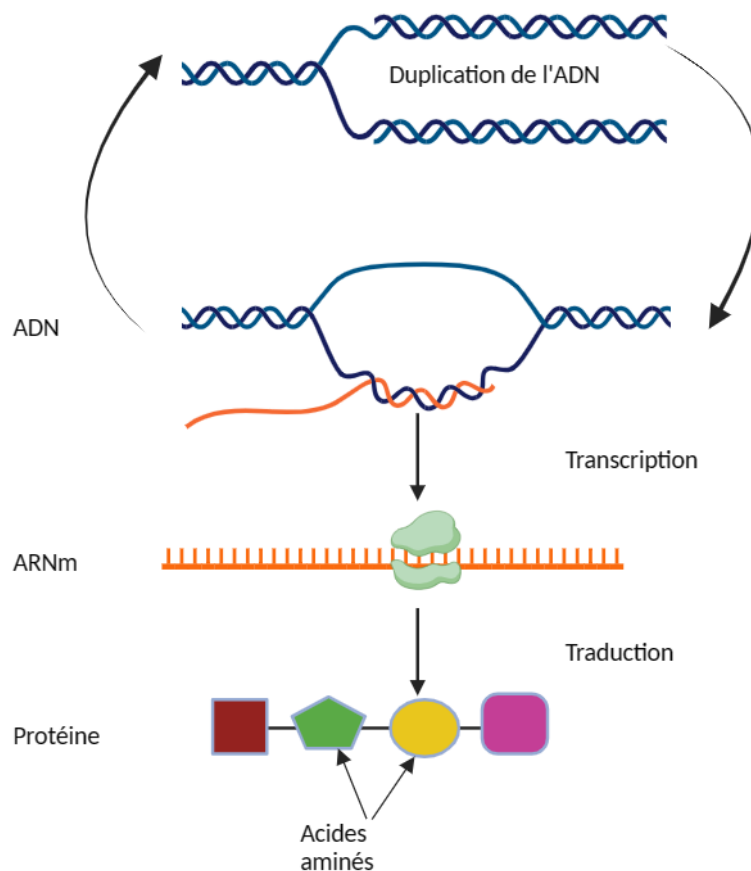


Figure 1.1. Le dogme central.

La duplication de l'ADN double brin lors de la réplication de l'ADN garantit que chaque cellule fille acquiert une copie exacte de l'information génétique. L'information codée dans l'ADN (nucléotides) est ensuite

transcrite en ARN messagers (ARNm). Ces ARNm sont ensuite traduits en acides aminés pour former une chaîne polypeptidique (protéine). Inspiré de Pukkila, 2001.

## 1.2 Les gènes multi codants

Dans les procaryotes et les virus, on connaît bien la possibilité que les ARNm soient multi codants (Nomark *et al.*, 1983). Chez les procaryotes, les ARNm peuvent être monocistroniques ou polycistroniques, ce dernier comportant plusieurs codons d'initiation et plusieurs codons stop, générant ainsi plusieurs cadres de lecture ouvert (ORF) et permettant la synthèse de diverses protéines à partir d'un seul ARNm (Kozak, 2005). Cependant, chez les eucaryotes, l'idée que l'ARNm puisse coder pour plusieurs protéines a longtemps été rejetée. Les ARNm des eucaryotes étaient considérés comme monocistroniques, capables de coder pour une seule protéine (Blumenthal, 2004). Malgré cela, des études de plus en plus nombreuses indiquent que les ARNm des eucaryotes pourraient avoir un potentiel multi codant (Vanderperre *et al.*, 2011; Bergeron *et al.*, 2013; Samandi *et al.*, 2017).

En général, un ORF est défini comme un segment de triplets de nucléotides (codons) conservés et non chevauchants, pouvant être traduit en une protéine fonctionnellement annotée (Couso *et al.*, 2017). Les ARNm eucaryotes contiennent généralement un ORF principal encodant une seule protéine. En effet, les règles génétiques traditionnelles ont limité l'identification de transcrits ayant des capacités de codage protéique non canoniques, telles que la conservation et l'homologie des acides aminés, l'exigence absolue d'un codon de départ (méthionine) ainsi que d'un contexte de Kozak nécessaire pour permettre la fixation du ribosome (Kozak, 1999) et la longueur de traduction minimale (100 codons) (Andrew et Rothnagel, 2014 ; Dong *et al.*, 2023).

Or, contrairement à ce dogme, le développement des nouvelles technologies et de techniques à grande échelle pour étudier le traductome (ARNm engagés dans le processus de traduction) et le protéome (protéines exprimées) a révélé des lacunes dans les annotations actuelles du génome (Brunet *et al.*, 2018). Des études de profilage des ribosomes indiquent que jusqu'à 40 % des ribosomes en initiation sont liés à des ORF précédemment non annotés (Erhard *et al.*, 2018). De même, des pipelines protéomiques adaptées ont démontré la présence de nouvelles protéines qui proviennent d'ORF non annotés (Brunet et Roucou, 2020). Ainsi, de nombreux ORF n'ont pas encore été annotés dans les génomes eucaryotes (Vanderperre *et al.*, 2013 ; Bergeron *et al.*, 2013 ; Anderson *et al.*, 2015). Par conséquent, il est apparu non seulement nécessaire de reconnaître la nature polycistronique des ARNm eucaryotes, mais aussi d'actualiser l'annotation des génomes (Mouilleron *et al.*, 2016 ; Brunet *et al.*, 2018).

### 1.3 Les cadres de lecture alternatifs (AltORF)

Historiquement, les cadres de lecture ouverts (ORF) de moins de 100 codons étaient exclus des annotations génomiques, à moins d'être préalablement caractérisés, car ils étaient considérés comme trop courts pour être fonctionnels (Cheng *et al.*, 2011). Ce critère de longueur, ainsi que l'exigence d'un codon d'initiation ATG et la restriction d'une seule séquence codante par transcrite, ont considérablement façonné et limité l'exploration du protéome (Hellens *et al.*, 2016; Olexiouk & Menschaert, 2016; Brunet *et al.*, 2018; Orr *et al.*, 2019). Cependant, l'accumulation des preuves fonctionnelles démontrant l'existence d'ORF précédemment non annotés dans les bactéries (Wadler et Vanderpool 2007; Hemm *et al.*, 2008, 2010; Baek *et al.*, 2017), les drosophiles (Kondo *et al.*, 2007; Reinhardt *et al.*, 2013; Aspden *et al.*, 2014; Albuquerque *et al.*, 2015), les plantes (Hanada *et al.*, 2013; Juntawong *et al.*, 2014; Hsu et Benfey 2017), ainsi que chez d'autres eucaryotes (Oyama *et al.*, 2007; Ingolia *et al.*, 2011; Vanderperre *et al.*, 2011, 2013; Ma *et al.*, 2014), a mis en lumière la nécessité d'une révision des annotations génomiques. En conséquence, plusieurs groupes de recherche ont approfondi l'exploration des ORF non annotés par des approches bio-informatiques à grande échelle afin de donner un aperçu plus précis des capacités codantes du génome (Xu *et al.*, 2010, Chung *et al.*, 2007, Ribrioux *et al.*, 2008, Vanderperre *et al.*, 2013). En effet, une réanalyse du transcriptome de différents organismes, en utilisant une taille limite inférieure de 30 codons pour limiter la taille de la base de données, a été réalisée et a permis de prédire des centaines de milliers de nouvelles séquences codantes en plus des séquences préalablement annotées (Samandi *et al.*, 2017; Brunet *et al.*, 2021). Les ORF préalablement identifiés sont désignés comme ORF de référence ou RefORF (aussi appelée séquence codante ou CDS), et codent pour des protéines de référence, tandis que les nouveaux ORF sont qualifiés d'alternatifs (AltORF) et pourraient potentiellement coder pour des protéines alternatives (AltPro).

Par convention, la protéine de référence est codée, à partir du CDS, dans le cadre de lecture +1. Les AltORF peuvent se trouver dans les régions non traduites 5' ou 3' UTR (également appelées *upstream* uORF et *downstream* dORF respectivement) dans tous les cadres de lecture possibles, ou bien chevaucher le RefORF (totalement ou partiellement) dans le cadre de lecture +2 ou +3. Des AltORF ont également été trouvés dans les ARN non codants (ARNnc), tels que les ARN non codants longs (ARNncl), les ARN circulaires (ARNcir), dans tous les cadres de lecture (Vanderperre *et al.*, 2013; Ji *et al.*, 2015; Xing *et al.*, 2021). L'émergence de cette perspective selon laquelle un seul ARNm mature pourrait coder pour plusieurs protéines entièrement distinctes, ainsi que des ARN que l'on pensait non-codant participent à la

traduction des protéines, pourrait contribuer à la diversité du protéome au sein des organismes eucaryotes (Figure 1.2).

### Dogme: ancienne annotation

Une seule séquence codante annotée (RéfORF/CDS)

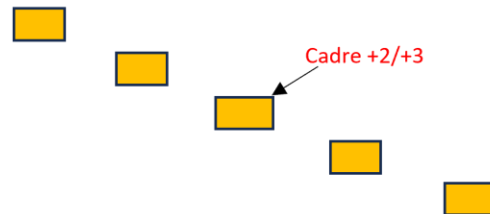
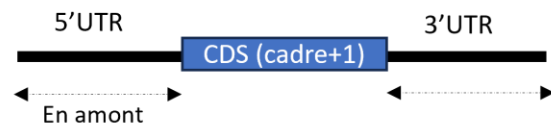


ARN non codant (ARNnc)



### Cadres de lecture ouverts alternatifs (AltORF)

ORF en amont (uORF)  
 uoORF chevauchant le CDS  
 intORF chevauchant le CDS  
 doORF chevauchant le CDS  
 ORF en aval (dORF)



Traduction de plusieurs protéines

ARN non codant (ARNnc-ORF)

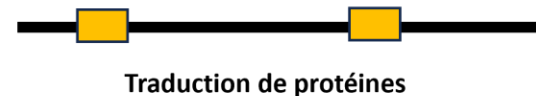


Figure 1.2. Les ORF alternatifs, une source de diversité protéique.

Ces AltORF peuvent être localisés dans des transcrits déjà annotés comme codants mais à partir des 5' UTR (régions 5' non traduites), des CDS (séquences codantes ou RefORF) dans un cadre de lecture différent +2/+3, des 3'UTR (3' régions non traduites), ou même être localisés dans des ARN non codants. Inspiré de Mudge *et al.*, 2022.

#### 1.4 Méthodes d'identification des AltORF

Au fil des dernières décennies, différentes méthodes ont émergé et ont prouvé leur efficacité dans l'identification et l'étude des AltORF, notamment en raison des limitations des techniques biochimiques classiques dans ce domaine (Vazquez-Laslop *et al.*, 2022). Ces approches impliquent l'identification des AltORF soit directement au niveau génomique (ADN) ou transcriptomique (ARN), soit par la détection de leurs produits de traduction. Il s'agit notamment d'approches computationnelles, de méthodes basées sur

le séquençage, la protéomique basée sur la spectrométrie de masse ainsi que l'édition de gènes par CRISPR-Cas9.

#### 1.4.1 Approche computationnelle

L'identification computationnelle permet des recherches systématiques d'AltORF, en utilisant des méthodes basées sur l'analyse de séquence ou la génomique comparative (Mackowiak *et al.*, 2015; Makarewich et Olson, 2017). Cette dernière est basée sur l'analyse des fréquences et la conservation de substitution des codons, à l'aide de l'algorithme PhyloCSF par exemple, entre les espèces (Lin *et al.*, 2011). La conservation des séquences a longtemps été associée au potentiel de codage et à la fonction des protéines, et avec l'augmentation exponentielle des séquences génomiques disponibles, ces outils deviennent plus précis et plus puissants. L'outil PhyloCSF a été utilisé pour identifier, depuis le génome de l'humain, 144 nouvelles régions codantes, 236 AltORF au sein de régions codantes connues et 169 pseudogènes ayant un potentiel codant (Mudge *et al.*, 2019). De plus, des méthodes pour prédire spécifiquement les AltORF dans le transcriptome ont été développées (Vanderperre *et al.*, 2012 ; 2013). Brunet et ses collègues ont également récupéré les annotations NCBI RefSeq (O'Leary *et al.*, 2016) et Ensembl (Zerbino *et al.*, 2018), qu'ils ont compilées en une base de données plus exhaustive. Ils ont ensuite effectué une traduction *in silico* sur trois cadres en utilisant EMBOSS Transeq (Li *et al.*, 2015) pour prédire tous les ORF possibles avec un codon de départ ATG et une longueur minimale de 30 codons, ce qui a conduit à la construction de l'ORFéome OpenProt (Brunet *et al.*, 2019). Ces analyses ont clairement détecté des AltORF potentiellement codants pour des protéines et ont révélé leur présence généralisée dans différents transcriptomes.

#### 1.4.2 Le profilage des ribosomes (RiboSeq)

Cette technique, connue sous le nom de RiboSeq ou d'empreinte des ribosomes, permet la détection des AltORF traduits. Elle a été introduite pour la première fois en 2009 par Ingolia et son équipe. Le principe est simple : lorsqu'ils sont liés à un transcrit, les ribosomes en traduction encapsulent environ 30 nucléotides (22-35), les préservant ainsi des nucléases. Ces empreintes des ribosomes témoignent d'une activité translationnelle. Cette méthode offre un aperçu de la position des ribosomes le long d'une molécule d'ARNm à un moment précis, permettant aux chercheurs d'analyser l'efficacité de la traduction, d'identifier les cadres de lecture ouverts (ORF) traduits et de mieux comprendre la régulation de l'expression des gènes au niveau de la traduction. Ingolia et son équipe 2009 ont réussi à détecter l'activité traductionnelle au niveau de 153 AltORF à l'intérieur de la région 5'UTR des ARNm. Par la suite, de

nombreuses études utilisant le RiboSeq ont révélé une activité traductionnelle dans des régions ne présentant pas de séquences codantes annotées précédemment, telles que les ARN non codants (ARNnc), les régions 5'UTR et les chevauchements de CDS (Ingolia *et al.*, 2011; Bazzini *et al.*, 2014; Fields *et al.*, 2015). En 2016, la Ribo-séquence à super-résolution a été développée, ce qui a permis d'améliorer la détection de l'activité de traduction à partir des alt-ORF dans les ARNm, les ncARN et des nouveaux ORF dans les transcrits dérivés de pseudogènes (Hsu *et al.*, 2016).

#### 1.4.3 La protéomique basée sur la spectrométrie de masse (SM)

La spectrométrie de masse (SM) est actuellement la meilleure méthode pour l'identification directe de nouvelles protéines. Cette technique sensible permet de détecter l'expression des protéines de manière directe et fiable. En revanche, les analyses RiboSeq et computationnelles fournissent des preuves indirectes de l'existence possible de la protéine (Lin *et al.*, 2011; Brunet *et al.*, 2019). Cependant, l'utilisation de SM nécessite une base de données de séquences protéiques, ce qui signifie que toute protéine absente de cette base de données ne peut être identifiée (Brunet et Roucou, 2019). L'exploitation des bases de données comprenant des AltProt basées sur des prédictions, *in silico*, a permis de démontrer l'expression à grande échelle de protéines alternatives par spectrométrie de masse (Vanderperre *et al.*, 2013; Samandi *et al.*, 2017). Toutefois, l'utilisation d'une base de données contenant tous les cadres de lecture ouverts (ORF) possibles à travers le génome entier ne constitue pas une stratégie judicieuse. En effet, il a été démontré que les bases de données volumineuses diminuent la sensibilité et le rapport signal/bruit (Nesvizhskii, 2010). De ce fait, le profilage ribosomique peut servir à générer des bases de données pour des analyses par SM, de manière spécifique aux échantillons (Menschaert *et al.*, 2013).

#### 1.4.4 Édition de gènes par CRISPR-Cas9

L'édition de gènes CRISPR-Cas9 (*clustered regularly interspaced short palindromic repeats-associated protein*) est une approche ciblée et pratique pour affirmer le potentiel codant des AltORF. Bien que cette technique ne puisse pas prédire les AltORF et qu'elle nécessite la connaissance de leur séquence nucléotidique avant l'expérience, elle constitue un outil unique pour valider l'expression, la localisation subcellulaire et la fonction des protéines (Delcourt *et al.*, 2018, Yeasmin *et al.*, 2018). Le système CRISPR de type II utilise une endonucléase ADN (Cas9) dirigée par un ARN guide unique de 20 nucléotides (ARNsg) pour créer des cassures double brin dans tout locus génomique correspondant à l'ARNsg (Jinek *et al.*, 2012; Jinek *et al.*, 2013). La cassure double brin induite par CRISPR-Cas9 peut être réparée soit par une jonction

d'extrémité non homologue (JENH), soit par recombinaison homologue (HDR). JENH est le mode de réparation prédominant en absence de gabarit homologue. Dans ce processus, les cellules tentent de reconnecter directement les deux extrémités brisées. Étant donné que ces extrémités peuvent avoir de courtes délétions et/ou insertions, cela peut conduire à des mutations ou des décalages de cadre de lecture dans le gène ciblé (Ran *et al.*, 2013). D'autre part, la HDR peut être utilisée pour introduire une étiquette d'épitope au locus ciblé afin de créer une protéine de fusion endogène, en fournissant un gabarit contenant cette étiquette (Dewari *et al.*, 2018). Ainsi, cette technique valide non seulement l'expression d'un AltORF en utilisant l'analyse *Western blot* sans avoir à développer un anticorps personnalisé, mais elle aide également à la caractérisation fonctionnelle (D'Lima *et al.*, 2017; Rathore *et al.*, 2018). Par exemple, Anderson et ses collègues ont utilisé le système CRISPR-Cas9 pour insérer un épitope FLAG à l'extrémité C-terminale de la protéine MLN afin de démontrer la présence de son expression endogène dans des myoblastes de souris (Anderson *et al.*, 2015). De plus, CRISPR-Cas9 est également utilisé pour des tests de criblage en générant des altération génétique définis. En effet, Guo et ses collègues ont ciblé 315 AltORF et ont constaté que 68 d'entre eux jouaient un rôle dans la forme physique de la levure sous divers stress environnementaux (Guo *et al.*, 2018).

### 1.5 La différence entre les sORF et les AltORF et leur base des données

Les sORF, ou petits ORF, présentent une définition distincte malgré leurs similitudes avec les AltORF. Les sORF sont généralement définis comme des ORF d'une longueur de 10 à 100 codons situés dans des régions annotées comme non codantes (ncARN et UTR des ARNm). Ils sont principalement initiés par des codons près-homologues (des codons qui diffèrent des AUG par un nucléotide) tels que CUG, GUG, UUG et ACG (Albuquerque *et al.*, 2015; Orr *et al.*, 2020). En se basant sur cette définition et en utilisant les données de RibSeq, la base de données des sORF a été répertoriée dans sORF.org (Olexiouk *et al.*, 2016).

En revanche, les AltORF sont définis comme des ORF de plus de 30 codons, sans limite maximale de 100 codons, présents dans n'importe quel ARN (ARNm et ncARN) et commençant par un codon d'initiation ATG. En utilisant cette définition et en se basant sur les données de MS et RiboSeq, la base de données correspondante a été répertoriée dans OpenProt (Brunet *et al.*, 2019).

### 1.6 L'importance des protéines alternatives

OpenProt a prédit environ 595 788 AltProt chez l'humain, soit plus de 2.4 fois le nombre de RefProt (247 352 protéines de référence avec les isoformes; Leblanc *et al.*, 2024). Ces prédictions contiennent



inévitablement des faux positifs (prédits mais non exprimés). Cependant, OpenProt répertorie également les preuves expérimentales d'expression des AltORF (par SM ou RiboSeq). Par conséquent, 45 % des AltProt sont répertoriées comme exprimées selon ce répertoire. L'ensemble de ces nouvelles protéines pourrait apporter de nouvelles perspectives de recherche et permettre d'expliquer, de façon directe ou indirecte, des mécanismes cellulaires encore incompris.

Tableau 1-1. Exemples de AltProt fonctionnels dans différentes localisations d'AltORF

Classe d'AltORF (localisation)	Gènes	Nom de AltORf	Espèce	Fonction	Source
5'UTR	MiD51	AltMiD51	Humain	Alternative-MiD51 est un facteur de fission des mitochondries et un régulateur en amont de la traduction de MiD51.	Delcourt <i>et al.</i> , 2018
CDS	FUS	AltFUS	Drosophile, Humain	Rôle dans la toxicité neuronal chez la Drosophile	Brunet <i>et al.</i> , 2020
3'ORF	RPP14	AltRPP14 (HsHTD2)	Humain	Déshydratase mitochondriale essentielle pour la biosynthèse des acides gras	Autio <i>et al.</i> , 2008 ; Vanderperre <i>et al.</i> , 2013 ; Kim <i>et al.</i> , 2014 ;

<b>ncARN</b>		CIP2A-BP est codée par le lncRNA LINC00665.	Humain	Il s'agit d'un marqueur pronostique du cancer du sein et il a été suggéré qu'il constitue une nouvelle cible thérapeutique.	Vitorino <i>et al.</i> , 2021
--------------	--	---	--------	---	-------------------------------

En dépit de la caractérisation de certains AltProt (Tableau 1) et l'intensification des études sur ces protéines au cours des dernières années, les fonctions de la plupart d'entre elles sont encore inconnues. De plus, comme l'abondance relative des AltProt change lors d'un stress (Khitun *et al.*, 2019; Moro *et al.*, 2021), il est légitime de se demander si elles jouent un rôle dans l'homéostasie de la cellule.

### 1.7 Le stress cellulaire

Une cellule et son environnement ont un lien vital qui détermine directement le destin de la cellule. La cellule est toujours en interaction avec son environnement interne et externe afin de garantir la santé cellulaire globale et la survie de l'organisme. L'environnement extérieur contient et permet l'accès aux macro- et micro-nutriments tandis que le milieu intérieur comme Claude Bernard l'a nommé (Bernard, 1865) assure le bon fonctionnement de la cellule. Toute anomalie ou tout déséquilibre dans ces environnements peut entraîner un dysfonctionnement de l'homéostasie cellulaire, et dans certains cas s'il est prolongé, une maladie (Fulda *et al.*, 2010).

En conséquence, la cellule vise à maintenir activement son homéostasie en réponse à tout changement dans son environnement interne ou externe. Effectivement, la cellule est continuellement exposée à des perturbations de son environnement. Par définition, tout ce qui peut être perçu comme ayant ou pouvant avoir pour effet de perturber l'homéostasie cellulaire est considéré comme un facteur de stress (Zhou *et al.*, 2019). En effet, lorsque les cellules sont soumises à un stress, elles déclenchent une réponse spécifique au facteur de stress en cause, appelée réponse intégrée au stress (Pakos-Zebrucka *et al.*, 2016).

### 1.8 La réponse intégrée au stress (RIS)

Les cellules eucaryotes ont développé des voies de signalisation complexes qui assurent la médiation de la réponse cellulaire aux stress, y compris les facteurs de stress externes tels que l'irradiation UV, l'hypoxie, les infections et la privation de nutriments, ou les facteurs de stress interne, y compris le stress du

réticulum endoplasmique (RE) et le stress oxydatif (Taniuchi *et al.*, 2016; Pakos-Zebrucka *et al.*, 2016). Ces voies déclenchent un large éventail de mécanismes d'adaptation. La cellule a deux réponses possibles au stress: elle peut activer des mécanismes de défense pour s'adapter et survivre au stress. Cependant, si le stress est trop sévère, soit en intensité ou en durée, elle peut choisir de subir une mort cellulaire programmée (apoptose) pour protéger les cellules voisines et les fonctions physiologiques vitales (Fulda *et al.*, 2010; Pakos-Zebrucka *et al.*, 2016). Ces mécanismes de défense impliquent une modification de l'expression des gènes au niveau de la transcription et de la traduction (Zinszner *et al.*, 1998; Harding *et al.*, 2003).

### 1.9 La synthèse des protéines : La traduction

La synthèse des protéines est le processus au cours duquel les ribosomes lisent les instructions génétiques contenues dans l'ARNm et assemblent une chaîne d'acides aminés. Dès que l'ARNm est transcrit à partir de l'ADN matrice, il subit de multiples modifications dans le noyau. Tout d'abord, un groupement 7-méthyl guanosine triphosphate (m7GTP) est ajouté à l'extrémité 5' de l'ARN formant une structure coiffe. Cette coiffe protège l'ARN de la dégradation par les exonucléases 5' à 3' (Ghosh *et al.*, 2010). Ensuite, les introns sont épissés (Scheres *et al.*, 2017) et l'extrémité 3' est polyadénylée (Eckmann *et al.*, 2011). Cela protège davantage l'ARN de la dégradation par les exonucléases 3' à 5' (Gagliardi et Dziembowski, 2018). L'ARNm est alors exporté hors du noyau pour être traduit (Hocine *et al.*, 2010). Chez les eucaryotes, la traduction comprend quatre phases principales : l'initiation, l'élongation, la terminaison et le recyclage (Jackson *et al.*, 2010; Aitken et Lorsch, 2012; Hinnebusch, 2014). La traduction nécessite l'assemblage organisé de ribosomes, de facteurs de traduction et d'ARN de transfert (ARNt). Au cours de la phase d'initiation, le complexe de pré-initiation 43S (CPI), composé de la sous-unité 40S du ribosome, d'un Met-ARNti et eIF2-GTP, est recruté dans la région 5'UTR de l'ARNm. Puis, il commence à balayer l'ARNm base par base à la recherche de la complémentarité avec l'anticodon du Met-ARNt, connue aussi sous le nom site d'initiation de la traduction (SIT). Une fois qu'il reconnaît le codon d'initiation AUG dans un contexte favorable (Kozak, 2002), la sous-unité 60S vient se lier avec le CPI pour produire un complexe d'initiation 80S compétent pour initier la synthèse polypeptidique ou l'élongation. L'élongation se fait par ajout d'acides aminés à la chaîne peptidique en croissance (Hinnebusch, 2014). Enfin, la terminaison se produit lorsque le ribosome rencontre un codon d'arrêt, ce qui entraîne son détachement de l'ARNm. La chaîne protéique synthétisée est alors libérée et les sous-unités ribosomiques sont alors recyclées.

### 1.10 La traduction coiffe-dépendante et coiffe-indépendante

La dérégulation de la machinerie de traduction peut conduire à divers états pathologiques (Holcik et Sonenberg 2005; Topisirovic et Sonenberg, 2011). La régulation de la traduction des ARNm intervient à toutes les étapes du processus (Sonenberg et Hinnebusch, 2009; Jackson *et al.*, 2010). Il est généralement admis que l'initiation de la traduction constitue l'étape limitante et la plus étroitement régulée (Sonenberg et Hinnebusch, 2009). En outre, l'initiation de la traduction de la plupart des ARNm eucaryotes se produit par un mécanisme appelé balayage dépendant de la coiffe (Hinnebusch et Lorsh, 2012). En effet, la plupart des ARNm contiennent une structure coiffe (m<sup>7</sup>GTP) au niveau de leur extrémité 5'. Ainsi, la traduction des ARNm s'initie par la reconnaissance de la structure coiffe par les complexes d'initiation de la traduction puis ces complexes commencent à balayer le transcrit à partir de son extrémité 5' jusqu'à atteindre le premier site d'initiation comme décrit dans la partie précédente.

