Cytoplasmic switch of ARS2 isoforms promotes nonsense-mediated mRNA decay and arsenic sensitivity

by

Monica Mesa Perez

B.Sc., University of Havana, 2014

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry & Microbiology

University of Victoria

We acknowledge and respect the lək^{w}əŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and W̲SÁNEĆ peoples whose historical relationships with the land continue to this day.

Supervisory Committee

Cytoplasmic switch of ARS2 isoforms promotes nonsense-mediated mRNA decay and arsenic sensitivity

by

Monica Mesa Perez

B.Sc., University of Havana, 2014

Dr. Perry L. Howard, Supervisor

Department of Biochemistry & Microbiology

Dr. Christopher J. Nelson, Departmental Member

Department of Biochemistry & Microbiology

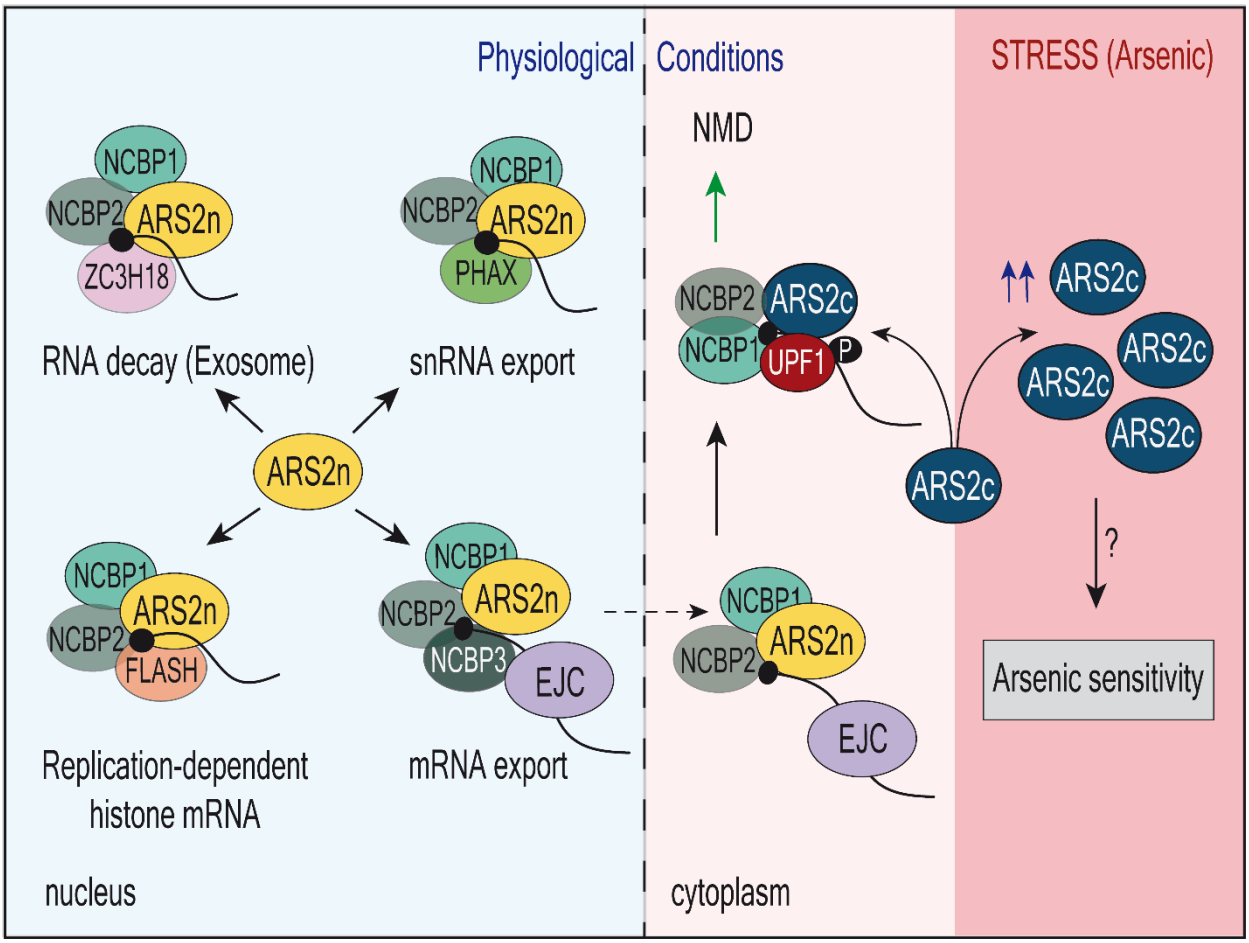
Dr. John E. Burke, Departmental Member

Department of Biochemistry & Microbiology

Dr. Robert L. Chow, Outside Member

Department of Biology

Abstract



The life of RNA polymerase II (RNAPII) transcripts is shaped by the dynamic formation of mutually exclusive ribonucleoprotein complexes (RNPs) that direct transcript biogenesis and turnover. A key regulator of RNA metabolism in the nucleus is the scaffold protein ARS2 (arsenic resistance protein 2), that binds to the cap binding complex (CBC) and regulates processing, degradation, and export of RNAPII transcripts.

We report here that alternative splicing of ARS2's intron 5, generates cytoplasmic isoforms that lack 270 amino acids from the N-terminal of the protein and are functionally distinct from nuclear ARS2. ARS2 isoforms distinctive roles are evidenced under

physiological conditions and stress. Under physiological conditions, ARS2 isoforms differentially regulate transcript degradation through nonsense mediated decay (NMD). Switching of ARS2 isoforms within the CBC in the cytoplasm has dramatic functional consequences, changing ARS2 from a NMD inhibitor to a NMD promoter that enhances the binding of UPF1 to CBP80 and ERF1, favouring SURF complex formation, SMG7 recruitment and transcript degradation. ARS2 isoform exchange is also relevant during arsenic stress. Cytoplasmic ARS2 is specifically induced during arsenic exposure. It is crucial for arsenic sensitivity, and promotes a global response to arsenic in a CBC independent manner. We propose that ARS2 isoform switching promotes the proper recruitment of RNP complexes during NMD and the cellular response to arsenic stress. The existence of non-redundant ARS2 isoforms is relevant for cell homeostasis, stress response and cancer treatment.

Table of Contents

List of Tables

List of Figures

Abbreviations

4E-BP/ EIF4EBP1	Eukaryotic translation initiation factor 4E-binding protein
m7G	7-methylguanosine
aa	Amino acids
ACINUS	Apoptotic chromatin condensation inducer in the nucleus
AGO	Argonaute
ANOVA	Analysis of Variance
AP-LC-MS	Affinity Purification-Liquid chromatography-Mass spectrometry
APL	Acute promyelocytic leukemia
ARS2	Arsenite resistance protein 2
As	Arsenic
ATF3/4/6	Activating transcription factor 3/4/6
ATO	Arsenic trioxide
BAK	Bcl-2 homologous antagonist/killer
BioID	Proximity-dependent biotin identification
CAT	Catalase
CBC	Cap-binding complex
CBP20	Cap-binding protein 20 kDa
CBP80	Cap-binding protein 80 kDa
CDS	Coding sequence

CEC	Capping enzyme complex
CFIm	Cleavage factor Im
CFIIm	Cleavage factor IIm
CHOP/ DDIT3	DNA damage-inducible transcript 3 protein
CNOT8	CCR4-NOT transcription complex subunit 8
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulation factor
CTD	RNA polymerase II C-terminal domain
DCL1	Dicer-Like 1
DCP1	mRNA-decapping enzyme 1A
DECID	Decay-inducing complex
DGCR8	DiGeorge syndrome chromosomal region 8
DMEM	Dulbecco's modified eagle's medium
DR5	Death receptor
DROSHA	Double stranded RNA-Specific Endoribonuclease
DSIF	DRB sensitivity-inducing factor
dsRNA	Double-stranded RNA
DUF	Domain of unknown function
eGFP	Enhanced green fluorescent protein
eIF2α	Eukaryotic translation initiation factor 2 subunit 1
EIF4A3	Eukaryotic initiation factor 4A-III

EIF4E	Eukaryotic translation initiation factor 4E
EJC	Exon junction complex
ER	Endoplasmic reticulum
eRF1/3	Eukaryotic peptide chain release factor 1/3
eRNA	Enhancer RNA
FLASH	FLICE-associated huge protein
FBS	Fetal bovine serum
FSD1L	Fibronectin type III and SPRY domain containing 1 like
GADD34/ PPP1R15A	Protein phosphatase 1 regulatory subunit 15A
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GO	Gene Ontology
GPx	Glutathione peroxidase
GSEA	Gene Set Enrichment Analysis
GSH	Glutathione synthetase
GTEx	Genotype-Tissue Expression Program
HC	Helical core
HCC	Histone pre-mRNA cleavage complex
HDE	Histone downstream element
HERP	Homocysteine-induced endoplasmic reticulum protein
HLBs	Histone locus bodies
hnRNP	Heterogeneous ribonucleoprotein

HS	Horse serum
HYL1	Hyponastic leaves 1
IP	Immunoprecipitation
IP3R	Inositol 1,4,5-triphosphate receptor
IRE1α	Inositol-requiring transmembrane kinase/endoribonuclease 1α
IRES	Internal Ribosome Entry Site
ITAF	IRES-Trans activating factor
JNK	Jun N-terminal kinases
kDa	kilo Dalton
KD	Knockdown
KIRC	Kidney Renal Clear Cell Carcinoma
LGG	Low Grade Glioma
LIHC	Liver Hepatocellular Carcinoma
lncRNA	Long noncoding RNA
LSM11	U7 snRNA-associated Sm-like protein LSm11
MAGOH	Protein mago nashi homolog
MAPK	Mitogen-activated protein kinase
MEFs	Mouse embryonic fibroblasts
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein
miRNA	MicroRNA
NCBP3	Nuclear cap-binding protein subunit 3

NCBI	National Center for Biotechnology Information
NELF	Negative elongation factor
NES	Nuclear export signal
NES	Normalized enrichment score
NEXT	Nuclear exosome targeting complex
NLS	Nuclear localization signal
NMD	Nonsense-mediated decay
nt	Nucleotide
NXF1	Nuclear RNA export factor 1
NXT1	NTF2-related export protein 1
ORF	Open reading frame
pA	Polyadenylation
PAAD	Pancreatic Adenocarcinoma
PABP	Poly (A) binding protein
PAP	Poly (A) polymerase
PAS	Polyadenylation signal
PAXT	Poly (A) tail exosome targeting
PERK/ EIF2AK3	Eukaryotic translation initiation factor 2 alpha kinase
PBS	Phosphate buffered saline
PFA	Paraformaldehyde
PHAX	Phosphorylated adaptor for RNA export

PI3K	Phosphoinositide 3-kinase
PNRC2	Pro-rich nuclear receptor co-activator 2
PP2A/ PPP2R1A	Serine/threonine-protein phosphatase 2A
Pre-miRNA	Precursor microRNA
Pri-miRNA	Primary microRNA
PROMPTs	Promoter upstream transcripts
PTBP1	Polypyrimidine tract-binding protein 1
PTC	Premature termination codon
RDH	Replication-dependent histone
RIDD	Regulated IRE-dependent decay
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNAPII	RNA polymerase II
RNP	Ribonucleoprotein
RNPS1	RNA-binding protein with serine-rich domain 1
ROS	Reactive oxygen species
RRM	RNA recognition motif
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
SAINT	Significance analysis of Interactome
SD	Standard deviation
SE	SERRATE

SEM	Standard error of the mean
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
SLBP	Stem-loop binding protein
SMG	Suppressor with morphological effect on genitalia
SN	Supernatant
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
snoRNA	Small nucleolar RNA
SODs	Superoxide dismutases
STRING	Search Tool for Retrieval of Interacting Genes/Proteins
SRSF	Serine/arginine-rich splicing factor
ssRNA	Single stranded RNA
SURF	Surveillance complex
TBP	TATA box-binding protein
TBST	Tris-buffered saline-Tween 20 (0.5%)
TCGA	The Cancer Genome Atlas
TRAF2	TNF receptor associated factor 2
TREX	Transcription export
uORF	Upstream open reading frame
UPF/RENT (1, 2, 3A/B)	Regulator of nonsense transcripts (1, 2, 3A/B)

UPR	Unfolded protein response
UTR	Untranslated region
WCL	Whole cell lysate
XBP1	X-box-binding protein 1
XPO5	Exportin 5
XRN1/2	5'-3' exoribonuclease 1/2
ZC3H4	Zinc finger CCCH domain-containing protein 4
ZC3H18	Zinc finger CCCH domain-containing protein 18
ZCCHC8	Zinc finger CCHC domain-containing protein 8
ZFC3H1	Zinc finger C3H1 domain-containing protein
ZnF	Zinc finger

Acknowledgements

To my supervisor: Dr. Perry Howard. Thank you for your support during all this time, it has been essential for my development as a scientist. I've learned a lot from you, and I've really enjoyed the process. I still don't know how we managed to build such a great relationship despite cultural differences, language barriers, personalities and COVID, but I'm glad we did, and I'll take it with me. Thanks for all the advice (in the lab and career wise), for the challenges, the patience, and the grammar corrections. Thank you for being a great supervisor and person, for believing in the project and in me.

To Dr. Chris Nelson. Thank you for all your help and support. You have been more a co-supervisor than a committee member. Thanks for the lab and career advice, the reagents, the paper revisions, and the hypothesis-breaking questions that always pushed the project in the right direction. Thank you for liking ARS2/Pir2 as much as Perry and I do, and being a great teammate in the development of this project. Thank you for bringing Fargus into the equation to remind me that nothing, not even science, is more important than food.

To my supervisory committee members: Dr. Robert Chow and Dr. John Burke. Your guidance over the years has been very useful for my research. I really appreciate the time and interest you have invested in my project.

To the BCMB department. Dr. Juan Ausio: for the long and passionate talks about science and life. It has been a truly, unexpected pleasure that I look forward to everyday. To Ryan Erdman: thanks for being an incredible problem solver and a true friend. To Dr. Lisa Reynolds, thanks you for all the advice and discussions. You and Stephen were like older friends (not too old), that guided my way through this PhD. Besides, you are the proud parents of the cutest therapy dog in the world. Mr. Murphy kept us sane and active for all these years. To BCMB department members: Cindy, Kat, Ladan, Manjinder, Richard, Alloysius, Ayman, Melinda, Janice, Adrienne and all the others I'm forgetting right now. Thanks for the talks, words of support and jokes.

To las chicas: Tara Brosschot and Aikaterini Tavri. Tarita, you have been a great inspiration of what a scientist could and should be. Katia, you are a true role model for everybody around you, you are fearless, passionate and loving. Chicas, I'm so glad that we shared this journey together, every downfall and success have been 10 times sweeter because you were there. To the gang here in Victoria: Rad, Arash, Milad, Somayeh. Thanks for the beers, the talks and the trips. To Jim and Susy, thank you for celebrating every small victory and make me feel at home. To my dear friends in Cuba and around the world: Gretchen, Nayi, Chini, Darien, Darel, Giselle, Alejo, Luisi, Kerlys and super Pichy. You have been my support and inspiration for more than 10 years, in the lab and outside of it.

To my family, the one that raised me and the one that found me. Everything I'm it is significantly associated to your influence. All my OCDs and flaws are definitely your fault. On the bright side, you are such a good reason to live for.

Finally to the three most important things in my life. Thank you for being always there, being imperfect, smelly, annoying and exactly what I needed.

Dedication

A mi familia, la que me tocó y la que me ha encontrado. A los que empezaron conmigo este viaje hace 13 años, a los que siguen conmigo y a los que no. Gracias por tanto amor. Nos vemos en la próxima aventura. Hasta la Victoria Siempre.

Chapter 1: Introduction

RNA polymerase II (RNAPII) transcribes DNA into several types of RNAs including, messenger RNA (mRNA), microRNA (miRNA), replication-dependent histone (RDH) mRNA, small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA)[1–3]. All of these RNA families contain a 7-methylguanosine cap (m7G) at the 5'-end, which is deposited by the capping enzyme complex (CEC), during the NELF-E/DSIF-mediated pausing of the polymerase shortly after the initiation of transcription[4,5]. Once capping has occurred the cap binding complex (CBC), composed by CBC20/CBC80, is recruited to the 5'cap. Interaction between the CBC and several multi-protein complexes promotes the processing, degradation and export of capped RNAs. Indeed, the CBC is critical during splicing, 3'-end formation, transcription termination, exosomal degradation, intranuclear transport and export to the cytoplasm[6–14]. How the CBC complex recognizes different classes of RNAPII transcripts and connects them with the correct processing machinery is not fully understood.

However, it is clear that the life of a RNAPII transcript is highly coordinated from its beginnings during transcriptional initiation to its inevitable degradation, and ARS2 has emerged as a crucial component of this coordination. ARS2 binds directly to CBP80/CBP20 and capped RNA to form the CBCA complex[13–15], and promotes the recruitment, assembly and disassembly of transient RNA-protein complexes. Interestingly, most of these transient complexes are mutually exclusive in their interaction with ARS2. Indeed, the formation of mutually exclusive complexes between CBCA and downstream processing machinery is a major theme in how RNAPII transcripts are properly processed and regulated. Since my thesis is focussed on understanding how ARS2 functions in both nuclear and cytoplasmic RNA metabolism, I will discuss ARS2 structure and will illustrate through examples how CBC bound ARS2, through the formation of mutually exclusive complexes, regulates RNA metabolism[12–18].

1.1 ARS2 structure

The first crystal structure of the ARS2 orthologue SERRATE was described in *Arabidopsis thaliana*[19]. In this study, Machida et al demonstrated that ARS2 core consisted of 3 domains: N-terminal, Mid and zinc finger (ZnF), arranged in a walking man-like conformation (Fig. 1A)[19]. The N and C-termini parts of the protein are unstructured. Mutagenesis of ARS2 domains showed that the unstructured N-terminus, C-terminus, and the ZnF domain mediated the interaction between ARS2, pri-miRNA and DICER-LIKE 1 (DCL1), promoting the nuclear cleavage of pri-miRNA during microRNA biogenesis in plants[19,20].

Based on the reported structure of ARS2 in *A. thaliana*, the Howard lab, modelled the structure of mouse ARS2 and identified two new domains: an N-terminal DUF3546 domain and an RRM domain, predicting that the core of metazoan ARS2 is formed by 4 domains: DUF3546, RRM, Mid and ZnF domains (Fig. 1B)[18]. Metazoans ARS2 also contains a putative nuclear localization signal (NLS), a nuclear export signal (NES) and an unstructured proline rich C-terminus (Fig. 1D)[18]. Systematic mutagenesis of these domains showed that the DUF3546 and ZnF domains are required for the association to miRNAs and RDH mRNA. The ZnF and RRM domains mediate ARS2 binding to FLASH (FLICE-associated huge protein), an essential component of histone pre-mRNA processing. The Mid domain is required for interaction with DROSHA and miRNA biogenesis while the unstructured C-terminus, specifically amino acids 743-875, binds to the CBC complex and promotes RDH mRNA interaction and miRNA biogenesis[18]. RNA and protein interactions for each domain are represented on Fig. 1D.

In 2018, the crystal structure of the folded core of human ARS2, corresponding to residues 147-270 and 408-763, was determined[21]. Similar to the walking man-like conformation, described for SERRATE, the human ARS2 core has a large helical body with N-terminal and C-terminal legs (Fig. 1C). In the human ARS2 structure, the N-terminal DUF3546 domain is referred as N-terminal leg, while the Mid domain corresponds to the large helical body, onto which the RRM domain is packed. Finally, the ZnF domain was renamed as C-terminal leg since it is missing one of the 2 conserved cysteines within the C2H2 motif and Zinc was not found in the structure[18,21]. The C-

terminal leg ends in a 'foot'-like structure that contains a basic patch. A second basic patch is formed by a β-sheet within the RRM domain[21]. Analysis of RNA and protein binding to the described domains, showed that the basic patch on the RRM domain is critical for ARS2 binding to ssRNA (Fig. 2E)[21].

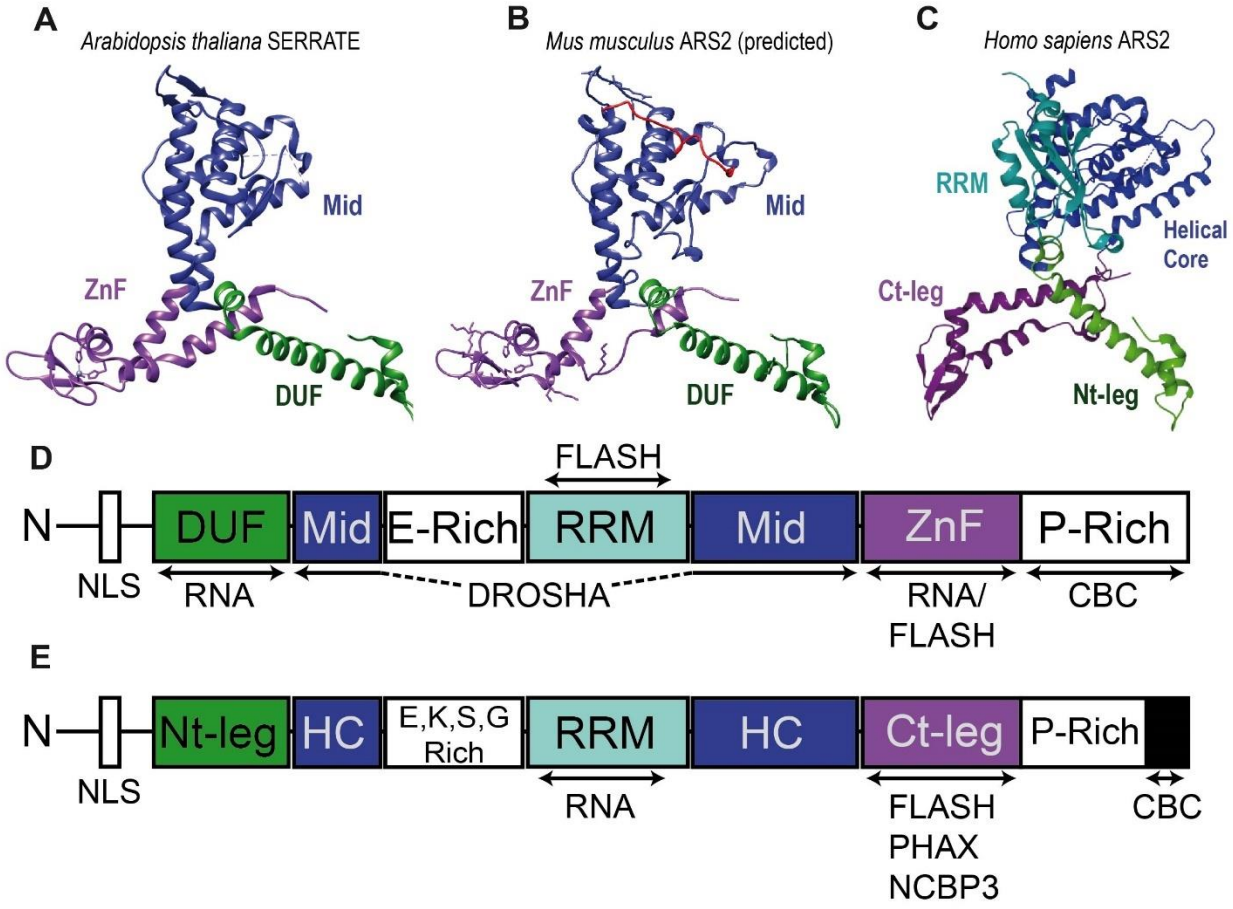


Figure 1: Structure of ARS2. A), C) Crystal structure ARS2 in *Arabidopsis thaliana* (PDB accession 3AX1) **(A)**[19], and *Homo sapiens* **(C)** (PDB accession 6F7S)[21]. **B)** Predicted ARS2 structure in *Mus musculus* (without the glutamate rich and RRM domain insert), modeled using *A. thaliana* ARS2 as a reference[18]. **D), E)** Schematic representation of ARS2 domains, showing their different nomenclature and reported protein and RNA interactions. **D)** As reported in Perry Howard's lab[18]. **E)** As reported in Stephen Cusack's lab[21]. DUF/Nt-leg domain is highlighted in green, Mid/HC domain in blue, RRM domain in teal, ZnF/Ct-leg domain in purple and CBC binding region in black.

The structural/function analysis of human ARS2 confirmed that the ZnF domain (C-terminal leg) mediated ARS2 interaction with FLASH, and it was also involved in the interaction with PHAX and NCBP3, but not with the CBC complex[21]. ARS2 binding to the

CBC complex was mapped to amino acids (845-871) within the unstructured C-terminus[22]. The existence of overlapping binding sites for FLASH, NCBP3, and PHAX within the C-terminal leg may provide a mechanism for the recruitment of mutually exclusive RNA-processing machineries, to the adjacent CBC. The formation of CBC–ARS2–NCBP3 or CBC–ARS2–PHAX complexes is thought to sort mRNAs or snRNA into different export pathways. Similarly, CBC-ARS2-FLASH complex formation directs RDH mRNAs into a RDH specific maturation pathway[21]. These results highlight the critical role of ARS2 as a platform that connects the CBC complex and capped RNAs with the correct processing machinery.

1.2 ARS2 role in RNA metabolism

ARS2, as part of the CBCA complex, works as a nexus between capped RNAPII transcripts and their correct processing machinery. In this way, ARS2 regulates the processing, degradation and export of quite dissimilar RNA families: mRNA, miRNA, snRNA, snoRNA and RDH mRNAs[12–16,18,21–23]. In this section I will explain in detail how this regulation takes place and when known, illustrate the formation of mutually exclusive complexes that determine RNA fate.

1.2.1 miRNA biogenesis

The canonical pathway of miRNA biogenesis in mammals starts with the transcription of primary miRNA (pri-miRNA) by RNAPII. pri-miRNAs are processed in the nucleus into precursor miRNA (pre-miRNAs) by the microprocessor complex, formed by DGCR8 (DiGeorge Syndrome Critical Region 8) and DROSHA[24,25]. pre-miRNAs are exported to the cytoplasm by exportin 5 (XPO5)/ RanGTP complex and processed by the RNAse III endonuclease Dicer[25,26]. Dicer processing generates a mature miRNA duplex, from which one of the strands is loaded onto a Ago-dependent RNA-induced silencing complex (RISC)[27].

In mammals ARS2 interacts with DROSHA in nucleus and influences DROSHA efficiency and specificity[15] (Fig. 2). ARS2 is required for the stability and efficient processing of pri-miRNA to pre-miRNA by the microprocessor complex, and deletion of ARS2 is sufficient to reduce pri-miRNA processing and mature miRNA levels[15]. Interestingly in *D. melanogaster*, ARS2 has nuclear and cytoplasmic roles in RNA interference. miRNA biogenesis in *Drosophila* is very similar to mammals, involving the microprocessor complex in nucleus, Dicer processing in the cytoplasm and RISC formation[25,28–30]. *Drosophila* ARS2, like mammalian ARS2, interacts with the CBC and microprocessor complex in nucleus, and is required for pri-miRNA processing and stability[31]. Insects also use RNA interference to control RNA viruses, as part of their innate immune response[32]. In *Drosophila,* endogenous or exogenous double stranded RNA (dsRNA) are processed by Dicer-2 in the cytoplasm and incorporated to Ago2-dependent RISC, where they lead the degradation of complementary mRNAs or viral RNAs[33]. *Drosophila* ARS2 interacts with cytoplasmic Dicer-2, modulating its activity in vitro. Furthermore, dARS2 is required for siRNA mediated RNA silencing, and deletion of ARS2 increases susceptibility to RNA viruses[31]. ARS2 role in cytoplasmic RNA processing in *Drosophila*, it's the only function described for ARS2 in metazoans that potentially occurs outside the nucleus.

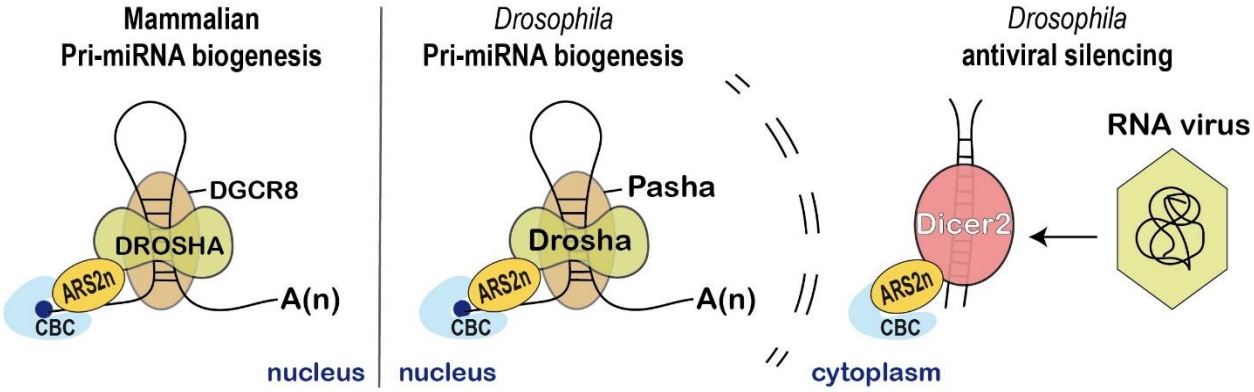


Figure 2: ARS2 role in miRNA biogenesis. In mammals and *Drosophila melanogaster*, ARS2 interacts with DROSHA in nucleus, and it's required for efficient processing of pri-miRNA by the Microprocessor. Drosophila ARS2, additionally interacts with Dicer2 in the cytoplasm and it's required for siRNA mediated RNA silencing.

1.2.2 Replication dependent histone (RDH) mRNA processing

RDH mRNAs are capped, intronless, RNAPII transcripts with conserved stem-loop at their 3' ends[34]. RDH mRNA processing involves an endonucleolytic cleavage between the conserved 3'-end stem-loop and a histone downstream element (HDE), mediated by CPSF73 (cleavage and polyadenylation specificity factor)[35–37]. RDH mRNA processing is cap-dependent and requires the coordination of several multimeric complexes such as: SLBP (stem-loop binding protein), U7 snRNA (part of the multi-subunit U7 small nuclear ribonucleoprotein (snRNP) complex) and histone pre-mRNA cleavage complex (HCC), formed by CPSF73, CPSF100 and symplekin[38–42]. NELF-E (negative elongation factor), is also required for RDH pre-mRNA processing at the 3'-end[10]. Misprocessing of RDH – transcripts results in aberrant read-through and polyadenylation[12,43].

The CBC interacts with SLBP and NELF-E[10,11], while ARS2 interacts with RDH mRNAs and FLASH[12,16] (Fig. 3). FLASH directly binds to the U7 snRNP component LSM11 and is required for RDH pre-mRNA endonucleolytic cleavage [44–47]. Deletion of either ARS2 or FLASH disrupts the formation of histone locus bodies (HLBs), increases aberrant polyadenylated histone transcripts and decreases histone protein levels, resulting in a delayed S-phase cell cycle progression[12,16,18,48]. The binding of ARS2 to FLASH, occurs through ARS2 C-terminal leg, and impedes ARS2 association to NCBP3 or PHAX, involved in mRNA and snRNA export, respectively[21]. CBC-bound ARS2 binds to RDH mRNAs and FLASH, creating a complex that excludes other processing pathways and allows the correct and specific processing of this RNA family.

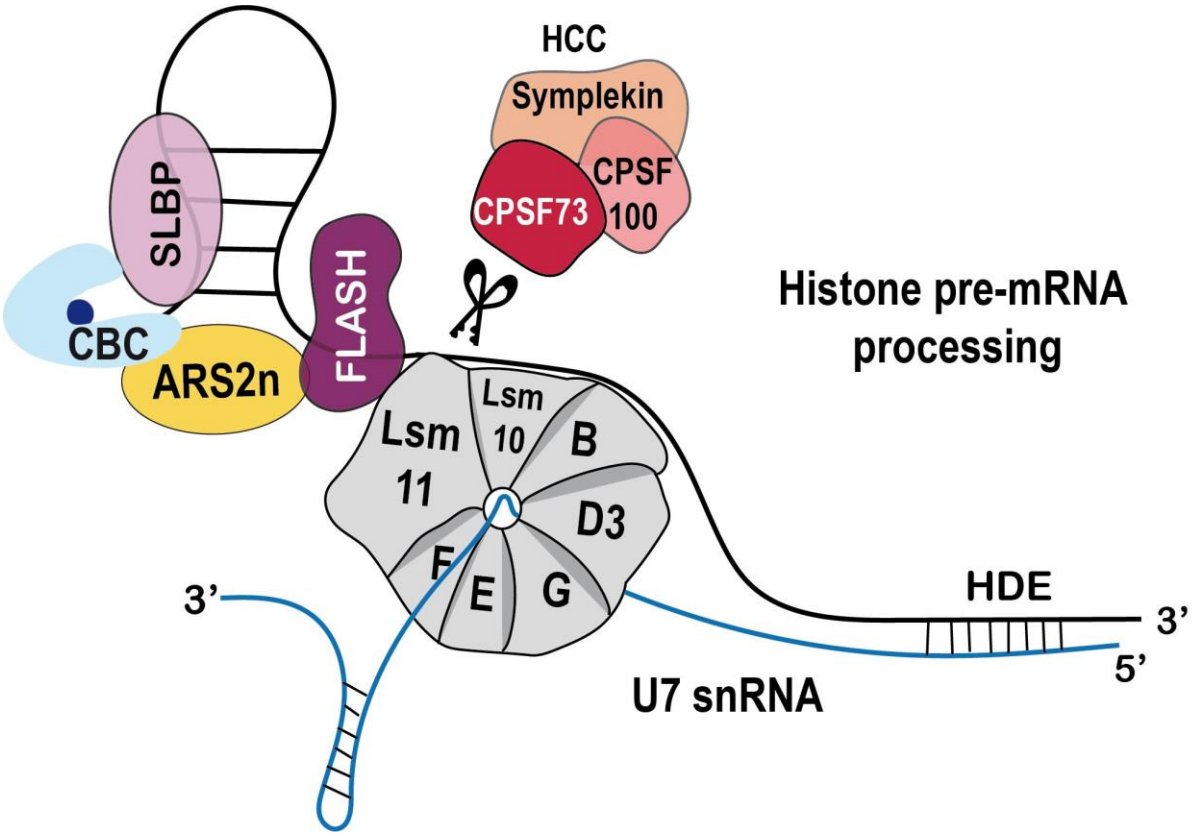


Figure 3: ARS2 role in RDH mRNA processing. ARS2 bound to the CBC, interacts with RDH mRNAs and FLASH, while the CBC interacts with NELF and SLBP. FLASH and the U7 snRNP are required for the endonucleolytic cleavage of RDH pre-mRNAs, mediated by the HCC complex component CPSF73.

1.2.3 Splicing

Pre-mRNA splicing is a two-step transesterification reaction that involves intron removal and exon-exon ligation. This process is carried out by the spliceosome complex, composed by five snRNPs (U1, 2, 4-6), along with many associated factors[49]. Detailed reviews on splicing can be found here[49–51].

CBC and ARS2 are required for cap-proximal splicing and alternative splicing in *Arabidopsis*, preferentially affecting the selection of 5' splice sites and retention of the first intron[11,52,53]. However, it is still unknown how these processes take place. Interestingly, a recent study demonstrated that ARS2 regulates the transcription of intronless genes in *Arabidopsis*[54]. ARS2, potentially through its interaction with the CBC complex, binds to

intronless genes and promotes their association with RNAPII complexes during transcriptional pausing and elongation.

In mammals, the CBC complex interacts with numerous components of the spliceosome, including U4/U5.U6 snRNPs during co-transcriptional spliceosome assembly. CBC is required for cap-proximal splicing and alternative splicing, linking pre-mRNA capping to transcription elongation and splicing[55–58]. Although mammalian ARS2 co-purifies with multiple splicing factors[14,21,59], the direct role of ARS2 in splicing hasn't been established for metazoans.

1.2.4 mRNA 3'end formation

3'-end processing of mRNAs takes places in a two-step process involving endonucleolytic cleavage followed by polyadenylation. Efficient mRNA 3' end formation is ensured by a network of interacting proteins that link 3' end formation to other steps of mRNA biosynthesis and processing[60,61]. Endonucleolytic cleavage and polyadenylation of pre-mRNA requires four multisubunit complexes CPSF, CstF, CFIm and CFIIm, as well as poly (A) polymerase (PAP)[61,62]. The cleavage and polyadenylation machinery are recruited co-transcriptionally through the CTD of RNAPII. The first step in mRNA 3' end processing involves the binding of CPSF and CstF at the polyadenylation signal (PAS). CPSF binds to the central hexameric AAUAAA/AUUAAA, while CstF binds to the U/GU-rich downstream sequence element (DSE)[61,63–68]. CFIIm bridges RNAPII and the nascent transcript to CPSF and CFIm[69], and is required for pre-mRNA cleavage by CPSF73[60]. Following cleavage and PAP mediated polyadenylation, the poly (A) binding protein (PABPN1) binds and stabilizes the processed mRNA for nuclear export. PABPN1 is exchanged in the cytoplasm by PABPC, which promotes translation and RNA stability[61].

CBCA interacts with pre-mRNA 3' end processing machinery and is required for efficient cleavage of short pre-mRNAs[9,13]. ARS2 interacts with CFIIm component CLP1, which binds to the RNAPII CTD-binding protein PCF11 and bridges CPSF with CFIm (Fig. 4). Depletion of either ARS2, CLP1 or PCF11, induces 3' read-through of luciferase reporters, supporting the functionality of ARS2-CLP1 interaction[13,70].

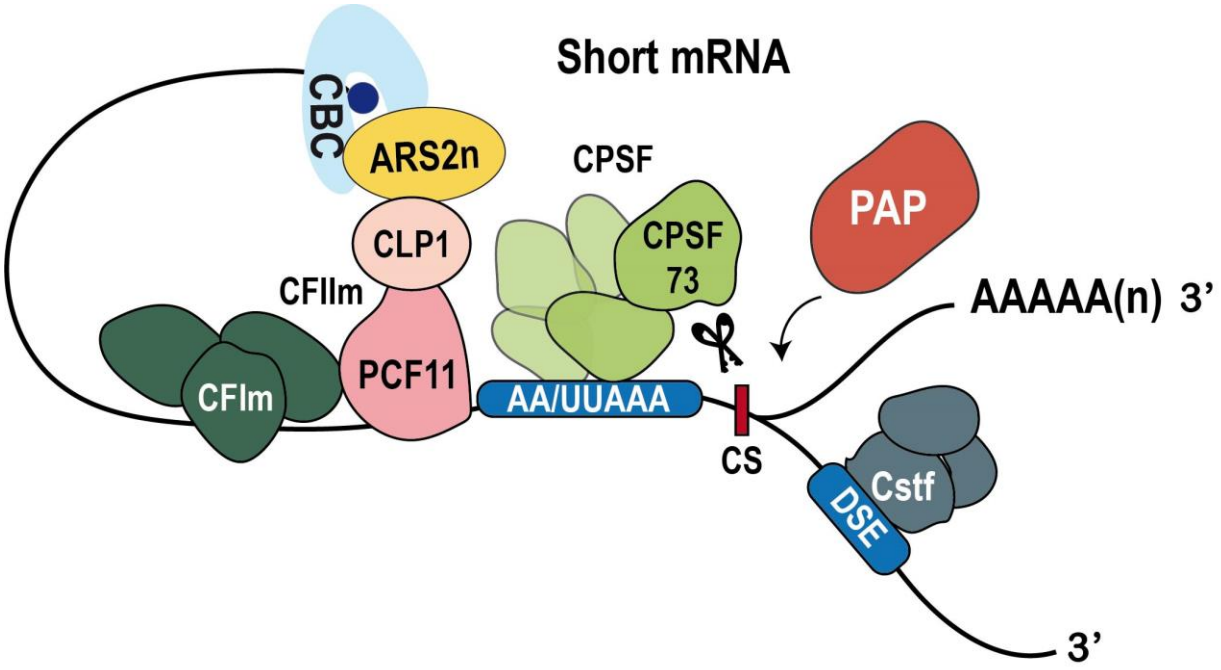


Figure 4: ARS2 role in mRNA 3' end processing. ARS2 interacts with CFIIm component CLP1 and it's required for CPSF73-mediated cleavage of short mRNAs and PROMPTs. The four multisubunit complexes CPSF, CstF, CFIm and CFIIm, and PAP are represented in the figure. CS: Cleavage site, DSE: Downstream sequence element.

1.2.5 snRNA 3' end processing

snRNA 3'-end formation is mediated by the Integrator complex[71]. NELF-E connects the capped CBC-bound snRNA to the Integrator, promoting 3'-end processing[10]. It was initially thought that ARS2 was also involved in this step of the process. However, is now known that NELF-E and ARS2 compete for the binding site at the interface between CBP20 and CBP80 and a CBC-ARS2-NELF-E complex does not occur[22]. Once NELF is substituted by ARS2 and the CBCA complex is formed, PHAX or the NEXT complex can be recruited to promote nuclear export or degradation, respectively[22].

1.2.6 Exosomal degradation

A major player in nuclear RNA processing and degradation is the RNA exosome. Human exosome consists of a nine-subunit large inert core that associates with active ribonucleases: RRP6 (EXOSC10) and RRP44 (DIS3), which provide the catalytic activity[72–74]. Besides these enzymatic co-factors, RNA binding adaptor proteins are

required to recruit the nuclear exosome to its many targets[73]. An essential co-factor of the nuclear exosome is the DExH/D box RNA helicase MTR4 (SKIV2L2). MTR4 unwind RNA substrates to facilitate their access to the central channel of the exosome core, and associates with specific RNA binding adaptor proteins (ZCCHC8 or ZFC3H1), providing RNA target specificity[75,76]. MTR4 associates with ZCCHC8 and RBM7 to form the nuclear exosome targeting (NEXT) complex, which promotes the degradation of rather short and immature RNAs such as: promoter upstream transcripts (PROMPTS), enhancer RNAs (eRNAs) and 3'extended snRNAs and snoRNAs[14,77–80]. On the other hand, MTR4 association with ZFC3H1, favors the poly (A) tail exosome targeting (PAXT) connection, which targets polyadenylated transcripts[78].

Through its C-terminal leg, ARS2 binds to ZC3H18 and MTR4, connecting the capped, processed, CBC bound transcripts to exosomal decay via NEXT or PAXT[14,78] (Fig. 5). Interestingly, ARS2 binding to ZC3H18 impedes ARS2-PHAX interaction and the formation of the CBCA-PHAX complex, that promotes the intranuclear transport of snoRNAs and nuclear export of snRNAs[13,23]. Alternatively, MTR4 competes with the mRNA export adaptor ALYREF for binding to ARS2, and this competition is critical for determining whether an mRNA is degraded or exported to the cytoplasm[81,82].