De plus, les cellules peuvent initier la traduction de certains ARNm de manière indépendante de la coiffe, en ignorant la coiffe 5' (Martínez-Salas *et al.*, 2012). Les complexes d'initiation de la traduction peuvent se fixer directement au transcrit à une position en aval de l'extrémité 5'UTR. Ce mécanisme est observé dans des conditions de stress, où la répression de la synthèse protéique globale est souvent accompagnée de la traduction sélective d'ARNm codant pour des protéines cruciales pour la survie cellulaire et la récupération après le stress (Holcik et Sonenberg, 2005 ; Jackson et Hellen, 2010). Dans les cellules eucaryotes, l'initiation de la traduction indépendante de la coiffe peut se faire par plusieurs mécanismes : par l'intermédiaire d'un site d'entrée ribosomique interne (IRES; Komar et Hatzoglou, 2011; Andreev *et al.*, 2012) ou traduction médiée par l'amplificateur de la traduction indépendant de la coiffe (CITE; Lacerda *et al.*, 2017; Shatsky *et al.*, 2018). De plus, la présence des séquences polyU et les séquences complémentaires à l'ARN ribosomique 18S dans les régions 5'UTR et 3'UTR, ainsi que la présence de 6-méthyladénosine (m<sup>6</sup>A) dans la région 5'UTR (Meyer *et al.*, 2015 ; Weingarten-Gabbay *et al.*, 2016), peuvent recruter directement la sous-unité 40S du ribosome ainsi que les facteurs nécessaires à la traduction sur un codon d'initiation. Par exemple, un seul m<sup>6</sup>A en 5'UTR est suffisant pour lier un facteur d'initiation eIF3 et recruter le complexe de pré-initiation 43S, permettant ainsi l'initiation de la traduction même en l'absence des facteurs eIF4 (présents sur la coiffe, Meyer *et al.*, 2015).

### 1.11 L'importance du facteur eIF2 $\alpha$ dans l'initiation de la traduction

Avant le commencement de la traduction, le facteur eIF2 $\alpha$  s'assemble avec une guanosine triphosphate (GTP) et un ARNt initiateur méthionylé (Met-ARNt) pour former le complexe ternaire (Hinnebusch, *et al.*,

2016). Le complexe ternaire se lie ensuite à la sous-unité 40S du ribosome et à plusieurs facteurs d'initiation pour former le complexe de pré-initiation 43S (CPI 43S; Hinnebusch *et al.*, 2014). Le CPI 43S est enfin recruté sur la coiffe 5' de l'ARNm, un processus facilité par le complexe de liaison à la coiffe eIF4F (Kong *et al.* 2012). Une fois fixé au transcrit, le CPI 43S parcourt l'ARNm de l'extrémité 5' à son extrémité 3' jusqu'à ce qu'il reconnaisse un codon d'initiation. Lorsqu'un codon d'initiation est rencontré, l'anticodon de Met-ARnt se lie au codon d'initiation (généralement un AUG). La reconnaissance du AUG entraîne l'hydrolyse du GTP en guanosine diphosphate (GDP) sous l'aide de l'eIF5. Cela entraîne la dissociation du facteur eIF1 et du complexe eIF2-GDP du complexe ribosomal 43S et produit un complexe de pré-initiation 48S stable. Ceci est immédiatement suivi par l'association de la grande sous-unité ribosomale 60S pour produire un complexe d'initiation 80S prêt à commencer la synthèse des protéines (Hinnebusch *et al.*, 2016). Enfin, le GDP est libéré et remplacé par une molécule de GTP par eIF2B, un facteur d'échange de guanosine (GEF). Cela restaure le facteur eIF2α à sa forme active, qui peut être utilisé à nouveau pour former un nouveau complexe ternaire (Hinnebusch *et al.*, 2016; **Figure 1.3**).

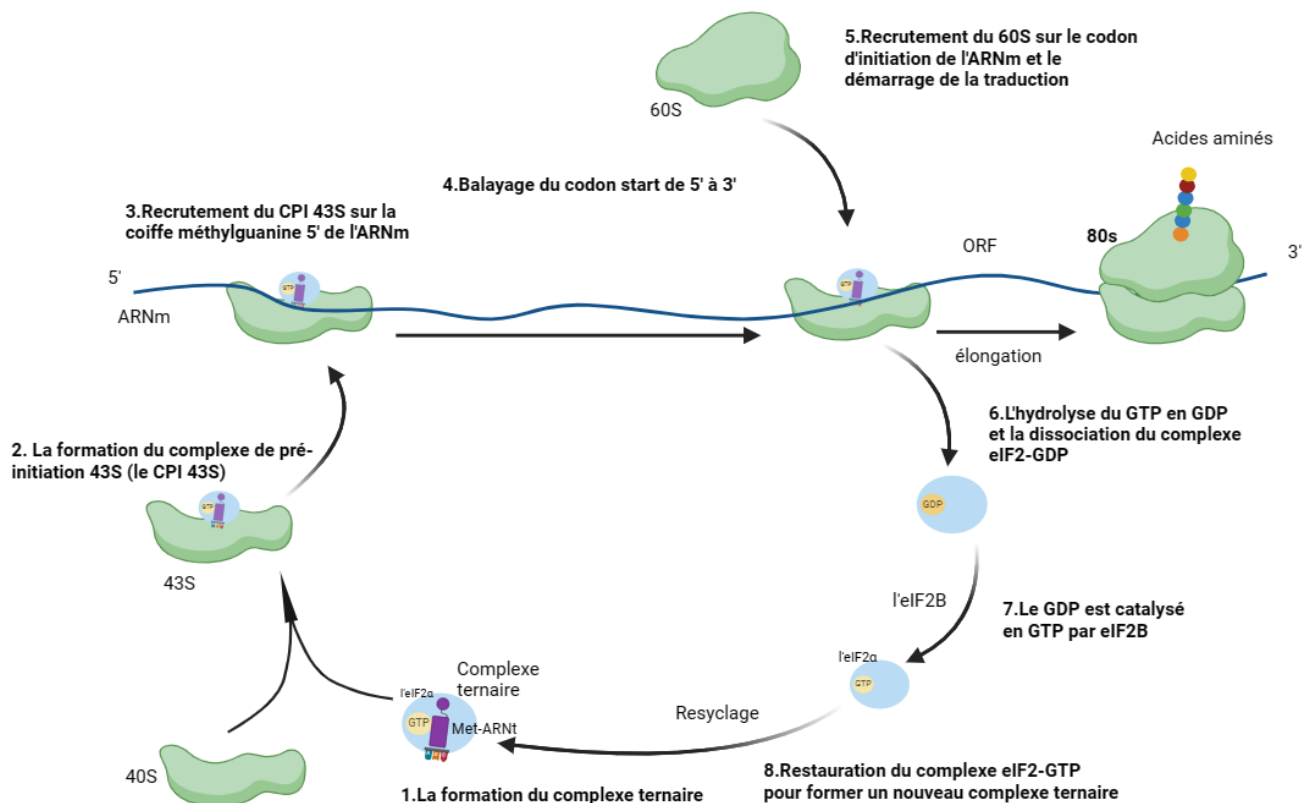


Figure 1.3. L'importance du facteur eIF2α dans la traduction en condition normale.

De 1 à 8, les différentes étapes qui mènent à l'initiation de la traduction dans des conditions normales sont décrites. Tiré de Choteau SA, 2022.

### 1.12 Régulation de la traduction : la phosphorylation d'eIF2 $\alpha$

Le principal mécanisme qui entrave l'initiation de la traduction est la phosphorylation d'eIF2 $\alpha$ . Cela empêche la formation du complexe ternaire et entraîne une inhibition globale de la synthèse des protéines dépendante de la coiffe (Pakos-Zebrucka *et al.*, 2016; Schoof *et al.*, 2021). Une telle phosphorylation d'eIF2 $\alpha$  est déclenchée par la réponse intégrée au stress (RIS), activée par la détection du stress par l'une des quatre kinases (GCN2, PKR, HRI ou PERK) présentant différents domaines de régulation et qui phosphorylent eIF2 $\alpha$  sur la sérine 51 (Ron, 2002). Cela entraîne une diminution de la formation du complexe ternaire nécessaire à l'initiation de la traduction. Bien qu'elle soit associée à une répression globale de la traduction, la phosphorylation de l'eIF2 $\alpha$  permet également la traduction de certains transcrits d'ARNm particuliers, tels que ceux codant pour les facteurs de transcription 4 (ATF4), ATF5, CHOP ou GADD34, qui sont préférentiellement traduits pendant la RIS (Andreev *et al.*, 2015).

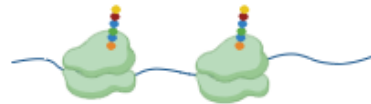
La RIS se termine par la déphosphorylation d'eIF2 $\alpha$  sous l'action de CREP ou de GADD34, deux protéines respectivement exprimées de manière constitutive dans les cellules non stressées et spécifiquement exprimées dans la dernière phase de la RIS (Pakos-Zebrucka *et al.*, 2016; Schoof *et al.*, 2021; Figure 1.4).

### 1.13 La régulation de la traduction par les uORF

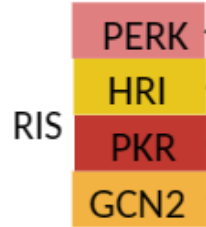
Les uORF sont les régions de l'ORF qui se situent en amont (*upstream*) du CDS (5'UTR). Il a été observé que les CDS les plus résistants au stress possèdent des uORF, et plusieurs études ont démontré l'importance de ces uORF dans cette résistance au stress (Andreev *et al.*, 2015; Andreev *et al.*, 2018). En effet, pendant la RIS, l'initiation générale de la traduction est effectivement inhibée. Cependant, la traduction de la plupart des ARNm, qui résistent à cette inhibition et contiennent des uORF, se produit par une traduction dépendante de la coiffe, à travers deux mécanismes : la fuite du balayage et la réinitiation. Dans la fuite du balayage, le ribosome a la capacité d'omettre le premier AUG au profit d'un autre situé plus en aval (Kozak *et al.*, 1991), tandis que dans la réinitiation, le ribosome traduit l'uORF. Après la terminaison, il reprend le balayage vers le CDS et traduit avec succès l'ORF situé plus en aval (Skabkin *et al.*, 2013). L'exemple standard de ce modèle est le gène ATF4 (facteur de transcription activateur 4), qui se positionne en tant que régulateur central exerçant un rôle essentiel dans l'adaptation au stress, orchestrant la régulation transcriptionnelle de multiples gènes (Harding *et al.*, 2000; Lu *et al.*, 2004; Wek *et al.*, 2006; B'chir *et al.*, 2013). L'ARNm d'ATF4 contient deux uORF, dont le second chevauche le CDS (Vattem *et al.*, 2013).

En condition normale

Initiation de la traduction



En condition de stress



La phosphorylation de eIF2α sous l'action de kinases

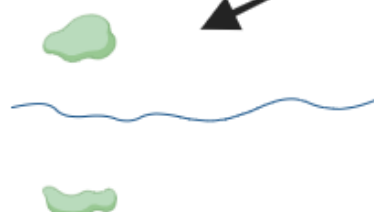
eIF2α  
GDP

eIF2α  
GDP  
P

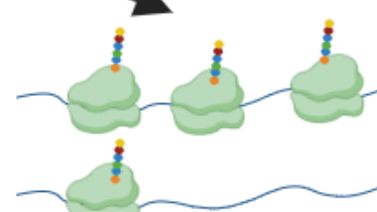
Diminution du complexe ternaire

Séquestration de eIF2B

eIF2B  
GDP  
P



Inhibition globale de la traduction



Traduction de certains ARNm spécifiques

Figure 1.4. L'importance du facteur eIF2α dans la traduction en condition de stress.

Lorsqu'un eIF2alpha est phosphorylé par l'une ou l'autre des kinases en condition de stress, il se lie plus étroitement au facteur d'échange de nucléotides guanine eIF2B et inhibe son activité. eIF2 lié au GDP ne peut plus échanger son GDP contre du GTP, ce qui limite la formation de complexes ternaires ternaires disponibles pour la traduction cap-dépendante. Ce processus engendre deux effets : l'inhibition globale de la traduction de la plupart des ARNm et la traduction de certains ARNm spécifiques. Inspiré de l'article Pakos-Zebrucka *et al.*, 2016.

2004). Dans des conditions normales, la traduction du second uORF inhibe l'expression de la protéine ATF4 en empêchant toute réinitiation du ribosome sur le CDS (Lu *et al.*, 2004). Dans des conditions de stress, la réduction de la disponibilité du complexe ternaire, due à la phosphorylation d'eIF2 alpha, permet la réinitiation retardée du ribosome sur le CDS. Donc, de contourner l'inhibition par le second uORF (Vattem et Wek, 2004).

En outre, ce modèle implique au moins deux uORF, alors que des exemples de régulation de la traduction par un seul uORF ont été décrits. En effet, dans le cas des gènes GADD34 et IFRED qui contiennent un seul uORF, la traduction du CDS est faite par une fuite de balayage qui augmente en condition de stress (Young *et al.*, 2015; Andreev *et al.*, 2015). Cependant, l'existence d'autres mécanismes de régulation de la traduction pendant le stress cellulaire est possible (Andreev *et al.*, 2015).

Enfin, si la plupart des ARNm résistants au stress possèdent des uORF, seule une petite fraction des ARNm contenant des uORF est effectivement résistante au stress, ce qui démontre que la seule présence d'un uORF sur le transcrit n'est pas suffisante pour garantir la résistance au stress de la CDS (Andreev *et al.*, 2015).

#### 1.14 La famille de eIF2 $\alpha$ kinase

Dans les cellules de mammifères, la phosphorylation de la sous-unité alpha du facteur d'initiation de la traduction eucaryote 2 (eIF2 $\alpha$ ) sur la sérine 51 (Ron, 2002) est catalysée par l'une des quatre membres de la famille des kinases eIF2 $\alpha$  à sérine/thréonine, à savoir la kinase PKR-like ER (PERK), la kinase double-stranded RNA-dependent protein kinase (PKR), la kinase heme-regulated eIF2 $\alpha$  (HRI), et la kinase general control nonderepressible 2 (GCN2). Les quatre kinases eIF2 $\alpha$  partagent une homologie significative dans leurs domaines catalytiques, bien qu'elles présentent des domaines régulateurs distincts (Harding *et al.*, 1999). Chaque kinase eIF2 $\alpha$  subit une dimérisation et une autophosphorylation pour atteindre une activation complète (Lavoie *et al.*, 2014). Cependant, chaque kinase réagit de manière spécifique à des stress environnementaux et physiologiques, reflétant ainsi leurs mécanismes de régulation uniques (Donnelly *et al.*, 2013).

##### 1.14.1 PKR-like ER kinase (PERK)

La PERK (également connue sous le nom de PEK, eIF2AK3) est trouvée dans la membrane du réticulum endoplasmique et est principalement activée par l'accumulation de protéines mal repliées dans le



réticulum endoplasmique (RE), un phénomène appelé stress de RE (Harding *et al.*, 1999). Par sa phosphorylation d'eIF2 $\alpha$ , PERK bloque la synthèse de nouveaux polypeptides, réduisant ainsi l'entrée de polypeptides naissants dans la lumière du réticulum endoplasmique. Cela donne au RE le temps de replier les protéines mal repliées et d'éliminer celles qui sont mal repliées en phase terminale, éléments importants de la "réponse aux protéines non repliées" (RPNR/UPR) de la cellule, qui cherche à rétablir l'homéostasie du RE (Szegezdi *et al.*, 2006). PERK a également été impliquée dans des pathologies telles que le cancer (Healy *et al.*, 2009). En effet, il est bien établi que les tumeurs solides sont sujettes à des régions d'hypoxie et l'hypoxie elle-même est un puissant inducteur de la phosphorylation d'eIF2 $\alpha$  dépendant de PERK (Rouschop *et al.*, 2013).

#### 1.14.2 Protein kinase Double-stranded RNA-dependent (PKR)

PKR (ou eIF2AK2) est principalement localisée dans le cytoplasme. Mais dans certaines conditions de stress cellulaire, PKR est transloquée au niveau du noyau (Kim *et al.*, 2014). La PKR mammifère est principalement activée par l'ARN double brin (ARNds) pendant une infection virale, entraînant ainsi l'inhibition consécutive de la synthèse protéique virale et cellulaire et favorisant l'apoptose par la phosphorylation de l'eIF2 $\alpha$ . (Balachandran *et al.*, 2000; Xu *et al.*, 2018). Cependant, PKR joue également un rôle plus général dans la physiologie cellulaire, elle peut être activée en réponse à des signaux aussi divers que le stress oxydatif et au RE (Onuki *et al.*, 2004). De plus, elle a été impliquée dans la pathologie de l'obésité (Nakamura *et al.*, 2010), ainsi que dans le cancer (Pataer *et al.*, 2009).

#### 1.14.3 General control non-derepressible 2 (GCN2)

GCN2 (ou eIF2AK4) est principalement un capteur de la disponibilité des acides aminés, en se liant à l'ARN de transfert (ARNt) non chargé, et un régulateur des changements dans l'expression des gènes en réponse à la privation d'acides aminés (Vazquez de Aldana *et al.*, 1994; Hinnebusch *et al.*, 2005). GCN2 peut également être activée par de nombreux autres facteurs de stress, comme la lumière UV, stress oxydatifs et des médicaments anticancéreux (Deng *et al.*, 2002; Peidis *et al.*, 2010; Grallert *et al.*, 2013). De plus, GCN2 joue des rôles importants dans divers processus biologiques et maladies (Chenxu Zhao *et al.*, 2023). En effet, GCN2 a été découvert pour réduire significativement les symptômes de la maladie auto-immune systémique (Ravishankar *et al.*, 2015), ce qui ouvre la voie au développement de médicaments ciblant GCN2 (Tang *et al.*, 2022).

#### 1.14.4 Heme-regulated eIF2a kinase (HRI)

HRI (ou EIF2AK1) est principalement exprimée dans les cellules érythroïdes, contrairement aux autres kinases eIF2 $\alpha$  qui présentent une large distribution tissulaire, où elle participe à la différenciation des érythrocytes au cours de l'érythropoïèse. HRI relie la traduction des ARNm de la globine à la disponibilité de l'hème pour la production d'hémoglobine et protège ainsi les cellules érythroïdes contre l'accumulation d'agrégats de globine toxiques en cas de carence en fer (Suragani *et al.*, 2012). HRI peut également être activée par d'autres facteurs de stress, notamment le stress oxydatif induit par l'arsénite, le choc thermique, le stress osmotique, l'inhibition du protéasome 26S et l'oxyde nitrique (McEwen *et al.*, 2005; Lu *et al.*, 2001, Ramos-Fernández *et al.*, 2016). Par ailleurs, le concept global du stress oxydatif est défini comme un déséquilibre entre les oxydants (espèces réactives de l'oxygène) et les antioxydants en faveur des oxydants, conduisant à une perturbation du contrôle et/ou à des dommages moléculaires (Sies *et al.*, 2017). Un stress oxydatif prolongé conduit à diverses conditions pathologiques telles que le vieillissement, l'inflammation, les dommages à l'ADN, et des maladies chroniques (Valko *et al.*, 2007; Kudryavtseva *et al.*, 2016; Hajam *et al.*, 2022). De manière intéressante, l'activation de HRI par ces divers stress est indépendante de l'hème et se produit avec l'aide des protéines de choc thermique HSP90 et HSP70 (Lu *et al.*, 2001).

#### 1.15 Modifications post-traductionnelles

La traduction est souvent considérée comme la "dernière étape" du "dogme central" de la biologie, où l'ADN est transformé en ARN puis en protéine. Cependant, il existe des étapes supplémentaires après la synthèse des protéines qui sont nécessaires pour qu'une cellule, un tissu et un organisme atteignent leur biologie fonctionnelle et leur diversité. Les modifications post-traductionnelles (MPT) sont des changements apportés aux protéines après leur synthèse, généralement médiés par des enzymes. Ces modifications comprennent la phosphorylation, la glycosylation, la sulfatation, la méthylation, la sumoylation, la citrullination, la farnésylation, la biotinylation, l'ubiquitination et le clivage protéolytique (Khoury *et al.*, 2011; Ramazi *et al.*, 2021; Zhou *et al.*, 2022). Les MPT jouent un rôle clé dans la signalisation cellulaire, la dégradation des protéines, la localisation cellulaire, et de nombreux autres processus biologiques. Grâce aux modifications post-traductionnelles, les cellules sont capables d'ajuster finement la fonction et l'activité des protéines pour s'adapter aux besoins environnementaux et physiologiques changeants (Stram and Payne, 2016; Xu *et al.*, 2018; Gupta *et al.*, 2021).

## 1.16 La glycosylation

La glycosylation des protéines est l'une des modifications post-traductionnelles les plus courantes et les plus complexes que l'on trouve dans les protéines. Le processus de glycosylation attache de manière covalente des sucres à l'extrémité croissante des chaînes de glycanes sur les protéines néosynthétisées, catalysé par les glycosyltransférases. Ces modifications complexes modulent les propriétés des protéines qu'elles décorent. Ces modifications jouent un rôle crucial dans tous les aspects de la biologie, notamment l'augmentation de la solubilité des protéines, l'accroissement de la stabilité structurelle et la protection contre la protéolyse, l'aide au repliement des protéines, la participation aux réponses immunitaires, la reconnaissance cellule-cellule et cellule-matrice extracellulaire (ECM) et le ciblage sélectif des protéines dans les destinations intra- ou extracellulaires (Nalivaeva *et al.*, 2001; Rudd *et al.*, 2001; Varki *et al.*, 2009; Hadley *et al.*, 2014; Ramazi *et al.*, 2021).

La diversité de la glycosylation chez les eucaryotes dépend des transporteurs de sucres nucléotidiques qui acheminent les sucres nucléotidiques spécifiques synthétisés dans le cytosol et le noyau vers le réticulum endoplasmique et l'appareil de Golgi où se déroulent les réactions de glycosylation par les glycosyltransférases (Hadley *et al.*, 2014). Ce processus de transport est réalisé par une famille de protéines transmembranaires appelées transporteurs de sucres nucléotidiques (NST; Gerardy-Schahn *et al.*, 2001; Caffaro *et al.*, 2006).

## 1.17 Les transporteurs de solutés SLC

Les membranes cellulaires, y compris celles qui entourent les organites, sont des barrières biologiques qui peuvent sélectivement permettre, inhiber, restreindre ou déterminer le débit d'une gamme de solutés, tels que des molécules chargées organiques ou inorganiques. Les transporteurs constituent une solution efficace pour le déplacement de solutés spécifiques dans ces barrières hydrophobes, qui seraient autrement repoussés. Les transporteurs de solutés (SLC) constituent la deuxième famille de protéines membranaires par ordre d'importance. Les SLC, qui constituent une classification des transporteurs humains, comprennent aujourd'hui 52 familles pouvant transporter une grande variété de solutés (Hediger *et al.*, 2004).

### 1.17.1 La famille de transporteurs de solutés SLC35

La famille de transporteurs de solutés SLC35 (HUGO Gene Nomenclature Committee) comprend des membres de la famille de NST humains conservés au cours de l'évolution. La famille de transporteurs de

solutés SLC35 des NST humaines est divisée en 7 sous-familles (SLC35A-G), identifiées sur la base de la similarité des séquences. Chaque sous-famille de NST est ensuite divisée pour différencier le type de substrat(s) supposé(s) être transporté(s) (Fredriksson *et al.*, 2008). De plus, plusieurs membres de la sous-famille SLC35 ont été qualifiés de transporteurs orphelins, car la spécificité de leur substrat et leurs fonctions physiologiques restent à déterminer (Hadley *et al.*, 2019).

À ce jour, la sous-famille SLC35A a été largement caractérisée. Il est postulé que SLC35A1 est responsable de l'acheminement de l'acide CMP-sialique et CDP-ribitol (Eckhardt *et al.*, 1998; Ury *et al.*, 2021), SLC35A2 est essentiel pour le transport de l'UDP-galactose (Brändli *et al.*, 1988) tandis que SLC35A3 est crucial pour l'absorption de l'UDP-N-acetylglucosamine (Guillen *et al.*, 1998). Quant à SLC35A4, elle interagit avec SLC35A2 et SLC35A3 (Wiktor *et al.*, 2021). De plus, Ury *et al.* ont trouvé que le SLC35A4 transporte le CDP-ribitol, avec SLC35A1, mais de manière redondante (Ury *et al.*, 2021).

#### 1.17.2 L'importance du CDP-ribitol

La glycoprotéine  $\alpha$ -dystroglycane joue un rôle crucial en reliant le cytosquelette intracellulaire à la matrice extracellulaire. Une structure de glycane spécifique est nécessaire pour son interaction avec des protéines de la matrice extracellulaire, comme la laminine. Il s'agit du seul glycan mammifère connu contenant des groupes ribitol phosphate. Les enzymes du Golgi utilisent le CDP-ribitol pour intégrer le phosphate de ribitol dans la chaîne de glycan de l' $\alpha$ -dystroglycane, synthétisé dans le cytoplasme et transporté par SLC35A1/4. Les anomalies dans les enzymes liées à la biogenèse de ce glycan entraînent des syndromes congénitaux, les dystroglycanopathies, caractérisés par des symptômes musculaires, cérébraux et oculaires, avec une gamme clinique allant d'une faiblesse musculaire légère tardive à une dystrophie musculaire congénitale sévère, comprenant des atteintes cérébrales et oculaires (maladie muscle-œil-cerveau, syndrome de Walker-Warburg et dystrophie musculaire congénitale de Fukuyama) (Michele *et al.*, 2002; Endo *et al.*, 2015; Live *et al.*, 2013; Yoshida-Moriguchi *et al.*, 2015).

#### 1.17.3 Le gène double codant *SLC35A4*

Le gène *SLC35A4* est localisé dans le chromosome 5q31.3. Il code pour deux protéines, (d'où le nom de gène double codant) : une protéine alternative dont la séquence codante est localisée dans la région 5'UTR (AltSLC35A4), en plus de la protéine de référence SLC35A4 codée à partir de la région codante de référence (CDS/RefORF) (Figure 1.5, Vanderperre *et al.*, 2013).

La protéine alternative AltSLC35A4 demeure non caractérisée. Elle est constituée de 103 acides aminés et est très abondante et conservée chez les vertébrés, suggérant qu'elle joue un rôle important dans l'organisme. Quant à sa localisation et sa fonction, les résultats préliminaires de notre laboratoire ont pu confirmer qu'elle est localisée au niveau de la membrane interne de la mitochondrie et est transmembranaire. Une analyse interactomique de AltSLC35A4 par la méthode de co-immunoprécipitation nous révèle qu'elle interagit principalement avec des sous-unités de myosine et d'actine.

La protéine de référence SLC35A4 est constituée de 324 acides aminés (Bechtel *et al.*, 2007). Elle est exprimée dans la membrane du Golgi et est composée de dix domaines transmembranaires. Son ARNm est largement exprimé dans une variété de cellules humaines (GeneCards). Elle partage une partie de la fonction avec le SLC35A1, qui consiste à importer le CDP-Ribitol dans la lumière du Golgi, mais de manière redondante, comme décrit précédemment (Sosicka *et al.*, 2017; Ury *et al.*, 2021).

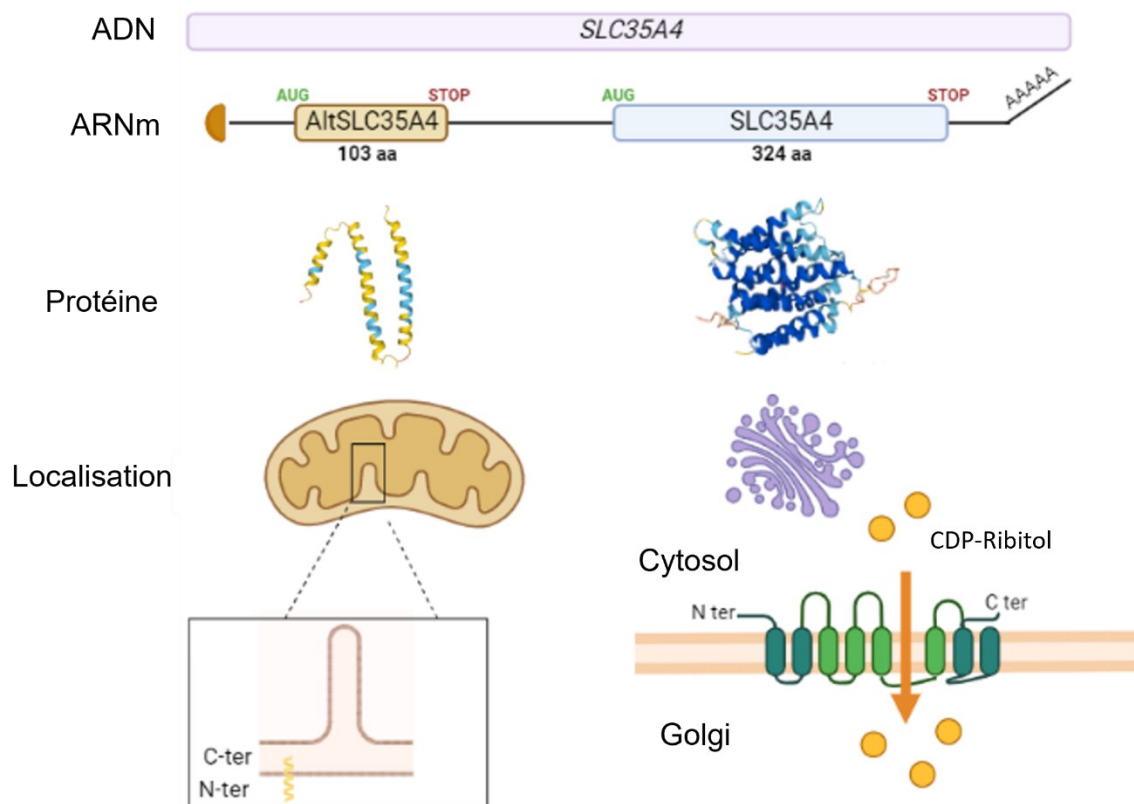


Figure 1.5. Le gène double codant *SLC35A4*.

### 1.18 Hypothèse et objectif

Des études précédentes, par profilage ribosomique, lors du stress oxydatif et du stress du réticulum endoplasmique induits respectivement par l'arsénite de sodium (AS) et la tunicamycine (Tn), indiquent que la traduction d'AltSLC35A4 est maintenue, et que la CDS de SLC35A4 subit la plus forte augmentation d'efficacité traductionnelle parmi tous les ARNm cellulaires (Andreev *et al.*, 2015 ; Paolini *et al.*, 2018; Figure 1.6). Ceci est un trait crucial pour les protéines essentielles dans l'adaptation de la cellule au stress, que ce soit de manière protectrice ou délétère (Fulda *et al.*, 2010). De plus, comme expliqué précédemment, les CDS traduits dans des conditions de stress possèdent des uORF (Andreev *et al.*, 2018) ce qui est le cas pour SLC35A4. Ces observations renforcent l'idée que la protéine SLC35A4 joue un rôle crucial pendant le stress. Cependant, aucune étude n'a encore décrit l'importance de la protéine SLC35A4 pendant le stress cellulaire et son lien avec la réponse intégrée au stress.

Notre hypothèse suggère une augmentation de l'expression de la protéine SLC35A4 pendant le stress oxydatif. De plus, les protéines codées par *SLC35A4* pourraient influencer l'aptitude cellulaire en situation de stress en conférant aux cellules une résistance. Par conséquent, notre objectif initial est de valider la dynamique d'expression de la protéine SLC35A4 pendant le stress oxydatif induit par l'arsénite de sodium. De plus, notre second objectif vise à évaluer l'importance des protéines codées par le gène *SLC35A4* lors du même type de stress cellulaire en comparant la résistance au stress entre les lignées cellulaires knock-out (KO) de SLC35A4 et de AltSLC35A4 avec des cellules de type sauvage (*Wild type*).

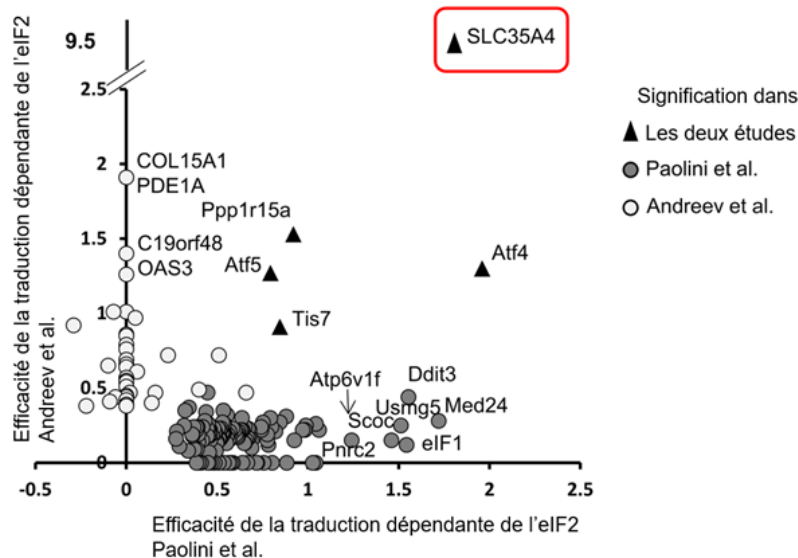


Figure 1.6. Comparaison de l'occupation des ribosomes en réponse à la phosphorylation d'eIF2 $\alpha$ .

Cette expérience a été faite dans les cellules HEK293 (Andreev *et al.*, 2015) et les érythroblastes de souris (Paolini *et al.*, 2018). Les triangles indiquent les transcrits dont la traduction est régulée de manière similaire lors de la phosphorylation du eIF2 dans les deux études. Les cercles blancs représentent les transcrits dont la traduction est augmentée (Z-score >3) dans les cellules HEK293 mais pas dans les érythroblastes de souris; les cercles gris foncé représentent les transcrits dont la traduction est augmentée dans les érythroblastes de souris (FDR (*False Discovery Rate*) <0,01) mais pas dans les cellules HEK293.

## CHAPITRE 2

### MATÉRIELS ET MÉTHODES

#### 2.1 Culture cellulaire et traitement

Les cellules 143B et HEK293T, ont été cultivées dans du milieu Dulbecco's Modified Eagle Medium (DMEM) (Wisent, #319-005-CL) contenant 10 % de sérum de veau fœtal (SVF) (Wisent, #080-450) et solution contenant 1 % pénicilline-streptomycine (Bio Basic, #PB0135, SB0494) dans des conditions standards (37°C et dans une atmosphère humide contenant 5 % de CO<sub>2</sub>).

Les cellules ont été traitées avec de l'arsénite de sodium (Sigma, #7784-46-5). La solution mère de 4 mM a été préparée avec de l'eau. Une concentration de 40 µM après dilution dans du DMEM complet a été appliquée aux cellules pendant les périodes spécifiées pour chaque expérience.

#### 2.2 Clonage

##### 2.2.1 Plasmide pLJM1\_SLC35A4-3xHA

La séquence SLC35A4 a été obtenue par amplification par PCR (amorce : 103/160) à partir du plasmide pLJM1\_SLC35A4-mNeonGreen. Ensuite, un amplicon contenant trois (3) étiquettes HA, obtenu par PCR (amorce :7/8) à partir du plasmide pLJM1\_hAltSLC35A4-3xHA, ont été ajouté en C-terminale sur la séquence par l'assemblage Gibson avec un ratio 1:2 (SLC35A4 : 3xHA). Après l'assemblage Gibson, nous avons amplifié la séquence SLC35A4-3xHA par PCR (amorce :103/8). Ensuite, le produit de PCR a été sous cloné dans le plasmide pLJM1 par clonage classique (digestion et ligation). En effet, la digestion du plasmide pLJM1-EGFP (Addgene #19319) et de l'insert SLC35A4-3xHA a été faite par les enzymes : EcoRI (TransGen, #JE201-01) et NheI (TransGen, #JN301-02) conformément aux instructions du fabricant, suivie par la ligation qui a été faite par un ratio de 6:1 (plasmide : insert), nous avons utilisé T4 ligase (TransGen, #FL101-01) conformément aux instructions du fabricant. Toutes les PCR ont été faites en utilisant l'ADN polymérase Q5 (Neb, #M0491S) en suivant les instructions du fabricant. Les produits de PCR ont été migrés sur un gel d'agarose de 1%, extraits et purifiés en utilisant le kit : EZ-10 Spin Column DNA Gel Extraction Minipreps Kit (BIO BASIC).



### 2.2.2 Plasmide pLJM1\_AltSLC35A4\_SLC35A4-3xHA

La séquence AltSLC35A4\_SLC35A4 a été obtenue par amplification par PCR (amorce : 207/160) à partir de l'ADNc humain fourni par le laboratoire de la Pre. Mounier. Ensuite, un amplicon contenant trois (3) étiquettes HA, obtenu par PCR (amorce : 7/8) à partir du plasmide pLJM1\_SLC35A4-3xHA, ont été ajouté en C-terminal sur la séquence par l'assemblage Gibson avec un ratio 1:2 (Alt\_Slc : 3xHA). Nous avons amplifié, par la suite, la séquence AltSLC35A4\_SLC35A4-3xHA par PCR (amorce :207/8). Ensuite, le produit de PCR a été sous clonées dans le plasmide pLJM1 par clonage classique (digestion et ligation). En effet, la digestion du plasmide pLJM1-EGFP (Addgene #19319) et de l'insert AltSLC35A4\_SLC35A4-3xHA a été faite par les enzymes : EcoRI (TransGen, #JE201-01) et NheI (TransGen, #JN301-02) conformément aux instructions du fabricant, suivie par la ligation qui a été faite par un ratio de 6:1 (plasmide : insert), nous avons utilisé T4 ligase (TransGen, #FL101-01) conformément aux instructions du fabricant. Toutes les PCR ont été faites en utilisant l'ADN polymérase Q5 (Neb, #M0491S) en suivant les instructions du fabricant. Les produits de PCR ont été migrés sur un gel d'agarose de 1%, extraits et purifiés en utilisant le kit : EZ-10 Spin Column DNA Gel Extraction Minipreps Kit (BIO BASIC).

Tableau 2-1. Amorces utilisées

Amorce	Nom de l'amorce	La séquence de l'amorce
7	3xHA_F	ggaggtggaggtagttatccc
8	pLJM1_Gibson_R	ccatttgtctcgaggtcgagaattc
103	F_SLC35A4-NG	tagtgaaccgtcagatccgctagcaccatgagtgtagaggatgggggatgc
160	HA_SLC35A4_R	gggataactacctccacctccGcggctgcatagtacagg
207	AltSLC35A4_F_Reg	Gtttagtgaaccgtcagatccgctagcaagttcggcgggaa

### 2.2.3 Transformations des bactéries

Le produit de la ligation a été ajouté dans les bactéries compétentes *Escherichia coli* XL-10 gold. Le mélange a été mis sur glace pendant 20 minutes, puis un choc thermique à 42°C/45 sec a été appliqué, suivi d'un refroidissement pendant 2 min sur glace. Du milieu LB a été ajouté, suivi d'une incubation d'une heure à 37°C avec agitation. Le tout a été centrifugé à 3000 g/2 min. Une bonne partie du surnageant a été enlevée et le reste (50µl) a été utilisé pour resuspendre le culot et l'étaler sur un pétri de LB agar avec de l'ampicilline à 37 °C pendant la nuit. Des colonies ont été isolées et déposées dans un volume de 3 ml du milieu LB contenant de l'ampicilline 0.1 mg/ml et incubées à 37°C pendant une nuit avec agitation.

### 2.2.4 Extraction des plasmides par Miniprep

Un Kit EZ-10 Spin Column Plasmide DNA Minipreps Kit, de la compagnie BIO BASIC, a été utilisé pour extraire les plasmides. La vérification du clonage a été faite par PCR et par séquençage Sanger chez Génome Québec.

## 2.3 Génération de populations cellulaires stables

### 2.3.1 Production des lentivirus

Des cellules HEK293T ont été cultivées dans une plaque de 6 puits dans le milieu DMEM avec 10% de SVF et solution contenant 1 % pénicilline-streptomycine. Une fois à 80% de confluence (1 millions de cellules), elles ont été co-transfectées avec les plasmides pRSV-Rev (Addgene, #12253), pMDLg/pRRE (Addgene, #12251), pMD2G (Addgene, #12259) et la construction SLC35A4-3xHA ou AltSLC35A4\_SLC35A4-HA avec les quantités d'ADN respectives de 0,5µg, 0,6µg, 1µg et 2 µg. Cette opération a été réalisée en utilisant la méthode de coprécipitation au phosphate de calcium, comme décrit précédemment (Jordan *et al.*, 1996).

Brièvement, les plasmides avec les bonnes concentrations ont été mélangés avec de l'eau tamponnée, puis le CaCl<sub>2</sub> 2.5M a été ajouté. Ensuite, le HeBS 2x (HEPES-buffered saline) a été ajouté goutte à goutte avec agitation par vortex. Après 10 min d'incubation à température ambiante tout le mélange a été ajouté goutte à goutte sur les cellules de la plaque de 6 puits. Les cellules ont été incubées pendant 8 h à 37°C. Après 8 h, le milieu a été remplacé par 2mL de DMEM avec 10% de SVF et 1 % pénicilline-streptomycine et NEAA (*Non Essential Amino Acid*) (Wisent, #321-011-EL) et les cellules ont été incubées pendant 12h à 37°C . Les virus ont donc été récoltés le lendemain en prélevant le milieu des HEK293T (contenant les

lentivirus en suspension). À l'aide d'une seringue de 10mL le milieu a été filtré avec un filtre stérile de 0,45µm (membrane polyéthersulfone), et les virus recueillis ont été congelés à -80°C.

### 2.3.2 Infections lentivirales

Le jour de l'infection, les cellules 143B, *Wild type* ou Knock-out (afin de générer des cellules sauvées) ont été préparées dans une plaque à 6 puits, de manière à ce que chaque puits contient 40 000 cellules dans 1 ml de DMEM complet (10% de SVF et 1 % pénicilline-streptomycine) et 1 µl de polybrene (Thermofisher, #TR1003G). Ensuite, différentes concentrations de lentivirus ont été ajoutées dans les puits de manière croissante, tout en laissant un puits sans virus (cellules contrôle non infectées), pour amener le volume final à 2 mL. Après 24 h d'incubation à 37 °C, le milieu de culture a été remplacé par du milieu DMEM complet. Après une autre incubation de 24 h à 37 °C, les cellules ont été passées avec un milieu de culture contenant 1,25 µg/mL de puromycine (Wisent, #450-162-XL) pour effectuer la sélection des cellules stables. Après 48 heures d'incubation à 37 °C, ou après la mort de toutes les cellules sauvages, on conserve les puits où il y a eu >50% de mort cellulaire (pour augmenter les chances qu'une copie du virus soit présente dans chaque cellule) dans un milieu DMEM avec 10% de SVF et 1 % pénicilline-streptomycine. Des *Western blot* et Immunofluorescences ont été réalisés afin de vérifier l'efficacité de la génération de ces populations stables.

### 2.4 Immunobuvardage (*Western blot*)

Les cellules sur la plaque à 6 puits ont été initialement rincées avec du PBS (phosphate buffered saline), puis traitées à la trypsine et recueillies dans des tubes à centrifugation. Par la suite, elles ont été centrifugées, le surnageant a été éliminé, et le culot a été rincé une seconde fois avec du PBS froid. Après une nouvelle centrifugation, le culot a été resuspendu soit dans le tampon Laemmli 1x avec du 2-Mercaptoéthanol (βME), soit dans le tampon RIPA (25 mM Tris, 1 M, pH 7.4, 0.1% SDS, 1% Sodium Deoxycholate, 1% de 10% NP40, 150 mM de 5 M NaCl, 1 mM EDTA et des inhibiteurs de protéases), puis incubé sur de la glace pendant 20 minutes avant une sonication à 20% pendant 20 secondes. Ensuite, le tampon de Laemmli, 5x contenant βME, a été ajouté aux échantillons de protéines pour atteindre une concentration finale 1x (dans le cas où la lyse a été effectuée dans du RIPA) ensuite ils ont été bouillis à 95 °C pendant 10 minutes. Ces échantillons ont été chargés sur un gel SDS-PAGE 12% pour l'électrophorèse, puis les protéines ont été transférées à 100 V pendant 1 h sur une membrane de nitrocellulose de 0,45 µm (Millipore Sigma). Après le blocage des membranes avec du lait à 5% dans une solution de TBS contenant 0,1% de Tween pendant une heure, celles-ci ont été incubées avec les anticorps primaires à 4 °C pendant

la nuit. Après trois lavages de 5 minutes chacun avec du TBST, les membranes ont été incubées avec des anticorps secondaires conjugués à la peroxydase pendant 1 h (*Goat anti-rabbit ou Goat anti-mouse IgG*, Jackson Immunoresearch). Les anticorps ont été dilués dans la solution de blocage (5% de lait, TBST). Enfin, après trois nouveaux lavages au TBST de 5 minutes chacun, les membranes ont été exposées à du *Clarity Western ECL* (chimioluminescence améliorée) puis révélées à l'aide de la machine Fusion Fx7 (MBI).

Tableau 2-2. Anticorps utilisés

Anticorps Primaire	Numéro de catalogue
ATF4	Cell Signaling #11815S
Étiquette HA	Cell Signaling (6E2) #2367
Golgin-97 (D8P2K)	Cell Signaling #13192S
SLC35A4	Sigma, #HPA062210
SLC35A4	Sigma #SAB1304268

## 2.5 Immunofluorescence (IF)

Des cellules 143B ont étéensemencées dans des plaques 24 puits contenant des lamelles. Le lendemain, les cellules ont été fixées avec une solution de PFA 4% (paraformaldéhyde) diluée dans du PBS pendant 15 min, puis lavées au trois fois, ensuite, perméabilisées avec 0,15% de Triton X-100 dans du PBS pendant 5 min, puis lavées encore 2 fois au PBS. Les cellules ont été incubées dans une solution de blocage (PBS, 5% de NGS, pour sérum normal de chèvre) pendant 30 min, le tout à température ambiante. Les lamelles ont ensuite été incubées toute la nuit à 4 °C ou 3 h à température ambiante avec les anticorps primaires anti-HA (dilution 1/500), anti-Golgin97 (dilution 1/500) préparés dans du NGS 2.5% dans du PBS. Après 3 lavages au NGS 2.5%/PBS, les cellules ont été incubées pendant 1 h, à température ambiante et dans le noir, avec les anticorps secondaires anti-lapin Alexa-Fluor647 (abcam, #ab150079) (dilution 1/1000), anti-souris DyLight480 (ImmunoReagents Inc #GTXMU-003-D488NHSX). Après 3 lavages au PBS, les noyaux des cellules ont ensuite été colorés avec du Hoechst (Invitrogen, #H3570) (dilution 1/10 000) dans du PBS pendant 10 min, toujours à température ambiante et dans le noir. Enfin, les lamelles ont été montées et

fixées sur des lames avec du Dabco. Les cellules ont été examinées à l'aide d'un microscope inversé EVOS M5000 et l'acquisition des images a été faite avec une unité confocale Nikon A1 (Objectif à l'huile 20x) et le logiciel NIS-Element (Nikon).

## 2.6 Génération des cellules SLC35A4 knock-out par CRISPR Cas9 (*Clustered regularly interspaced short palindromic repeats*)

### 2.6.1 Designs et protocole

Nous avons utilisé la technique d'édition des gènes CRISPR/Cas9 pour l'inactivation du SLC35A4. Les petits ARN guide (ARNsg ou *sgRNA*) ont été conçus suivant l'outil de conception CHOPCHOP, ils ont été vérifiés pour les coupures hors cible à l'aide de la même plateforme, et ont été commandés par Synthego. En effet, nous avons utilisé deux ARNsg de 20 nucléotides afin de pouvoir déléter théoriquement 275 paires de bases (pb) de la séquence codant pour SLC35A4 du gène *SLC35A4* (Figure 2.1).

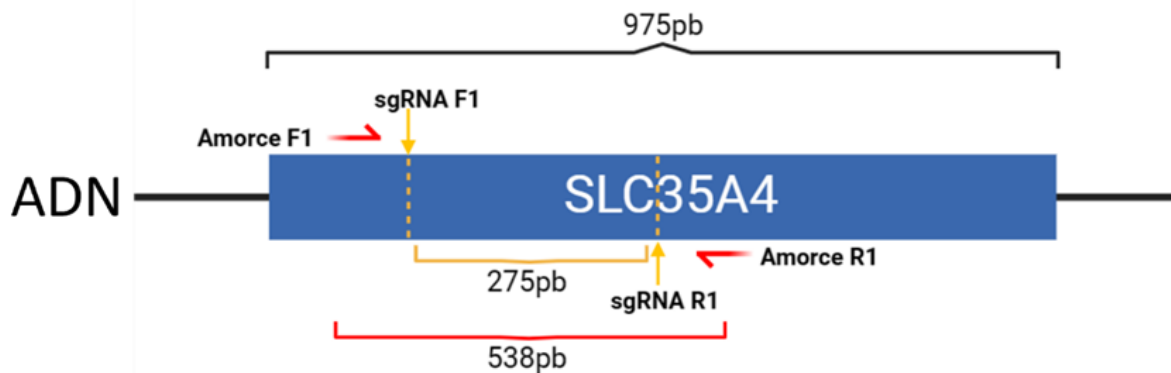


Figure 2.1. Édition de la séquence codante de SLC35A4.

### 2.6.2 Transfection

En résumé, dans une plaque de 24 puits, 15,5 µl de chaque ARNsg ont été ajoutés à partir d'une solution de stockage de 3 µM. Puis un mélange de l'agent de transfection Hiperfect (Qiagen #301705) et de DMEM (sans SVF) a été ajouté. Après une incubation de 10 minutes à température ambiante, 13 750 cellules 143B exprimant Cas9 de manière stable ont été ajoutées. Après 24 h, le milieu de culture a été remplacé par du DMEM complet. Quatre jours plus tard, lorsque les cellules polyclonales ont atteint une confluence approximative de 70%, une partie des cellules a été soumise à la cytométrie de flux (FACS, pour tri cellulaire activé par fluorescence), une partie a été congelée et mise en banque à -80 °C, et le reste de la population polyclonale a été testé pour évaluer l'efficacité de l'édition génique. En effet, la région englobant le site de clivage prédit a été amplifiée par PCR sur ADN génomique.

### 2.6.3 Cytométrie en flux (FACS, *Fluorescence-Activated Cell Sorting*)

Cette technique permet d'analyser et de trier des populations cellulaires en fonction de divers paramètres, tels que la taille, la granularité cellulaire et la fluorescence (Picot *et al.*, 2012). Nous avons utilisé cette technique pour générer des populations de cellules monoclonales. En effet, une seule cellule de la population polyclonale a été déposée dans chaque puits d'une plaque de 96 puits. Deux semaines plus tard, l'ADN génomique a été isolé (QuickExtract, #LGN-QE09050) à partir d'une portion des cellules de chaque puits qui contient une colonie. Les produits de PCR amplifiés à partir de la région englobant le site de clivage prédit ont d'abord été analysés par électrophorèse sur gel d'agarose.

### 2.6.4 Validation de l'édition du gène

Ensuite, afin de vérifier l'édition de chaque allèle dans le génome (qu'il soit homozygote ou hétérozygote) et compte tenu du fait que le produit amplifié par PCR contient les deux allèles, nous avons cloné le produit amplifié par PCR dans des plasmides pBluescript KS (-) préalablement digérés. Cette étape a été réalisée dans l'attente que chaque allèle puisse être cloné individuellement dans un plasmide. En séquençant suffisamment de clones plasmidiques, les différentes versions alléliques peuvent toutes être déterminées. Le clonage du produit de PCR dans le plasmide a été effectué par ligation à l'aide de la ligase T4 (TransGen, #FL101-01), conformément aux instructions du fabricant. La transformation des bactéries et l'extraction des plasmides ont été réalisées comme décrit dans les sections 2.2.3 et 2.2.4 mais la souche bactérienne utilisée est DH5Alpha. La confirmation du clonage a été effectuée par PCR en utilisant les amorces T7 et T3. Enfin, les candidats prometteurs ont été envoyés à Génome Québec pour séquençage Sanger en utilisant l'amorce T7.

Extraction d'ADN par QuickExtract : les cellules ont été bien mélangées avec la solution QuickExtract, puis placées dans la machine PCR à 65 °C pendant 20 minutes, suivies de 98 °C pendant 5 minutes.

Toutes les PCR ont été faites en utilisant l'ADN polymérase Q5 (Neb, #M0491S) en suivant les instructions du fabricant.

Tableau 2-3. Liste des ARNsg et des amorces utilisées pour l'édition de SLC35A4 par CRISPR/Cas9

	Nom	Séquence

sgRNA 5' > 3'	sgRNA 1F	tggacggccgagtgcccttc
	sgRNA 1R	ctaaccctgacgcacagag
Génotype amorces	F1	gatgctactcctatccactgcc
	R1	gcctgagatgaggcagtac

## 2.7 Mesure de la prolifération cellulaire

Pour évaluer la prolifération cellulaire, nous avons utilisé une machine de comptage automatique (LUNA 2). Un nombre égal de cellules ( $0,05 \times 10^6$ ) a étéensemencé dans des plaques de 12 puits avec du DMEM complet et incubé à 37 °C pendant 72 h. Pendant cette période, une plaque a été prélevée chaque 24 h, et les cellules ont été comptées conformément aux instructions du fabricant. Le nombre de cellules déposé à temps 0 h a été défini comme 1, ensuite le nombre à chaque autre temps (24 h, 48 h, 72 h) a été normalisé par rapport à celui de 0 h. Les données ont été analysées par GraphPad PRISM.

## 2.8 Mesure de la viabilité cellulaire : essai CellTiter-Glo® 2.0

L'essai CellTiter-Glo® 2.0 (Pomega, #G9242) offre une mesure de viabilité des cellules en culture en quantifiant l'ATP, ce qui indique la présence de cellules métaboliquement actives. La lecture de la luminescence est directement proportionnelle au nombre de cellules viables en culture.

En effet, nous avons réalisé une expérience en séquence temporelle au cours de laquelle différents groupes de cellules ont été traités par l'arsénite de sodium (AS) pendant des durées différentes (0 h pour le contrôle, 1 h, 2 h, 4 h, 8 h). Brièvement, la veille de l'expérience, un nombre égal de cellules a été cultivé dans une plaque à 96 puits, compatible avec l'expérience, avec du DMEM complet, puis incubé à 37 °C. Le jour suivant, et à chaque moment prédéterminé (0 h, 4 h, 6 h, 7 h), la plaque a été retirée de l'incubateur, et le milieu de culture de chaque groupe de cellules a été remplacé par du DMEM contenant 40 µM d'AS (sauf pour le groupe contrôle). À la fin de la 8<sup>e</sup> heure, 100 µL de la solution CellTiter-Glo 2.0 ont été ajoutés dans tous les groupes cellulaires selon les instructions du fabricant, et la plaque a été analysée par une machine de lecture de luminescence. Chaque luminescence de contrôle (temps 0 h) pour chaque type

cellulaire a été définie comme 100%. Ensuite, la luminescence à chaque autre temps a été normalisée par rapport au contrôle (0 h). Les résultats ont été analysés par GraphPad PRISM.