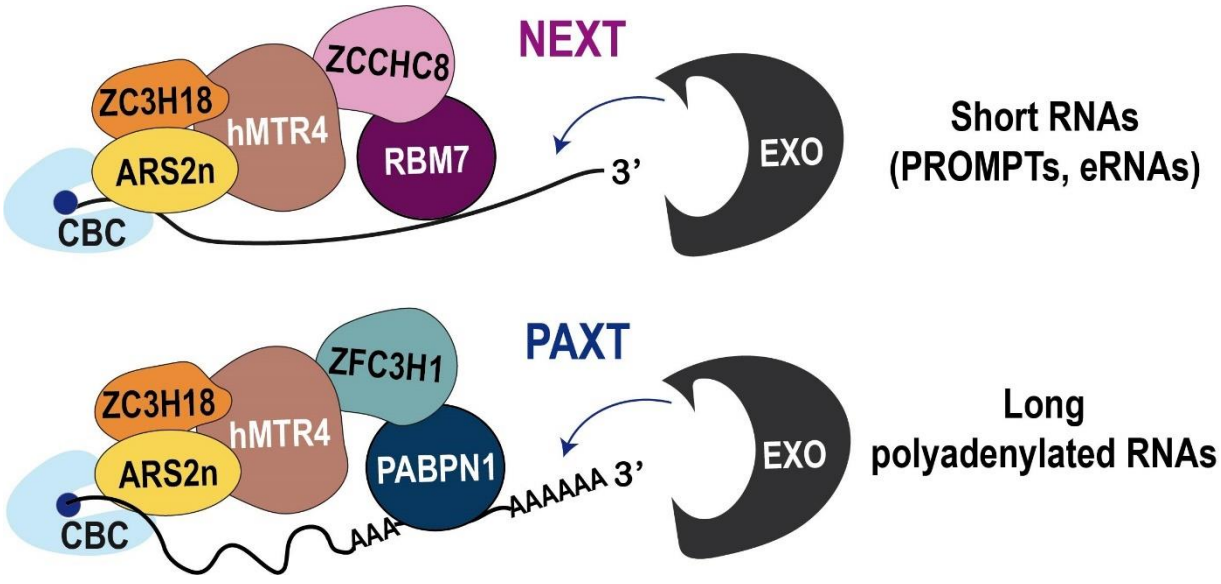


Figure 5: ARS2 role in exosomal degradation. ARS2 binds to ZC3H18 and MTR4, connecting the capped, CBC-bound transcripts to exosomal decay via NEXT or PAXT. Short and immature

RNAs such as PROMPTS, 3'extended snRNAs and snoRNAs are predominantly degraded through NEXT, while PAXT targets polyadenylated transcripts. Adapted from[78].

1.2.7 RNA export

snRNA export in metazoans is mediated by the CBCAP complex, formed by CBP20, CBP80, ARS2 and PHAX[13]. ARS2 interacts with PHAX though its C-terminal leg and promotes the binding of PHAX to the CBC[22,83] (Fig. 6). ARS2 binding to PHAX impedes ZC3H18 association to the CBC, favoring snRNA export rather than exosomal degradation[23]. Formation of the CBCAP complex is followed by the recruitment of the export receptor CRM1/XPO1 and RanGTP, which leads to snRNA export[84,85]. Interestingly PHAX also competes with hnRNP C for the binding to the CBCA complex. hnRNP C tetramer directly interacts with CBP80, binds to transcripts longer than 200-300nt and inhibits PHAX binding[86]. Association of the CBCA complex to either PHAX or hnRNP C, ensures that U snRNAs (typically < 200nt), are exported via PHAX-CRM1-RanGTP and excludes longer RNA transcripts from using this export mechanism[86,87].

mRNA export is largely mediated by the multi-subunit transcription export (TREX) complex constituted, in humans, by the THO complex (THOC1, THOC2, THOC5, THOC6, THOC7, Tex1), CIP29, UAP56 and ALY/REF[88–90]. The TREX complex is recruited to the mRNA co-transcriptionally, stimulated by the presence an exon junction complex (EJC), deposited on exon-exon junctions after splicing[81,90–93]. Binding of TREX to the mRNA is followed by the interaction between ALY/REF, THOC5 and the export receptor TAP/NXF1, that transports the mRNA through the nuclear pore[94–97].

CBCA interacts with several components of the TREX complex and such interactions are required for efficient mRNA export[98,99]. ALY/REF is recruited to the 5' end of capped mRNA by the CBC, while ARS2 additionally interacts with THOC2, UAP56 and CIP29[89,98,100,101]. Interestingly a newly described protein called NCBP3, binds to CBC-ARS2 and interacts with the EJC and TREX complexes (Fig. 6). NCBP3 aids the TREX complex in mRNA export and contributes to licensing polyadenylated RNPs for nuclear export[102,103]. Binding of NCBP3 to CBCA, prevents the recruitment of ZC3H18 or PHAX,

and the formation of CBCA-ZC3H18 or CBCA-PHAX complexes respectively[102]. In this way, CBCA-NCBP3 bound spliced mRNAs are distinguished from snRNAs (CBCA-PHAX) or potentially defective mRNAs (CBCA-ZC3H18) and are exported to the cytoplasm.

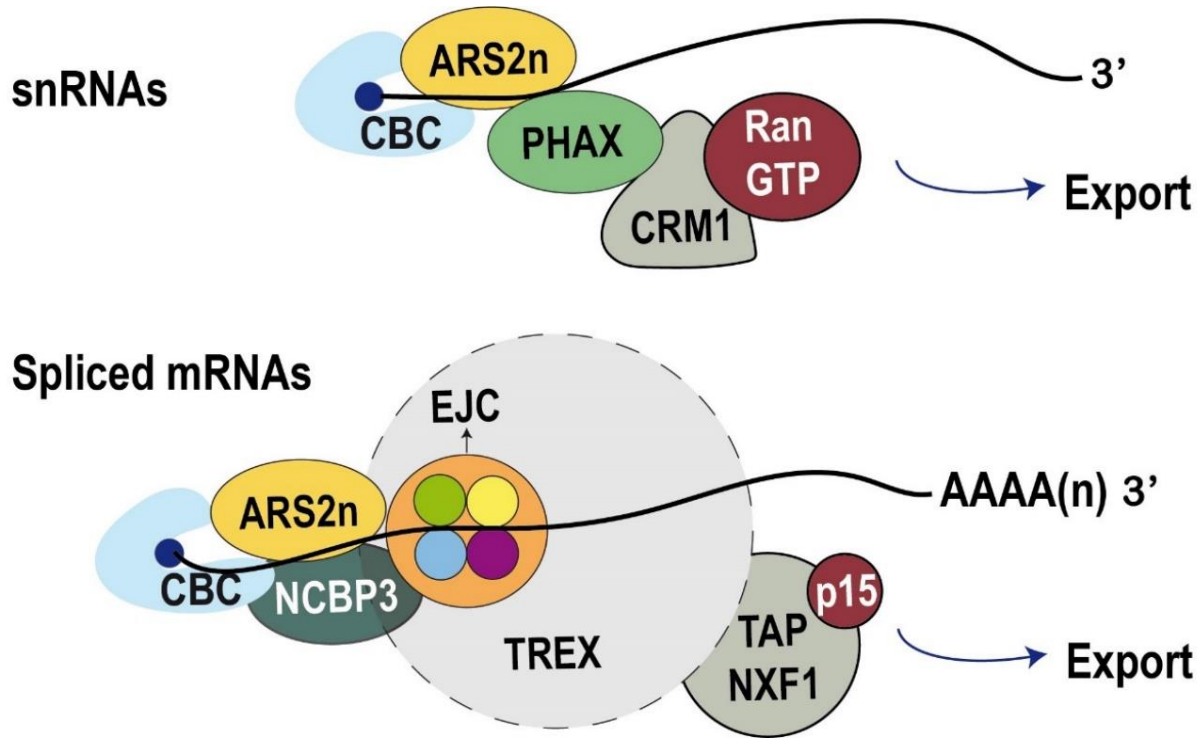


Figure 6: ARS2 and export. CBC bound ARS2, interacts with PHAX to form the CBCAP complex. Recruitment of CRM1/XPO1 and RanGTP, leads to snRNA export. mRNA export is mediated by the TREX complex, which interacts with the export receptor TAP/NXF1 to transport mRNA through the nuclear pore. NCBP3 binds to CBC-ARS2 and interacts with the EJC and TREX complexes, favoring the export of capped-spliced mRNAs.

1.2.8 ARS2 regulates RNA metabolism through the formation of mutually exclusive complexes

Capped RNAPII transcripts have different processing requirements. How the CBC complex recognizes different classes of RNAs and connect them with the correct processing machinery is still under study, however a key element in this process is ARS2.

Since the CBC complex is highly specific for the m7-G cap but not for individual RNA sequences, an important question in the field is how different classes of capped RNAs are distinguished[7,23,104,105]. ARS2, and other CBC effectors like PHAX and ZC3H18 have broad target specificities and interact with a large range of capped RNAs such as snRNAs, PROMPTS, RDH mRNAs and mRNAs[23]. Therefore, a simple sequence recognition model, in which CBC effectors recognize specific classes of RNAs and regulate their processing, can't be used to explain the RNA family-specific activities of the CBC complex[23]. The current model proposes the existence of short-lived, mutually exclusive CBC-containing complexes that determine RNA fate by reacting to molecular cues imposed at specific times[23]. As part of this model, ribonucleoprotein (RNP) complexes are continuously remodeled, and transcript fate is the result of "locking" decisive complexes at particular checkpoints during RNA processing. The model proposes that the RNAPII "does not know" which type of transcription unit is engaged until relevant cues are instigated[23]. For example, the appearance of a 3'-end stem-loop and a histone downstream element (HDE), identifies RDH mRNAs and promotes their processing through CBC-SLBP, CBC-NELF and CBC-ARS2-FLASH interactions[10,12,14,16,18]. Another example is the recognition of the splicing sequence on mRNAs, which leads to splicing and EJC deposition, favoring NCBP3 binding to CBC-ARS2 and export to the cytoplasm[93,102,103]. Termination signals are another example of an RNA processing checkpoint[23]. Cryptic, cap-proximal 3'-end termination signals (found in PROMPTS) promote the interaction of CBC-ARS2 with the 3' end NEXT complex, via ARS2-ZC3H18 interaction. ARS2-ZC3H18 binding stabilizes CBCAN complex, while impedes ARS2-PHAX binding and formation of the CBCAP complex, leading to RNA exosomal degradation (CBCAN) rather than export (CBCAP)[14,23,70]. In contrast, an snRNA 3'-end processing sequence favors the recruitment of the Integrator complex[71], ARS2-PHAX interaction and CBCAP complex formation[13], promoting export rather than ARS2-ZC3H18 interaction and exosomal degradation[23].

ARS2, through its binding to the CBC complex, acts as a dynamic platform that recruits protein machineries to regulate RNA fate (Fig. 7). The C-terminal leg of ARS2 mediates the binding to FLASH[18,21], PHAX[21,83], NCBP3[21], ZC3H18[83] and MTR4[83] promoting the formation of CBCA-FLASH, CBCAP, CBCA-NCBP3 and CBCAN complexes, that favor

RDH processing, snRNA export, mRNA export and RNA degradation, respectively[12–16,18,21–23]. ARS2 is a versatile effector of the CBC, and through the formation of mutually exclusive complexes, plays an active role in the regulation of RNA metabolism.

1.2.9 Unexplained functions of ARS2.

1.2.9.1 Cytoplasmic processing of mRNAs

As described, ARS2 plays an important scaffolding role within the CBC by helping to synchronize the dynamic assembly and disassembly of RNA-protein complexes that regulate mRNA splicing, degradation, and export[12–16,18,21–23]. Once the CBC-ARS2 bound mRNAs are exported to the cytoplasm, ARS2 and the CBC are recycled back to nucleus[15]. However, the CBC is not just a carrier for mRNAs to the cytoplasm, it is also involved in cytoplasmic RNA metabolism. For example, the CBC is required for the pioneering round of translation and for promoting transcript degradation through nonsense mediated decay (NMD)[106–108]. Although ARS2 is translocated to the cytoplasm, it is unknown whether ARS2 participates in CBC-related functions in the cytoplasm.

1.2.9.2 ARS2 role in arsenic sensitivity

Another unanswered question of ARS2 functions relates to its known role in the cellular response to arsenic. The initial study of ARS2 showed that overexpression conferred resistance to arsenic[109]. However, the cDNA used in this study encoded only a small portion of the extreme C-terminus of full-length ARS2, and it was subsequently found through knockdown studies that full length ARS2 conferred arsenic sensitivity instead[15]. Arsenic is not only a widespread environmental contaminant, but is also used to treat certain cancers such as acute promyelocytic leukemia (APL)[110]. Despite the significant role of ARS2 in arsenic sensitivity, the mechanistic details of this effect are still unknown and haven’t been related with ARS2 role in RNA metabolism.

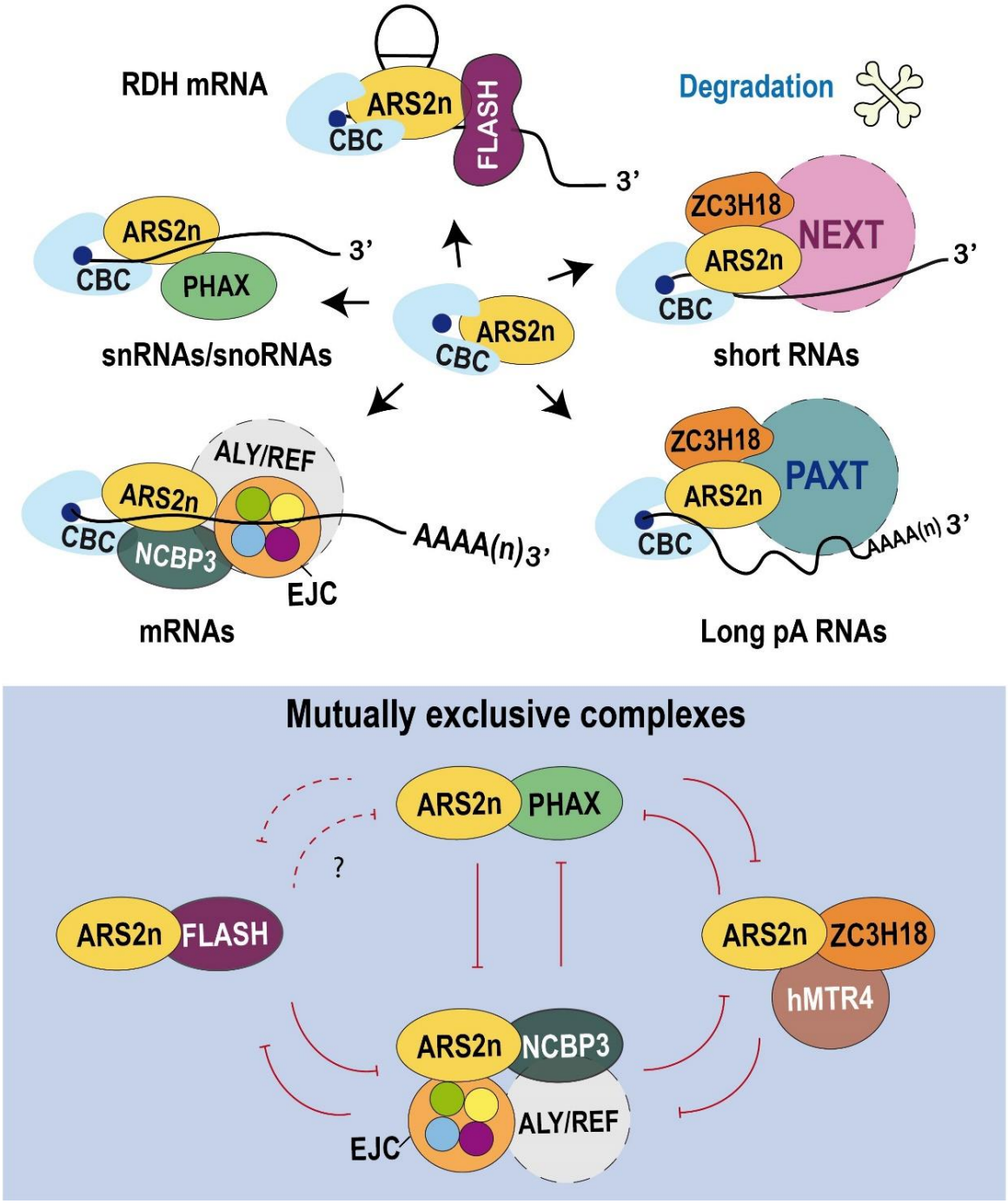


Figure 7: Mutually exclusive CBC-ARS2 complexes. ARS2, through its binding to the CBC complex acts as a dynamic platform that promotes the formation of mutually exclusive complexes that determine RNA fate. The C-terminal leg of ARS2 mediates the binding to FLASH, PHAX, NCBP3, ZC3H18 and MTR4 promoting the formation of the mutually exclusive complexes CBCA-FLASH, CBCAP, CBCA-NCBP3 and CBCAN that favor RDH processing, snRNA export, mRNA export and RNA degradation, respectively.

1.3 A layer of complexity: *Ars2* isoforms

In the original characterization of the *Ars2* gene[111], northern blot analysis of mouse tissues showed a predominant transcript size of ~3kb, which corresponded to the full-length mouse cDNA sequence (BC066831), and codified the now well studied ARS2 protein[111]. However, additional weaker bands (~2kb and 1kb) were observed with a 3' probe. This result, in combination with the unexpectedly high degree of genomic conservation observed in the intron 5 of the *Ars2* gene (Fig. 8A)[111,112], suggested the existence of *Ars2* isoforms produced from an internal promoter or alternative splicing.

To examine the possibility of *Ars2* isoforms in mouse cells, Jennifer Christie, a previous master student in the lab, performed RT-PCR using a forward primer located within the intron 5 and reverse primers located in exon 6 or exon 7 (Fig. 8B)[112]. Product bands corresponding in size to spliced mRNA that retained intron 5, were obtained, demonstrating that an isoform containing the conserved intron 5 was indeed transcribed in mouse cells (Fig. 8C)[112]. To amplify the full sequence of *Ars2* isoform, Jennifer used the same intron 5 forward primer with an additional reverse primer (*Ars2*_R3'UTR) that was complementary to the 3'UTR of canonical *Ars2* mRNA (Fig. 8B). Two bands of approximately 700bp and 2.1kb were obtained (Fig. 8D)[112]. After sequencing, it was demonstrated that the 2.1kb transcript contained a single ORF and encoded a predicted protein that was identical to amino acids 228-875 of canonical ARS2.

At the time, it was known that canonical (full-length) ARS2 was a nuclear protein, and that the nuclear localization signal was located on the first 200 amino acids of the protein[18]. Therefore, the absence of the first 227aa of canonical ARS2 in the intron 5 containing isoform, suggested a cytoplasmic localization. To test this hypothesis, the cDNA of the intron 5 containing isoform was fused in-frame to GFP and compared in localization to canonical ARS2. As expected, canonical ARS2 localized predominantly on the nucleus, while intron 5 containing isoform was enriched in the cytoplasm (Fig. 8E)[112]. For simplicity, from now on canonical ARS2 will be referred as nuclear ARS2 (ARS2n), and the intron 5 containing isoform will be called cytoplasmic ARS2 (ARS2c).

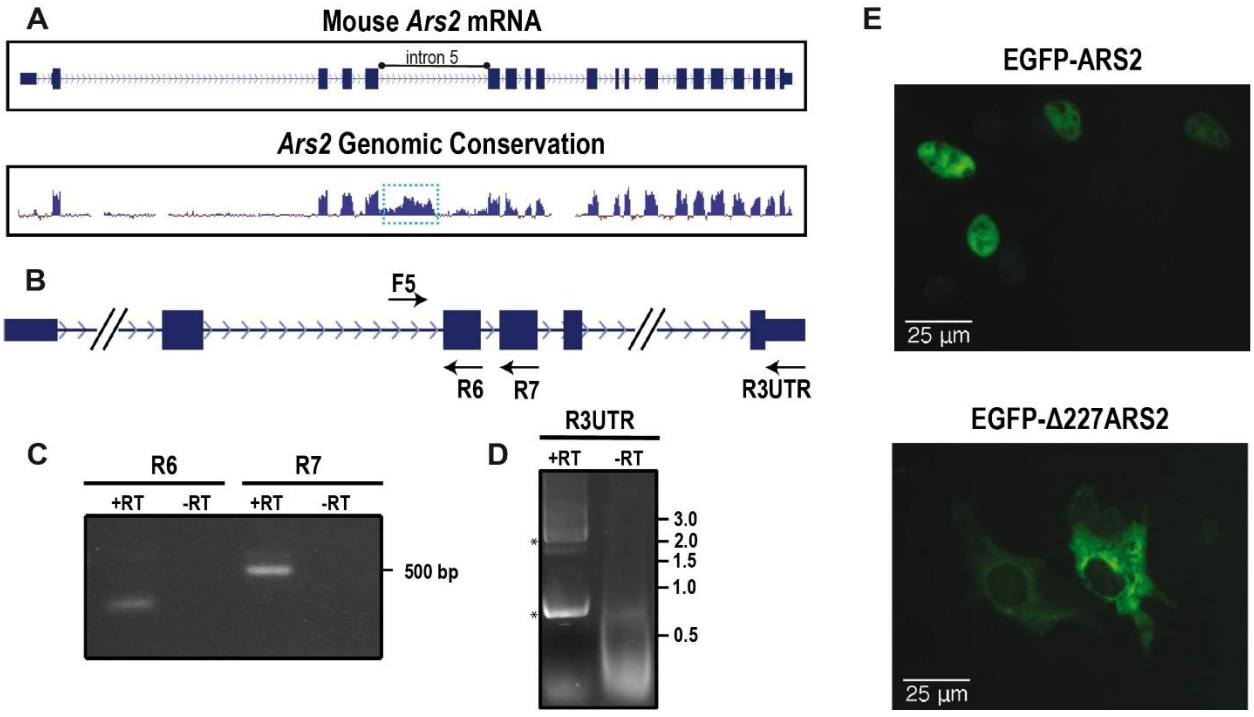


Figure 8: Intron 5 containing ARS2 isoform. A) Conservation analysis between mouse *Ars2* and additional 39 placental mammals. The alignments were generated using Multiz and UCSC/Penn State Bioinformatics comparative genomics alignment pipeline. Evolutionary conservation was measured using PhyloP. Dotted box indicates region in intron 5 with high conservation. **B)** Diagram of primers used in **C)** and **D)** to amplify intron 5 **(C)** or the full coding sequence of the potential *Ars2* isoform **(D)**, in C_2C_{12} cells. Asterisks indicate bands that were excised and sent for sequencing. Sequenced data is included in[112]. **E)** C_2C_{12} myoblasts transfected with EGFP-ARS2 or EGFP-Δ227-ARS2 and imaged live. Adapted from Jennifer Christie's thesis[112]. Predicted amplicon sizes: R6: 268pb, R7: 428bp (expected unspliced pre-mRNA: 520bp), R3UTR: 2192 (expected unspliced pre-mRNA: 4732bp).

Further characterization of ARS2 isoforms is necessary to understand their biological roles. However, based on the functions of the CBC in cytoplasmic RNA metabolism, the potential existence ARS2 isoforms with different locations, suggest a role for ARS2 in both nuclear and cytoplasmic processes. Moreover, cytoplasmic ARS2 isoforms could have unique, distinct functions from nuclear ARS2, including previously unexplained or undescribed functions like ARS2 potential role in NMD or arsenic sensitivity. My thesis focussed on understanding the role of ARS2 isoforms in NMD and arsenic sensitivity.

Therefore, I will discuss the mechanisms of NMD pathway and cellular response to arsenic stress in the next sections.

1.4 Nonsense mediated decay

Nonsense mediated decay (NMD) is a translation-dependent RNA decay pathway that targets mRNA containing premature termination codons (PTC) within the ORF[113,114]. PTCs can arise from spontaneous mutations, faulty transcription or splicing errors, and degradation of PTC containing mRNAs by NMD, prevents the synthesis of truncated, potentially harmful proteins[115–120] (Fig. 9A). In addition to its quality control function, NMD also serves a regulatory role by degrading a significant fraction of mutation-free physiological transcripts[121–123] (Fig. 9B). In these cases, PTCs are generated due to naturally occurring events like alternative splicing and alternative translation (uORF)[118]. Furthermore, normal mRNAs with long 3' untranslated regions (3'UTRs) can also be targeted by NMD[118,124,125], although 3'UTR sequence composition, structure, and associated proteins contribute to their fate[115,118]. In mammals, it is estimated 10% of the transcriptome is normally regulated by NMD and these NMD-regulated error-free transcripts are often positioned at central nodes of biochemical pathways, and may be used to tune pathway activity[118,126–129]. Due to the important regulatory role of NMD, NMD factors are essential for embryonic development in vertebrates[130,131], maintenance of hematopoietic stem and progenitor cells[130], maturation of T cells[130], stress response[127,132,133], as well as for liver development, function and regeneration in mice[134].

NMD is evolutionary conserved and has acquired an increased importance from yeast to vertebrates. *Saccharomyces cerevisiae* and *Caenorhabditis elegans* are viable without a functional NMD pathway[135,136], but vertebrates are not[130,131,137]. The gain of essential function over the course of evolution underscores the expansion of NMD functions from a quality control or surveillance mechanism to a more sophisticated regulatory mechanism[115].

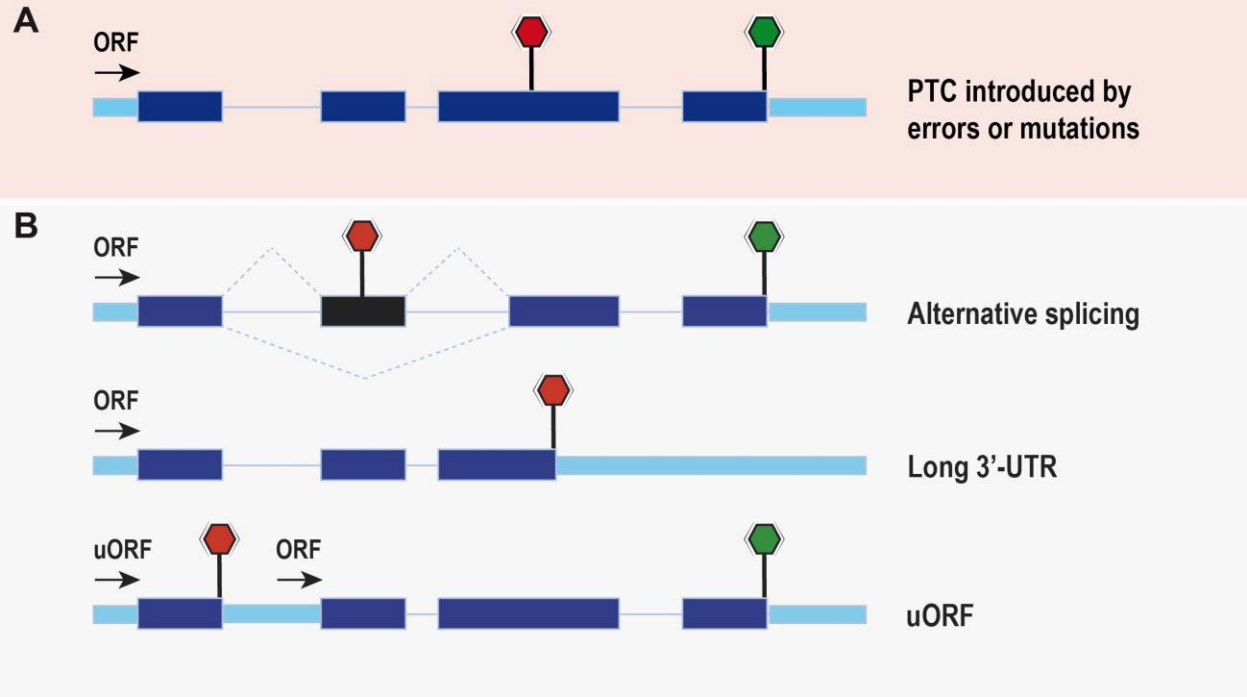


Figure 9: NMD targets erroneous and mutations-free physiological transcripts. **A)** In erroneous transcripts, PTCs (red) can arise from mutations in the ORF, faulty transcription or aberrant splicing. **B)** In mutations-free physiological transcripts, PTCs are generated due naturally occurring events like alternative splicing and alternative translation (uORFs). In all these cases, the presence of an EJC >55nt downstream of the PTC, constitutes a strong signal for transcript degradation through NMD. In addition, long 3'UTR (>1000nt in mammalian cells) can be targeted by NMD. The spatial distance between the terminating ribosome and the mRNA termini makes translation termination inefficient and allows the interaction of UPF1- eRF3 and the recruitment of NMD factors. Adapted from[115].

1.4.1 NMD components

NMD components and complexity varies from yeast to vertebrates. This document will focus on NMD in mammals, only referring to other groups of organisms for comparisons.

1.4.1.1 Exon junction complex (EJC)

The EJC is not a NMD component *per se*, but it is a key element on the major branch of the NMD pathway in mammals, the EJC-dependent NMD. The EJC is a multiprotein complex organized around a central core of four proteins: eIF4A3, MAGOH, Y14 (also

known as RBM8A) and CASC3 (also known as MLN51), which are recruited and assembled on the spliced exon junctions by the splicing machinery[138–140]. Once locked onto exon junctions, the EJC complexes remain on the mRNA until translation, where they are removed by the ribosome[140]. The composition of the EJC complexes is very dynamic as EJC peripheral factors transiently interact with the EJC core during the transit of the mRNPs from nucleus to cytoplasm[140]. The EJC peripheral factors include splicing factors like RNPS1, ACINUS, PININ, SAP18 and mRNA export factors such as UAP56, Aly/REF, NXF1, NXT1[140,141]. NMD factors UPF3A, UPF3B, UPF2, UPF1 and SMG6 also join the EJC after splicing in the nucleus or in the cytoplasm[140]. UPF3B binds to the composite binding site formed by eIF4A3, MAGOH and Y14[142–144]. Once in the cytoplasm, UPF2 joins the EJC through direct binding to UPF3B[93,143,145]. UPF1, which is an essential NMD factor, joins the complex only in the context of NMD[145], while SMG6 is mainly cytoplasmic and probably binds the EJC after mRNA export[146]. Although the remodelling of EJC and the mRNPs in the transition from nucleus to cytoplasm is not fully understood, it is known that some peripheral factors are restricted to nucleus (PININ and ACINUS) or cytoplasm (CASC3), and some factors cannot coexist in the same EJC (UPF3A, UPF3B and SMG6)[140].

By recruiting several factors involved in splicing, transport, translation and NMD, the EJC modulates these post-transcriptional processes and establishes a molecular link between nuclear splicing and cytoplasmic NMD[93,140,147]. The role of the EJC in NMD is one of the best documented function of the complex[140], and in mammals, most of NMD transcript degradation is EJC dependent[115].

1.4.1.2 NMD core factors: UPF1, UPF2 and UPF3

UPF1 is a monomeric helicase, highly regulated and conserved in eukaryotic evolution from fungi to humans[148]. UPF1 has RNA binding, ATPase and RNA helicase activities, which are essential for NMD[116,117,149]. UPF1 translocates on the RNA to unwind double-stranded RNA structures, remodelling the 3'UTR RNP of NMD targets[150]. The ATPase activity promotes UPF1 dissociation from non-NMD targets, mediating target

discrimination[151], and its also required for the release of UPF1 from cleaved RNA fragments during transcript degradation[152].

Recruitment of UPF1 to NMD targets during translation termination, is a crucial step in the NMD pathway. Currently, two models describe how UPF1 differentiates between transcripts going through normal translation termination vs aberrant termination that leads to NMD[115]. The first model proposes that the presence of a PTC prevents the displacement of RNA-bound UPF1 from a stretch of the coding sequence during translation, leading to UPF1 accumulation on the RNA, increased engagement of UPF1 with the termination machinery, and abnormal translation termination that eventually concludes in NMD[153–155]. The second model argues that efficient translation termination requires the interaction between PABP, bound to the poly (A) tail, and ribosome-bound release factor 3 (eRF3). The presence of a long 3'UTR disfavors PABP-eRF3 interaction and 3'UTR-bound UPF1 outcompetes PABP for eRF3 interactions, thereby affecting translation termination (Fig. 10)[124,156–160]. Nevertheless, in both models UPF1 interacts with the terminating ribosome and it's activated by UPF2 and UPF3[115]. Phosphorylation of UPF1 in several serine/threonine-glutamine motifs [(S/T) Q], is mediated by SMG1 and licenses the mRNA for degradation[161–163]. Phosphorylated UPF1 serves as a binding platform for NMD effectors (SMG5/6/7) that initiate mRNA decay and promote UPF1 dephosphorylation[107,146].

Across all eukaryotic organisms UPF1 is the only NMD factor that is essential for NMD and is a core NMD player. In vertebrates, the NMD pathway can be conceptualized as a branched network that converges at UPF1[115].

UPF2 is conserved across eukaryotes[148]. UPF2 interacts with UPF1 through its C-terminal domain and activates UPF1 ATPase and helicase activities[143,164]. UPF2 functions as a bridge between UPF1 and UPF3, forming the UPF1-UPF2-UPF3 complex[115]. Interestingly UPF2 can also interact directly with eRF3, although the significance of this interaction for NMD remains unclear[115,165].

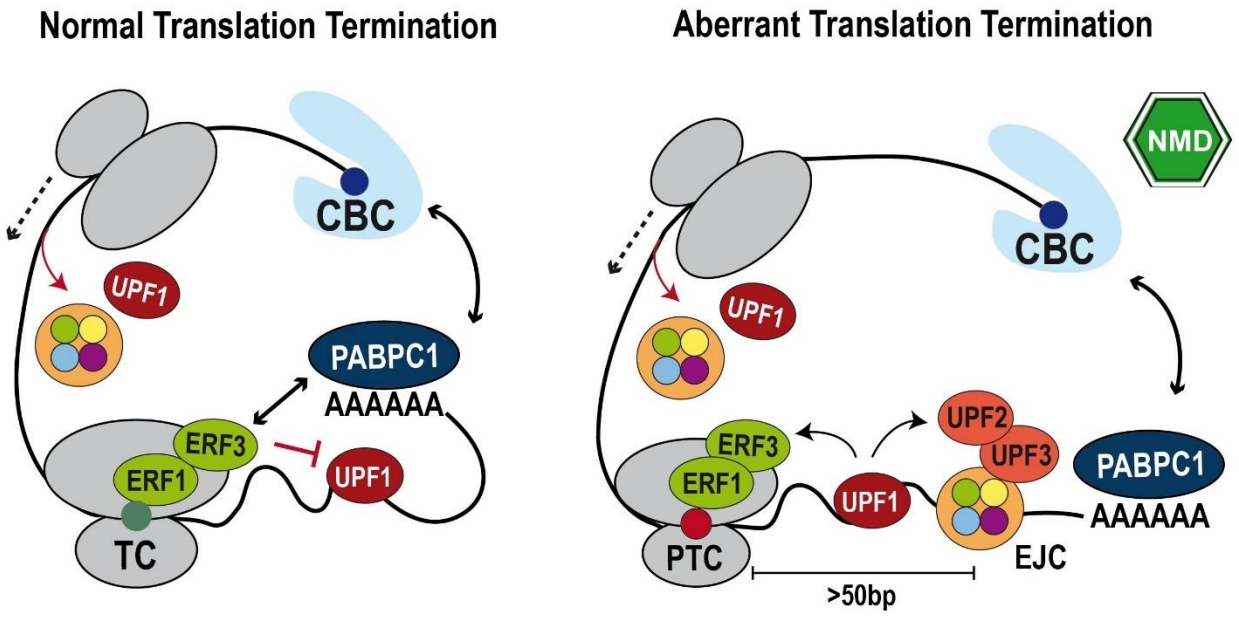


Figure 10: Normal vs aberrant translation termination. UPF1 and EJCs bound to the mRNA are displaced by translating ribosomes. During normal translation termination PABPC1-eRF3 interaction is dominant and inhibits UPF1-eRF3 interaction. The presence of a long 3'UTR and/or EJC downstream of the termination codon decreases PABPC1-eRF3 interaction and stabilizes UPF1-eRF3 and UPF1-EJC interaction, thus promoting transcript degradation through NMD.

UPF3 is the last NMD core factor and it's the least conserved of the three[148]. Mammalian genomes encode two UPF3 paralogs, UPF3A and UPF3B. Both UPF3A/B are capable of binding UPF2 and the EJC, connecting UPF proteins to the EJC and linking translation termination with the downstream EJC[115,142,144]. In human cells, UPF3B provides the major NMD-activating functions, while UPF3A acts as a weaker activator[166,167]. Furthermore, UPF3A can behave as a NMD suppressor, potentially sequestering UPF2 from UPF3B and the NMD[166]. Like UPF2, UPF3B stimulates UPF1's ATPase and helicase activity and can interact directly with eRF3, the latter with unknown consequences for NMD[143,168].

1.4.1.3 SMGs proteins: SMG1/8/9 and SMG5/6/7

SMG1 is a PIK3-like kinase that forms a complex with SMG8 and SMG9, which regulate the activity and substrate specificity of SMG1[169–173]. SMG1 binds to UPF1 on the stalled ribosome, where UPF1 binding to UPF2 and UPF3B, stimulates the phosphorylation of

UPF1 by SMG1[115,140,145]. Phosphorylation of UPF1 on several serine/threonine-glutamine motifs [(S/T) Q] by SMG1 provides a binding platform for SMG6 and SMG5/7, recruiting them to NMD targets[174–176].

UPF1 phosphorylation marks the beginning of the mRNA degradation phase, mainly mediated by SMG6 and SMG5/SMG7. SMG5/6/7 bind to phosphorylated UPF1 but can also interact with UPF1 in a phosphorylation independent manner[175,177]. SMG6 has an endonuclease activity and cleaves the mRNA at the PTC, stimulated by the presence of an EJC[178,179]. SMG5/SMG7 are recruited to UPF1 as a heterodimer. SMG7-CNOT8 interaction connects the target mRNA with the CCR4/NOT deadenylation complex, which in turn recruits the decapping complex DCP1/DCP2 and 5-3 exonuclease XRN1[180,181]. Initial reports suggested that SMG6 and SMG5/7 work in a partially redundant and independent manner[115,182]. However, the current view is that they have intertwined functions in the NMD pathway and SMG6 function depends, at some degree, on SMG5/SMG7[115].

1.4.1.4 NMD activators and inhibitors

Several reports suggest that canonical SR proteins can enhance NMD[115]. Except for SRSF2, all canonical SR proteins interact with the EJC[183], specifically the RNPS1-EJC[184], and can act in concert with the EJC to promote NMD[115]. Additionally, SRSF1 interacts with UPF1 in nucleus and cytoplasm, thus enhancing NMD independently of the EJC or even UPF2 and UPF3B[185]. SRSF1 potentially enhances UPF1 activity by recruiting phosphatase PP2A[185], which dephosphorylates UPF1. Overexpression of SRSF1 can change the location of NMD and the pioneer round of translation from cytoplasm to nucleus-associated, which suggest another potential mechanism of NMD regulation by enhancing translation[115,186].

HNRNP factors such as PTBP1 and HNRNPL bind in the vicinity of the stop codon and inhibit NMD[187,188]. Binding of PTBP1 close to stop codons, promotes UPF1 dissociation from target mRNAs, inhibiting NMD[189]. Similarly, HNRNPL binding to 3'UTR can override NMD induction even by downstream EJCs[188]. Interestingly, mRNAs with reduced UPF1 binding, have 3'UTRs with an increased density of binding sites for other HNRNP factors,

suggesting that this NMD-protective mechanism is a widespread function of HNRNP proteins[115,188]. Since HNRNPs are well-known regulators of cell and tissue specific mRNA processing and translation, it is possible that they may also regulate NMD in context-dependent manner[115,190].

1.4.2 NMD mechanism

When a ribosome encounters a termination codon during translation termination, release factors eRF1 and eRF3 recognize the codon and promote polypeptide release to accomplish termination[115]. Mammalian-cell NMD generally occurs when translation terminates more than 50–55 nucleotides upstream of an EJC complex[191]. In this scenario, the termination codon is recognized as a PTC, which triggers RNP recruitment and remodelling. First, UPF1 is recruited to the stalled ribosome by eRF1, eRF3 and CBP80. Subsequently, SMG1 kinase joins the complex to generate the SURF complex (SMG1–UPF1– ERF1–ERF3). Next, in a step dependent on DHX34 and CBP80, UPF1 interacts with the downstream UPF2/3 within the EJC to form the decay-inducing (DECID) complex, triggering SMG1 activation. The phosphorylation of UPF1 by SMG1 signals the 'point of no return' in the NMD process, which leads to translation inhibition and mRNA degradation through the recruitment of SMG6, SMG5–SMG7 and mRNA decay factors[140] (Fig. 11).

Although the previously described mechanism, corresponding to the EJC-dependent NMD, is the major branch of NMD in mammals, there are other branches of the pathway which are independent of UPF2, UPF3B and the EJC. Loss-of-function genetic studies of UPF2 and UPF3B support their dispensability during NMD[192–194]. However, the precise mechanism, mRNA targets, and biological context of these NMD branches, remains to be defined[115]. The EJC-independent NMD branch, on the other hand, explains the degradation of NMD transcripts with long 3'UTRs[124]. Long 3'UTRs increase the spatial distance between the terminating ribosome and the mRNA termini, impeding the interaction between the polyadenylate-binding protein (PABPC1) and the terminating ribosome, which leads to inefficient translation termination. In this context, UPF1 unspecifically bound to the 3'UTR, interacts with eRF3 in the terminating ribosome and

subsequently recruits additional NMD factors to promote NMD[115]. However, not all long 3'UTRs trigger NMD and 3'UTR sequence composition, structure and associated proteins can override NMD degradation[115,118]. Although the EJC-independent degradation of NMD transcripts is not as common as the EJC dependent branch in mammals, well known physiological NMD targets contain long 3'UTRs, including those involved in the cellular response to stress.

1.4.3 The role of the CBC complex in NMD

Mammalian NMD occurs both during the pioneer round of translation, where transcripts are associated with the CBC complex, and once the CBC has been replaced by the cytoplasmic cap-binding protein EIF4E[195–197]. Indeed, a recent study in human cells using single-molecule microscopy of NMD reporter RNAs, showed that each ribosome terminating translation at a PTC has an equal probability of inducing NMD, thus confirming the hypothesis that NMD transcripts could be bound to CBC or EIF4E[198]. However, since the CBC complex is the major interacting partner of ARS2, I will focus on the NMD mechanistic description that involves this complex.