## 2.9 Immunoprécipitations (IP)

Des cellules 143B de la lignée stable SLC35A4-3xHA et *Wild type* ont été cultivées dans des boîtes de 100 mm. Lorsqu'elles ont atteint une confluence de 90-100 %, les cellules ont été rincées avec du PBS, puis directement lysées et récoltées dans 1,5 mL de tampon d'immunoprécipitation (IP) frais (0,5 % d'Igepal, 50 mM de Tris pH 7,5, 250 mM de NaCl, 1 mM d'EDTA, des inhibiteurs de protéases et un inhibiteur de phosphatase) avant d'être transférées dans un tube de 1,5 mL. Une centrifugation de 30 minutes (18 000 × g, 4 °C) a été réalisée. Le surnageant a été conservé, dont 50 µL ont été mis de côté pour une analyse par *Western blot* (entrée). Les billes magnétiques couplées aux anticorps anti-HA (Pierce #88836) ont été préparées par deux rinçages avec 1 mL de tampon IP. La totalité du surnageant a été ajoutée à 50 µL de billes magnétiques. Les mélanges ont ensuite été placés sur un agitateur rotatif pendant 2 heures à 4 °C. Après 2 heures, à l'aide d'un portoir magnétique, le surnageant a été récupéré (fraction non liée). Par la même méthode, les billes ont ensuite été lavées trois fois avec 1 mL de tampon IP et deux fois avec du tampon IP sans Igepal. Un aliquot de 5µL de billes a été mis de côté pour chaque échantillon afin de vérifier par *Western Blot* (sortie) l'efficacité des immunoprécipitations (IP) avant d'envoyer les échantillons pour analyse par spectrométrie de masse. L'analyse des échantillons a été faite par deux différentes plateformes CERMO-FC et RI-MUHC. La première utilise la technique de spectrométrie de masse QToF (*Quadrupole Time-of-Flight*), tandis que la seconde utilise la technique Orbitrap.

Préparation des figures : Toutes les figures ont été créés par BioRender (BioRender.com) et Microsoft Powerpoint.



## CHAPITRE 3

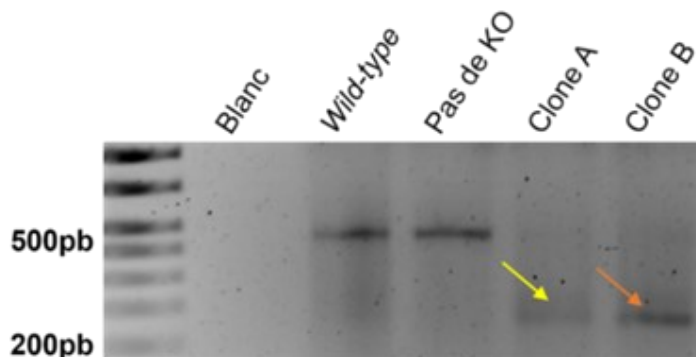
### RÉSULTATS

#### 3.1 La génération des cellules knock-out (KO) SLC35A4 avec la technique CRISPR/Cas9

Afin de valider la spécificité de la détection de la protéine endogène par *Western blot* (WB), et de réaliser des essais fonctionnels visant à évaluer l'importance de la protéine SLC35A4 lors du stress oxydatif, nous avons généré des cellules 143B SLC35A4 KO en utilisant la technique CRISPR/Cas9. La figure 3.1A illustre le succès de l'excision d'une grande partie de la séquence codante de SLC35A4 dans les clones A et B. Le séquençage des deux clones candidats a montré que le clone A est hétérozygote, où l'allèle 1 (clone A1) et l'allèle 2 (clone A2) ont subi une excision de 288 pb et 276 pb, respectivement, tandis que le clone B est homozygote avec une excision de 282 pb (Figure 3.1B). La différence entre les deux allèles de clone A est de 12 pb, ce qui explique la difficulté de voir deux bandes distinctes dans le gel d'agarose. Le séquençage a montré également qu'il n'y avait pas de décalage de cadre de lecture (*frameshift*). Ainsi, le cadre de lecture est maintenu, et la protéine résultante après le KO pourrait être produite normalement. En effet, la protéine originale contient 324 acides aminés (AA), tandis que celle produite à partir du clone B, homozygote, en contient 230 AA avec une délétion de 94 AA, et celle produite du clone A, en contient 232 ou 228 AA avec une délétion de 92 ou 96 AA respectivement.

De plus, mes collègues de laboratoire, Ikram Ajala et Damien Lipuma, ont réussi à générer des cellules 143B KO pour AltSLC35A4. Il s'agit d'une excision de 117 paires de bases (39 AA) sur un total de 309 paires de bases (103 AA) dans un premier allèle, et d'un décalage de cadre de lecture dans le deuxième allèle (*frameshift*).

A



**B**

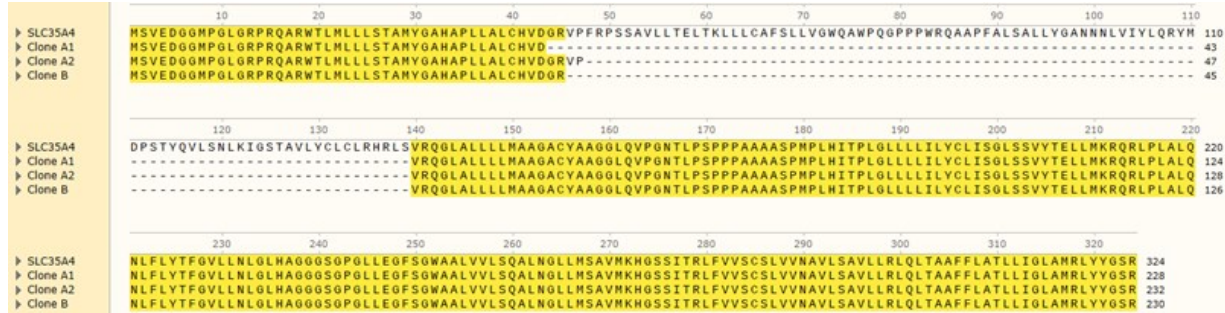


Figure 3.1. Génération des lignées 143B SLC35A4 KO par CRISPR/Cas9

(A) Un gel d'agarose montre la réussite de l'excision d'un fragment de la séquence SLC35A4. La piste WT représente la taille originale de l'amplicon (538 pb). Les flèches jaune et orange indiquent une bande plus basse d'environ de 230 pb et la disparition de la bande originale. La piste pas de KO montre que la bande amplifiée présente une taille semblable à la taille originale (538 pb), donc il n'est pas un candidat optimal pour le séquençage Sanger. (B) Un alignement de séquences traduites in silico à partir des séquences codantes WT et éditées identifiées par Séquençage Sanger. L'alignement montre que le clone A, désigné par la flèche jaune, est hétérozygote (clone A1=allèle 1, clone A2=allèle 2), tandis que le clone B, désigné par la flèche orange, est homozygote. L'alignement et les résultats de séquençage ont été analysés par SnapGene.

### 3.2 Évaluation de l'expression de SLC35A4 durant le stress cellulaire

#### 3.2.1 Résultats négatifs

##### 3.2.1.1 Expression de la protéine SLC35A4 endogène

Nous avons utilisé l'anticorps commercial (Sigma #SAB1304268, #HPA062210) pour évaluer l'expression du SLC35A4 endogène lors du stress oxydatif provoqué par l'arsénite de sodium durant différent temps de traitement, à l'aide de la technique du *Western blot*. En effet, il est important de tester différents temps de traitement, étant donné la non-traduction potentielle de SLC35A4 à l'état basal. Nous cherchons à savoir si l'expression de la protéine est précoce ou tardive, ce qui nous donne un aperçu du type de réponse au stress.

Les résultats du *Western Blot* ont montré des signaux détectés au-dessus de 40 kDa (Figure 3.2A), alors que le poids moléculaire exact de la protéine SLC35A4 est de 34 593 Da. Ces observations ont été faites tant pour les cellules sauvages (*Wild type*), avec différents temps de traitement à l'arsénite de sodium (AS),

que pour des cellules Knock-out SLC35A4 (Figure 3.2B). Cela suggère que les signaux observés sont non spécifiques. De plus, des IF ont été réalisées en utilisant l'anticorps anti-SLC35A4 (#HPA062210, voir l'Annexe A). En effet, les signaux observés sont également non spécifiques, ce qui rend ces anticorps non spécifiques. Le *Western blot* dirigé contre ATF4 a été utilisé comme contrôle positif de l'induction du stress cellulaire. Nous observons que le stress est bien induit à partir du temps 1 h.

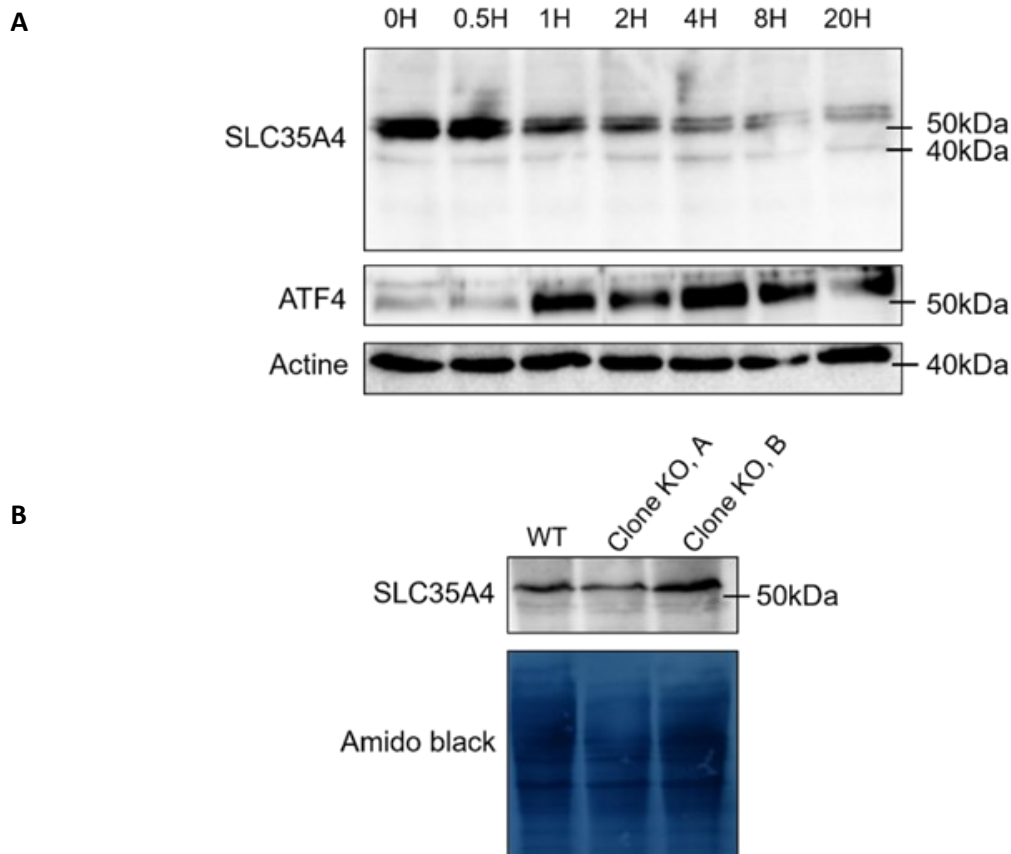


Figure 3.2. Analyse de l'expression de la protéine endogène SLC35A4.

(A) Un *Western Blot* des extraits protéiques des cellules 143B *Wild type* (20µg), obtenus à partir de différents temps de traitement à l'arsénite de sodium à 40µM (0H sans traitement, 0.5H, 1H, 2H, 4H, 8H, 20H), a été réalisé en utilisant un anticorps anti-SLC35A4 (35 kDa, Sigma #SAB1304268). Des anticorps anti-ATF4 et anti-actine ont également été utilisés. (B) Un *Western Blot* des extraits protéiques des cellules 143B Knock-out (KO) SLC35A4 et *Wild type* (WT).

### 3.2.1.2 Détection de la protéine SLC35A4 par Spectromètre de masse (SM)

La non-spécificité de l'anticorps anti-SLC35A4 nous a poussés à rechercher une autre méthode pour détecter et quantifier la protéine SLC35A4, à savoir la spectrométrie de masse (SM). Nous avons soumis différents échantillons, après divers essais de détection, notamment le fractionnement (Golgi), le sodium

carbonate (protéines membranaires), les lysats totaux avec ou sans stress, à la plateforme du CERMO. Malgré nos efforts, aucune détection convaincante de la protéine SLC35A4, qu'elle soit endogène, marquée avec HA ou marquée avec NeonGreen, n'a été obtenue. Cela nous a conduit à effectuer un dernier essai par Immunoprécipitation (IP) afin d'enrichir au maximum la protéine avant la SM et d'optimiser les chances de détection.

Des expériences d'Immunoprécipitation ont été réalisées afin de détecter la protéine SLC35A4 après enrichissement, afin d'établir si SLC35A4 est détectable ou non par SM. Pour ce faire, des IP ont été faites à partir de lysats totaux de cellules 143B de la lignée cellulaire stable SLC35A4-3xHA.

Effectivement, la construction de SLC35A4-3xHA a été réalisée en ajoutant l'étiquette 3xHA au C-terminal de la séquence de SLC35A4 (Figure 3.3A), puis la lignée stable a été générée par transduction lentivirale. Nous avons réussi à détecter la protéine SLC35A4-3xHA au niveau du Golgi, conforme à nos attentes (Figure 3.3A). De plus, environ 60% des cellules présentent une positivité pour HA. Ainsi, cette lignée cellulaire était prête à être utilisée pour l'immunoprécipitation (IP) suivie de la spectrométrie de masse (SM).

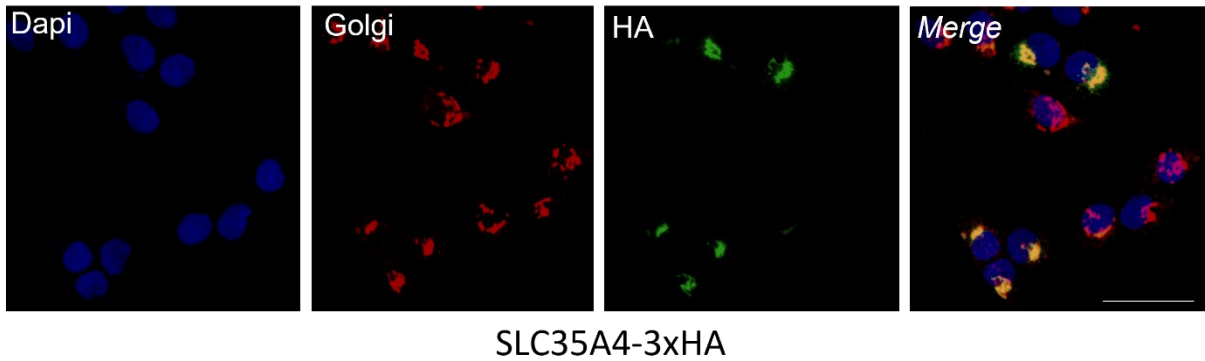
Des WB ont été effectués pour donner suite aux IP. Dans les cellules SLC35A4-3xHA, la protéine SLC35A4-3xHA est détectée aussi bien dans les pistes entrées (*input*) que dans les pistes sorties (*output*) à leurs tailles attendues d'environ 38.5 kDa (avec étiquette), tandis que aucun signal n'a été détecté dans les cellules WT. Cela nous confirme que le SLC35A4-3xHA a bien été immunoprécipité avant l'envoi des échantillons pour l'analyse SM.

Un WB dirigé contre SLC35A4 a également été réalisé à partir des mêmes échantillons afin de tester une fois de plus l'anticorps anti-SLC35A4 endogène (Sigma #SAB1304268, Sigma #HPA062210). Le signal détecté correspond à une bande à environ 50 kDa, sachant que la bande attendue est de 38 kDa (SLC35A4 marquée). Aucun signal n'a été détecté dans les pistes sorties (*output*). Le signal dans la lignée SLC35A4-3xHA est moins important que les cellules WT, ce qui suggère que moins de protéines ont été déposées. Cela pourrait être confirmée avec une coloration de la membrane avec l'amido black (Figure 3.3B).

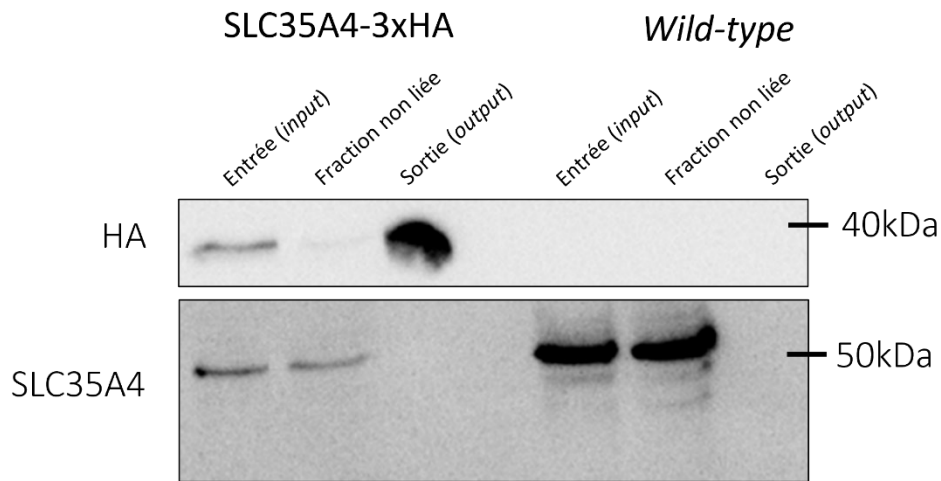
Après avoir confirmé que l'IP a bien fonctionné, les échantillons des IP ont été analysés par spectrométrie de masse sur deux plateformes distinctes. Cependant, les résultats de la spectrométrie de masse étaient négatifs. Sur une plateforme, nous n'avons pas pu détecter de peptides de la protéine SLC35A4, tandis que

sur l'autre plateforme, un seul peptide a été détecté dans les cellules enrichies, mais le résultat n'était pas significatif par rapport à l'IP sur cellules WT (Figure 3.3C).

A



B



C

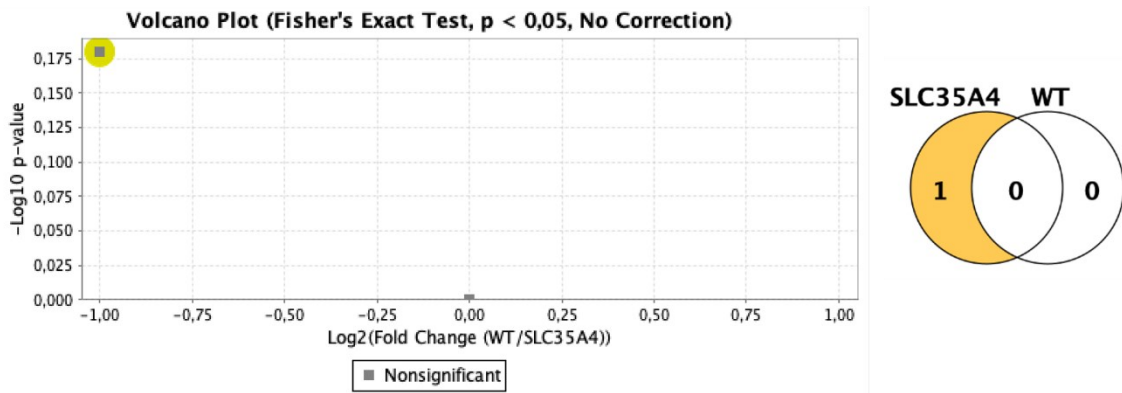


Figure 3.3. Absence de détection de la protéine SLC35A4 par IP-SM

(A) IF des cellules 143B stable exprimant SLC35A4-3xHA. La protéine SLC35A4-3xHA a été marquée avec un anticorps anti-HA (vert). Les Golgi ont été marqués avec un anticorps anti-Golgin 97 (rouge). Les noyaux des cellules ont été colorés au DAPI (bleu). La couleur orange sur la photo merge, indique la colocalisation

du signal HA avec celui de Golgi. Ces images ont été prises par un microscope confocal. Barre d'échelle: 60µm. (B) Validation de l'efficacité des IP par WB avant l'envoi des échantillons pour l'analyse SM. Les IP ont été réalisées sur des lysats totaux de cellules 143B, une lignée cellulaire stable exprimant SLC35A4-3xHA et des cellules *Wild type*, avec des billes magnétiques couplées à des anticorps anti-HA d'espèce rat. Un anticorps anti-HA d'espèce lapin a été utilisé pour la révélation. Ce WB est représentatif de trois expériences indépendantes. Un anticorps anti-SLC35A4 a aussi été utilisé pour vérifier la spécificité de l'anticorps à détecter la protéine SLC35A4 (SLC). Les panneaux (C) correspondent au résultat du IP-SM représentés sous forme de Volcano plot et un diagramme de Venn à l'aide du logiciel Scaffold. Un seul peptide par un seul spectre est détecté et seulement dans les cellules exprimant le SLC35A4-3xHA.

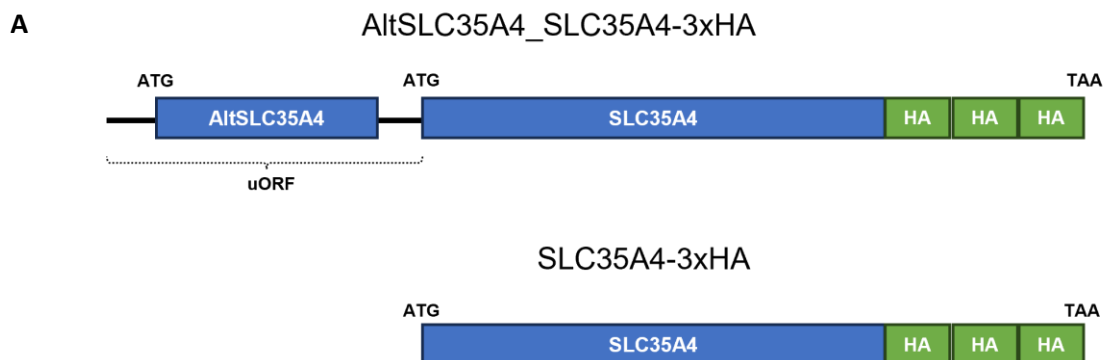
### 3.2.2 Résultats Positifs

#### 3.2.2.1 Évaluation de l'expression de SLC35A4 par WB lors du stress oxydatif

Afin de valider au niveau protéique que l'expression de SLC35A4 augmente lors du stress oxydatif, comme suggéré par l'augmentation de son efficacité traductionnelle (Andreev *et al.*, 2015; Paolini *et al.*, 2018). Nous avons procédé à la construction d'une lignée cellulaire stable, AltSLC35A4\_SLC35A4-3xHA.

Il s'agit de la reconstruction de l'ADNc du gène SLC35A4, couvrant la séquence jusqu'au codon stop de SLC35A4, excluant ainsi la région 3'UTR. L'objectif est d'exprimer une version détectable (3xHA) de SLC35A4 à partir d'un ARNm présentant une structure très similaire à l'ARNm endogène (Figure 3.4A). Cette approche revêt une importance particulière, car nous estimons que la présence d'un uORF (celui codant pour AltSCL35A4) est une condition préalable à l'augmentation de la traduction en réponse au stress (Andreev *et al.*, 2015; 2018).

L'étiquette HA a été ajoutée en C-terminal directement après la séquence codante de SLC35A4, facilitant ainsi la détection de l'expression de la protéine SLC35A4 par *Western blot* (WB) et immunofluorescence (IF) à l'aide d'un anticorps anti-HA. Cette démarche est nécessaire étant donné l'absence d'un anticorps spécifique contre cette protéine.



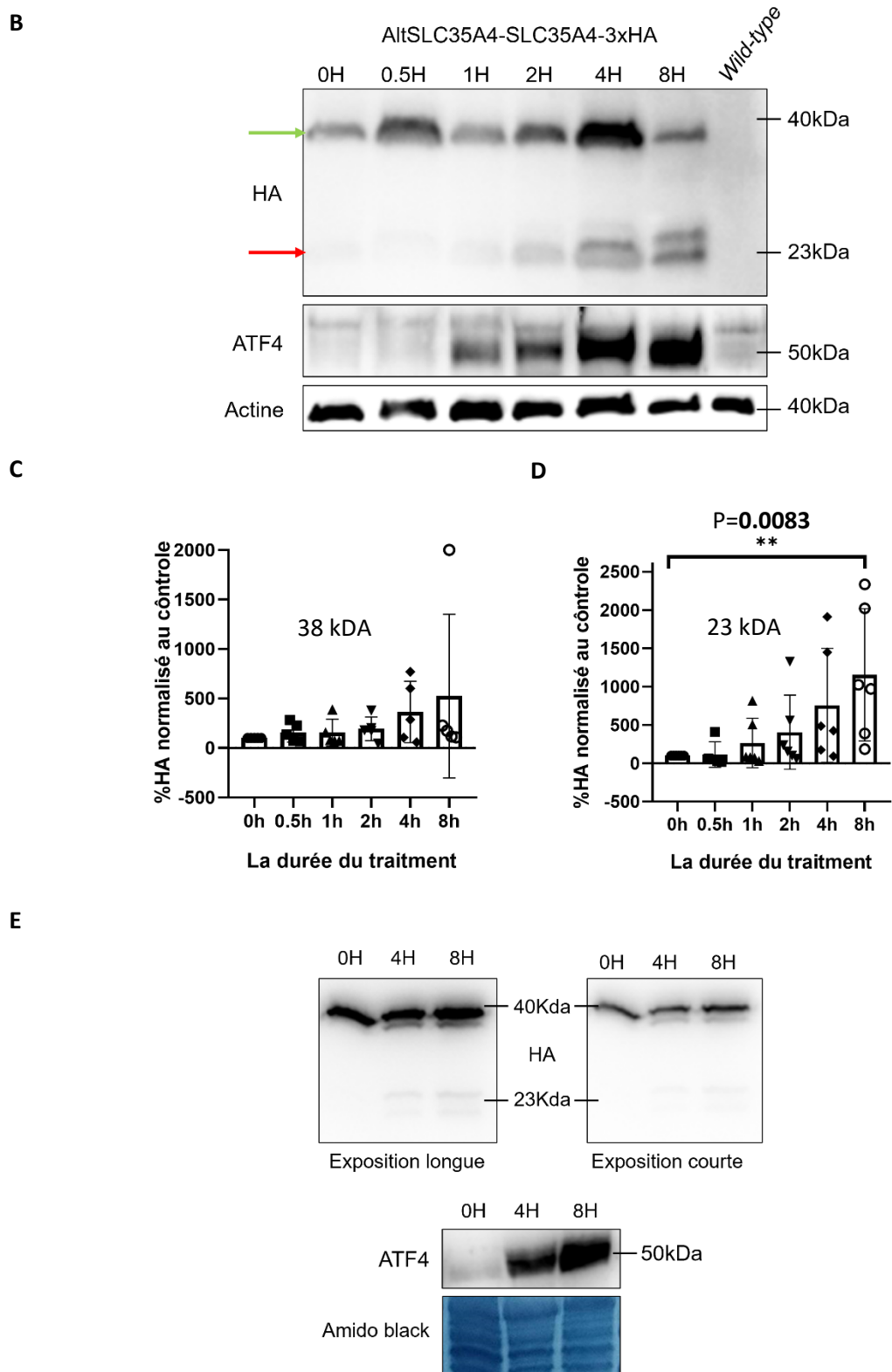


Figure 3.4. Évaluation de l'expression de la protéine SLC35A4 durant le stress cellulaire par *Western blot*

Le panneau A schématise les différentes constructions utilisées pour mesurer la variation d'expression de SLC35A4 lors du stress cellulaire. (B) Des WB ont été réalisés avec des lysats totaux de cellules stables 143B exprimant AltSLC35A4\_SLC35A4-3xHA obtenus à partir de différents temps de traitement à l'arsénite de sodium à 40uM (0H sans traitement, 0.5H, 1H, 2H, 4H, 8H, 20H) et des cellules WT (dernier puit). L'expression de SLC35A4 a été révélée avec l'anticorps anti-HA. La flèche verte représente la taille attendue de la protéine SLC35A4. La flèche rouge représente les nouvelles bandes qui sont apparues autour de 23 kDa suite au stress. Chaque *Western blot* est représentatif de 6 expériences indépendantes. (C), (D) Des quantifications issues des *western blot* (B) ont été normalisées au temps 0H (contrôle), la bande de 38.5 kDa étant représentée par (C) et la bande de 23.5 kDa par (D). La significativité des résultats a été déterminée en utilisant un test ANOVA unidirectionnel en comparaison multiple. \*\*  $P \leq 0.01$ . (E) Des WB ont été réalisés avec des lysats totaux de cellules stable 143B exprimant SLC35A4-3xHA. L'expression de SLC35A4 a été révélée avec l'anticorps anti-HA. Deux temps d'exposition ont été montrés (longue/courte). Chaque *Western blot* est représentatif de 3 expériences indépendantes.