The CBC component CBP80, promotes NMD at two distinct steps: First CBP80 favors the formation of the SURF complex at the PTC by increasing the association of SMG1 and UPF1 with the two eukaryotic release factors (eRF1/eRF3). Secondly, CBP80 promotes the subsequent association of SMG1 and UPF1 with the downstream EJC complex[108](Fig. 11). Although several studies have reported the interaction between ARS2 and components of NMD pathway, including the four core components of the EJC complex: eIF4A3, MAGOH, RBM8A and CASC3[21,59,102], the question of whether ARS2 participates in NMD has not been addressed. The existence of cytoplasmic ARS2 isoforms that could additionally modulate this pathway makes this area of research intriguing.

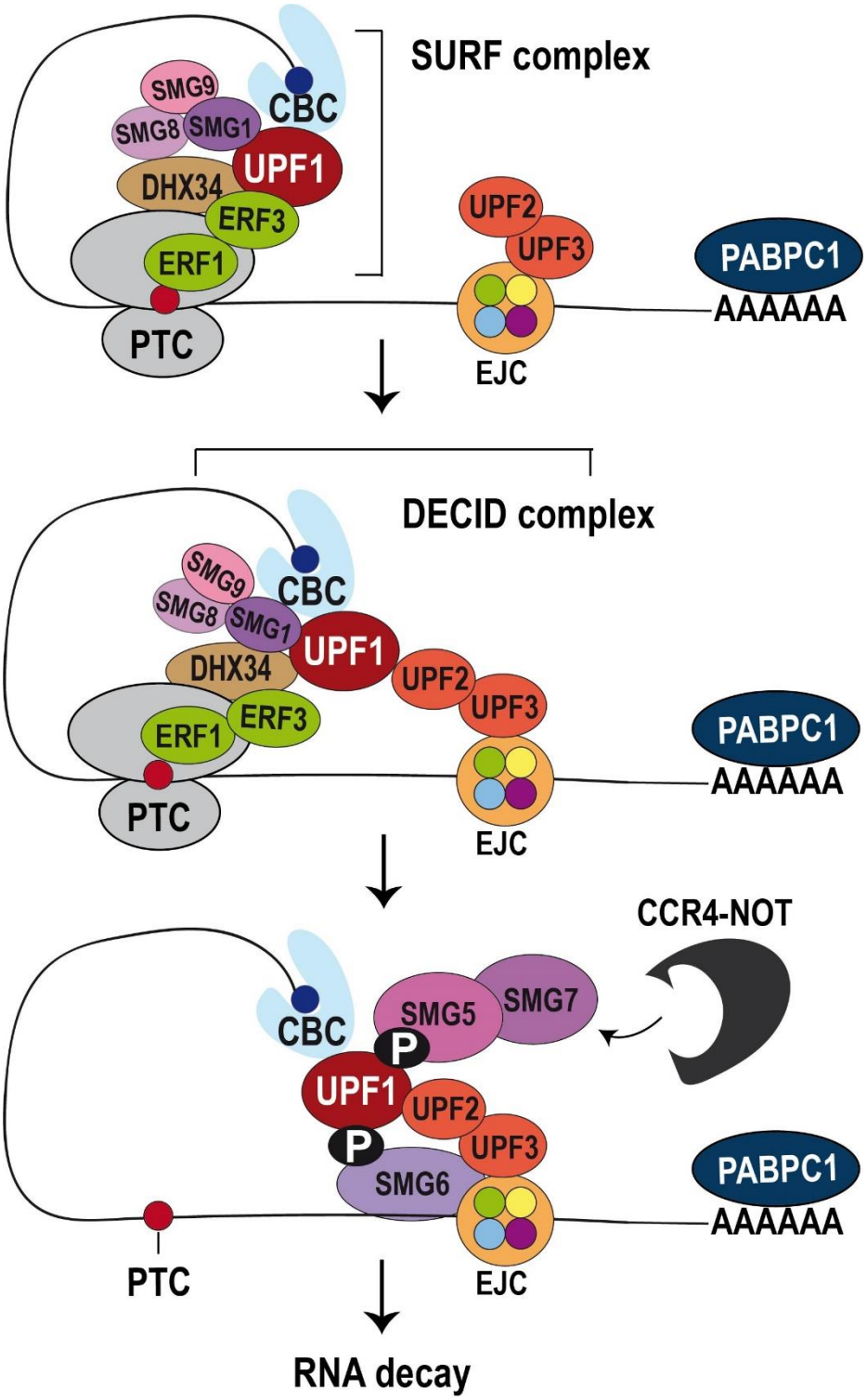


Figure 11: Mechanism of NMD. A) A Ribosome stalled at the PTC induces the formation of the SURF complex (eRF1-eRF3-UPF1-SMG1), accompanied by DHX34, CBC, SMG8 and SMG9. An EJC, containing UPF2 and UPF3 is located more than 50bp downstream of the PTC. **B)** Interaction of UPF1 with UPF2-UPF3, induces the remodelling of the SURF complex to form the DECID complex and activation of SMG1. **C)** Phosphorylation of UPF1 by SMG1 induces mRNP remodelling, and the recruitment of SMG6 and SMG5-SMG7. **D)** SMG6 cleaves the RNA in the vicinity of the PTC. SMG5-SMG7 recruit the CCR4–NOT deadenylase complex that mediates mRNA deadenylation. Decapping complex (DCPC) binds via Pro-rich nuclear receptor co-activator 2 (PNRC2) and mediates decapping. Endonucleolytic cleavage, decapping and deadenylation are followed by complete mRNA degradation by general cellular 5´–3´ and 3´–5´exonucleolytic activities (not shown). Adapted from[140].

1.4.4 NMD and the stress response.

As previously discussed, the NMD pathway is more than a quality control mechanism, it has a critical regulatory role in several biological processes. One of the best studied processes regulated by NMD is the unfolded protein response (UPR). UPR is activated in response to ER (endoplasmic reticulum) stress and it's tightly regulated so that cells can respond to stress and recover homeostasis, or if the stress cannot be resolved in a timely manner, cell death is triggered.

Recent studies have demonstrated that NMD plays a critical role in the regulation of the UPR. Several mRNAs encoding UPR components are direct physiological targets of NMD. Example of these NMD targets are: IRE1α, TRAF2, FSD1L, TNRC5 and HERP, all of which have relative long 3'UTRs, and ATF3, ATF4 and PERK, which have uORFs as NMD activating features[127,132,199]. Consistent with the regulatory role of NMD in the UPR, depletion of NMD component UPF3B, induced UPR activation in response to normally innocuous levels of the ER stressor tunicamycin. Additionally, UPF3B depleted cells or UPF3b-null mouse hepatocytes exhibited a prolonged stress response when exposed to moderate doses of tunicamycin[127]. These results suggest that NMD prevents the inappropriate activation of the UPR, and promotes its timely termination to protect cells from prolonged ER stress[127].

Since NMD constantly degrades transcripts encoding UPR components and prevents the activation of the UPR, a full-magnitude UPR response requires the inhibition of NMD. During strong ER stress, activation of the UPR leads to phosphorylation of eIF2α and NMD inhibition, in a mechanism that probably involves translation inhibition and/or the sequestration of NMD factors in stress granules[132,199–201]. If the stress is resolved, global translation is recovered, NMD is restored, and mRNA encoding UPR components return to basal levels though NMD degradation. NMD modulates and is modulated by the UPR, ensuring the activation of the stress response under the correct stimulus, at the right magnitude, and resolved in a timely manner[118] (Fig. 12). NMD also regulate responses to non-ER stressors such as hypoxia, amino-acid depravation, reactive oxygen species

(ROS), autophagy, and pathogen infection[132]. However, the precise mechanisms of this regulation are not well defined.

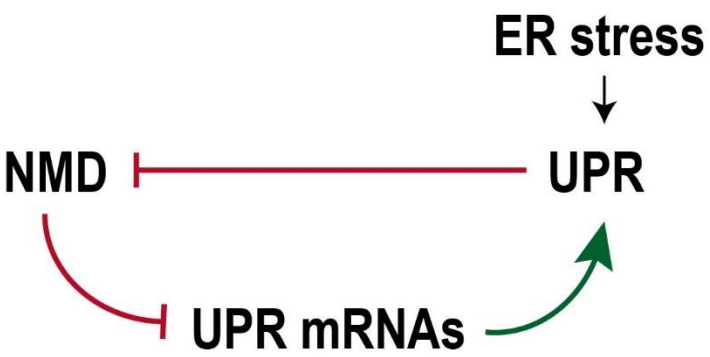


Figure 12: NMD modulates, and it's modulated by the UPR. NMD constantly degrades transcripts encoding UPR components and prevents the activation of the UPR. Strong ER stress and activation of the UPR, leads to phosphorylation of eIF2α and NMD inhibition. If the stress is resolved, global translation is recovered, NMD is restored, and mRNA encoding UPR components return to basal levels though NMD degradation.

1.5 Cellular stress response

Cellular stress responses are conserved mechanisms that protect organisms from changes in the internal or external environments and act to maintain homeostasis[202]. Depending on the type and the level of stress, different stress pathways are activated and prosurvival strategies are mounted. If the stress pathway resolves the stress and homeostasis is rescued, the cell returns to normal operations. When the stress is overly severe or cannot be resolved in a timely manner, cell death programs are elicited. Physiological cellular stressors include hypoxia, osmotic stress, pathogen-induced stress, ER stress and oxidative stress[202,203]. Since the last two are induced during arsenic exposure, for the purpose of this thesis, they will be discussed in detail in the following sections.

1.5.1 ER stress and the unfolded protein response

The UPR is a set of intracellular signalling pathways, activated in response to ER stress, and functions to re-establish cellular homeostasis. As part of this process, the UPR reduces global protein synthesis, increases ER protein-folding capacity and promotes the degradation of misfolded proteins[202,204]. Three principal branches of the UPR have been identified (Fig. 13). Each branch is defined by transmembrane ER-resident signalling

molecules: IRE1α (inositol requiring enzyme 1), ATF6 (activating transcription factor 6) and PERK (double–stranded RNA-activated protein kinase (PKR)-like kinase)[204]. The activation of the UPR sensors (IRE1α, PERK and ATF6), depends on dissociation from the ER chaperone BiP, which is sequestered from the sensors upon unfolded protein accumulation[205]. In the case of IRE1α a direct recognition model has also been proposed, where unfolded proteins bind directly to the luminal domains of IRE1α, facilitating the assembly of IRE1α clusters and IRE1α activation[202]. Arsenic exposure mainly triggers the PERK branch of the UPR[206]. However, since all three branches are deeply interconnected in the resolution of the stress or the induction of apoptosis, a brief description of each branch is included in this section.

1.5.1.1 IRE1α branch

Under ER stress, oligomerization and autophosphorylation of IRE1α, elicits its RNAse activity[207,208]. In metazoans, IRE1α excises a 26nt intron from the mRNA encoding the transcription factor X-box-binding protein 1 (XBP1), shifting its open reading frame and allowing the expression of an active XBP1 transcription factor termed XBP1s[209–211]. XBP1s induces genes involved in ER protein translocation, folding, secretion and degradation of misfolded proteins[202]. IRE1α also cleaves a small set of mRNAs or precursor miRNAs, leading to their degradation. This process is known as regulated IRE-dependent decay (RIDD), and it contributes to decrease mRNA abundance and hence protein folding load in the ER[212–214]. Finally, IRE1α associates with adapter proteins to favor the crosstalk with other stress pathways such as macroautophagy and the MAPK pathway[215].

1.5.1.2 ATF6 branch

During ER stress, a full-length ATF6 translocates from the ER to the Golgi apparatus, where it's cleaved by the proteases S1P and S2P, releasing the ATF6p50 transcription factor. ATF6p50 translocates to nucleus to induce the expression of genes encoding ER chaperones and enzymes that promote the resolution of the ER stress[202,216]. Furthermore, in conjunction with XBP1s, ATF6p50 promotes ER and Golgi biogenesis to increase the secretory capacity of the cell[202,217–219].

1.5.1.3 PERK branch

PERK kinase activation leads to the phosphorylation of the alpha subunit of eukaryotic translation factor 2 (eIF2α), which transiently attenuates protein synthesis, preventing the influx of newly synthesized proteins into the ER[220,221]. Concomitantly, phosphorylated eIF2α initiates the translation of specific mRNAs that harbour one or more upstream open reading frames (uORFs) in their 5'UTR regions[222–224]. One of these uORF-containing mRNAs encodes ATF4. The ATF4 transcript contains two uORFs in its 5'UTR region and the last uORF (uORF2) overlaps with the main ORF in a different reading frame. In unstressed cells, in which eIF2α is hypophosphorylated and eIF2-GTP is abundant, the ribosome initiates translation at uORF1 and after uORF1 translation is completed, reinitiates at uORF2. Since uORF2 overlaps with the main, functional ORF, translation initiation at the uORF2 generates a premature termination codon that leads to ATF4 transcript degradation though NMD[118]. Phosphorylation of eIF2α under ER stress, leads to a reduction in eIF2-GTP levels and a failure to reinitiate at uORF2. As a result, scanning ribosomes bypass the inhibitory uORF2 and reinitiate translation at the main ORF, producing a functional ATF4 protein [199,223,225]. ATF4 is a crucial transcription factor in the PERK branch that activates the expression of genes involved redox homoeostasis, amino acids metabolism, protein synthesis, apoptosis, and autophagy[202]. Additionally, ATF4 participates in a feedback loop to dephosphorylate eIF2α and restore protein synthesis[226,227].

1.5.1.4 Apoptosis

The UPR signalling promotes the resolution of the ER stress and cell survival. However, all three branches of the UPR induce apoptosis in the event of excessive or sustained stress, thus preventing the deleterious consequences of persistent stress (Fig. 13). The PERK and ATF6 branch induce the proapoptotic factor CHOP, which promotes ER stress-induced apoptosis by modulating GADD34, death receptor (DR5), and the proapoptotic members of the BCL-2 family including NOXA, BIM and PUMA[202,228,229]. The IRE1α branch is involved in caspase 2-dependent, caspase-8 dependent or BAX/BAK-dependent apoptosis, activated though RIDD or TRAF2-JNK signalling[202,230,231]. Finally, Ca^{2+} release from the ER, via inositol 1,4,5-triphosphate receptor (IP3R), increases

mitochondrial Ca^{2+} uptake, contributing to the release of reactive oxygen species (ROS) and the activation of the BAX/BAK-dependent apoptosome[202,232,233].

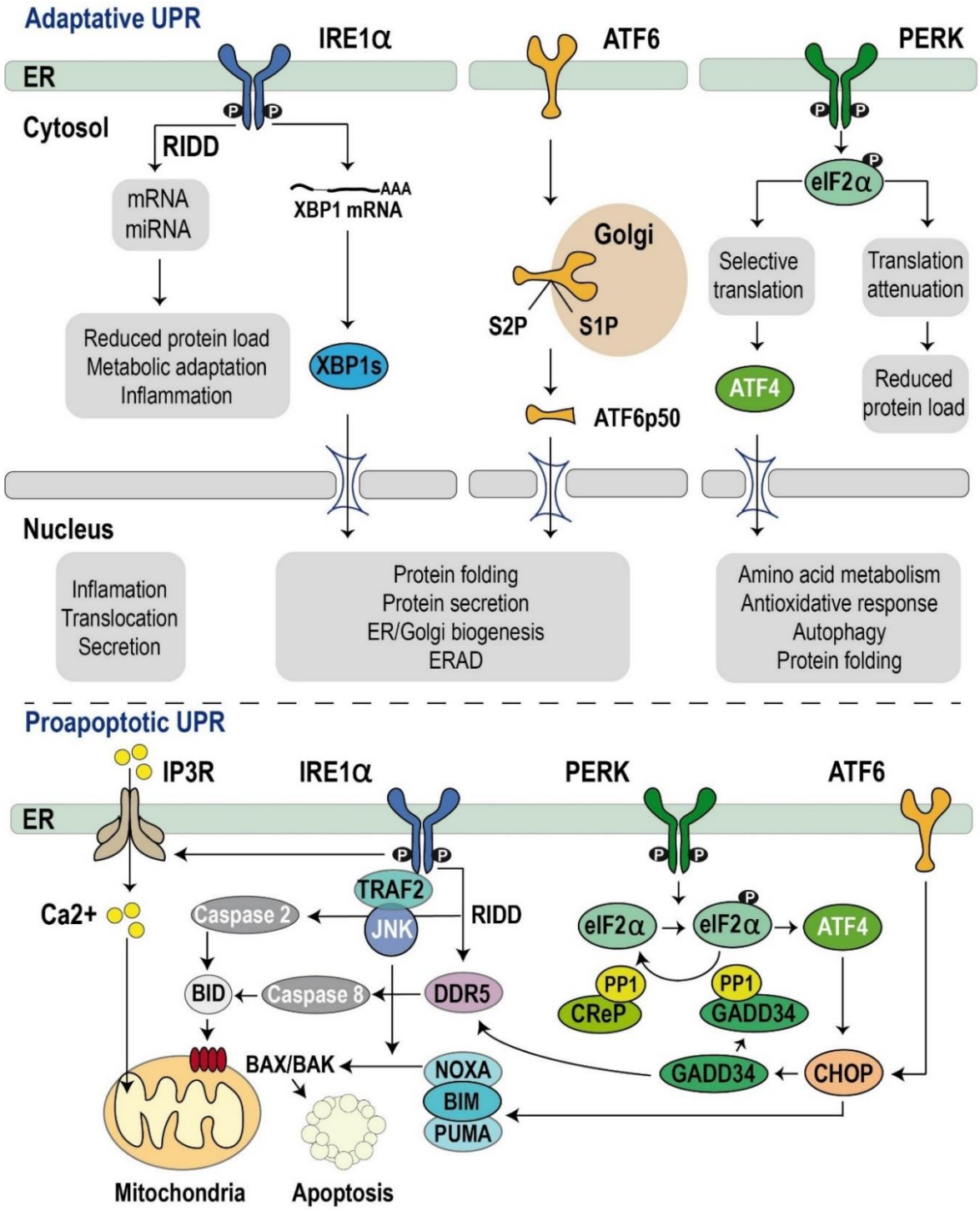


Figure 13: Branches of the UPR. Adaptative UPR. ER stress activates the three branches of the UPR. 1) IRE1α autophosphorylates and splices XBP1 mRNA, allowing the expression of

transcription factor XBP1s that induces the transcription of UPR target genes. IRE1α RNAse also cleaves ER-associated mRNAs or non-coding functional RNAs, leading to their degradation through RIDD, modulating protein folding load, cell metabolism and inflammation. 2) ATF6 translocates from the ER to the Golgi apparatus, where it's cleaved by the proteases S1P and S2P, releasing the ATF6p50 transcription factor. ATF6p50 migrates to nucleus to induce the expression of UPR target genes involved in ER protein folding homeostasis and cell physiology. In conjunction with XBP1s, ATF6p50 promotes the degradation of unfolded/misfolded proteins accumulated in the ER through the proteasome-based ER-associated protein degradation (ERAD) machinery and induces ER and Golgi biogenesis to increase the secretory capacity of the cell. 3) PERK kinase phosphorylates (eIF2α), transiently reducing protein synthesis. Selective mRNAs, such as ATF4 mRNA, are preferentially translated in the presence of phosphorylated eIF2α. Transcription factor ATF4, activates the transcription of UPR target genes involved in the antioxidative response, amino acids metabolism, protein synthesis, apoptosis and autophagy. **Proapoptotic UPR.** Under ER stress. The PERK and ATF6 branch induce the proapoptotic factor CHOP, which promotes ER stress-induced apoptosis by modulating GADD34, death receptor (DR5), and the proapoptotic members of the BCL-2 family including NOXA, BIM and PUMA. The IRE1α UPR branch is involved in caspase 2-dependent, caspase 8-dependent or BAX/BAK-dependent apoptosis through RIDD or activation of TRAF2–JUN N-terminal kinase (JNK) signalling. Ca 2+ release from the ER via IP3R, increases mitochondrial Ca^{2+} uptake, contributing to the release of reactive oxygen species (ROS) and the activation of the BAX/BAK-dependent apoptosome. Adapted from[202].

1.5.2 Oxidative stress

ROS are one of the most potent and omnipresent threats faced by the cells. Examples of ROS include the superoxide anion (O_2^{*-}), hydrogen peroxide (H_2O_2), singlet oxygen, hydroxyl radical (OH*), peroxy radical and nitric oxide (NO*) which can react with O_2^{*-} to form peroxynitrite (ONOO-)[203]. ROS are counteracted in the cells by antioxidant defense mechanisms that include ROS-metabolizing enzymes such as catalase (CAT), glutathione peroxidase (GPx) and superoxide dismutases (SODs), and other antioxidant proteins such as glutathione synthetase (GSH)[203]. Cellular survival depends on the maintenance of the equilibrium between pro-oxidant species and antioxidant defense mechanisms and the disturbance of this balance leads to oxidative stress[234]. In most cases, oxidative stress can be resolved by the cell's natural defenses, however sustained stress leads to cell death[234–236]. ROS can emanate from intracellular or extracellular sources. Intracellular ROS sources include the oxidative respiration in the mitochondria or oxidative protein folding at the ER[203]. In addition to physiological sources of ROS, diverse exogenous agents can contribute to the intracellular production of free radicals

and cause the generation of $O2^{*-}$ and H_2O_2[203,237]. Since ROS can cause damage to all the major classes of biological macromolecules including nucleic acids, proteins, carbohydrates, and lipids, when the excess of ROS surpasses the cell's antioxidant defenses, cell death is triggered[203].