Dans la construction AltSLC35A4\_SLC35A4-3xHA, nous avons traité les cellules par l'arsénite de sodium pendant différentes durées afin d'induire un stress cellulaire, puis nous avons examiné l'expression de SLC35A4 par WB. La protéine SLC35A4 est détectée à sa taille attendue 38 kDa (avec étiquette HA) dans toutes les durées de traitement avec de l'arsénite de sodium (Figure 3.4B\_flèche verte). Cependant l'analyse statistique de la quantification de cette protéine n'a pas montré une différence significative entre les différents temps de traitement (Figure 3.4C). De plus, deux nouvelles bandes sont apparues, d'une taille environ de 23 kDa (Figure 3.4B\_flèche rouge), plus prononcées aux temps 2 h, 4 h, 8 h. La quantification de ces bandes a montré une différence significative entre le temps de contrôle 0 h (sans traitement) et le temps de 8 h. L'absence de signal dans la piste de WT signifie que les signaux HA observés sont spécifiques. Nous suggérons que ces nouvelles bandes représentent des isoformes de la protéines SLC35A4. Quant à la construction SLC35A4-3xHA, l'expression de la nouvelle bande est beaucoup moins importante dans l'exposition longue et presque absente dans l'exposition courte. En effet, cette dernière exposition est réalisée afin d'obtenir un niveau d'exposition similaire à celui des expositions effectuées à partir de la lignée AltSLC35A4\_SLC35A4-3xHA (Figure 3.4E). Le WB contre l'ATF4 a été fait comme contrôle positif de l'induction du stress cellulaire.

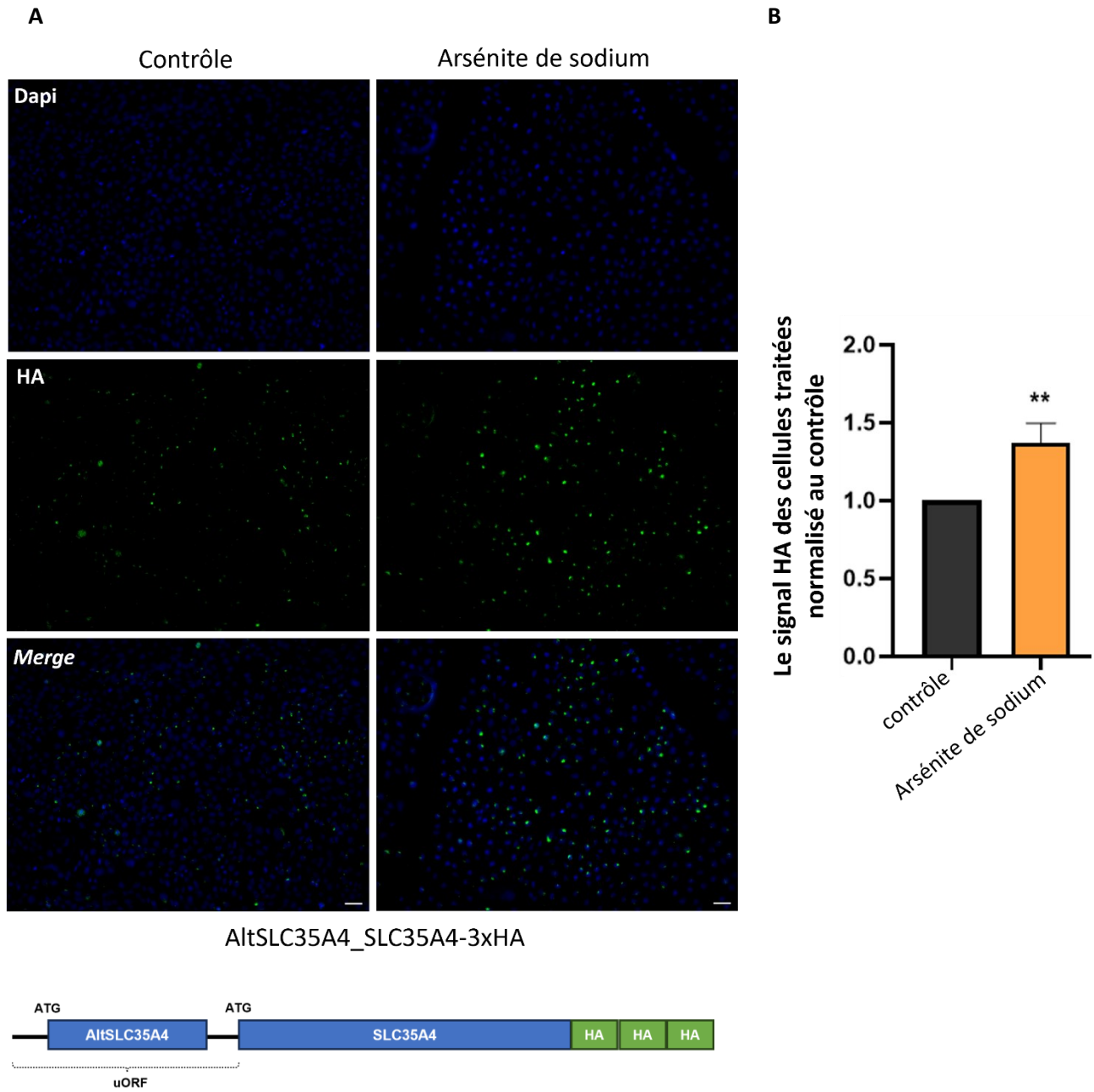
### 3.2.2.2 Évaluation de l'expression de SLC35A4 par IF lors du stress cellulaire

Afin de valider les résultats du WB, nous avons examiné l'expression de SLC35A4 dans les constructions AltSLC35A4\_SLC35A4-3xHA et SLC35A4-3xHA par l'immunofluorescence (IF) en utilisant un anticorps anti-HA, tant avant qu'après une induction de stress cellulaire par l'arsénite de sodium.

Tout d'abord, le signal HA est dans le Golgi car montrant un patron similaire au marqueur de Golgi utilisé. De plus, nous avons observé que le traitement des cellules stables AltSLC35A4\_SLC35A4-3xHA à l'AS a



conduit à une augmentation du signal HA (Figure 3.5A). Cette observation a été confirmée par l'analyse statistique, montrant une différence significative entre les cellules contrôles et les cellules traitées (Figure 3.5B). Quant aux cellules stables SLC35A4-3xHA, dépourvues de la séquence en amont 5'UTR, aucune différence significative entre les cellules contrôles et les cellules traitées a été notée (Figure 3.5C, D).



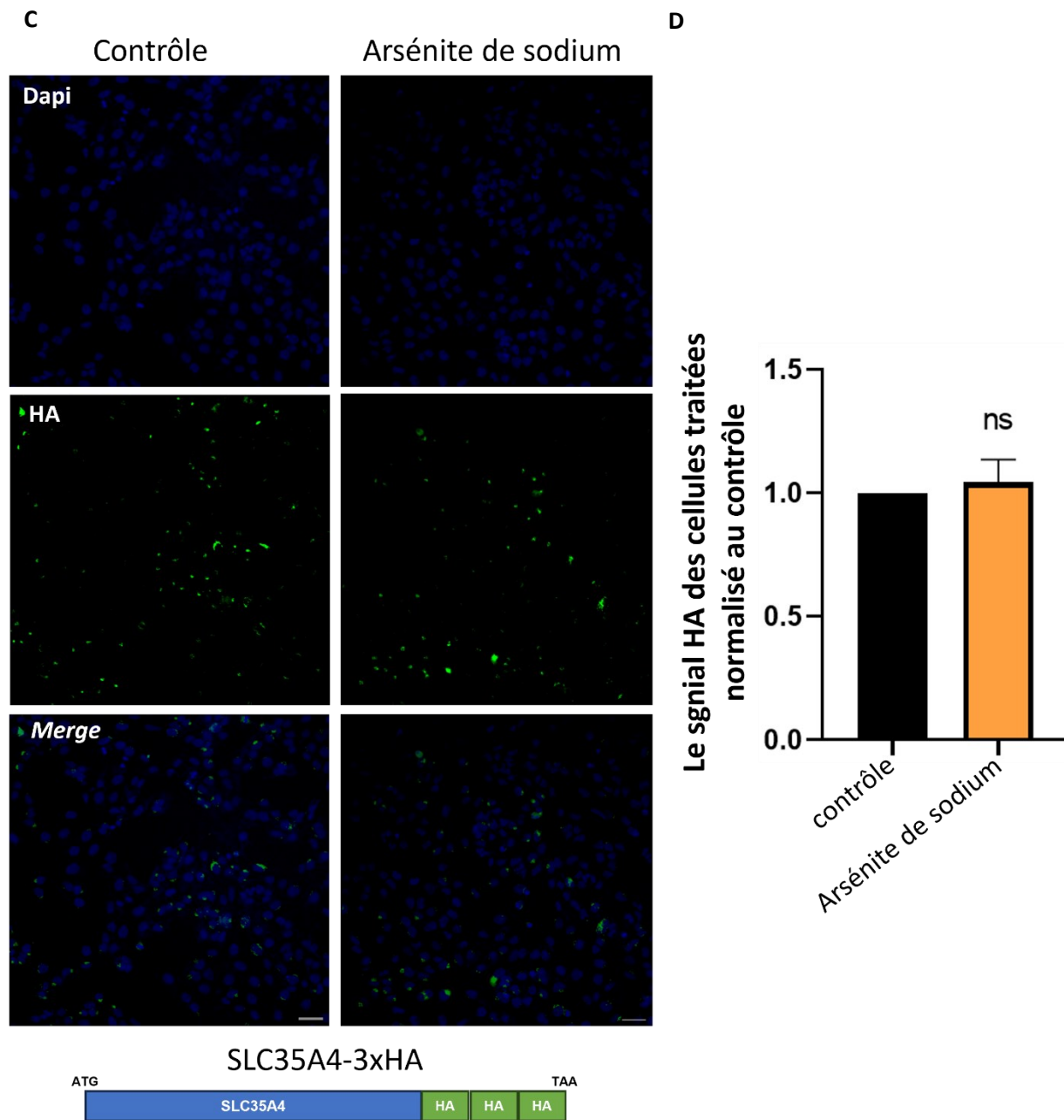


Figure 3.5. Évaluation de l'expression de la protéine SLC35A4 durant le stress cellulaire par Immunofluorescence.

(A) IF des cellules 143B stable exprimant AltSLC35A4\_SLC35A4-3xHA avec le schéma de la construction au-dessous. Les cellules traitées ont été fixées après 3h de traitement avec l'arsénite de sodium à 40µM. La protéine SLC35A4-3xHA a été marquée avec un anticorps anti-HA (vert). Les noyaux des cellules ont été colorés au DAPI (bleu). Ces images ont été prises par un microscope confocal et sont représentatives de trois expériences indépendantes. Barre d'échelle: 50 µm (C) IF des cellules 143B stable exprimant SLC35A4-3xHA avec le schéma de la construction au-dessous. Les cellules traitées ont été fixées après 3h de

traitement avec l'arsénite de sodium à 40mM. La protéine SLC35A4-3xHA a été marquée avec un anticorps anti-HA (vert). Les noyaux des cellules ont été colorés au DAPI (bleu). Ces images ont été prises par un microscope confocal et sont représentatives de deux expériences indépendantes. Barre d'échelle: 60  $\mu$ m (B), (D) Des analyses quantitatives ont été effectuées en se basant sur le signal HA des IF (A) ou (C) des cellules traitées, normalisé par rapport au signal HA des cellules contrôles. La significativité des résultats a été évaluée à l'aide d'un test t non apparié. \*\*  $P \leq 0.01$ .

### 3.3 Évaluation du rôle de SLC35A4 dans la réponse au stress cellulaire

#### 3.3.1 Analyse comparative des cellules knock-out SLC35A4, des cellules *Wild type* et des cellules knock-out AltSLC35A4

##### 3.3.1.1 Analyse de la prolifération cellulaire en condition normales

Afin de vérifier si les protéines SLC35A4 et AltSLC35A4 influent sur la prolifération cellulaire dans des conditions normales, nous avons comparé la prolifération des trois génotypes cellulaires (*Wild type*, SLC35A4 KO - Clone A, Clone B, et AltSLC35A4 KO - AltKO) sur une période de 3 jours, sous des conditions normales. L'analyse statistique a révélé l'absence de différence significative de la prolifération entre les différents phénotypes au cours de la période spécifiée (Figure 3.6). Ces résultats indiquent que les protéines SLC35A4 et AltSLC35A4 ne jouent pas un rôle sur la prolifération dans les conditions expérimentales sur lesquelles elles ont été testées.

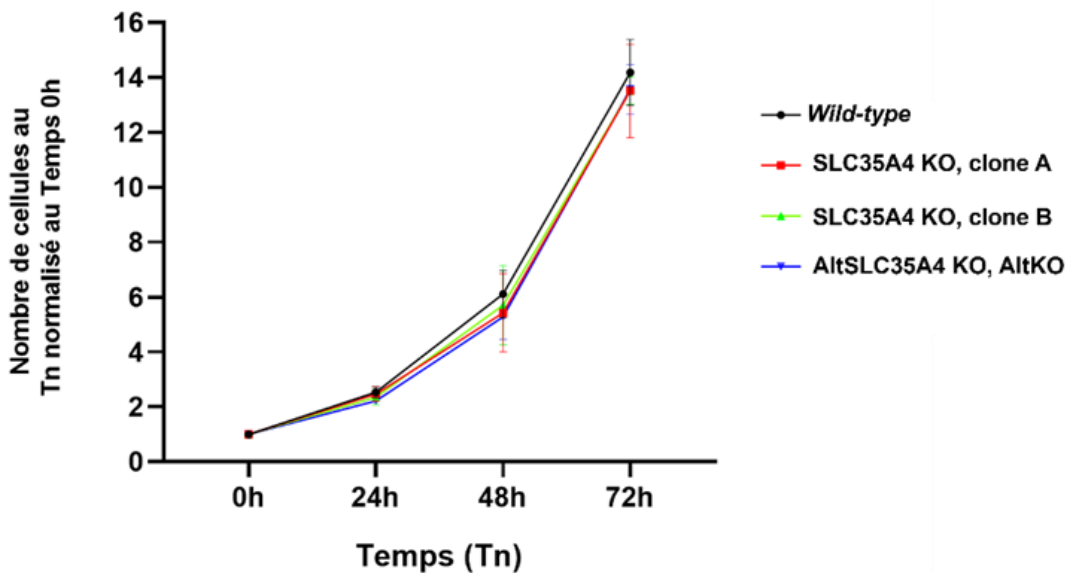


Figure 3.6. Évaluation de l'importance de SLC35A4 par un test de prolifération

Courbe de prolifération des cellules 143B WT et des 143B de phénotype SLC35A4 KO (clone A, B) et AltSLC35A4 KO (AltKO). Les cellules ont été ensemencées à temps 0 avec un nombre fixe de cellule ( $0.05 \cdot 10^6$ ) puis compter chaque jour sur une durée de 3 jours, par un compteur automatique. Les

résultats ont été normalisés au temps 0h pour chaque phénotype. La significativité des résultats a été évaluée à l'aide d'une analyse à effet mixtes. Le nombre de réplicats de cette expérience est 3.

### 3.3.1.2 L'analyse de la viabilité cellulaire lors du stress cellulaire

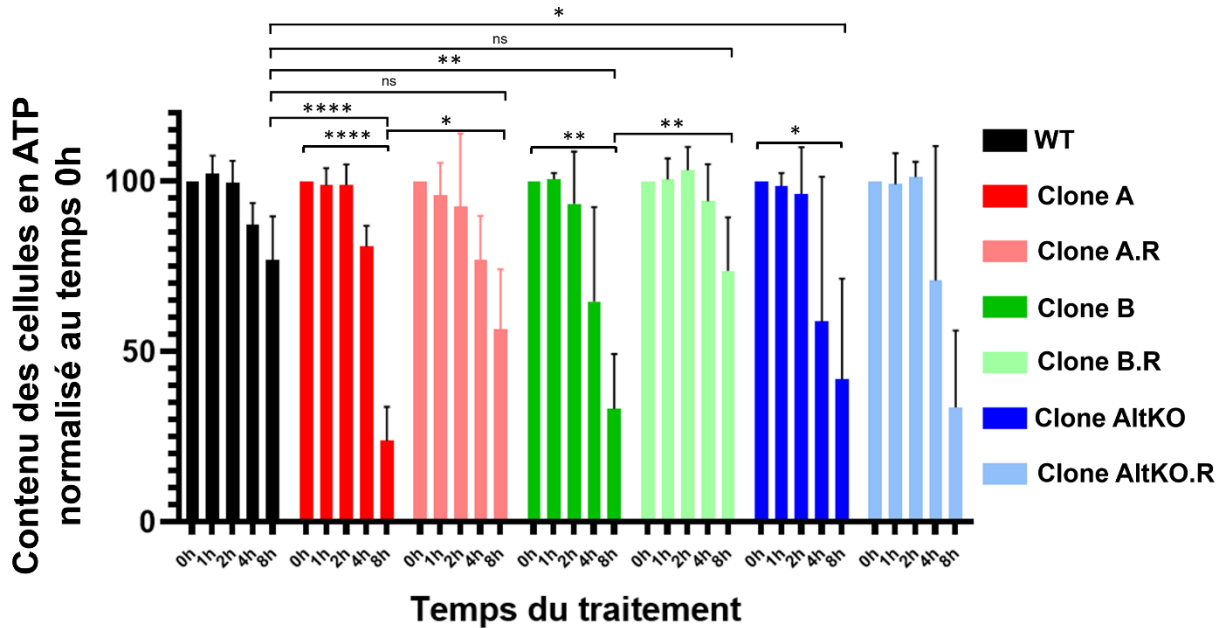
Afin de vérifier si les protéines SLC35A4 et AltSLC35A4 influent sur l'aptitude cellulaire lors du stress cellulaire, nous avons utilisé le kit CellTiter-Glo 2.0 pour comparer la viabilité cellulaire entre les différentes lignées : *Wild type*, SLC35A4 knock-out et leurs contreparties sauvées avec la construction SLC35A4-3xHA (Clone A.R, Clone B.R), AltSLC35A4 knock-out et sa contrepartie sauvée avec la construction 3xHA-AltSLC35A4 (AltKO.R), dans des conditions de stress oxydatif. Cette comparaison a été réalisée en analysant la teneur en ATP, qui reflète la viabilité cellulaire, aux moments spécifiques (0 h sans traitement, 1 h, 2 h, 4 h, 8 h) après un traitement à l'arsénite de sodium à 40  $\mu$ M (Figure 3.7A). Ensuite, la viabilité à 0 h de chaque génotype a été définie comme étant de 100%. Enfin, la viabilité à chaque autre temps a été normalisée par rapport à celle à 0 h, permettant finalement d'obtenir un pourcentage de viabilité par rapport au temps 0 h.

Tout d'abord, nous avons examiné les variations du pourcentage de viabilité cellulaire au sein de chaque génotype. En effet, les cellules *Wild type* ne présentent aucune différence significative entre les différents points temporels. Cela suggère que nous sommes dans des conditions où le stress n'est pas trop intense, et par conséquent, une sensibilité potentielle pourrait être révélée dans les autres génotypes. En revanche, nous observons une différence significative entre les temps 0 et 8 heures au sein des cellules KO, indiquant que les génotypes KO sont sensibles au stress.

Ensuite, nous avons analysé, pour chaque temps de traitement, la différence de viabilité cellulaire entre le groupe *Wild type* et tous les autres groupes. Nous avons observé une différence significative à un seul point temporel entre le groupe *Wild type* et les génotypes SLC35A4 KO (les deux clones A et B) et AltSLC35A4 KO, qui est à 8 heures. De plus, la comparaison entre le groupe WT à 8 heures avec les clones sauvés de SLC35A4 KO indique la disparition des différences significatives. Cela montre que les clones sauvés ont pu regagner la résistance au stress perdue dans le SLC35A4 KO. Toutefois, le clone AltSLC35A4 sauvé n'a pas pu rétablir la résistance au stress.

Par ailleurs, nous notons également que la disparité entre les clones KO/sauvés (A/A.R, B/B.R) à temps 8h, est plus significative dans le cas du clone B. Ceci pourrait s'expliquer par une meilleure expression de la protéine SLC35A4 dans le clone B.R (sauvé, Figure 3.7B).

A



B

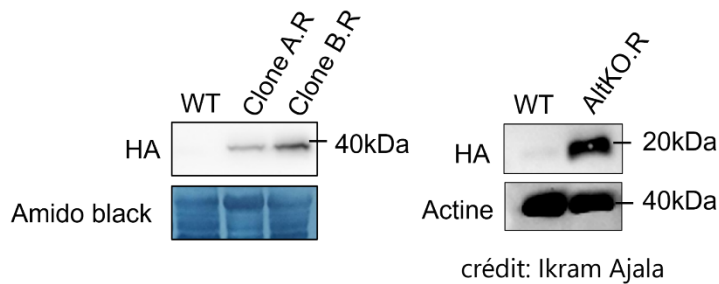


Figure 3.7. Évaluation de l'importance de SLC35A4 par un test de viabilité cellulaire.

(A) La contenance en ATP des cellules (qui reflète la viabilité cellulaire) a été quantifiée à l'aide du kit CellTiter-Glo 2.0. Les cellules 143B, WT, SLC35A4 KO (clones A et B), SLC35A4 *Rescue* (clones A.R et B.R), AltSLC35A4 (AltKO), et AltSLC35A4 *Rescue* (AltKO.R) ont été ensemencées à une densité fixe ( $0,03 \times 10^6$ ) dans une plaque de 96 puits. Ensuite, elles ont été traitées avec de l'arsénite de sodium, à  $40 \mu\text{M}$ , pendant différentes durées (0 h sans traitement, 1 h, 2 h, 4 h, 8 h) avant d'être lues avec une machine de lecture de luminescence. La luminescence à temps 0 h, pour chaque génotype cellulaire, a été définie comme 100%. La significativité des résultats a été déterminée en utilisant un test ANOVA à 2 voies en comparaison multiple. \* $P \leq 0.05$ . \*\*  $P \leq 0.01$ . \*\*\*\*  $\leq 0.0001$ . Le nombre de réplicats de cette expérience est 4 (B) WB a été réalisé avec des lysats totaux de cellules 143B WT, clone A.R et clone B.R. Expression de SLC35A4 a été révélée avec l'anticorps anti-HA.

### Investigation du rôle du gène *SLC35A4* lors du stress oxydatif

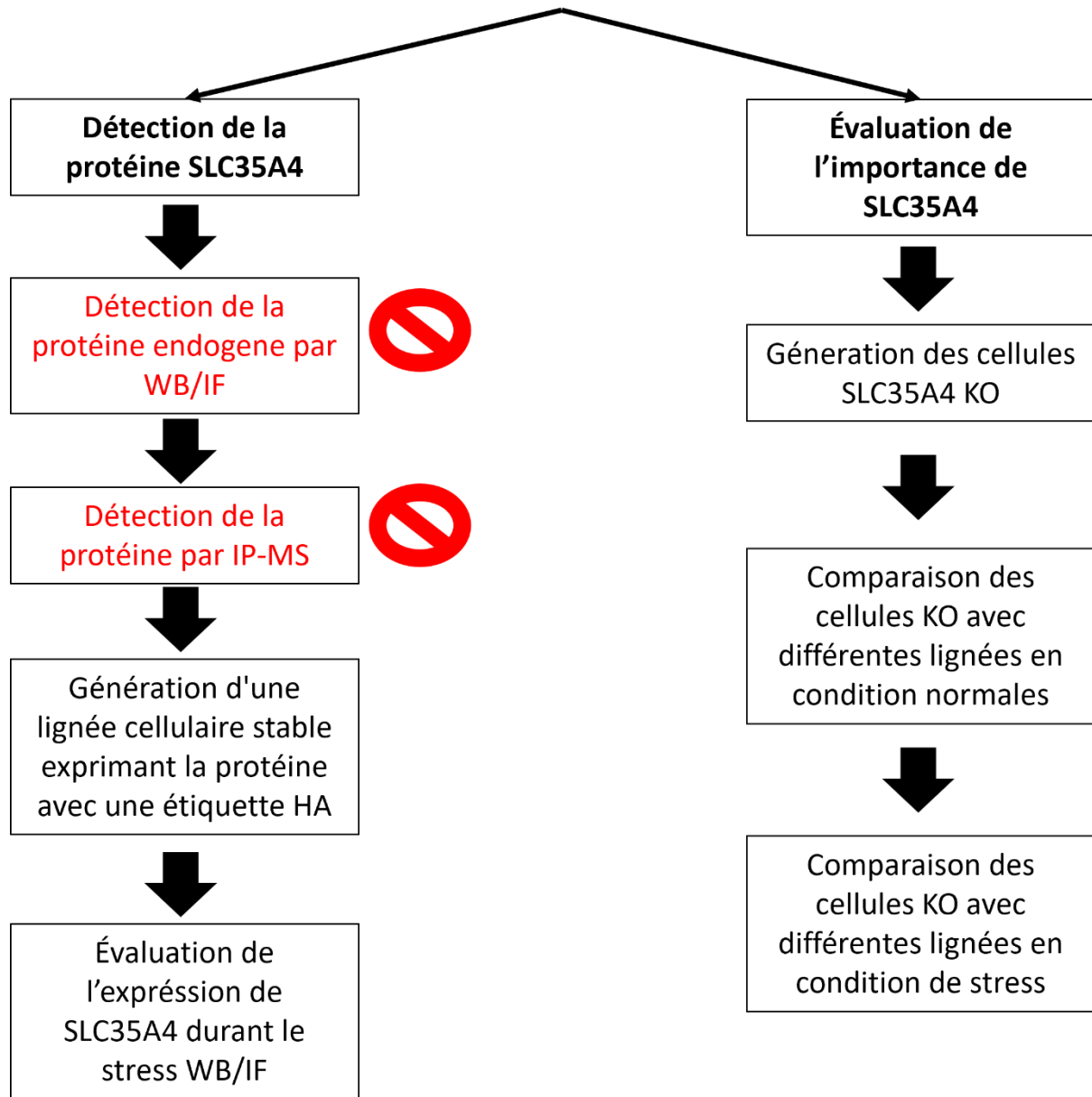


Figure 3.8. Diagramme résumant la partie résultats. Les approches en rouge sont celles qui ont échoué.

## CHAPITRE 4

### DISCUSSION

#### 4.1 Évaluation de l'expression de SLC35A4 durant le stress cellulaire

##### 4.1.1 Détection de la protéine endogène

Des études antérieures, par profilage ribosomique, ont montré que la traduction du CDS de SLC35A4, durant un stress oxydatif et un stress du réticulum endoplasmique, subit la plus forte augmentation de l'efficacité traductionnelle parmi tous les ARNm cellulaires (Andreev *et al.*, 2015 ; Paolini *et al.*, 2018). Nous souhaitons étudier la dynamique de l'expression de la protéine SLC35A4 à l'échelle protéique. La protéine de référence SLC35A4 est composée de 324 acides aminés. Il s'agit d'une protéine à dix domaines transmembranaires, principalement exprimée dans le Golgi (Sosicka *et al.*, 2017 ; Ury *et al.*, 2021).