1.5.3 Arsenic stress

Arsenic (As) is ubiquitous in the environment and human actions such as smelting, the burning of fossil fuels and the use of arsenic in pesticides, contribute to its further spread in soil, water, air, agricultural and aquatic food. The major route of human exposure for inorganic arsenic is contaminated drinking water and developing countries such as Bangladesh, India, Mexico and Taiwan are highly impacted by arsenic contamination in groundwater[238–240]. Chronic arsenic exposure has been associated several types of cancer (skin, bladder, liver, kidney, and lung)[241], diabetes[242], arteriosclerosis and cardiovascular diseases[243], hypertension[244] and neurological diseases (Alzheimer and Parkinson)[245]. Interestingly, arsenic has been historically used in the treatment of several diseases. Arsenic trioxide is actively studied as an anticancer drug and has been effectively used in the treatment of acute promyelocytic leukemia (APL). When used as single agent, arsenic trioxide triggers complete remission and prolongs the survival in up to 70% of the otherwise treatment resistant APL patients. Combination of arsenic trioxide with retinoic acid reach a stunning 90% cure rate[110,246,247]. However, the current use of arsenic trioxide as an anticancer agent is limited by its toxic nature and the lack of mechanistic studies in different tumour scenarios.

The primary mechanism of arsenic induced toxicity is the generation of ROS, which leads to oxidative stress (Fig. 14). The main species produced in arsenic-induced oxidative stress is superoxide anion which is transformed in the more reactive oxygenated species H_2O_2 and OH^{*}[238]. Arsenic-induced ROS are mainly generated in the mitochondria. Arsenic exposure induces a rapid decline of mitochondrial membrane potential, promoting dramatic morphologic changes and the loss of mitochondrial internal organization, while altering the activity of mitochondrial enzymes[238,248]. Other ROS sources during arsenic

exposure include the generation of H_2O_2 during the oxidation of As (III) to As (V), or the decoupling of nitric oxide (NO) synthase enzyme system. Additionally, the expression of various antioxidant enzymes that counteract oxidative stress (SODs, CAT and GPx) is decreased during arsenic exposure, thus promoting further the accumulation of ROS[238,240]. ROS interaction with macromolecules leads to DNA damage, alterations in DNA methylation and lipid peroxidation, triggering oxidative stress and cell death.

A recent study demonstrated that arsenic exposure also induces ER stress, which activates the UPR and affects Ca^{2+} homeostasis and mitochondrial integrity, leading to apoptosis of mouse embryonic fibroblasts (MEFs)[206] (Fig. 14). This study showed that arsenic trioxide activates the PERK/ATF4 branch of the UPR, increasing the phosphorylation of PERK, eIF2α and promoting the translocation to nucleus of transcription factors ATF4 and CHOP. Activation of PERK/ATF4 branch increased Ca^{2+} release from ER lumen to the cytosol, via the IP3R receptor, inducing Ca^{2+}/calpain-dependent apoptosis. Furthermore, Ca^{2+} imbalance induced mitochondrial-mediated apoptosis, evidenced by the increased phosphorylation of p53, mitochondrial ROS production, BAX translocation from cytoplasm to mitochondria and loss of the mitochondrial membrane potential integrity. Although more mechanistic studies are required to understand arsenic toxicity, the current literature suggest that ER-mitochondrial crosstalk is important for arsenic cytotoxicity, and both oxidative and ER stress individually or in combination, can trigger cell death.

As previously discussed, *Ars2* was originally described as a gene that conferred arsenic resistance to Chinese hamster ovary cells[109]. It was later discovered that the resistant phenotype was a dominant negative effect resulting from a truncated cDNA, and full-length ARS2 was related to arsenic sensitivity[15]. However, it is still unknown how ARS2's role in RNA metabolism relates to the cellular response to arsenic. Furthermore, the arsenic stress response involves the generation of ROS and UPR activation, which trigger cytoplasmic pathways that could be modulated by the potentially cytoplasmic ARS2 isoform.

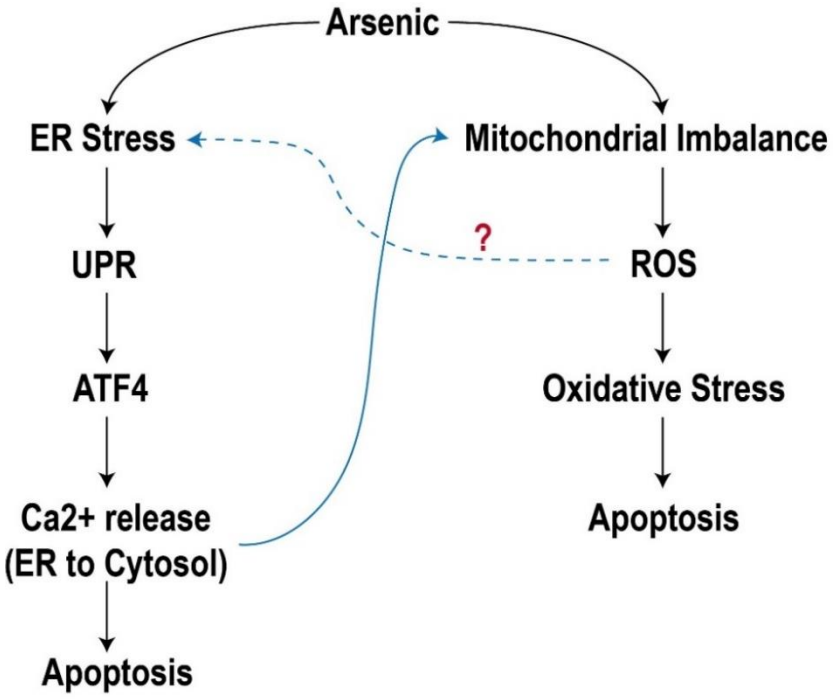


Figure 14: Arsenic trioxide mechanism of action. Arsenic exposure induces affectations in the mitochondria, increasing the production and release of ROS, which can trigger oxidative stress and apoptosis. Arsenic trioxide also binds to protein thiol groups, disturbing normal protein folding and function, which can lead to ER stress. Arsenic-mediated ER stress induces the PERK/ATF4 branch of the UPR, which increases the release of Ca^{2+} from the ER to the cytosol leading to Ca^{2+}/calpain-dependent apoptosis. Increased Ca^{2+} concentration of the cytosol, promotes mitochondrial Ca^{2+} uptake, which further increases mitochondrial imbalance and ROS production. ROS accumulation can also induce ER stress. However, this hasn't been studied during arsenic exposure, thus the interrogation symbol.

Research Objectives

CBC bound ARS2 acts as a multifunctional platform for protein and RNA interactions, and helps to synchronize the dynamic assembly and disassembly of mutually exclusive complexes that regulate RNA processing, degradation, and export[12–16,18,21–23]. ARS2n has a crucial role in nuclear RNA metabolism, and the potential existence of cytoplasmic isoforms of ARS2 suggests: 1) ARS2c may regulate RNA metabolism in the cytoplasm. 2) ARS2 isoforms may function to connect nuclear and cytoplasmic RNA metabolic processes. 3) There is an undiscovered cytoplasmic interactome for ARS2c that could impart new roles for the protein.

The guiding hypothesis for my thesis is that there are cytoplasmic isoform(s) of ARS2 that regulate RNA metabolism and are functionally distinct from ARS2n. In **Chapter 3**, I characterize the expression of ARS2 isoforms in human and mouse cell lines and develop RNAi strategies to specifically target *ARS2c* isoform mRNAs. In **Chapters 4** and **5**, my

aim is to identify and characterize differential and shared biological roles of ARS2 isoforms, focusing on previously unexplained or predicted functions of ARS2n: Arsenic stress response and NMD.

Chapter 2: Materials and Methods

Chapter disclaimer: The information in this chapter was published in:

Mesa-Perez, M., Hamilton, P. T., Miranda, A., Brodie, N., O'Sullivan, C., Christie, J., Ryan, B., Chow, R., Goodlett, D., Nelson, C. and Howard, P. Cytoplasmic Switch of ARS2 Isoforms Promotes Nonsense-Mediated mRNA Decay and Arsenic Sensitivity, *Nucleic Acids Research, 2022*;, gkac033, https://doi.org/10.1093/nar/gkac033

2.1 Cell Culture

C_2C_{12} (ATCC: CRL-1772), HeLa (ATCC: CCL-2), HEK 293T (ATCC: CRL-11268) or Flp-In T-REx (Invitrogen: R78007) cell lines were cultured in Dulbecco's Modified Eagle's Medium High Glucose (DMEM- high glucose, Hyclone) and supplemented with 10% fetal bovine serum (FBS, Hyclone) at 37°C in 5% CO_2 buffered incubators. Cell lines were passaged or harvested for use in experiments when they reached 80-90% confluence. Cells were washed once with Ca2+, Mg2+ and phenol red- free phosphate buffered saline (PBS) (HyClone) and then trypsinized for 3-5 minutes with 0.05% Trypsin-EDTA (Gibco) at 37°C. Trypsin was quenched with growth media and cells were centrifuged for 5 minutes at 1200 rpm to remove residual enzyme. The cells were then resuspended in growth media and plated at the desired density.

2.2 Generation of stable cell lines

To generate stable cell lines expressing Biotin ligase-ARS2n/ARS2c1/ARS2c2 or control, Flp-In T-REx cells were transfected with a 9:1 ratio of pOG44 (Thermo Fisher) to pcDNA5 integration vector and allowed to recover for 48h prior to selection with Hygromycin (InvivoGen). Stable clones were pooled and tested for expression with the addition of 1ug/ml Tetracycline (Sigma) for 24h.

2.3 Transfection

Transfections were performed using Lipofectamine 3000 (Invitrogen) in Figure 17B, but it was switched to JetPrime reagent (VWR) for the rest of the experiments due to toxicity concerns. Small interfering RNAs (siRNAs) were purchased from Integrated DNA Technologies (IDT) and AllStars negative-control siRNA (Qiagen) was used as a control siRNA. shRNAs were purchased from the HuSH™ shRNA collection at Origene in the vector pGFP-V-RS and scramble control were used as controls. All siRNA/shRNA sequences are reported in Table 1.

2.4 Plasmids

Expression vectors for stable integration and tetracycline-inducible expression, were generated by subcloning synthetized ARS2n/ARS2c1/ARS2c2-3xFLAG (GenScript) into a modified pcDNA5/FRT/TO vector containing BirA, kindly provided by C. Nelson[249]. BioID2-ARS2n/ARS2c1/ARS2c2-3xFLAG vectors were generated by synthetizing BioID2 and substituting BirA on the previous constructs (GenScript). BioID2-nlsKO-ARS2n mutant was generated by mutagenesis of BioID2-ARS2n and subcloned in pcDNA3.1 (+)-N-eGFP (GenScript). All constructs were validated by sequencing. Enhanced GFP-C1 (eGFP-C1) was used as a green fluorescent protein (GFP) expression control, eGFP-ARS2n and eGFP-ARS2c1 were generated as described previously[18]. The Firefly luciferase 5xbox b reporter (FLuc-5Xboxb) and firefly luciferase control (FLuc) were generated by subcloning FLuc-5Xbox b from plasmid pAc5.1C-Fluc-STOP-5boxb (Addgene) into pcDNA3.1 (+) (Life Technologies). Renilla luciferase control plasmid (RLuc) was obtained from Promega. λN was synthetized and cloned into pcDNA3.1 (-) ARS2[111] or pcDNA3.1 (-) alone generating ARS2-λN and λN control respectively. λN-RNPS1 plasmid was generated by synthesis of RNPS1 and cloning on λN pcDNA3.1 (-) plasmid. pNMD+ and pNMD- reporter plasmids were kindly provided by K. Lukyanov[250]. RNT1-GFP was a gift from H. Dietz (Addgene plasmid # 17708)[251]. pAc5.1C-FLuc-Stop-5BoxB was a gift from Elisa Izaurralde (Addgene plasmid # 21301)[252]. CBP20-3flag was a gift from Torben Heick Jensen and John LaCava[253].

2.5 RNA isolation, cDNA generation, quantitative real-time PCR and real-time PCR

To ensure data robustness and consistency, three individually treated wells from a 6 well plate were pooled into 1 RNA column. This RNA column was defined as “biological replicate”. Three technical replicates were run per biological replicate. For every analyzed gene, 3 to 6 biological replicates were included, leading to 9-18 data points per condition that represent 9-18 individually treated wells. Total RNA was isolated using RNeasy plus kit (Qiagen). To eliminate genomic DNA, RNA samples were additionally treated with DNAseI (Thermofisher). 1ug of total RNA was reverse transcribed using High-Capacity cDNA Reverse Transcription kit (Thermofisher). PCR was performed using Q5 High-Fidelity DNA polymerase (NEB) while qPCR products were amplified using Ssofast EvaGreen Supermix (BioRad) on a Stratagene MX3000P qPCR system. All primer sequences are reported in Table 1.

2.6 Confocal microscopy

HEK 293T cells were seeded on glass coverslips (neuVitro) and transfected with BioID2-ARS2n/ARS2c1/ARS2c2-3xFLAG or control for 48 hours in complete DMEM media. Cells were washed with PBS and fixed for 15 minutes with 4% paraformaldehyde in PBS at RT. Next, cells were permeabilized in 0.25% Triton X-100 for 15 minutes at RT and blocked with 1% BSA in PBS/Tween (0.01%), 1 hour at RT. Primary (anti-flag mouse 1:200, Cell Signalling) and secondary antibody (anti-mouse Alexa 488 1:500, Thermofisher) were diluted in blocking buffer solution and incubated overnight at 4˚C or 1 hour at RT, respectively. Hoescht 3334 (1:2000) was added at RT for 5 minutes to label nuclei. Images were acquired using a confocal microscope (Nikon) and processed in ImageJ.

2.7 Western Blot

Cell lysates, quantified by Pierce BCA assay kit (Thermofisher) and resuspended in Laemmli sample buffer, were resolved by SDS-PAGE and transferred to polyvinylidene

difluoride (PVDF) membranes. Membranes were blocked in Intercept (TBS) Blocking Buffer (Licor) for 1 hour at 37°C. Primary and secondary antibodies were diluted in Blocking Buffer/Tween (0.1%) and membranes were incubated overnight at 4°C or 1 hour at RT, respectively. Images were revealed and analyzed using Odyssey CLx (Licor) and Image Studio Lite software. Total proteins were detected using Revert 700 Total protein Stain kit (Licor).

2.8 Antibodies

The following antibodies were used in this study: ARS2 (XL14.1, 1:2000) generously provided by the Ludwig Institute for Cancer Research, Flag (1:1000, Cell signalling), Actin (1:4000, Sigma), TBP (1:1000, Cell Signalling), Tubulin (1:1000, Cell Signalling), GFP (1:1000, Cell Signalling), eRF1 (1:1000, Thermofisher), eIF3E (1:1000, Abclonal), eIF3F (1:1000, Abclonal), eIF3K (1:1000, Abclonal), Magoh (1:1000, Abclonal), MagohB (1:1000, Abclonal), eIF4A3 (1:1000, Abclonal), SMG7 (1:1000, Thermofisher), SMG1 (1:500, Santa Cruz), DHX34 (1:1000, Cedarlane), Streptavidin 680 (1:15000, Invitrogen), UPF1 (1:500, Santa Cruz), UPF1 (1:1000, Cell Signalling), CBP80 (1:1000, Cell Signalling), CBP20 (1:500, Novus Biologicals) Phospho-(Ser/Thr) ATM/ATR Substrate (1:1000, Cell Signalling).

2.9 Protein fractionation

C_2C_{12} cells were washed with buffer H (20mM HEPES pH:8, 2mM $MgCl_2$, 0.1mM EGTA, 1mM EDTA) and lysed for 10 minutes at 4°C in buffer I (2x buffer H, 0.2% NP-20, protease inhibitor cocktail). Cell's nuclei were pelleted at 16000 x g for 10 minutes and supernatant was stored as cytoplasmic fraction. Cold RIPA buffer with protease inhibitor cocktail was added to the pelleted nuclei and samples were incubated 30 minutes on ice. Cell debris was pelleted at 16000 x g for 20 minutes and supernatant was stored as nuclear fraction. Nuclear and cytoplasmic fractions were quantified by BCA Protein Assay kit (Thermofisher). Equal amounts of total protein (30ug) were added from each fraction. To

quantify ARS2n/ARS2c enrichment, ARS2n or ARS2c intensities were normalized to the total ARS2 (ARS2n+ARS2c) intensity in each fraction to minimize effects of differential loading.

2.10 Affinity capture of biotinylated proteins

Purification of biotinylated proteins was performed as outlined by[254]. Flp-In T-REx cells stably expressing tetracycline inducible ARS2n-BirA/control or BioID2-ARS2n/ARS2c1/ARS2c2/control, were treated for 24h with 10-200ng/ml of tetracycline. HEK 293T cells were transfected with BioID2-ARS2n/ARS2c1/ARS2c2/-3xFLAG or control and incubated for 24h. All the conditions were incubated an additional 24h in complete DMEM media supplemented with 50uM of biotin (Sigma). Five individually treated 10 cm^2 dishes, corresponding to approximately $50x10^6$ total cells were pooled together and considered 1 biological replicate. One biological replicate ($5x10cm^2$ dishes) was used for tetracycline inducible BioID2-ARS2n/ARS2c1/ARS2c2/control, and three biological replicates ($15x10cm^2$ dishes) were used for the rest of the evaluated conditions (tetracycline inducible ARS2n-BirA/control and BioID2-ARS2n/ARS2c1/ARS2c2/-3xFLAG/control overexpression). Cells were washed three times in PBS and lysed directly on the plate with 500ul Lysis Buffer (50mM Tris pH 7.4, 500mM NaCl, 0.2% SDS, 1mM DTT, and protease inhibitors cocktail). Triton X-100 was added to a final volume of 2%. Lysates were sonicated on 30% amplitude for 2 cycles of 30 seconds with 2 minutes rest, using a sonic dismembrator (Fisher Scientific). Lysates were diluted in chilled 50mMTris pH7.4 and subjected to a final round of sonication. Insoluble cellular debris was cleared by centrifugation at 16 000 x g for 20 minutes at 4°C. Cleared extracts were incubated with 200ul of Dynabeads MyOne Streptavidin C1 (Thermo Fisher) and incubated overnight at 4°C. The following morning, beads were washed two times with Wash Buffer 1 (2% SDS), once with Wash Buffer 2 (0.1% deoxycholic acid, 500mM NaCl, 1mM EDTA, 50mM HEPES ph7.5 and 1% Triton X-100), once in Wash Buffer 3 (10mM Tris pH 7.4, 250mM LiCl, 0.5% NP-40, 0.5% deoxycholic acid, and 1mM EDTA), and finally two times with Wash Buffer 4 (50mM Tris pH 7.4 and 50mM NaCl), pelleting beads using a magnetic rack. For western blotting, beads were resuspended in Laemmli sample

buffer supplemented with 50uM biotin, incubated for 10 minutes, and boiled for 10 minutes before loading on an SDS-PAGE gel. Total protein was detected using Silver Stain kit (Thermofisher) or Coomasie stain, and BioID2-ARS2n/ARS2c1/ARS2c2/control concentration was estimated by extrapolation on a BSA curve ran on the same gel. Equal amounts of proteins were used as input for LC-MS/MS.

2.11 Protein identification by mass spectrometry

BioID samples were processed for mass spectrometry as outlined by[249]. Briefly, streptavidin beads were washed with 50mM NH4HCO3, resuspended in 50mM NH_4HCO_3 containing 5mM dithiothreitol, and heated at 75°C for 10 minutes. Iodoacetamide was added to a final concentration of 10mM to each sample followed by incubation in the dark at RT for 1 hour. Afterwards, 1mM $CaCl_2$ and 1ug of sequence-grade trypsin (Promega) were added and incubation continued overnight. The following morning, trifluoroacetic acid (TFA) was added to a final concentration of 0.5% (v/v). Beads were pelleted using a magnetic rack and the supernatant removed. A second elution of digested peptides with 0.5% TFA was performed and the supernatant pooled. Digested peptides were passed over ZipTips (Millipore) and eluted with 0.5% formic acid/ 80% acetonitrile. Samples were diluted ½ in water, lyophilized and store at -80°C until use.

2.12 Mass spectrometry acquisition using Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific)

Digested samples (6 µL) were separated by on-line reverse phase chromatography using a Thermo Scientific EASY-nLC 1000 system with a reverse-phase pre-column Magic C18-AQ (100 µm I.D., 2.5 cm length, 5 µm, 100 Å) and an in-house prepared reverse phase nano-analytical column Magic C-18AQ (75 µm I.D., 15 cm length, 5 µm, 100 Å, Michrom BioResources Inc, Auburn, CA), at a flow rate of 300nl/min. The chromatography system was coupled on-line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a Nanospray Flex NG source (Thermo

Fisher Scientific). Solvents were A: 2% Acetonitrile, 0.1% Formic acid; B: 90% Acetonitrile, 0.1% Formic acid. After a 348 bar (~ 3 µL) pre-column equilibration and 348 bar (~ 3 µL) nanocolumn equilibration, samples were separated by a 60-minute gradient (0 min: 0%B; 52 min: 45%B; 2 min: 100%B; hold 6 min: 100%B). Data-dependent acquisition Orbitrap survey spectra were scheduled at least every 3 seconds, with the software determining "Automatic" number of MS/MS acquisitions during this period[255].