Nous avons initialement cherché à détecter l'expression de la protéine endogène à l'aide de la technique du *Western blot* et Immunofluorescence. Cependant, nos résultats ont révélé la non-spécificité des anticorps anti-SLC35A4 (3.2.1). Cette conclusion repose d'abord sur la taille observée des bandes, qui était supérieure à 40 kDa, alors que la taille attendue de la bande aurait dû être inférieure, étant donné que le poids moléculaire de la protéine endogène est de 34593 Da. Deuxièmement, l'utilisation de ces anticorps contre une lignée cellulaire stable surexprimant la protéine marquée avec HA n'a pas montré de différence de taille, tandis que l'anticorps anti-HA a donné la taille attendue de 38 kDa (SLC35A4 marquée au HA; Figure 3.3B). Nous avons tenté également d'optimiser les conditions de l'expérience en utilisant différentes méthodes d'extraction de la protéine (Annexe B) mais la conclusion reste la même. Finalement, des signaux non spécifiques ont été observés sur les images d'IF (Annexe A). Il est important de noter que les anticorps utilisés sont polyclonaux et n'ont pas été validés par un Knock-out. De plus, dans la littérature, une seule étude (Sosicka *et al.*, 2017) a utilisé un anticorps anti-SLC35A4. Cependant, cet anticorps n'était plus disponible pendant la réalisation de ce projet et au moment de la rédaction de ce mémoire.

Nous avons également envisagé une autre méthode pour détecter et quantifier la protéine endogène. Cette méthode s'appelle la quantification absolue des protéines par spectrométrie de masse. En résumé, la technique implique l'ajout de quantités connues de peptides standards (Peptides synthétiques produits à partir de la séquence d'un peptide endogène, marqués avec des isotopes stables) à l'échantillon, suivi de la comparaison des intensités de signal des peptides endogènes (soumis à différents traitements) avec

celles des peptides standards par spectrométrie de masse (Schmidt et Urlaub, 2012). Le choix du peptide standard revêt une importance particulière dans cette expérience. À cette fin, il est essentiel d'analyser la protéine endogène par spectrométrie de masse et de sélectionner le peptide le plus approprié (le plus robuste). Ainsi, nous avons généré des cellules stables surexprimant la protéine SLC35A4, dans un 1er temps avec une fusion de l'étiquette fluorescente NeonGreen en C-terminal. L'analyse directe de cet échantillon par SM s'est révélée négative malgré une expression de la construction validée par microscopie à fluorescence. Par la suite, nous avons généré des cellules stables SLC35A4-3xHA compatibles avec l'immunoprécipitation (IP) afin d'isoler et enrichir la protéine avant analyse (Figure 3.3). Cependant, nous n'avons pas pu la détecter. Nous attribuons ce problème à la structure de la protéine, qui est composée de 10 domaines transmembranaires qui ajoutent une complexité à sa structure, rendant ainsi difficile sa digestion et sa détection par des protocoles standard de spectrométrie de masse (Yang *et al.*, 2023). Bien que plusieurs protéases aient été testées et que les sites de coupure prédits de ces protéases indiquent la génération théorique de peptides analysables par MS, le défi persiste en pratique.

#### 4.1.2 L'apparition de deux bandes additionnelles dans WB lors du stress

À la suite des difficultés pour détecter la protéine endogène, nous avons décidé de générer des cellules stables mimant la traduction de la protéine à partir de son ARNm endogène. Pour ce faire, nous avons cloné l'ADNc du gène SLC35A4, couvrant l'intégralité de la séquence jusqu'au codon stop de SLC35A4, excluant ainsi la région 3' UTR. Cela nous permet de conserver l'uORF, englobant la séquence de la protéine alternative (lignée AltSLC35A4\_SLC35A4-3xHA). La raison de ce choix réside dans le fait que, lors du stress cellulaire, la présence d'un uORF a démontré un potentiel régulateur pour la protéine de référence (Andreev *et al.*, 2012, 2018). De plus, la traduction d'AltSLC35A4 est maintenue durant le stress (Andreev *et al.*, 2015 ; Paolini *et al.*, 2018).

Des WB ont été faits sur les cellules AltSLC35A4\_SLC35A4-3xHA traitées durant différents temps par l'arsénite de sodium (AS). Il est important de noter que l'AS (NaAsO<sub>2</sub>) est généralement considéré comme un réactif cytotoxique induisant la production d'espèces réactives de l'oxygène (stress oxydatif), des dommages à l'ADN, voire l'apoptose (Ruiz-Ramos *et al.*, 2009 ; Zhu *et al.*, 2014). Un des mécanismes par lesquels l'arsenic exerce son effet toxique est en perturbant la respiration cellulaire par l'inhibition de diverses enzymes mitochondriales et le découplage de la phosphorylation oxydative. La plupart de la toxicité de l'arsenic résulte de sa capacité à interagir avec les groupes sulfhydryles des protéines et des enzymes, ainsi qu'à substituer le phosphore dans une variété de réactions biochimiques (Zhang *et al.*,



2007). La réponse de la cellule se produit par la phosphorylation de eIF2alpha, qui est effectuée par eIF2AK1 également connu sous le nom de HRI, lui-même activé par l'AS. (McEwen *et al.*, 2005). Étonnamment, nos résultats révèlent l'apparition de deux bandes additionnelles d'une taille d'environ 23 kDa après un certain temps de traitement (Figure 3.4). La quantification montre une augmentation significative avec le temps de traitement. Cependant, la protéine de référence (38 kDa) indique une augmentation, mais non significative avec le même temps de traitement.

De plus, dans les cellules dépourvues de la séquence de la région 5'UTR (uORF), nous constatons l'émergence de ces nouvelles bandes, bien que de manière beaucoup moins prononcée par rapport aux cellules conservant uORF (Figure 3.4E). Ceci suggère que la présence d'un uORF pourrait influencer l'expression de la séquence de référence (RefORF).

#### 4.1.3 Nos hypothèses sur l'origine des bandes additionnelles

Pour expliquer l'apparition de ces nouvelles bandes, nous proposons deux origines possibles. La première est une initiation, plus en aval, de la traduction de l'ARNm par le ribosome, résultant d'une fuite de balayage causée par le stress oxydatif induit. En effet, dans certains cas, les ribosomes peuvent ignorer le premier codon d'initiation et commencer la traduction à un codon d'initiation en aval, permettant ainsi la production de différentes isoformes de la protéine lorsque les deux codons d'initiation sont dans le même cadre de lecture (Rossi *et al.*, 2010). La fuite de balayage est favorisée par la présence d'uORF et en cas de stress cellulaire, pendant lequel eIF2 $\alpha$  est phosphorylé (Baird et Wek, 2012; Andreev *et al.*, 2015, 2018). Nous avons examiné l'existence des codons ATG candidats en aval de la séquence, susceptibles de donner une taille d'environ 23 kDa correspondant potentiellement au codon d'initiation de cette bande (Annexe C). Pour confirmer cette hypothèse, nous proposons de concevoir une nouvelle construction dans laquelle nous introduisons une mutation ponctuelle au niveau de ces codons ATG candidats (comme AGG). Cela permettra d'éviter les mutations produisant des codons d'initiation alternatifs tout en respectant le cadre de lecture. Nous procéderons ensuite à des essais expérimentaux sur les cellules ainsi modifiées. Nous nous attendons à ce que les bandes supplémentaires, correspondant à chaque codon d'initiation muté, disparaissent sur le blot.

La deuxième origine possible consiste en ce que la protéine SLC35A4 soit clivée par des protéases intramembranaires activées en condition du stress cellulaire. En effet, les cellules ont développé plusieurs mécanismes de protection régulant l'homéostasie des protéines (protéostasie), tels que le système

ubiquitine-protéasome (Ciechanover et Kwon, 2015) et les lysosomes (Winckler *et al.*, 2018). Les protéases intramembranaires représentent des régulateurs clés supplémentaires du protéome organellaire (Avci et Lemberg, 2015). La protéolyse intramembranaire faisait référence au clivage de protéines transmembranaires à l'intérieur du noyau hydrophobe de la bicouche lipidique (Brown *et al.*, 2000). Ce clivage est réalisé par des protéases intramembranaires. De manière intéressante, le complexe de Golgi héberge trois protéases intramembranaires, à savoir la protéase de site-2 (S2P, également connue sous le nom de MBTPS2, Kinch *et al.*, 2006), la protéase aspartyl SPP-like 3 (SPPL3; Nyborg *et al.*, 2006), et la protéase sérine rhomboïde rhomboïde-like protein 1 (RHBDL1) de fonction inconnue (Lemberg and Freeman, 2007). Parmi les conséquences de ce clivage, nous pouvons citer l'activation et la dégradation des protéines. D'une part, la protéolyse qui génère un fragment protéique stable et soluble avec une fonction spécifique est une étape clé dans l'activation des protéines, telles que la protéine transmembranaire *cyclic AMP-dependent transcription factor* (ATF6) clivé par S2P durant le stress du réticulum endoplasmique (Ye *et al.*, 2000). D'autre part, la protéolyse intramembranaire peut également générer des fragments qui sont ensuite éliminés par des systèmes de dégradation cellulaire tels que le protéasome (Kühnle *et al.*, 2019). Pour tester cette hypothèse, nous proposons de traiter les cellules avec des inhibiteurs généraux des classes de protéases auxquelles appartiennent les protéases mentionnées ci-dessus (S2P, SPPL3, RHBDL1). Nous nous attendons à ce que les bandes supplémentaires disparaissent. Ensuite, nous pourrions chercher à déterminer précisément la protéase responsable de ce clivage.

En outre, nous postulons que la fuite du balayage est plus probable que le clivage, étant donné que la fuite du balayage est plus fréquente, lors du stress, dans les séquences possédant des uORF (Andreev *et al.*, 2015). Ceci est corroboré par l'observation dans la lignée cellulaire SLC35A4-3xHA, dépourvue d'uORF, où les deux bandes supplémentaires (isoformes de SLC35A4) sont presque absentes. Cela signifie que l'induction des isoformes est diminuée. De plus, le clivage en deux endroits distincts pour donner deux isoformes différentes est moins probable qu'un clivage unique.

#### 4.1.4 L'augmentation de l'expression de la protéine SLC35A4 par IF

En cherchant à mieux comprendre la dynamique d'expression de la protéine SLC35A4 lors du stress cellulaire, nous avons réalisé des IF pour comparer le signal du marquage HA avant et après le traitement, ce qui est représentatif de l'expression de SLC35A4 et cela dans lignées cellulaires AltSLC35A4\_SLC35A4-3xHA et SLC35A4-3xHA. Nos résultats montrent que, seulement dans la lignée cellulaire exprimant SLC35A4 depuis un ARNm possédant les séquences endogènes en amont (incluant l'uORF), l'expression de

la protéine est augmentée (Figure 3.5). Cela renforce notre hypothèse selon laquelle la présence d'un uORF a un impact sur l'expression de la protéine de référence.

De surcroît, ces résultats suscitent une question : pourquoi n'avons-nous pas pu observer une augmentation claire de l'expression de la protéine SLC35A4 pleine longueur par immunobuvardage (WB) ?

Une hypothèse repose sur la structure de la protéine SLC35A4, qui comporte dix domaines transmembranaires. En effet, les protéines transmembranaires sont souvent difficiles à solubiliser et à dénaturer en raison de leur nature hydrophobe et de leurs interactions avec la bicouche lipidique. Plus une protéine comporte de domaines transmembranaires, plus il est probable qu'elle forme des agrégats ou des polymères, ce qui peut interférer avec la résolution et la détection de la protéine par immunobuvardage (Tsuji, 2020). Par conséquent, certaines protéines transmembranaires nécessitent des méthodes spéciales de préparation d'échantillons, comme l'évitement de la dénaturation thermique ou l'utilisation de tampons et de détergents différents, pour améliorer leur solubilité et leur stabilité. Ceci est bien connu dans la famille des transporteurs de solutés SLC (SLC11A2/SLC40A1, Tsuji, 2020).

L'autre hypothèse repose sur la présomption que ce qui a été détecté par immunofluorescence (IF) correspond effectivement aux isoformes de la protéine SLC35A4 identifiées par WB et qu'elles n'ont pas quitté le Golgi. Nous pouvons vérifier ces hypothèses en déterminant l'origine des isoformes. Ainsi, en interrompant la traduction de ces dernières, nous pourrions refaire les IF afin d'observer la dynamique d'expression de la protéine SLC35A4 pleine longueur.

Il convient de souligner que la construction élaborée à partir du CDS seul, sans uORF, est conçue pour permettre une surexpression autonome de la séquence clonée, grâce à un promoteur puissant, indépendamment de la séquence 5'UTR (uORF), comme illustré par la construction SLC35A4-3xHA. Néanmoins, il est essentiel de noter qu'en situation de stress, la co-surexpression de l'uORF et CDS ensemble sur la même séquence (AltSLC35A4\_SLC35A4-3xHA) favorise l'expression d'une version étiquetée HA (protéine SLC35A4 ou/et isoformes, résultat de l'IF) ainsi que potentiellement la fuite de balayage (isoformes, résultat du WB) par rapport à la surexpression exclusive du CDS seul. Cela démontre davantage l'importance de l'uORF sur son CDS, même en condition de clonage.

## 4.2 Tester l'importance de SLC35A4 dans le stress cellulaire

### 4.2.1 SLC35A4 est invalidée fonctionnellement dans les cellules knock-out

Afin d'étudier l'importance de la protéine SLC35A4 pendant le stress, nous avons généré des cellules knock-out. Les résultats de séquençage des clones knock-out ont révélé que nous avons réussi à éliminer 282 paires de bases (correspondant à 94 acides aminés) dans le clone B homozygote et 288/276 paires de bases (96/92 AA) pour les deux allèles du clone A hétérozygote. Cependant, la nouvelle séquence résultante du knockout a conservé son cadre de lecture. Par conséquent, une protéine pourrait toujours être produite à partir de cette séquence de taille de 230 AA pour le clone B et taille de 228/232 AA pour le clone A. Nous nous sommes interrogés sur la fonctionnalité potentielle de la protéine ainsi produite. Pour répondre à cette question, nous avons analysé la littérature concernant la structure de SLC35A4 et examiné l'importance de la séquence supprimée (voir annexe D). Cette dernière comporte trois domaines transmembranaires (2, 3, 4) et le seul domaine alpha-hélice. Ce dernier est crucial dans la structure et la fonction des protéines transmembranaires (De Marothy et Elofsson, 2015). Nous avons également noté la suppression de 8/15 (53.3%) des acides aminés prédits comme importants dans la formation du site de liaison du substrat CDP-Ribitol (annexe D). Nous notons également la suppression du seul site de phosphorylation prédit à la position Y109, lors de la modification post-traductionnelle (PhosPhositePlus). Ces observations renforcent l'idée que la protéine produite est largement affectée, ce qui la rendrait dysfonctionnelle. Par conséquent, les clones générés sont forts probablement des knock-out fonctionnels.

Pour confirmer cette conclusion, nous reproduirons l'expérience réalisée par Ury *et al.*, 2021, étant donné que la protéine SLC35A4 partage la fonction de transport du CDP-Ribitol dans le Golgi avec SLC35A1. Ainsi, l'évaluation de la disponibilité du CDP-Ribitol dans le Golgi, tout en inhibant le SLC35A1 dans les cellules SLC35A4 KO, fournira des indications sur la fonction de la protéine produite.

### 4.2.2 L'absence d'impact de SLC35A4 sur la prolifération cellulaire sous condition normale

Après avoir réussi à générer des cellules SLC35A4 KO, nous avons cherché à déterminer si la protéine SLC35A4 joue un rôle dans l'aptitude cellulaire à l'état basal (Figure 3.6). Le résultat montrant l'absence d'impact sur la prolifération cellulaire n'était pas surprenant. Comme décrit dans la section (1.17.1), SLC35A4 transporte une partie du CDP-ribitol, tandis que le reste est principalement transporté par SLC35A1 (Ury *et al.*, 2021). Ainsi, le Golgi reste accessible au CDP-ribitol. Une autre raison est que l'expression de SLC35A4 à l'état basal est probablement faible si l'on se fie à la très faible efficacité

traductionnelle (Andreev *et al.*, 2015). Par conséquent, l'impact de l'absence de cette protéine à l'état basal est prédit pour être minime.

Nous remarquons également l'absence d'impact de l'absence de la protéine AltSLC35A4 sur la prolifération cellulaire. En effet, la protéine AltSLC35A4 est exprimée dans la mitochondrie, mais son rôle exact dans la cellule reste à déterminer. Nous suggérons que AltSLC35A4 n'a pas d'impact sur la prolifération cellulaire dans les conditions dans lesquelles elle a été testée.

#### 4.2.3 L'impact de SLC35A4 sur l'aptitude cellulaire durant le stress

La réponse intégrée au stress (RIS) est une réponse cellulaire activée en réaction aux stress internes ou environnementaux. Il s'agit d'une réponse conservée dans les cellules eucaryotes qui régule à la baisse la synthèse des protéines et régule à la hausse l'expression de gènes spécifiques nécessaire dans la réponse au stress (Pakos-Zebrucka *et al.*, 2016). L'ARNm de SLC35A4 a montré, par profilage ribosomique, une augmentation de la traduction durant le stress cellulaire (Andreev *et al.*, 2015; Paolini *et al.*, 2018). Dans cette expérience, nous cherchons à déterminer si les cellules knock-out (KO) conservent leur viabilité pendant le stress cellulaire induit par l'arsénite de sodium, qui provoque un stress oxydatif. La viabilité cellulaire a été évaluée en mesurant l'ATP.

L'inactivation de *SLC35A4* a eu un impact sur la viabilité cellulaire. L'induction du stress sur une durée de 8 heures a révélé une diminution significative de l'ATP, reflétant la viabilité cellulaire, dans les cellules KO (*SLC35A4* et *AltSLC35A4*) par rapport aux cellules *Wild type*. De plus, les cellules sauvées *SLC35A4* ont pu récupérer leur viabilité cellulaire, contrairement aux cellules sauvées *AltSLC35A4* qui n'ont pas montré de signe de récupération (Figure 3.7). Par ailleurs, afin d'étudier plus en détail les changements dans la survie cellulaire, nous proposons de réaliser un test d'apoptose. En utilisant des marqueurs spécifiques tels que l'annexine V, sur les cellules KO et WT après un traitement à l'AS, nous pourrions analyser les cellules marquées par cytométrie en flux. Cela nous permettra de quantifier les cellules en apoptose par rapport aux cellules totales.

Ces données délimitent le rôle central joué par la protéine de référence *SLC35A4* dans le maintien de la viabilité cellulaire en conditions de stress cellulaire. Cette observation est conforme au récit élaboré dans les résultats du profilage ribosomique présentés par Andreev *et al.*, 2015 et Paolini *et al.*, 2018.

En effet, le transport du CDP-ribitol est principalement effectué par le SLC35A1 en condition basale (Ury *et al.*, 2021). Parallèlement, dans le contexte de la réponse intégrée au stress, la traduction générale des ARNm est inhibée, englobant l'activité traductionnelle de SLC35A1. Par conséquent, nous émettons l'hypothèse que SLC35A4, en échappant à la régulation négative de la traduction dans son ARNm, assure la fonction de transport du CDP-ribitol en conditions de stress cellulaire. En outre, il a été démontré que le ribitol améliore la biosynthèse des nucléotides et augmente les niveaux de glutathion réduit (Tucker *et al.*, 2022), impliqué dans la synthèse et la réparation de l'ADN, ainsi que dans l'amélioration de l'activité antioxydante de la vitamine C, le transport des acides aminés et la détoxification des composés nocifs (Montecinos *et al.*, 2007). De plus, la supplémentation en CDP-ribitol a été associée à une réduction de l'inflammation au niveau musculaire (Wu *et al.*, 2022), suggérant ainsi que le CDP-ribitol possède un potentiel antioxydant et anti-inflammatoire, conférant ainsi une résistance cellulaire au stress. Par ailleurs, le CDP-ribitol sert de substrat critique dans la cellule pour la glycosylation de protéines spécifiques (Kanagawa *et al.*, 2018; Marcela *et al.*, 2018; Wu *et al.*, 2022). De surcroît, des études ont démontré que la glycosylation affecte le stress oxydatif (Medina-Navarro *et al.*, 2010; Gerlach *et al.*, 2012; Xiao *et al.*, 2021), et ce processus se déroule dans le Golgi (Hadley *et al.*, 2014). Une hypothèse plausible est que le stress oxydatif puisse interférer avec la glycosylation en cas de déficience en CDP-ribitol, mais des études supplémentaires seraient nécessaires pour établir un lien direct et prouver cette relation. Pour valider que le transport du CDP-ribitol vers le Golgi par SLC35A4 est important pour résister au stress, nous proposons d'évaluer la dynamique du CDP-Ribitol dans le Golgi, lors du stress, dans les cellules KO, WT et sauvées. Ces dernières seront générées avec une version mutée de SLC35A4, en modifiant les nucléotides codant pour les acides aminés importants pour la liaison au substrat CDP-Ribitol (voir Annexe D), de manière à ce que SLC35A4 soit déficiente en capacité de transport mais serait globalement similaire au WT.

De surcroît, nous suggérons une autre hypothèse plausible, selon laquelle les isoformes détectées dans le *Western blot* (fuite de balayage/protéine clivée) pourraient exercer une influence substantielle sur la survie cellulaire, indépendamment de la fonction principale attribuée à la protéine complète, à savoir le transport du CDP-ribitol. Pour vérifier cette dernière hypothèse, en nous basant sur les données obtenues sur l'origine des isoformes, nous interrompons la traduction complète de ces dernières. Ensuite, les tests de viabilité nous fourniront des indications sur leur importance. Cependant, il est à noter que cette hypothèse semble moins probable, étant donné que les cellules sauvées ont été générées avec SLC35A4-3xHA, une construction qui a montré, par *Western blot*, une induction faible des isoformes lors du stress.

En outre, concernant les cellules d'AltSLC35A4 sauvées, des analyses de génétique des populations effectuées sur plusieurs espèces ont permis de vérifier que les mutations abolissant les uATG (codons de départ des uORF) sont fortement délétères, perturbant la traduction du CDS (Xiao *et al.*, 2022). Cela souligne la signification fonctionnelle marquée attribuée aux uORF. Par conséquent, la récupération compromise de la viabilité dans les cellules sauvées AltSLC35A4 par rapport aux cellules KO peut être rationalisée. De surcroît, la modification de la taille de l'uORF (AltSLC35A4 KO) est aussi un facteur qui peut affecter les mécanismes de fuite de balayage, entravant ainsi l'efficacité optimale de l'induction traductionnelle du CDS (SLC35A4 et ses isoformes) lors du stress (Andreev *et al.*, 2015). Donc, la surexpression isolée d'une portion de l'uORF (AltSLC35A4 sauvées) n'était pas efficace et suffisante pour induire la traduction du CDS endogène, étant donné qu'ils ne sont pas sur la même séquence.

Une interrogation pourrait être soulevée à cet égard : pourquoi n'y avait-il pas d'impact sur la prolifération cellulaire dans les cellules AltSLC35A4 KO ? La réponse à cette question réside dans le fait que l'expression de SLC35A4 à l'état basal est basse (Andreev *et al.*, 2015), ce qui pourrait ne pas exercer un effet significatif sur la cellule.

En se basant sur ces données, la prochaine étape proposée pour examiner le rôle de l'uORF dans l'induction de la traduction du CDS, implique la génération d'une nouvelle lignée stable, dont la séquence exprimée est similaire à celle d'AltSLC35A4 KO, avec une modification de la taille de l'uORF (AltSLC35A4KO\_SLC35A4-3xHA). Ensuite, en procédant à des analyses WB pour évaluer les isoformes, une perte d'induction des isoformes serait une forte indication que la taille de l'uORF joue un rôle crucial, mettant ainsi en évidence l'implication de la fuite de balayage dans ce processus.

En parallèle, afin de valider de manière concluante le rôle de SLC35A4 dans la résistance cellulaire au stress, ainsi que l'importance de la présence de uORF, nous préconisons d'exprimer la séquence SLC35A4-3xHA dans les cellules AltSLC35A4 KO, suivie de la reproduction des tests de viabilité cellulaire. Cette approche nous permettra de vérifier si le rétablissement d'une expression appropriée de SLC35A4 compense l'absence d'induction traductionnelle de l'endogène. Nous anticipons que cette nouvelle lignée cellulaire sera en mesure de rétablir la résistance au stress compromise dans les cellules AltSLC35A4 KO.

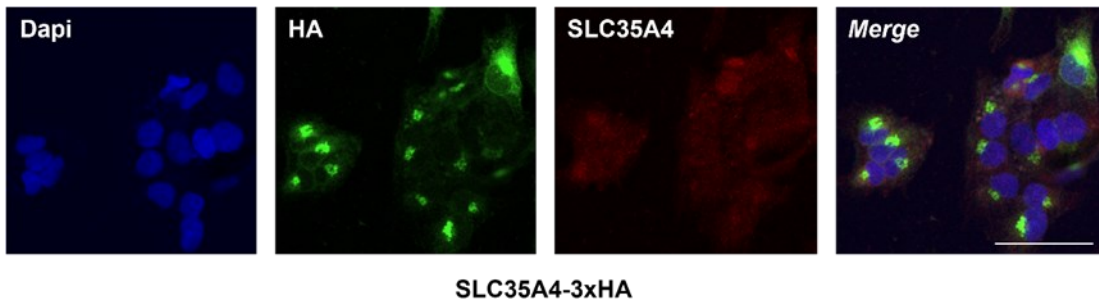
## CONCLUSION

L'objectif principal de ce mémoire était de valider la dynamique de l'expression de la protéine SLC35A4 lors du stress oxydatif induit par l'arsénite de sodium et d'évaluer l'importance des protéines codées par *SLC35A4* lors du même type de stress. Nos travaux ont tout d'abord permis de constater l'émergence des isoformes de la protéine SLC35A4 lors du stress, avec une augmentation proportionnelle de leur expression en fonction de la durée du stress. De plus, notre étude a mis en évidence une augmentation de l'expression pendant le stress d'une version étiquetée HA exprimée depuis une construction mimant le contexte endogène de l'ARNm. Nos résultats ont montré que cette augmentation est dépendante de la présence d'un uORF. Enfin, nos travaux ont permis de souligner l'importance cruciale de SLC35A4 dans la résistance cellulaire au stress oxydatif, ainsi que le rôle potentiel de son uORF dans cette résistance. En perspective à court terme, il est impératif de compléter les résultats obtenus. En effet, afin de valider le mécanisme par lequel les isoformes sont traduites, deux approches possibles doivent être explorées : la fuite de balayage et le clivage de la protéine de pleine longueur par les protéases. Le premier mécanisme pourra être testé par des mutations ponctuelles des ATG susceptibles d'avoir initié les isoformes. Quant au deuxième mécanisme, des traitements cellulaires avec des inhibiteurs de protéases candidats pourraient être réalisés. De plus, il est crucial de bien confirmer l'implication de l'uORF dans la traduction du CDS de SLC35A4 lors du stress. Pour ce faire, exprimer une séquence mimant celle d'AltSLC35A4 KO (AltKO\_SLC-3xHA), permettra d'évaluer, par WB, l'induction des isoformes dans une lignée où la taille de l'uORF est modifiée. En parallèle, l'expression de SLC35A4-3xHA dans les cellules AltSLC35A4 KO permettra de vérifier, par un test de survie, si cette nouvelle lignée est capable de rétablir la résistance au stress. Il est également important de valider ces résultats dans d'autres types de stress, tels que le stress du réticulum endoplasmique. À long terme, il est essentiel de comprendre le mécanisme par lequel SLC35A4 confère une résistance cellulaire au stress. Pour ce faire, nous commençons nos investigations en étudiant la dynamique du CDP-ribitol dans le Golgi lors du stress, à la fois dans les cellules WT et KO. Cette approche nous permettra de vérifier si SLC35A4 assume le rôle de SLC35A1 et facilite exclusivement le transport de CDP-ribitol dans des conditions de stress. En outre, l'utilisation de la technique TurboID pourrait être envisagée pour explorer l'interactome de SLC35A4 et de ses isoformes en condition de stress. Cela nous fournira des informations précieuses sur les protéines interagissant avec SLC35A4 et ses isoformes, offrant ainsi des indications sur leur rôle potentiel. Enfin, étudier ce gène permettra de mieux comprendre les mécanismes en jeu dans la résistance au stress, et lien avec des pathologies potentielles impliquant ces stress.