2.13 MS data analysis

Raw files were created by XCalibur 4.3.73.11 (Thermo Scientific) software and analyzed with PEAKS Client 7.0 software suite (Bioinformatics Solutions Inc.). Database search parameters as follows: precursor tolerance 5 ppm; MS/MS tolerance 0.035 Da; Trypsin enzyme 2 missed cleavages; Orbi-TRAP instrument type; fixed modification: Carbamidomethylation (C); variable modifications: Oxidation (M), Acetylation (K, N-term) Biotinylation. SwissProt_20200305 Database (561911Sequences /202173710 residues)[256]. To identify the significant interactors of the different ARS2 isoforms, we tested for differences in unique peptide counts across different baits using triplicate replicates for each bait condition, as well as additional negative controls from[257,258]. Unique peptide counts were used as input for SAINTexpress (v3.6.3; https://www.sciencedirect.com/science/article/abs/pii/S1874391913005381). For each bait, significant prey were defined as those with a SAINT score > 0.7 and > 2 average unique peptides across conditions. Over-representation analysis of prey for each bait was conducted using the enricher function of ClusterProfiler (https://pubmed.ncbi.nlm.nih.gov/22455463/). Data visualizations were generated in R (v > 4.0), using the ComplexHeatmap (https://academic.oup.com/bioinformatics/article/32/18/2847/1743594) package. Functional annotation maps of ARS2 isoform interactomes were generated by mining for enriched GO ''Biological Process'' terms using the ClueGO plugin[259] within the Cytoscape framework[260]. ClueGO is a user-friendly plugin that allows the decoding and visualization of functionally grouped GO terms in the form of networks. The size of the nodes shows the term significance after Bonferroni correction. Only GO terms with a P-value < 0.01

were considered significant. A kappa score was calculated reflecting the relationships between the terms based on the similarity of their associated genes, which was set to 0.5 as the threshold in this study. The Organic algorithm that determines node positions based on their connectivity was used for laying out the networks. Protein-Protein interaction networks were generated using STRING[261]. Saint analyses are reported in Tables 2-4. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier (PXD026453).

2.14 Stress treatments

C_2C_{12} and HeLa cells were cultured in Dulbecco's Modified Eagle's Medium High Glucose (DMEM- high glucose, Hyclone) and supplemented with 10% fetal bovine serum (FBS, Hyclone) at 37°C in 5% CO_2 buffered incubators. Once they reached 70% confluence, growth media (DMEM-high glucose+ 10%FBS) was removed, cells were washed twice with DMEM-high glucose and incubated with the stress treatment for the indicated times. Starvation: 0.1% FBS in DMEM-high glucose media. Differentiation induction: 2% Horse serum (Fisher) in DMEM-High glucose media. ER stress: Tunicamycin 0.5ug/ml in growth media. Translation inhibition: 2-5ug/ml Puromycin in growth media. Arsenic stress: 35uM or 40uM of arsenic trioxide in growth media as indicated.

2.15 Protein Immunoprecipitation

GFP or Flag IP: Cells co-transfected with BioID2-ARS2n/ΔNLS-ARS2n/ARS2c/control vectors and GFP-UPF1 or GFP control were lysed in RIPA lysis buffer for 30 minutes at 4°C. Cell debris was removed by centrifugation (20 minutes at 16 000 x g). Extracts were quantified by Pierce BCA Protein quantification kit (Thermofisher) and incubated with anti-FLAG beads or GFP Trap magnetic beads (Chromotek) overnight at 4°C. Beads were washed three times with lysis buffer-PBS at 1:4 and eluted with 3xFLAG peptide (Sigma-Aldrich) or Laemmli sample buffer, respectively.

GFP-ARS2 IP: Cells were transfected with GFP-ARS2n/ΔNLS-ARS2n/ARS2c/control plasmids and lysed in RIPA lysis buffer for 30 minutes at 4°C. Cell debris was removed by centrifugation (20 minutes at 16 000 x g). Extracts were quantified by Pierce BCA Protein quantification kit (Thermofisher) and incubated with GFP Trap magnetic beads (Chromotek) overnight at 4°C. Beads were washed three times with lysis buffer-PBS at 1:4 and eluted with Laemmli sample buffer.

Endogenous UPF1 or ARS2 IP: Cells were lysed in either RIPA or nuclear/cytoplasm fractionation buffers. After protein quantification, cell extracts were incubated overnight at 4°C, with 1ug of anti-UPF1, anti-ARS2 or IgG control antibodies and protein A/G magnetic beads (GenScript). The next day, supernatants were stored, beads were washed with PBS and resuspended in Laemmli sample buffer. Samples were heated at 95°C for 10 minutes and interactors were detected with specific antibodies. For all the performed IPs, 500ug of total protein were used.

Quantification: Four types of quantifications were performed, as indicated in the figure legends. **1) Enrichment analysis:** Represented as the ratio of ARS2 isoforms expression in the UPF1 IP versus ARS2 isoforms expression in the input. This analysis was performed to exclude the effects of differential plasmid expression. **2) Depletion analysis:** Represented as the ratio of ARS2 isoform expression in the input versus ARS2 isoform expression in the supernatant (SN), obtained after UPF1 IP. **3) Prey/Bait normalization:** Represented as the ratio of UPF1 interactor versus UPF1 in UPF1 pull downs, to demonstrate that results are not due to differential pull-down or expression of the bait (UPF1). **4) Prey/Bait normalization and control normalization:** Prey/Bait ratio is additionally normalized against the control sample (BioID2-3xflag empty vector) to allow the comparison between independent experiments.

2.16 Cell survival assays: WST1 and Crystal Violet

C_2C_{12} cells were transfected with *Ars2c*/*Ars2*(all)/*Cbp80*/*Ars2c+Cbp80* or control siRNA for 24 hours. Media was changed to arsenic (35uM/40uM) or puromycin 2ug/ml and cells were incubated for 24, 48, 72, 96, or 120 hours after the arsenic treatment. Cells

transfected with the described siRNA for 48 hours but untreated with arsenic or puromycin were defined as point zero. Arsenic/puromycin media was changed every 3 days. For WST1 assays, WST1 solution (Sigma) was added to the media of the cells and incubated for 4 hours at 37°C in 5% CO_2 buffered incubator. Absorbance was measured at 450 nm on BioTek Epoch 2 microplate spectrophotometer, normalizing samples against a blank well. For crystal violet assays, cells were washed with PBS and fixed in 10% formalin for 10 minutes at RT. After washing the plates with water, crystal violet solution was added, and plates were incubated on a rocker for 20 minutes at RT. Crystal violet solution was removed with several washes of water and plates were dried 24 hours at RT. After taking pictures, 10% acetic acid was added to the plates and absorbance was measured at 570 nm on BioTek Epoch 2 microplate spectrophotometer, normalizing samples against a blank well.

2.17 NMD tethering assay

HeLa cells were transfected with FLuc-5xboxb or Fluc; either ARS2, ARS2-λN, RNPS1, λN-RNPS1, λN, or empty pcDNA (+); and RLuc. Firefly and Renilla luminescence were analyzed 24 hours post-transfection using the Dual-Glo Luciferase Assay System (Promega) and a Perkin Elmer Victor3V 1420 multi-label plate reader. Firefly luminescence was normalized to Renilla luminescence. For the RNAi co-transfection experiments, HeLa cells were transfected with either control siRNA or *UPF1*si for two consecutive days, and on day 3 were co-transfected with NMD-firefly (FLuc-5xboxb) and Renilla luciferase (RLuc) and either λN or λN-ARS2n. Firefly and Renilla luminescence were analyzed 72h after the first transfection with the previously mentioned kit and plate reader.

2.18 NMD reporter assay

HeLa and C_2C_{12} cells were transfected with *ARS2*si/*ARS2c*si/*UPF1*si or control. Media was changed 16 hours post-transfection and each condition was transfected again with

either pNMD+ or pNMD- vectors. Flow cytometry and RT-qPCR analysis were performed 24 or 48h after the second transfection. For flow cytometry $1.0x10^5$ events were acquired using a BD FACSCalibur. Cells expressing either Katushka (pTurboFP635-N vector, Evrogen) or TagGFP2 (pTagGFP2-N vector, Evrogen) were used as controls for the crosstalk of the TagGFP2 signal into the red channel and the Katushka signal into the green channel**.** Data analysis was performed as described by the developers of the assay[262]. Flow cytometry data was deposited in the FlowRepository with the identifier FR-FCM-Z3V5.

2.19 Immunofluorescence

Cells were transfected with the indicated GFP-ARS2n/ARS2c/Δnls-ARS2n/control constructs. Twenty-four hours post-transfection, cells were fixed and imaged using a 20x objective on a Leica DMIRE2 inverted fluorescence microscope. Anti tubulin mouse (1:500, GenScript) and anti-mouse Texas Red-X (1:500, Invitrogen), were used to label the cytoplasm. Hoescht 3334 (1:2000) was added at RT for 5 minutes to label nuclei. Images were processed in ImageJ. GFP intensity in nuclei and cytoplasm was quantified for fifty individual cells across 5 image panels.

2.20 Statistical analysis

Statistical analyses were performed using GraphPad – Prism Version 9, or R v > 4.0. Statistical details can be found in the figure legends of corresponding experiments. For bioinformatic data processing, statistical analysis are detailed within the method description.

Chapter 3: Characterization of cytoplasmic ARS2 isoforms in human and murine cell lines

Contributions: MMP performed all experiments and data analysis except for the first RNAi design and western blot detection of ARS2c isoforms (Fig. 17A, B), performed by Jennifer Christie, Perry Howard's lab. Perry Howard captured the images presented on Figure 20B. Figures 21A, B were obtained from the Human Protein Atlas[263].

The data in this chapter contributed to the publication:

Mesa-Perez, M., Hamilton, P. T., Miranda, A., Brodie, N., O'Sullivan, C., Christie, J., Ryan, B., Chow, R., Goodlett, D., Nelson, C. and Howard, P. Cytoplasmic Switch of ARS2 Isoforms Promotes Nonsense-Mediated mRNA Decay and Arsenic Sensitivity, *Nucleic Acids Research, 2022*;, gkac033, https://doi.org/10.1093/nar/gkac033

3.1 Introduction

ARS2 is an essential gene with orthologs from yeast to vertebrates, except for *Saccharomyces cerevisiae*[111,264–268]. The NCBI database reports that in mouse and human, the ARS2 locus produces several different transcripts that codify for ARS2 protein isoforms of variable length (644-906 amino acids). Based on ARS2 structure[21], these protein isoforms should have different locations, where larger ARS2 forms (791-906 aa) are located in the nucleus and smaller forms (644-710 aa) are located in the cytoplasm. However, all reports so far in the literature study ARS2 as a single nuclear protein of approximately 875 amino acids. This chapter addresses this gap in the literature and explores the existence of cytoplasmic isoforms of ARS2 in both mouse and human cells.

The canonical *Ars2* sequence is composed by twenty exons, which are shared by most *Ars2* transcripts in mouse and human. Smaller ARS2 isoforms (644-710 aa) retain a portion of the intron 5 that is highly conserved in mammals[111], and introduces stop codons in all 5' open reading frames. Thus, in smaller ARS2 isoforms, translation is predicted to

re-initiate in intron 5 or exon 6, leading to ARS2 forms that lack the N-terminus of the protein and the nuclear localization signal. Previous studies in our lab, suggested that intron 5 containing *Ars2* isoforms are expressed in mouse cells, but further characterization of the isoforms is required.

In this chapter, I demonstrate that intron 5 containing ARS2 isoforms are expressed in both mouse and human cells, at transcript and protein levels. As predicted by their coding sequence, intron 5 containing ARS2 isoforms are enriched in the cytoplasm and can be specifically targeted using RNAis. Furthermore, expression levels of ARS2 isoforms varies between tissues, transfection conditions and cell lines. The existence of cytoplasmic ARS2 isoforms, different in localization and structure from canonical nuclear ARS2, opens the field to the discovery of new roles of this essential protein in the biology of the cell.

3.2 Results

3.2.1 *Ars2* isoforms in mouse and human, predicted on the NCBI database

The *Ars2* (*Srrt)* locus produces twelve transcripts in both mouse and human, some of which encode different proteins isoforms (NCBI Gene database Gene ID: 83701; Gene ID: 51593 respectively). In mouse, the transcripts can be organized into 3 groups based on the potential major protein products. For simplicity, only representative transcripts from each group are shown in Figure 15A. The first group represents canonical ARS2, which initiates translation within exon 2, encodes 4 similar isoforms of 864-875 amino acids, and has a molecular weight of ~130kDa (Fig. 15A). This group of isoforms, which we designate ARS2n due to its nuclear localization, is well studied as a component of the CBC and has a role in RNA polymerase II (RNAPII) transcript biogenesis and turnover. The second group (2*) comprises 1 isoform, initiates translation within exon 4 and encodes a protein of 778 amino acids. This *Ars2* isoform was reported in the NCBI database recently, when all the isoform-based experimentation was concluded, and it was not included in our analyses.

The third group (3*) is comprised of 7 transcripts that retain a portion or the whole intron 5. Conservation analysis between mouse and additional 39 placental mammals showed that intron 5 is highly conserved and it is striking that 7 of the 12 *Ars2* isoforms retain portions of this intron (Fig. 15A). The high genomic conservation of intron 5 is not observed for any other introns in the sequence and it's comparable to the conservation showed by the exons, which suggests that retention of intron 5 during alternative splicing could be functionally important. Intron 5 containing transcripts are predicted to initiate translation either in intron 5 or exon 6 and encode putative isoforms of 710 or 648 amino acids (named here ARS2c2 and ARS2c1 respectively) with a predicted molecular weight ~100kDa, based on ARS2n size. This group of ARS2 isoforms are especially intriguing since they are missing up to 227 amino acids from the functionally important N-terminus of ARS2n, which contains a structural N-terminal helix-turn-helix and conserved tyrosine phosphorylation sites. The N-terminus region also contains the nuclear localization signal, thus group 3* ARS2 isoforms were designated as ARS2c for their potential to encode cytoplasmic isoforms of unknown function. Starting in exon 6, all 12 *Ars2* transcripts share almost 100% sequence homology, as reflected on Figure 15A diagram.

In humans, the twelve *ARS2* transcripts can also be divided into 3 groups, based on the potential major protein products. Only representative transcripts from each group are shown in Figure 15B. The first group (1*) include 9 similar transcripts that initiate translation in either exon 1 or exon 2, encoding isoforms of 872-906 amino acids with a molecular weight of ~130kDa. As with the mouse transcripts, group 1 isoforms represent canonical, nuclear ARS2. The second group (2*) comprises 2 isoforms that have an extra non-conserved exon, initiate translation within exon 4 (canonical *ARS2* exon 3) and encode a protein of 792-793 amino acids. Similar to group 2* transcripts in mouse, these isoforms were recently reported and therefore were not included in our analyses. The third group (3*) is comprised of 1 transcript that retains a portion of the highly conserved intron 5. This isoform is the human analogous of mouse group 3* isoforms, it's predicted to start translation in exon 6 and encode a cytoplasmic, 644 amino acids long ARS2 isoform. As reflected on Figure 15B diagram, starting in exon 6, human *ARS2* transcripts share almost 100% sequence homology.

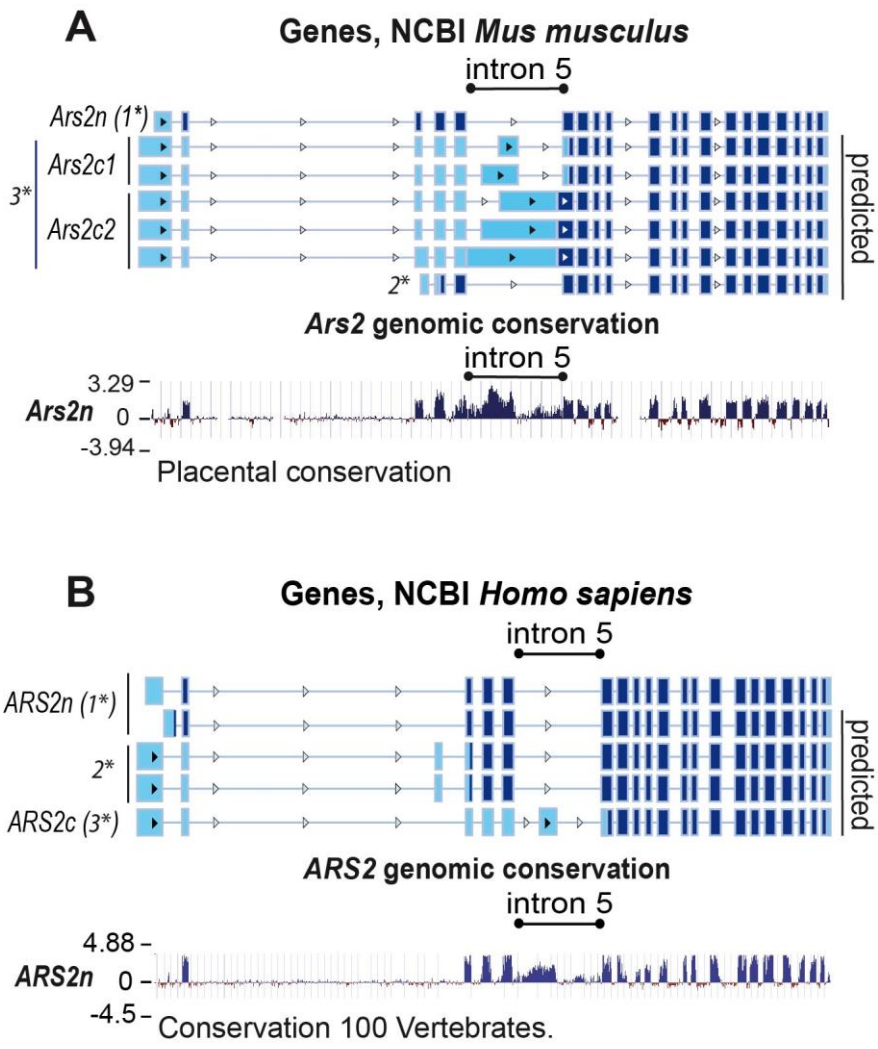


Figure 15: *Ars2* isoforms in mouse and human (NCBI database). A, B) Representative *Ars2* sequences of major isoforms in NCBI Gene database (*Mus musculus*) **(A)** and *Homo sapiens* **(B)**, exons are represented in dark blue boxes. **Bottom:** Conservation analysis of *Ars2*, between mouse and 39 placental mammals or between human and 100 vertebrates. The alignments were generated using Multiz and UCSC/Penn State Bioinformatics comparative genomics alignment pipeline. Evolutionary conservation was measured using PhyloP and PhastCons.

3.2.2 ARS2c isoforms are expressed in mouse and human cells and can be specifically targeted with interference RNAs.

To investigate the expression of the group 3* transcripts in mouse and human cells, RT-PCR was performed. Diagrams showing the targeted regions are included in Figure 16A, B. Primer sequences are included on Table 1. Using a combination of primers flanking the conserved intron 5, amplification products were detected in mouse C_2C_{12} cells, and human HeLa cells (Fig. 16C, D). Controls lacking reverse transcriptase did not show any amplification products, confirming that these results are not a consequence of genomic contamination. Using a forward primer within intron 5 and reverse primer within the 3'UTR

of *Ars2*, a full-length amplicon was amplified in mouse and in human cells (Fig.16 E, F). Sequencing of the mouse transcript showed that it corresponded to variants XM_030254927.2, XM_006504631.3 and XM_036165632.1, which retain the second half of intron 5 and share the rest of the *Ars2* sequence from exon 6 to the end of the 3'UTR. Similarly, the amplified human transcript corresponded to XM-024446794.1, which retains a part of intron 5 and shares the rest of the *ARS2* sequence from exon 6. Amplicons detected in Fig.16C, D, E and F were sequenced, and sequencing data is included in the appendix.

We next determined if mouse and human *Ars2c* isoforms are translated into proteins and could be targeted with shRNA or siRNAs. siRNAs targeting a common region of mouse *Ars2* isoforms were designed (Figure 17A). ARS2 isoforms were detected with an anti-ARS2 antibody (XL14.1), raised against the C-terminus of ARS2 (676-871), which is shared between ARS2n and ARS2c isoforms. This anti-ARS2 antibody, previously validated for ARS2n detection by western blot[18,111], detected approximately 130kDa and 100kDa products in lysates from C_2C_{12} mouse cells. Although both products were knocked down with two of the three siRNAs, the lower MW product showed increased sensitivity to the siRNAs, suggesting it is not a protein degradation product of the larger isoform. A combination of the three siRNAs strongly knocked down both protein products (Fig. 17B).