## ANNEXE A

### Analyse de l'expression de la protéine endogène SLC35A4 avec l'anticorps SLC35A4



Analyse de l'expression de la protéine endogène SLC35A4 avec l'anticorps SLC35A4

IF des cellules 143B stable exprimant SLC35A4-3xHA. La protéine SLC35A4-3xHA a été marquée avec un anticorps anti-HA (vert). La protéine SLC35A4 a été marquée avec un anticorps anti-SLC35A4 (Sigma, #HPA062210) avec une concentration finale de 1:50. Les noyaux des cellules ont été colorés au DAPI (bleu). Ces images ont été prises par un microscope confocal. Barre d'échelle: 60  $\mu$ m.

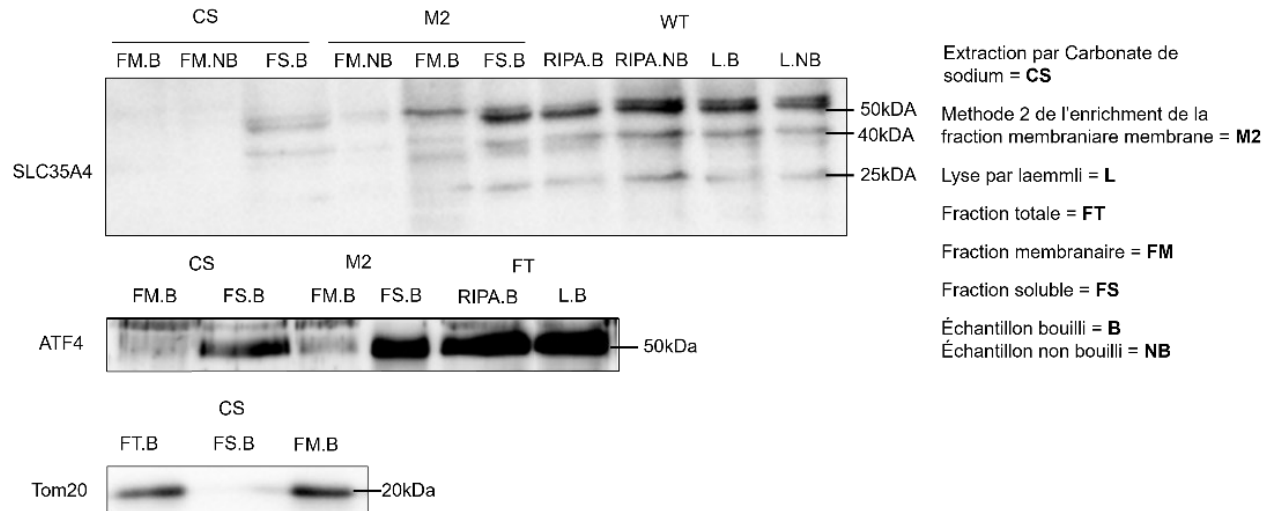
## ANNEXE B

### Test de différentes méthodes d'extraction

Afin d'obtenir une compréhension approfondie de la spécificité de l'anticorps, nous avons cherché à optimiser la technique du *Western blot*. Étant donné que la protéine SLC35A4 possède 10 domaines transmembranaires, nous avons envisagé tout d'abord d'enrichir la fraction membranaire par différentes méthodes d'extraction. De plus, nous avons cherché à améliorer la solubilité de la protéine en évitant l'ébullition.

L'extraction par carbonate de sodium : La solution de carbonate alcalin (0.1M Na<sub>2</sub>CO<sub>3</sub>, pH 11.5) perturbe les interactions protéine-protéine, en libérant les protéines périphériques faiblement associées aux membranes, sans altérer la membrane. Ensuite, l'ultracentrifugation (55 000 RPM, 30 minutes à 4°C) est utilisée pour générer une fraction de culot, contenant les protéines membranaires intégrales, et une fraction de surnageant contenant des protéines solubles et associées à la membrane. Ces fractions peuvent ensuite être analysées (après resuspension dans le Laemmli) par immunobuvardage. L'autre méthode d'enrichissement de la fraction membranaire (M2) consiste à appliquer une ultracentrifugation directe sur les cellules suspendues dans le PBS frais (45 000 RPM, 30 minutes à 4°C) afin d'isoler les membranes (avec leurs protéines intégrales) présentes dans la fraction du culot.

Les résultats montrent tout d'abord plusieurs signaux (25, 40, 50 kDa), et les tailles des signaux obtenus sont différentes de celle attendu pour SLC35A4 (35 kDa). Ainsi, l'absence de signal dans la fraction membranaire de carbonate de sodium.



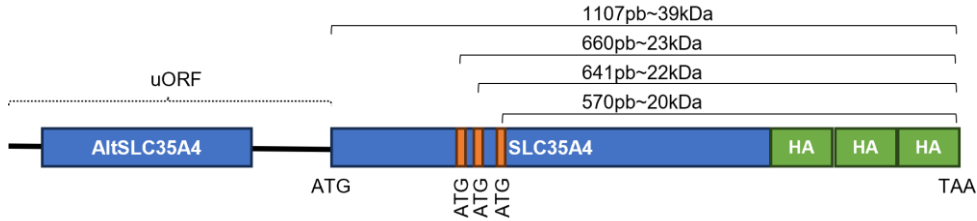
#### Test de différentes méthodes d'extraction.

WB des extraits protéiques des cellules 143B *Wild type*, obtenus après un traitement de 4h avec l'arsénite de sodium. L'extraction des protéines a été faite en utilisant différentes méthodes d'extraction (extraction par carbonate de sodium, enrichissement de la fraction membraniare, RIPA, Laemmli) et avec ou sans ébullition de l'échantillon après extraction. Des anticorps, anti-SLC35A4 (Sigma #SAB1304268), anti-ATF4 (marqueur de stress et protéine soluble) et anti-Tom20 (protéine mitochondriale membraniare) ont été utilisés.

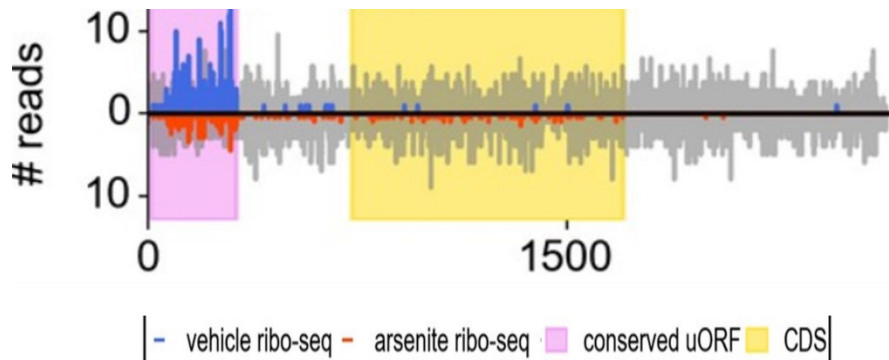
## ANNEXE C

### Origine potentielle des isoformes

A



B



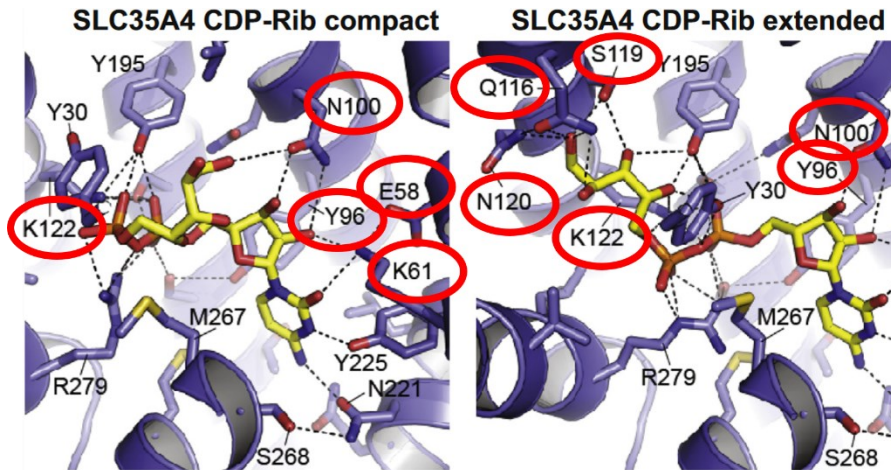
Origine potentielle de nouvelles isoformes.

(A) Le schéma représente les positions des différents codons ATG qui pourraient être à l'origine de l'initiation de la traduction retardée de la deuxième protéine. Le calcul de ces positions est estimé à l'aide d'une calculatrice en ligne qui prédit le poids moléculaire d'une séquence en acides aminés (BCCM/GeneCorner). Cependant, cette estimation reste relative. (B) La densité et l'occupation des ribosomes sur la séquence de l'ARNm du gène SLC35A4 a été obtenue par profilage ribosomique avant et après un traitement de 30min avec l'arsénite de sodium. En gris, l'abondance de l'ARNm (en nombre de « reads » de séquençage d'ARN). L'abondance des empreintes ribosomiques est montrée en bleu (véhicule) et rouge (arsénite de sodium). La région CDS annotée est mise en évidence en jaune. L'ORF conservé dans le 5'UTR (correspondant à AltSLC35A4) est mis en évidence en rose. Ce schéma est tiré de Andreev et al., 2015.

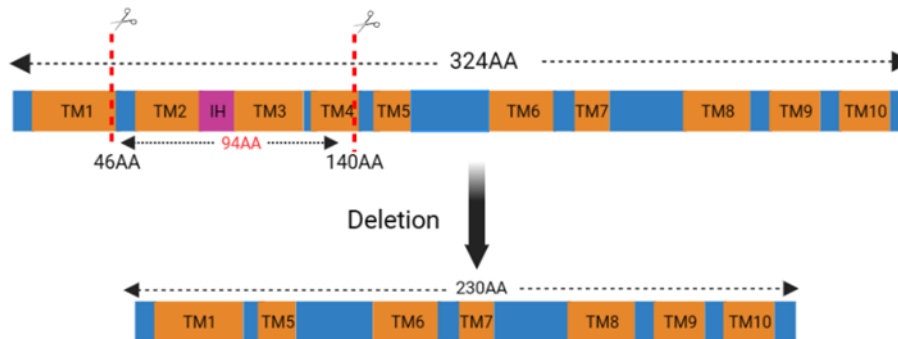
## ANNEXE D

### Les cellules SLC35A4 KO produisent une protéine non-fonctionnelle

A



B



Les cellules SLC35A4 KO produisent une protéine non-fonctionnelle.

(A) Sites de liaison au substrat (CDP-Ribitol) dans SLC35A4. Les cercles en rouge représentent les acides aminés supprimés dans les KO clone. Ce schéma est pris de Ury et al., 2021. (B) Schéma illustratif de la protéine produite à partir des cellules SLC35A4 KO, clone B homozygote. TM : domaines transmembranaires, l'hélice alpha interfaciale (IH) connectant les TM 2 et 3.

## RÉFÉRENCES

- Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., Costello, C. E., Cravatt, B. F., Fenselau, C., Garcia, B. A., Ge, Y., Gunawardena, J., Hendrickson, R. C., Hergenrother, P. J., Huber, C. G., Ivanov, A. R., Jensen, O. N., Jewett, M. C., Kelleher, N. L., Kiessling, L. L., ... Zhang, B. (2018). How many human proteoforms are there?. *Nature Chemical Biology*, 14(3), 206–214.
- Aitken, C. E., & Lorsch, J. R. (2012). A mechanistic overview of translation initiation in eukaryotes. *Nature Structural & Molecular Biology*, 19(6), 568–576.
- Albuquerque, J. P., Tobias-Santos, V., Rodrigues, A. C., Mury, F. B., & Fonseca, R. N. D. (2015). Small ORFs: a new class of essential genes for development. *Genetics and Molecular Biology*, 38, 278–283.
- Amos-Fernández, E., Tajés, M., Ill-Raga, G., Vargas, L., Busquets-García, A., Bosch-Morató, M., Guivernau, B., Valls-Comamala, V., Gomis, M., Grau, C., Fandos, C., Rosen, M. D., Rabinowitz, M. H., Inestrosa, N., Maldonado, R., Altafaj, X., Ozaita, A., Alvarez, A., Vicente, R., Valverde, M. A., ... Muñoz, F. J. (2016). Glutamatergic stimulation induces GluN2B translation by the nitric oxide-Heme-Regulated eIF2 $\alpha$  kinase in cortical neurons. *Oncotarget*, 7(37), 58876–58892.
- Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., ... & Olson, E. N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell*, 160(4), 595–606.
- Anderson, D. M., Anderson, K. M., Chang, C.-L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R. et Olson, E. N. (2015). A Micropeptide encoded by a putative long noncoding rna regulates muscle performance. *Cell*, 160(4), 595 606.
- Andreev, D. E., Dmitriev, S. E., Zinovkin, R., Terenin, I. M., & Shatsky, I. N. (2012). The 5' untranslated region of Apaf-1 mRNA directs translation under apoptosis conditions via a 5' end-dependent scanning mechanism. *FEBS Letters*, 586(23), 4139–4143.
- Andreev, D. E., Hirnet, J., Terenin, I. M., Dmitriev, S. E., Niepmann, M., & Shatsky, I. N. (2012). Glycyl-tRNA synthetase specifically binds to the poliovirus IRES to activate translation initiation. *Nucleic Acids Research*, 40(12), 5602–5614.
- Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., Cormican, P., Morris, D. W., Shatsky, I. N., & Baranov, P. V. (2015). Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *eLife*, 4, e03971.
- Andrews, S. J., & Rothnagel, J. A. (2014). Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews. Genetics*, 15(3), 193–204.

- Aspden, J. L., Eyre-Walker, Y. C., Phillips, R. J., Amin, U., Mumtaz, M. A., Brocard, M., & Couso, J. P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *eLife*, 3, e03528.
- Avci, D., & Lemberg, M. K. (2015). Clipping or extracting: two ways to membrane protein degradation. *Trends in Cell Biology*, 25(10), 611–622.
- Baek, J., Lee, J., Yoon, K., & Lee, H. (2017). Identification of unannotated small genes in salmonella. *G3* (Bethesda, Md.), 7(3), 983–989.
- Baird, T. D., & Wek, R. C. (2012). Eukaryotic initiation factor 2 phosphorylation and translational control in metabolism. *Advances in Nutrition* (Bethesda, Md.), 3(3), 307–321.
- Balachandran, S., Roberts, P. C., Brown, L. E., Truong, H., Pattnaik, A. K., Archer, D. R., & Barber, G. N. (2000). Essential role for the dsRNA-dependent protein kinase PKR in innate immunity to viral infection. *Immunity*, 13(1), 129–141.
- Bazzini, A. A., Johnstone, T. G., Christiano, R., Mackowiak, S. D., Obermayer, B., Fleming, E. S., Vejnar, C. E., Lee, M. T., Rajewsky, N., Walther, T. C., & Giraldez, A. J. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *The EMBO Journal*, 33(9), 981–993.
- B'chir, W., Maurin, A. C., Carraro, V., Averous, J., Jousse, C., Muranishi, Y., Parry, L., Stepien, G., Fafournoux, P., & Bruhat, A. (2013). The eIF2 $\alpha$ /ATF4 pathway is essential for stress-induced autophagy gene expression. *Nucleic Acids Research*, 41(16), 7683–7699.
- Bechtel, S., Rosenfelder, H., Duda, A., Schmidt, C. P., Ernst, U., Wellenreuther, R., Mehrle, A., Schuster, C., Bahr, A., Blöcker, H., Heubner, D., Hoerlein, A., Michel, G., Wedler, H., Köhrer, K., Ottenwälder, B., Poustka, A., Wiemann, S., & Schupp, I. (2007). The full-ORF clone resource of the German cDNA Consortium. *BMC Genomics*, 8, 399. Bergeron, D., Lapointe, C., Bissonnette, C., Tremblay, G., Motard, J. et Roucou, X. (2013). An out-of-frame overlapping reading frame in the ataxin-1 coding sequence encodes a novel ataxin-1 interacting protein. *Journal of Biological Chemistry*, 288(30), 21824-21835
- Bernard, C. (1957). An introduction to the study of experimental medicine (Vol. 400). *Courier Corporation*.
- Blumenthal T. (2004). Operons in eukaryotes. *Briefings In Functional Genomics & Proteomics*, 3(3), 199–211.
- Brändli, A. W., Hansson, G. C., Rodriguez-Boulan, E., & Simons, K. (1988). A polarized epithelial cell mutant deficient in translocation of UDP-galactose into the Golgi complex. *Journal of Biological Chemistry*, 263(31), 16283-16290.
- Brown, M. S., Ye, J., Rawson, R. B., & Goldstein, J. L. (2000). Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell*, 100(4), 391–398.
- Brunet, M. A., Brunelle, M., Lucier, J. F., Delcourt, V., Levesque, M., Grenier, F., Samandi, S., Leblanc, S., Aguilar, J. D., Dufour, P., Jacques, J. F., Fournier, I., Ouangraoua, A., Scott, M. S., Boisvert, F. M., &

- Roucou, X. (2019). OpenProt: a more comprehensive guide to explore eukaryotic coding potential and proteomes. *Nucleic Acids Research*, 47(D1), D403–D410.
- Brunet, M. A., Leblanc, S., & Roucou, X. (2020). Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. *Experimental Cell Research*, 393(1), 112057.
- Brunet, M. A., Levesque, S. A., Hunting, D. J., Cohen, A. A., & Roucou, X. (2018). Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Research*, 28(5), 609–624.
- Caffaro, C. E., & Hirschberg, C. B. (2006). Nucleotide sugar transporters of the Golgi apparatus: from basic science to diseases. *Accounts of Chemical Research*, 39(11), 805-812.
- Cataldi, M. P., Lu, P., Blaeser, A., & Lu, Q. L. (2018). Ribitol restores functionally glycosylated  $\alpha$ -dystroglycan and improves muscle function in dystrophic FKRP-mutant mice. *Nature Communications*, 9(1), 3448.
- Cataldi, M. P., Lu, P., Blaeser, A., & Lu, Q. L. (2018). Ribitol restores functionally glycosylated  $\alpha$ -dystroglycan and improves muscle function in dystrophic FKRP-mutant mice. *Nature Communications*, 9(1), 3448.
- Cheng, H., Chan, W. S., Li, Z., Wang, D., Liu, S., & Zhou, Y. (2011). Small open reading frames: current prediction techniques and future prospect. *Current Protein & Peptide Science*, 12(6), 503–507.
- Choteau, S. (2022). Exploration of the Functions of Short Open Reading Frames in Monocytes.
- Ciechanover, A., & Kwon, Y. T. (2015). Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. *Experimental & Molecular Medicine*, 47(3), e147.
- Clancy, S. & Brown, W. (2008) Translation: DNA to mRNA to Protein. *Nature Education* 1(1), 101.
- Costa-Mattioli, M., & Walter, P. (2020). The integrated stress response: From mechanism to disease. *Science (New York, N.Y.)*, 368(6489), eaat5314.
- Couso, J. P., & Patraquim, P. (2017). Classification and function of small open reading frames. *Nature Reviews Molecular Cell Biology*, 18(9), 575-589.
- De Marothy, M. T., & Elofsson, A. (2015). Marginally hydrophobic transmembrane  $\alpha$ -helices shaping membrane protein folding. *Protein Science: a Publication of the Protein Society*, 24(7), 1057–1074.
- Delcourt, V., Brunelle, M., Roy, A. V., Jacques, J. F., Salzet, M., Fournier, I., & Roucou, X. (2018). The protein coded by a short open reading frame, not by the annotated coding sequence, is the main gene product of the dual-coding gene MIEF1. *Molecular & Cellular Proteomics*, 17(12), 2402-2411.



- Deng, J., Harding, H. P., Raught, B., Gingras, A. C., Berlanga, J. J., Scheuner, D., Kaufman, R. J., Ron, D., & Sonenberg, N. (2002). Activation of GCN2 in UV-irradiated cells inhibits translation. *Current Biology: CB*, 12(15), 1279–1286.
- D'Lima, N., Ma, J., Winkler, L. et al. (2017). A human microprotein that interacts with the mRNA decapping complex. *Nature Chemical Biology*, 13, 174–180.
- Dong, X., Zhang, K., Xun, C., Chu, T., Liang, S., Zeng, Y., & Liu, Z. (2023). Small Open Reading Frame-Encoded Micro-Peptides: An Emerging Protein World. *International Journal of Molecular Sciences*, 24(13), 10562.
- Donnelly, N., Gorman, A. M., Gupta, S., & Samali, A. (2013). The eIF2 $\alpha$  kinases: their structures and functions. *Cellular and Molecular Life Sciences*, 70, 3493-3511.
- Eckhardt, M., Gotza, B., & Gerardy-Schahn, R. (1998). Mutants of the CMP-sialic acid transporter causing the Lec2 phenotype. *The Journal of Biological Chemistry*, 273(32), 20189–20195.
- Endo, T. (2015) Glycobiology of alpha-dystroglycan and muscular dystrophy. *Journal of Biochemistry*. 157,1–12.
- Erhard, F., Halenius, A., Zimmermann, C., L'Hernault, A., Kowalewski, D. J., Weekes, M. P., ... & Dölken, L. (2018). Improved Ribo-seq enables identification of cryptic translation events. *Nature Methods*, 15(5), 363-366.
- Fields, A. P., Rodriguez, E. H., Jovanovic, M., Stern-Ginossar, N., Haas, B. J., Mertins, P., Raychowdhury, R., Hacohen, N., Carr, S. A., Ingolia, N. T., Regev, A. et Weissman, J. S. (2015). A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. *Molecular Cell*, 60(5), 816 827.
- Fredriksson, R., Nordström, K. J., Stephansson, O., Hägglund, M. G., & Schiöth, H. B. (2008). The solute carrier (SLC) complement of the human genome: phylogenetic classification reveals four major families. *FEBS Letters*, 582(27), 3811-3816.
- Fulda, S., Gorman, A. M., Hori, O. et Samali, A. (2010). Cellular Stress Responses: Cell Survival and Cell Death. *International Journal of Cell Biology*, 2010, e214074.
- Gagliardi, D., & Dziembowski, A. (2018). 5' and 3' modifications controlling RNA degradation: from safeguards to executioners. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, 373(1762), 20180160.
- Gerardy-Schahn, R., Oelmann, S., & Bakker, H. (2001). Nucleotide sugar transporters: biological and functional aspects. *Biochimie*, 83(8), 775-782.
- Gerlach, J., Sharma, S., Leister, K., & Joshi, L. (2012). A Tight-Knit Group: Protein Glycosylation, Endoplasmic Reticulum Stress and the Unfolded Protein Response. In P. Agostinis & S. Afshin (Eds.), *Endoplasmic Reticulum Stress in Health and Disease*. Springer.
- Ghosh, A., & Lima, C. D. (2010). Enzymology of RNA cap synthesis. *Wiley Interdisciplinary Reviews: RNA*, 1(1), 152-172.

- Grallert, B., & Boye, E. (2013). GCN2, an old dog with new tricks. *Biochemical Society Transactions*, 41(6), 1687–1691.
- Guillen, E., Abeijon, C., & Hirschberg, C. B. (1998). Mammalian Golgi apparatus UDP-N-acetylglucosamine transporter: molecular cloning by phenotypic correction of a yeast mutant. *Proceedings of the National Academy of Sciences*, 95(14), 7888-7892.
- Guo, X., Chavez, A., Tung, A., Chan, Y., Kaas, C., Yin, Y., ... & Church, G. M. (2018). High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast. *Nature Biotechnology*, 36(6), 540-546.
- Gupta, R., Sahu, M., Srivastava, D., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Post-translational modifications: Regulators of neurodegenerative proteinopathies. *Ageing Research Reviews*, 68, 101336.
- Hadley, B., Litfin, T., Day, C. J., Haselhorst, T., Zhou, Y. et Tiralongo, J. (2019). Nucleotide sugar transporter slc35 family structure and function. *Computational and Structural Biotechnology Journal*, 17, 1123 1134.
- Hadley, B., Maggioni, A., Ashikov, A., Day, C. J., Haselhorst, T., & Tiralongo, J. (2014). Structure and function of nucleotide sugar transporters: Current progress. *Computational and Structural Biotechnology Journal*, 10(16), 23-32.
- Hajam, Y. A., Rani, R., Ganie, S. Y., Sheikh, T. A., Javaid, D., Qadri, S. S., Pramodh, S., Alsulimani, A., Alkhanani, M. F., Harakeh, S., Hussain, A., Haque, S., & Reshi, M. S. (2022). Oxidative stress in human pathology and aging: molecular mechanisms and perspectives. *Cells*, 11(3), 552.
- Han, A. P., Yu, C., Lu, L., Fujiwara, Y., Browne, C., Chin, G., Fleming, M., Leboulch, P., Orkin, S. H., & Chen, J. J. (2001). Heme-regulated eIF2 $\alpha$  kinase (HRI) is required for translational regulation and survival of erythroid precursors in iron deficiency. *The EMBO Journal*, 20(23), 6909–6918.
- Hanada, K., Higuchi-Takeuchi, M., Okamoto, M., Yoshizumi, T., Shimizu, M., Nakaminami, K., Nishi, R., Ohashi, C., Iida, K., Tanaka, M., Horii, Y., Kawashima, M., Matsui, K., Toyoda, T., Shinozaki, K., Seki, M., & Matsui, M. (2013). Small open reading frames associated with morphogenesis are hidden in plant genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 110(6), 2395–2400.
- Harding, H. P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., & Ron, D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Molecular Cell*, 6(5), 1099-1108.
- Harding, H. P., Zhang, Y., & Ron, D. (1999). Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase. *Nature*, 397(6716), 271-274.
- Harding, H. P., Zhang, Y., Zeng, H., Novoa, I., Lu, P. D., Calton, M., Sadri, N., Yun, C., Popko, B., Paules, R., Stojdl, D. F., Bell, J. C., Hettmann, T., Leiden, J. M., & Ron, D. (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Molecular Cell*, 11(3), 619–633.

- Healy, S. J., Gorman, A. M., Mousavi-Shafaei, P., Gupta, S., & Samali, A. (2009). Targeting the endoplasmic reticulum-stress response as an anticancer strategy. *European Journal of Pharmacology*, 625(1-3), 234-246.
- Hediger, M. A., Romero, M. F., Peng, J. B., Rolfs, A., Takanaga, H., & Bruford, E. A. (2004). The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins Introduction. *Pflügers Archiv : European Journal of Physiology*, 447(5), 465–468.
- Hellens, R. P., Brown, C. M., Chisnall, M. A. W., Waterhouse, P. M., & Macknight, R. C. (2016). The emerging world of small orfs. *Trends in Plant Science*, 21(4), 317–328.
- Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G., & Rudd, K. E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. *Molecular Microbiology*, 70(6), 1487–1501.
- Hinnebusch A. G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology*, 59, 407–450.
- Hinnebusch A. G. (2014). The scanning mechanism of eukaryotic translation initiation. *Annual Review of Biochemistry*, 83, 779–812.
- Hinnebusch AG, Lorsch JR (2012) The mechanism of eukaryotic translation initiation: new insights and challenges. *Cold Spring Harb Perspect Biol* 4:a011544–a011544.
- Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science (New York, N.Y.)*, 352(6292), 1413–1416.
- Hocine, S., Singer, R. H., & Grünwald, D. (2010). RNA processing and export. *Cold Spring Harbor perspectives in biology*, 2(12), a000752.
- Holcik M, Sonenberg N (2005) Translational control in stress and apoptosis. *Nature Reviews. Molecular Cell Biology*, 6:318–327.
- Hsu, P. Y., & Benfey, P. N. (2018). Small but mighty: functional peptides encoded by small orfs in plants. *Proteomics*, 18(10), e1700038.
- Hsu, P. Y., Calviello, L., Wu, H. Y. L., Li, F. W., Rothfels, C. J., Ohler, U., & Benfey, P. N. (2016). Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proceedings of the National Academy of Sciences*, 113(45), E7126-E7135.
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (New York, N.Y.)*, 324(5924).
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (New York, N.Y.)*, 324(5924), 218–223.

- Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), 789–802.
- Jackson, R. J., Hellen, C. U., & Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature reviews. Molecular Cell Biology*, 11(2), 113–127.
- Jensen O. N. (2006). Interpreting the protein language using proteomics. *Nature Reviews. Molecular Cell Biology*, 7(6), 391–403.
- Ji, Z., Song, R., Regev, A., & Struhl, K. (2015). Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife*, 4, e08890.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096), 816–821. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife*, 2, e00471.
- Jordan, M., Schallhorn, A., & Wurm, F. M. (1996). Transfecting mammalian cells: optimization of critical parameters affecting calcium-phosphate precipitate formation. *Nucleic Acids Research*, 24(4), 596–601.
- Juntawong, P., Girke, T., Bazin, J., & Bailey-Serres, J. (2014). Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1), E203–E212.
- Kanagawa, M., & Toda, T. (2018). Ribitol-phosphate-a newly identified posttranslational glycosylation unit in mammals: structure, modification enzymes and relationship to human diseases. *Journal Of Biochemistry*, 163(5), 359–369.
- Khitun, A., Ness, T. J., & Slavoff, S. A. (2019). Small open reading frames and cellular stress responses. *Molecular Omics*, 15(2), 108–116.
- Khoury, G. A., Baliban, R. C., & Floudas, C. A. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Scientific Reports*, 1, 90.
- Kim, Y., Lee, J. H., Park, J., Cho, J., Yi, H., and Kim, V. N. (2014). PKR is activated by cellular dsRNAs during mitosis and acts as a mitotic regulator. *Genes Dev.* 28, 1310–1322.
- Kinch, L. N., Ginalski, K., & Grishin, N. V. (2006). Site-2 protease regulated intramembrane proteolysis: sequence homologs suggest an ancient signaling cascade. *Protein Science: a publication of the Protein Society*, 15(1), 84–93.
- Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., & Kageyama, Y. (2007). Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nature Cell Biology*, 9(6), 660–665.
- Kondrashov F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings. Biological Sciences*, 279(1749), 5048–5057.

- Kozak M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene*, 234(2), 187–208.
- Kozak, M. (1991). An analysis of vertebrate mRNA sequences: intimations of translational control. *The Journal of Cell Biology*, 115(4), 887-903.
- Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361, 13–37.
- Kudryavtseva, A. V., Krasnov, G. S., Dmitriev, A. A., Alekseev, B. Y., Kardymon, O. L., Sadritdinova, A. F., Fedorova, M. S., Pokrovsky, A. V., Melnikova, N. V., Kaprin, A. D., Moskalev, A. A., & Snezhkina, A. V. (2016). Mitochondrial dysfunction and oxidative stress in aging and cancer. *Oncotarget*, 7(29), 44879–44905.
- Kühnle, N., Dederer, V., & Lemberg, M. K. (2019). Intramembrane proteolysis at a glance: from signalling to protein degradation. *Journal of Cell Science*, 132(16), jcs217745.
- Lacerda, R., Menezes, J. et Romão, L. (2017). More than just scanning: the importance of cap-independent mRNA translation initiation for cellular stress response and cancer. *Cellular and Molecular Life Sciences*, 74(9), 1659–1680.
- Lavoie, H., Li, J. J., Thevakumaran, N., Therrien, M., & Sicheri, F. (2014). Dimerization-induced allostery in protein kinase regulation. *Trends in Biochemical Sciences*, 39(10), 475-486.
- Leblanc, S., Yala, F., Provencher, N., Lucier, J. F., Levesque, M., Lapointe, X., Jacques, J. F., Fournier, I., Salzet, M., Ouangraoua, A., Scott, M. S., Boisvert, F. M., Brunet, M. A., & Roucou, X. (2024). OpenProt 2.0 builds a path to the functional characterization of alternative proteins. *Nucleic Acids Research*, 52(D1), D522–D528.
- Lemberg, M. K., & Freeman, M. (2007). Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Research*, 17(11), 1634–1646.
- Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N., & Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*, 43(W1), W580–W584.
- Lin, M. F., Jungreis, I., & Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics (Oxford, England)*, 27(13), i275–i282.
- Live, D., Wells, L., and Boons, G. J. (2013) Dissecting the molecular basis of the role of the O-mannosylation pathway in disease: Alpha- dystroglycan and forms of muscular dystrophy. *Chembiochem* 14, 23922402.
- Lo, H. S., Wang, Z., Hu, Y., Yang, H. H., Gere, S., Buetow, K. H., & Lee, M. P. (2003). Allelic variation in gene expression is common in the human genome. *Genome Research*, 13(8), 1855–1862.
- Lu L, Han AP, Chen JJ (2001) Translation initiation control by heme - regulated eukaryotic initiation factor 2alpha kinase in erythroid cells under cytoplasmic stresses. *Molecular and Cellular Biology* 21: 7971 - 7980.

- Lu, P. D., Harding, H. P., & Ron, D. (2004). Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *The Journal of Cell Biology*, 167(1), 27-33.
- Ma, J., Ward, C. C., Jungreis, I., Slavoff, S. A., Schwaid, A. G., Neveu, J., Budnik, B. A., Kellis, M., & Saghatelian, A. (2014). Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. *Journal of Proteome Research*, 13(3), 1757–1765.
- Mackowiak, S.D., Zauber, H., Bielow, C. et al. (2015). Extensive identification and analysis of conserved small ORFs in animals. *Genome Biology*, 16, 179.
- Makarewich, C. A., & Olson, E. N. (2017). Mining for Micropeptides. *Trends in cell biology*, 27(9), 685–696.
- Mann, M., & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nature Biotechnology*, 21(3), 255–261.
- Martínez-Salas E, Piñeiro D, Fernández N (2012) Alternative mechanisms to initiate translation in eukaryotic mRNAs. *Comparative and Functional Genomics*, 2012:391546.
- McEwen, E., Kedersha, N., Song, B., Scheuner, D., Gilks, N., Han, A., Chen, J. J., Anderson, P., & Kaufman, R. J. (2005). Heme-regulated inhibitor kinase-mediated phosphorylation of eukaryotic translation initiation factor 2 inhibits translation, induces stress granule formation, and mediates survival upon arsenite exposure. *The Journal of Biological Chemistry*, 280(17), 16925–16933.
- Medina-Navarro, R., Durán-Reyes, G., Díaz-Flores, M., & Vilar-Rojas, C. (2010). Protein antioxidant response to the stress and the relationship between molecular structure and antioxidant function. *Plos One*, 5(1), e8971.
- Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., Skabkin, M.A., Elemento, O., Pestova, T.V., Qian, S.-B., et al. (2015) 5'UTR m6A promotes cap-independent translation. *Cell* 163(4):999–1010.
- Michele, D. E., Barresi, R., Kanagawa, M., Saito, F., Cohn, R. D., Satz, J. S., Dollar, J., Nishino, I., Kelley, R. I., Somer, H., Straub, V., Mathews, K. D., Moore, S. A., and Campbell, K. P. (2002) Post-translational disruption of dystroglycan-ligand in interactions in congenital muscular dystrophies. *Nature* 418, 417–422.
- Montecinos, V., Guzmán, P., Barra, V., Villagrán, M., Muñoz-Montesino, C., Sotomayor, K., Escobar, E., Godoy, A., Mardones, L., Sotomayor, P., Guzmán, C., Vásquez, O., Gallardo, V., van Zundert, B., Bono, M. R., Oñate, S. A., Bustamante, M., Cárcamo, J. G., Rivas, C. I., & Vera, J. C. (2007). Vitamin C is an essential antioxidant that enhances survival of oxidatively stressed human vascular endothelial cells in the presence of a vast molar excess of glutathione. *The Journal of Biological Chemistry*, 282(21), 15506–15515.
- Moro, S. G., Hermans, C., Ruiz-Orera, J., & Albà, M. M. (2021). Impact of uORFs in mediating regulation of translation in stress conditions. *BMC Molecular and Cell Biology*, 22(1), 29.
- Moulleron, H., Delcourt, V., & Roucou, X. (2016). Death of a dogma: eukaryotic mRNAs can code for more than one protein. *Nucleic Acids Research*, 44(1), 14–23.

- Mudge, J. M., Jungreis, I., Hunt, T., Gonzalez, J. M., Wright, J. C., Kay, M., Davidson, C., Fitzgerald, S., Seal, R., Tweedie, S., He, L., Waterhouse, R. M., Li, Y., Bruford, E., Choudhary, J. S., Frankish, A., & Kellis, M. (2019). Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. *Genome Research*, 29(12), 2073–2087.
- 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., Yang, Y. T., Albà, M. M., Aspden, J. L., Baranov, P. V., Bazzini, A. A., Bruford, E., Martin, M. J., Calviello, L., Carvunis, A. R., Chen, J., ... van Heesch, S. (2022). Standardized annotation of translated open reading frames. *Nature Biotechnology*, 40(7), 994–999.
- Muñoz, J., & Heck, A. J. (2014). From the human genome to the human proteome. *Angewandte Chemie (International ed. in English)*, 53(41), 10864–10866.
- Nakamura, T., Furuhashi, M., Li, P., Cao, H., Tuncman, G., Sonenberg, N., Gorgun, C. Z., & Hotamisligil, G. S. (2010). Double-stranded RNA-dependent protein kinase links pathogen sensing with stress and metabolic homeostasis. *Cell*, 140(3), 338–348.
- Nalivaeva, N. N., & Turner, A. J. (2001). Post-translational modifications of proteins: acetylcholinesterase as a model system. *Proteomics*, 1(6), 735–747.
- Nikolov, M., Schmidt, C., & Urlaub, H. (2012). Quantitative mass spectrometry-based proteomics: an overview. *Methods in Molecular Biology* (Clifton, N.J.), 893, 85–100.
- Normark, S., Bergström, S., Edlund, T., Grundström, T., Jaurin, B., Lindberg, F. P., & Olsson, O. (1983). Overlapping genes. *Annual Review of Genetics*, 17, 499–525.
- Nyborg, A. C., Ladd, T. B., Jansen, K., Kukar, T. et Golde, T. E. (2006). Intramembrane proteolytic cleavage by human signal peptide peptidase like 3 and malaria signal peptide peptidase. *The FASEB Journal*, 20(10), 1671–1679.
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., ... Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1), D733–D745.
- Olexiouk, V., & Menschaert, G. (2016). Identification of small novel coding sequences, a proteogenomics endeavor. *Advances in Experimental Medicine and Biology*, 926, 49–64.
- Olexiouk, V., Crappé, J., Verbruggen, S., Verhegen, K., Martens, L., & Menschaert, G. (2016). sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Research*, 44(D1), D324–D329.
- Onuki, R., Bando, Y., Suyama, E., Katayama, T., Kawasaki, H., Baba, T., Tohyama, M., & Taira, K. (2004). An RNA-dependent protein kinase is involved in tunicamycin-induced apoptosis and Alzheimer's disease. *The EMBO Journal*, 23(4), 959–968.

- Orr, M. W., Mao, Y., Storz, G., & Qian, S. B. (2020). Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Research*, 48(3), 1029-1042.
- Oyama, M., Kozuka-Hata, H., Suzuki, Y., Semba, K., Yamamoto, T., & Sugano, S. (2007). Diversity of translation start sites may define increased complexity of the human short ORFeome. *Molecular & Cellular Proteomics: MCP*, 6(6), 1000–1006.
- Pakos-Zebrucka, K., Koryga, I., Mnich, K., Ljujic, M., Samali, A., & Gorman, A. M. (2016). The integrated stress response. *EMBO Reports*, 17(10), 1374–1395.
- Paolini, N. A., Moore, K. S., di Summa, F. M., Fokkema, I. F. A. C., 't Hoen, P. A. C., & von Lindern, M. (2018). Ribosome profiling uncovers selective mRNA translation associated with eIF2 phosphorylation in erythroid progenitors. *PLoS One*, 13(4), e0193790.
- Pataer, A., Swisher, S. G., Roth, J. A., Logothetis, C. J., & Corn, P. G. (2009). Inhibition of RNA-dependent protein kinase (PKR) leads to cancer cell death and increases chemosensitivity. *Cancer Biology & Therapy*, 8(3), 245–252.
- Peidis, P., Papadakis, A. I., Rajesh, K., & Koromilas, A. E. (2010). HDAC pharmacological inhibition promotes cell death through the eIF2 $\alpha$  kinases PKR and GCN2. *Aging (Albany NY)*, 2(10), 669.
- Picot, J., Guerin, C. L., Le Van Kim, C., & Boulanger, C. M. (2012). Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology*, 64(2), 109–130.
- Ponomarenko, E. A., Poverennaya, E. V., Ilgisonis, E. V., Pyatnitskiy, M. A., Kopylov, A. T., Zgoda, V. G., Lisitsa, A. V., & Archakov, A. I. (2016). The size of the human proteome: the width and depth. *International Journal of Analytical Chemistry*, 2016, 7436849.
- Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 35(Database issue), D61–D65.
- Pukkila, P. J. (2001). Molecular Biology: The central dogma. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd.
- Ramazi, S., & Zahiri, J. (2021). Posttranslational modifications in proteins: resources, tools and prediction methods. *Database : The Journal of Biological Databases and Curation*, 2021, baab012.
- Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6), 1380–1389.
- Rathore, A., Chu, Q., Tan, D., Martinez, T. F., Donaldson, C. J., Diedrich, J. K., Yates, J. R., 3rd, & Saghatelian, A. (2018). MIEF1 microprotein regulates mitochondrial translation. *Biochemistry*, 57(38), 5564–5575.
- Ravishankar, B., Liu, H., Shinde, R., Chaudhary, K., Xiao, W., Bradley, J., Koritzinsky, M., Madaio, M. P., & McGaha, T. L. (2015). The amino acid sensor GCN2 inhibits inflammatory responses to apoptotic



- cells promoting tolerance and suppressing systemic autoimmunity. *Proceedings of the National Academy of Sciences of the United States of America*, 112(34), 10774–10779.
- Reinhardt, J. A., Wanjiru, B. M., Brant, A. T., Saelao, P., Begun, D. J., & Jones, C. D. (2013). De novo ORFs in *Drosophila* are important to organismal fitness and evolved rapidly from previously non-coding sequences. *PLoS genetics*, 9(10), e1003860.
- Ron, D. (2002). Translational control in the endoplasmic reticulum stress response. *The Journal of clinical investigation*, 110(10), 1383-1388.
- Rossi, A., Pisani, F., Nicchia, G. P., Svelto, M., & Frigeri, A. (2010). Evidences for a leaky scanning mechanism for the synthesis of the shorter M23 protein isoform of aquaporin-4: implication in orthogonal array formation and neuromyelitis optica antibody interaction. *The Journal of biological chemistry*, 285(7), 4562–4569.
- Rouschop, K. M., Dubois, L. J., Keulers, T. G., van den Beucken, T., Lambin, P., Bussink, J., van der Kogel, A. J., Koritzinsky, M., & Wouters, B. G. (2013). PERK/eIF2 $\alpha$  signaling protects therapy resistant hypoxic cells through induction of glutathione synthesis and protection against ROS. *Proceedings of the National Academy of Sciences of the United States of America*, 110(12), 4622–4627.
- Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., & Dwek, R. A. (2001). Glycosylation and the immune system. *Science (New York, N.Y.)*, 291(5512), 2370–2376.
- Ruiz-Ramos, R., Lopez-Carrillo, L., Rios-Perez, A. D., De Vizcaya-Ruiz, A., & Cebrian, M. E. (2009). Sodium arsenite induces ROS generation, DNA oxidative damage, HO-1 and c-Myc proteins, NF-kappaB activation and cell proliferation in human breast cancer MCF-7 cells. *Mutation research*, 674(1-2), 109–115.
- Samandi, S., Roy, A. V., Delcourt, V., Lucier, J.-F., Gagnon, J., Beaudoin, M. C., Vanderperre, B., Breton, M.-A., Motard, J., Jacques, J.-F., Brunelle, M., Gagnon-Arsenault, I., Fournier, I., Ouangraoua, A., Hunting, D. J., Cohen, A. A., Landry, C. R., Scott, M. S. et Roucou, X. (2017). Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins. *eLife*, 6, e27860.
- Scheres, S. H., & Nagai, K. (2017). CryoEM structures of spliceosomal complexes reveal the molecular mechanism of pre-mRNA splicing. *Current opinion in structural biology*, 46, 130-139.
- Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 106 Suppl 1(Suppl 1), 9955–9962.
- Schoof, M., Boone, M., Wang, L., Lawrence, R., Frost, A., & Walter, P. (2021). eIF2B conformation and assembly state regulate the integrated stress response. *eLife*, 10, e65703.
- Shatsky, I. N., Terenin, I. M., Smirnova, V. V., & Andreev, D. E. (2018). Cap-Independent Translation: What's in a Name?. *Trends in biochemical sciences*, 43(11), 882–895.
- Sies H. (2017). Hydrogen peroxide as a central redox signaling molecule in physiological oxidative stress: Oxidative eustress. *Redox biology*, 11, 613–619.

- Skabkin, M. A., Skabkina, O. V., Hellen, C. U., & Pestova, T. V. (2013). Reinitiation and other unconventional posttermination events during eukaryotic translation. *Molecular cell*, 51(2), 249–264.
- Smith, L. M., & Kelleher, N. L. (2018). Proteoforms as the next proteomics currency. *Science (New York, N.Y.)*, 359(6380), 1106–1107.
- Sonenberg N, Hinnebusch AG (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136:731–745.
- Sosicka, P., Maszczak-Seneczko, D., Bazan, B., Shauchuk, Y., Kaczmarek, B., & Olczak, M. (2017). An insight into the orphan nucleotide sugar transporter SLC35A4. *Biochimica et biophysica acta. Molecular cell research*, 1864(5), 825–838.
- Stram, A. R., & Payne, R. M. (2016). Post-translational modifications in mitochondria: protein signaling in the powerhouse. *Cellular and molecular life sciences: CMLS*, 73(21), 4063–4073.
- Suragani, R. N., Zachariah, R. S., Velazquez, J. G., Liu, S., Sun, C. W., Townes, T. M., & Chen, J. J. (2012). Heme-regulated eIF2 $\alpha$  kinase activated Atf4 signaling pathway in oxidative stress and erythropoiesis. *Blood*, 119(22), 5276–5284.
- Szegezdi, E., Logue, S. E., Gorman, A. M., & Samali, A. (2006). Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Reports*, 7(9), 880–885.
- Tang, C. P., Clark, O., Ferrarone, J. R., Campos, C., Lalani, A. S., Chodera, J. D., ... & Mellingshoff, I. K. (2022). GCN2 kinase activation by ATP-competitive kinase inhibitors. *Nature Chemical Biology*, 18(2), 207–215.
- Tang, C. P., Clark, O., Ferrarone, J. R., Campos, C., Lalani, A. S., Chodera, J. D., Intlekofer, A. M., Elemento, O., & Mellingshoff, I. K. (2022). GCN2 kinase activation by ATP-competitive kinase inhibitors. *Nature chemical biology*, 18(2), 207–215.
- Taniuchi, S., Miyake, M., Tsugawa, K., Oyadomari, M. et Oyadomari, S. (2016). Integrated stress response of vertebrates is regulated by four eIF2 $\alpha$  kinases. *Scientific Reports*, 6(1), 32886.
- Terenin IM, Andreev DE, Dmitriev SE, Shatsky IN (2013) A novel mechanism of eukaryotic translation initiation that is neither m7G-cap-, nor IRES-dependent. *Nucleic Acids Res* 41:1807–1816.
- Topisirovic I, Sonenberg N (2011) mRNA translation and energy metabolism in cancer: the role of the MAPK and mTORC1 pathways. *Cold Spring Harb Symp Quant Biol* 76:355–367.
- Tsuji Y. (2020). Transmembrane protein western blotting: Impact of sample preparation on detection of SLC11A2 (DMT1) and SLC40A1 (ferroportin). *PLoS one*, 15(7), e0235563.
- Tucker, J. D., Doddapaneni, R., Lu, P. J., & Lu, Q. L. (2022). Ribitol alters multiple metabolic pathways of central carbon metabolism with enhanced glycolysis: A metabolomics and transcriptomics profiling of breast cancer. *PLoS one*, 17(12), e0278711.

- Ury, B., Potelle, S., Caligiore, F., Whorton, M. R., & Bommer, G. T. (2021). The promiscuous binding pocket of SLC35A1 ensures redundant transport of CDP-ribitol to the Golgi. *The Journal of Biological Chemistry*, 296, 100789.
- Valko, M., Leibfritz, D., Moncol, J., Cronin, M. T., Mazur, M., & Telser, J. (2007). Free radicals and antioxidants in normal physiological functions and human disease. *The International Journal of Biochemistry & Cell Biology*, 39(1), 44–84.
- Vanderperre, B., Lucier, J.-F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F.-M. et Roucou, X. (2013). Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS ONE*, 8(8), e70698.
- Vanderperre, B., Staskevicius, A. B., Tremblay, G., McCoy, M., O'Neill, M. A., Cashman, N. R., & Roucou, X. (2011). An overlapping reading frame in the PRNP gene encodes a novel polypeptide distinct from the prion protein. *The FASEB Journal*, 25(7), 2373-2386.
- Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., & Etzler, M. E. (Eds.). (2009). *Essentials of Glycobiology*. (2nd ed.). Cold Spring Harbor Laboratory Press.
- Vattem, K. M., & Wek, R. C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proceedings of the National Academy of Sciences*, 101(31), 11269-11274.
- Vazquez De Aldana, C. R., Wek, R. C., Segundo, P. S., Truesdell, A. G., & Hinnebusch, A. G. (1994). Multicopy tRNA genes functionally suppress mutations in yeast eIF-2 $\alpha$  kinase GCN2: evidence for separate pathways coupling GCN4 expression to uncharged tRNA. *Molecular and Cellular Biology*, 14(12), 7920-7932.
- Vazquez-Laslop, N., Sharma, C. M., Mankin, A., & Buskirk, A. R. (2022). Identifying Small Open Reading Frames in Prokaryotes with Ribosome Profiling. *Journal of Bacteriology*, 204(1), e0029421.
- Wadler, C. S., & Vanderpool, C. K. (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51), 20454–20459.
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221), 470–476.
- Weingarten-Gabbay, S., Elias-Kirma, S., Nir, R., Gritsenko, A. A., Stern-Ginossar, N., Yakhini, Z., Weinberger, A., & Segal, E. (2016). Comparative genetics. Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science (New York, N.Y.)*, 351(6270), aad4939.
- Wek, R. C., Jiang, H. Y., & Anthony, T. G. (2006). Coping with stress: eIF2 kinases and translational control. *Biochemical Society Transactions*, 34(1), 7-11.

- Wiktor, M., Wiertelak, W., Maszszak-Seneczko, D., Balwierz, P. J., Szulc, B., & Olczak, M. (2021). Identification of novel potential interaction partners of UDP-galactose (SLC35A2), UDP-N-acetylglucosamine (SLC35A3) and an orphan (SLC35A4) nucleotide sugar transporters. *Journal of Proteomics*, 249, 104321.
- Winckler, B., Faundez, V., Maday, S., Cai, Q., Almeida, C. G., & Zhang, H. (2018). The endolysosomal system and proteostasis: from development to degeneration. *Journal of Neuroscience*, 38(44), 9364-9374.
- Wu, B., Drains, M., Shah, S. N., Lu, P. J., Leroy, V., Killilee, J., Rawls, R., Tucker, J. D., Blaeser, A., & Lu, Q. L. (2022). Ribitol dose-dependently enhances matriglycan expression and improves muscle function with prolonged life span in limb girdle muscular dystrophy 2I mouse model. *PLoS One*, 17(12), e0278482.
- Xiao, W., Sun, Y., Xu, J., Zhang, N., & Dong, L. (2022). uORF-Mediated translational regulation of atf4 serves as an evolutionarily conserved mechanism contributing to non-small-cell lung cancer (nscl) and stress response. *Journal of Molecular Evolution*, 90(5), 375–388.
- Xing, J., Liu, H., Jiang, W., & Wang, L. (2021). LncRNA-Encoded Peptide: functions and predicting methods. *Frontiers in Oncology*, 10, 622294.
- Xu, C., Gamil, A. A. A., Munang'andu, H. M., & Evensen, Ø. (2018). Apoptosis Induction by dsRNA-Dependent Protein Kinase R (PKR) in EPC Cells via Caspase 8 and 9 Pathways. *Viruses*, 10(10), 526.
- Xu, H., Wang, Y., Lin, S., Deng, W., Peng, D., Cui, Q., & Xue, Y. (2018). PTMD: A Database of Human Disease-associated Post-translational Modifications. *Genomics, Proteomics & Bioinformatics*, 16(4), 244–251.
- Yang, H. C., Li, W., Sun, J., & Gross, M. L. (2023). Advances in mass spectrometry on membrane proteins. *Membranes*, 13(5), 457.
- Ye, J., Rawson, R. B., Komuro, R., Chen, X., Davé, U. P., Prywes, R., Brown, M. S., & Goldstein, J. L. (2000). ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Molecular Cell*, 6(6), 1355–1364.
- Yeasmin, F., Yada, T., & Akimitsu, N. (2018). Micropeptides encoded in transcripts previously identified as long noncoding RNAs: a new chapter in transcriptomics and proteomics. *Frontiers in Genetics*, 9, 144.
- Yoshida-Moriguchi, T., and Campbell, K. P. (2015) Matriglycan: A novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology* 25, 702–713.
- Young, S. K., Willy, J. A., Wu, C., Sachs, M. S., & Wek, R. C. (2015). Ribosome reinitiation directs gene-specific translation and regulates the integrated stress response. *Journal of Biological Chemistry*, 290(47), 28257-28271.
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H.,

- Juettemann, T., To, J. K., Laird, M. R., ... Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, 46(D1), D754–D761.
- Zhao, C., Guo, H., Hou, Y., Lei, T., Wei, D., & Zhao, Y. (2023). Multiple roles of the stress sensor gcn2 in immune cells. *International Journal of Molecular Sciences*, 24(5), 4285.
- Zhao, Y., & Jensen, O. N. (2009). Modification - specific proteomics: strategies for characterization of post - translational modifications using enrichment techniques. *Proteomics*, 9(20), 4632-4641.
- Zhou, D. R., Eid, R., Boucher, E., Miller, K. A., Mandato, C. A. et Greenwood, M. T. (2019). Stress is an agonist for the induction of programmed cell death: A review. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1866(4), 699 712.
- Zhu, X.-X., Yao, X.-F., Jiang, L.-P., Geng, C.-Y., Zhong, L.-F., Yang, G., Zheng, B.-L. et Sun, X.-C. (2014). Sodium arsenite induces ROS-dependent autophagic cell death in pancreatic  $\beta$ -cells. *Food and Chemical Toxicology*, 70, 144 150.
- Zinszner, H., Kuroda, M., Wang, X., Batchvarova, N., Lightfoot, R. T., Remotti, H., Stevens, J. L., & Ron, D. (1998). CHOP is implicated in programmed cell death in response to impaired function of the endoplasmic reticulum. *Genes & Development*, 12(7), 982–9