To specifically modulate ARS2c isoforms and distinguish them from the rest of ARS2 isoforms, shRNA/siRNAs specifically targeting intron 5 were designed (Fig.17C, 18A). Taking advantage of the sequence homology between mouse and human *ARS2* isoforms I designed a set shRNA/siRNAs that target both mouse and human isoforms. The nomenclature change of *Ars2* or *ARS2* refers to its use in mouse or human cells respectively. *Ars2/ARS2* sh/si (all) target a sequence in exon 4, which is common to all *Ars2* isoforms in mouse and human (Fig.17C, 18A). *Ars2/ARS2c* sh/si targets the middle sequence in intron 5, present in all *Ars2c* isoforms in mouse and human, but absent from *Ars2n* isoforms (Fig.17C, 18A). Finally, *Ars2c2* sh/si, targets the last region of intron 5, which is only retained in the mouse *Ars2c2* isoforms (Fig.17C). I validated this set of shRNA/siRNAs in C_2C_{12} (Fig. 17D-F), HeLa (Fig.18B-D) and Flp-In TRex cells (Fig.19A-

D). In C_2C_{12}, *Ars2 (all)*, *Ars2c* and *Ars2c2* sh/si were validated by RT-qPCR (Fig 17D-F). In these experiments, *Ars2c* isoforms were detected using primers that amplified intron 5, while the rest of *Ars2* isoforms were detected with primers that amplify the exon 5/6 junction, thus excluding intron 5. As expected, *Ars2 (all)* si targets both *Ars2n* and *Ars2c* (Fig.17D, E), while *Ars2c* si and *Ars2c2* si are specific for intron 5 containing isoforms and do not affect *Ars2n* isoforms expression (Fig.17D-F). Similarly in HeLa cells, western blot validation of the shRNAs showed that *ARS2sh* (all) decreased both ARS2n and ARS2c bands, detected with an anti-ARS2 antibody (XL14.1), while *ARS2c* sh was specific for the lower band corresponding to ARS2c (Fig.18B, C). *ARS2c* sh was additionally validated by RT-qPCR, showing the specific targeting of the intron 5 containing *ARS2c* isoform (Fig.18D). To demonstrate that the effects observed by western blot in HeLa cells, were specific for ARS2, I generated Flp-In-TRex cell lines in which the expression of Flag tagged ARS2n and ARS2c2 was induced with tetracycline and detected with anti-flag antibodies (Fig.19A-D). Confirming the HeLa and C_2C_{12} results, *ARS2* sh (all) targets ARS2n (Fig.19A, B), while *Ars2c2* sh targets the ARS2c (Fig.19C, D). The effect of *ARS2* sh (all) on ARS2c2 expression was not evaluated, since only the coding sequence of *Ars2c2* was stably transfected on the cells, thus missing the sequence on exon 4 targeted by *ARS2* sh (all). Our combined results, strongly suggest that ARS2c isoforms are expressed in mouse and human cells and can be specifically modulated using shRNA/siRNAs that target intron 5.

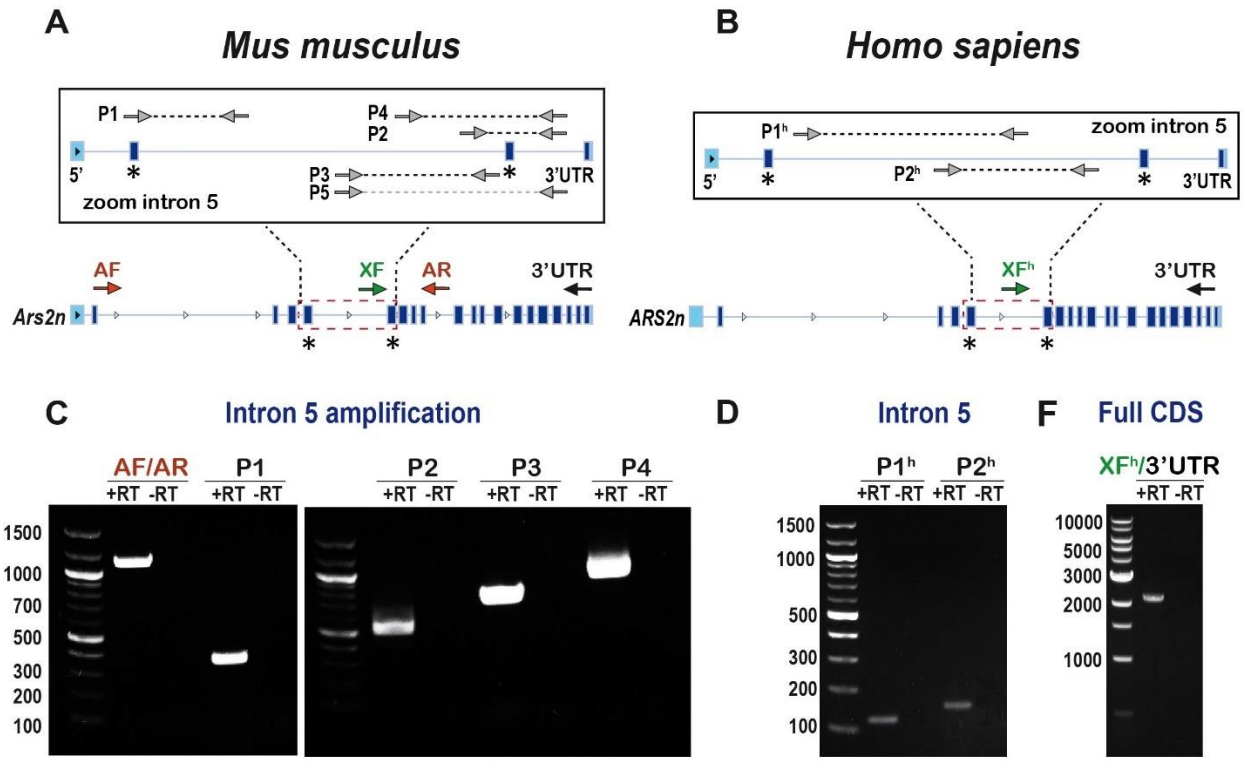


Figure 16: *Ars2c* isoforms are transcribed in mouse and human cell lines. A) Diagram of primers used in **C)** and **E)** to amplify intron 5 **(C)** or the full CDS of the new isoforms **(E)**, in C_2C_{12} cells. AF/AR primer set is showed as positive control. **B)** Diagram of primers used in **D)** and **F)** to amplify intron 5 **(D)** or the full CDS of the new isoforms **(F)**, in HeLa cells. Predicted amplicon sizes: AF/AR:1100bp, P1: 332bp, P2: 545bp, P3: 750bp, P4: 915bp, XF/3'UTR: 2150bp. $P1^h$: 123bp, $P2^h$: 152bp, XF^h/3'UTR: 2217bp. P5 primer set was used only for sequencing. Asterisks highlight exons 5 and 6.

.

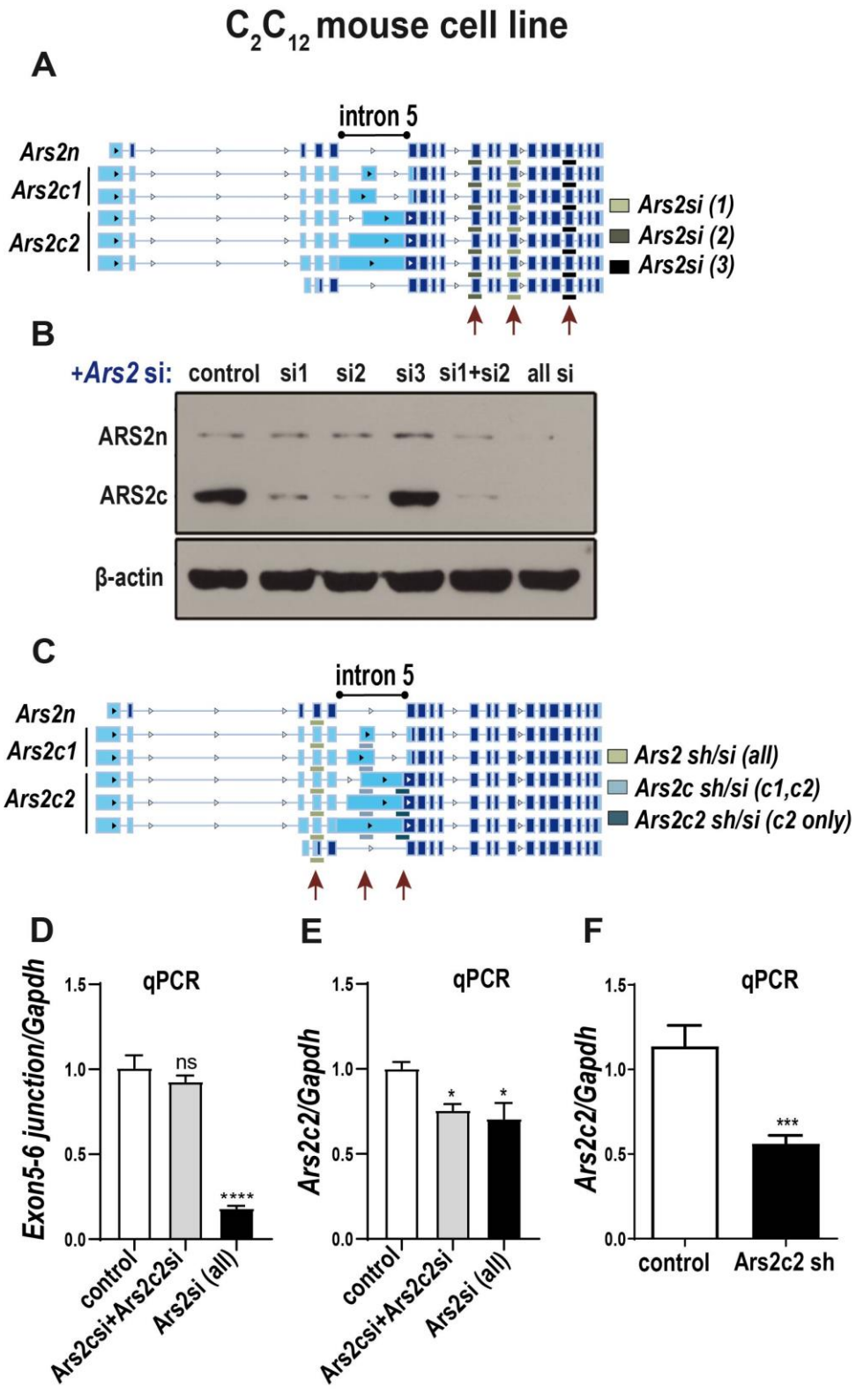


Figure 17: ARS2c isoforms are translated in mouse cells and can be specifically targeted with shRNA/siRNAs. A) Diagram of shRNA/siRNAs used in **(B)**. *Ars2 si* (1-3) target regions in

common to all *Ars2* isoforms. **B)** Western Blot of C_2C_{12} whole cell lysates transfected with *Ars2* si (1-3) or control. ARS2 was detected using an anti-ARS2 antibody (XL14.1) and β-actin is shown as loading control. **C)** Diagram of shRNA/siRNAs used in **(D-F)**. *Ars2* sh/si targets all *Ars2* isoforms, *Ars2c* sh/si targets all cytoplasmic isoforms and *Ars2c2* sh/si is specific for the *Ars2c2* isoform. **D, E, F)** C_2C_{12} cells were transfected with A*rs2c*/*Ars2c2*/*Ars2*(all)/control siRNAs for 48h. *Ars2* isoforms were detected by RT-qPCR, using a primer amplifying intron 5 (*Ars2c*) or the exon5/exon6 junction (*Ars2n*). Each gene was normalized against *Gapdh.* Data are represented as mean ± SEM for n=3 biologically independent samples**.** Statistical analysis: One way ANOVA, post-test Dunnett's **(D), (E)** or two-tail unpaired t-test **(F)** (*p≤0.05, ***p≤0.001, ****p≤0.0001).

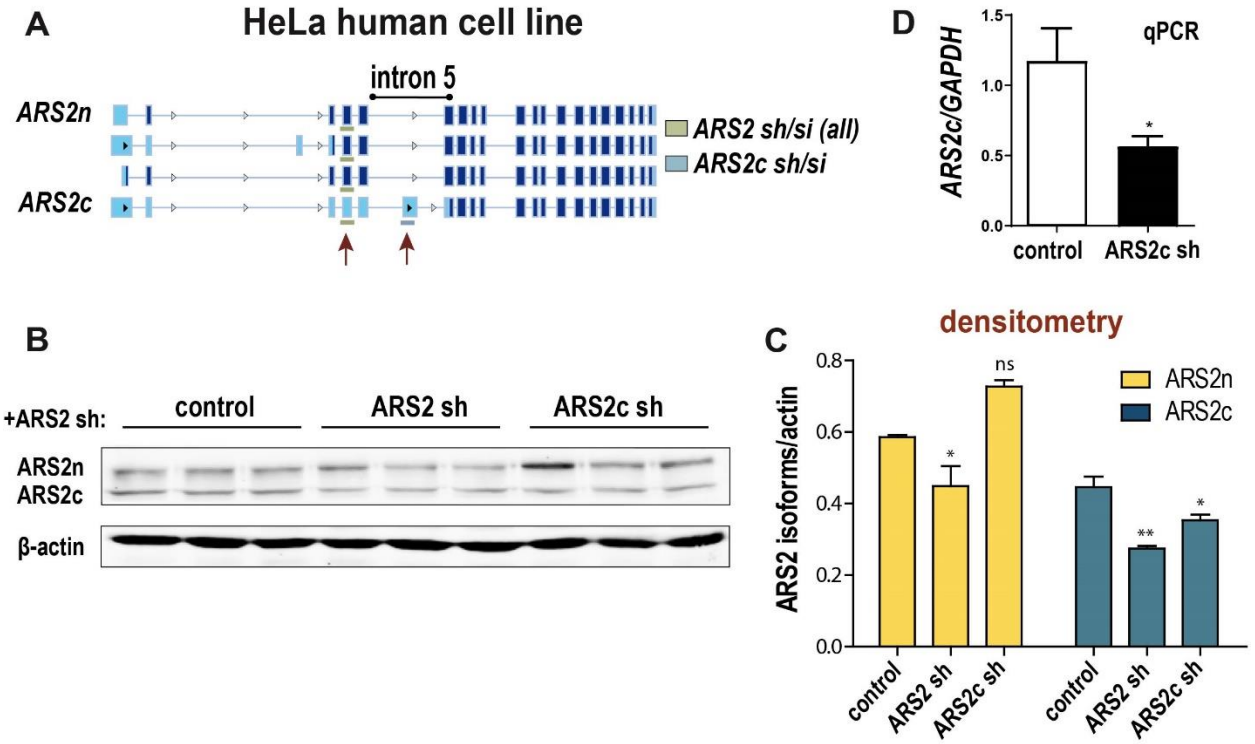


Figure 18: ARS2c isoforms are translated in human cells and can be specifically targeted with shRNA/siRNAs. A) Diagram of shRNA/siRNAs used in **(B-D).** *ARS2* sh targets all *ARS2* isoforms, while *ARS2c* sh is specific for the *ARS2c* isoform. **B)** Western Blot of HeLa whole cell lysate transfected with *ARS2*(all)/*ARS2c* sh or control. ARS2 was detected using an anti-ARS2 antibody (XL14.1) and β-actin is shown as loading control. **C)** Densitometry analysis of **(B)**. **D)** HeLa cells were transfected with *ARS2c*/control siRNA for 48h. *ARS2c* isoforms were detected by RT-qPCR, using a primer amplifying intron 5. Each gene was normalized against *GAPDH.* Data are represented as mean ± SEM for n=3 biologically independent samples**.** Statistical analysis: One way ANOVA, post-test Dunnett's **(C)** or two-tail unpaired t-test **(D)** (*p≤0.05, **p≤0.01).

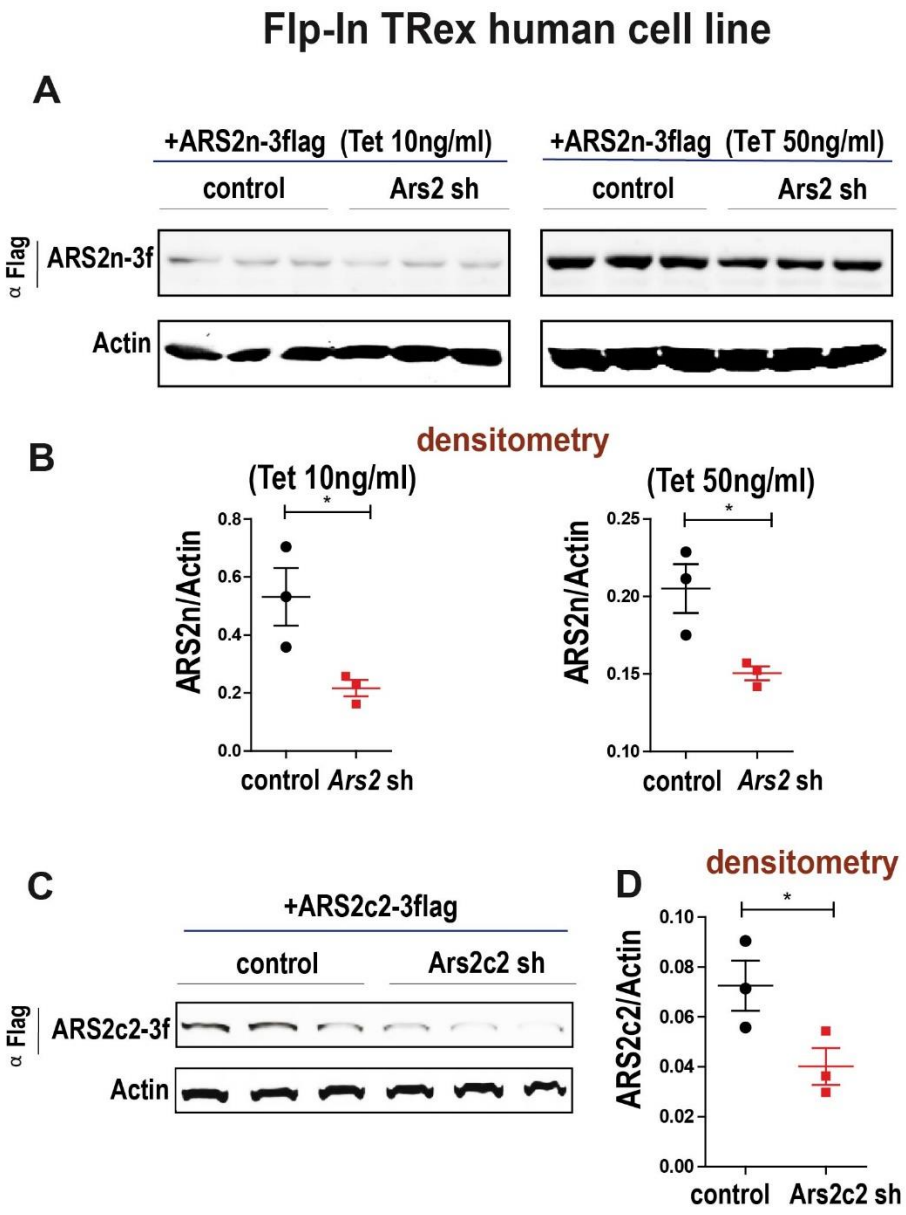


Figure 19: Validation of the used shRNAs in Flp-In TRex inducible cell line. A), C) Flp-In T-REx 293 inducible cell lines were transfected with *Ars2*(all) sh/*Ars2c2* sh/ control for 48h. After transfection, expression of ARS2n **(A)** or ARS2c2 **(C)** was induced with 10 or 50 ng/ml of Tetracycline for 36h. **A), C)** Flag-tagged ARS2n and ARS2c were detected by Western Blot using anti-flag antibodies. Densitometry is shown in **(B)** and **(D)**. Actin is used as a loading control. Data are represented as mean ± SEM for n=3 biologically independent samples. Statistical analysis: One-tail unpaired T-test (*p≤0.05).

3.2.3 ARS2c isoforms are cytoplasmic

ARS2c isoforms lack up to 227 amino acids from ARS2n N-terminus, and as shown in Fig 20A cartoon, this region includes ARS2n N-terminal leg and the nuclear localization signal. Thus, we predicted that ARS2c isoforms would be located in the cytoplasm. Alignment of ARS2 isoforms amino acids sequences are included in the Appendix (Figure 47). To determine the localization of ARS2c isoforms, I expressed the cDNAs fused in-frame with a GFP tag (Fig. 20B). ARS2n isoforms have been shown to be predominantly

nuclear but capable of shuttling between the nucleus and the cytoplasm[15]. As expected, ARS2n was predominantly nuclear in localization. In contrast, ARS2c2 localized predominantly to the cytoplasm (Fig. 20B). Quantification of GFP-ARS2n or GFP-ARS2c intensity in nucleus and cytoplasm for fifty individual cells across 5 image panels is shown in Figure 20C and D. Similarly, nuclear/cytoplasmic fractionation and immunoprecipitation from C_2C_{12} cell lysates showed that endogenous ARS2n and ARS2c are preferentially enriched in the nuclear and cytoplasm fractions, respectively (Fig. 20E-H). Taken together, our data supports the distinct localization of ARS2 isoforms, where ARS2n is preferentially located in the nucleus, while ARS2c isoforms are enriched in the cytoplasm.

3.2.4 Expression levels of ARS2 isoforms are variable

Expression of the ARS2c isoforms varies considerably between tissues, cell lines, passage number, and transfection reagents. As observed in Figure 21A and B, the Human Protein Atlas shows that cytoplasmic expression of ARS2 is high in skeletal muscle samples, but barely apparent in adrenal gland and seminal vesicle. Similarly in our experiments, when western blots are performed with early passages of the C_2C_{12} cell line (<5), using Lipofectamine as a transfection agent, the expression levels of ARS2c are higher than ARS2n (Fig 17B). Since Lipofectamine was toxic to cells, I switched reagents to JetPrime, which solved the toxicity problem but also resulted in less expression of the ARS2c isoform (Fig 21C). The cell passage number in Figure 21C was 15-20. In both experiments (Fig 17B and 21C) ARS2 isoforms were detected using the same anti-ARS2 antibody (XL14.1), and Figure 21C results were confirmed with the anti-ARS2 antibody (HPA042858), validated by the Human Protein Atlas (Fig 21D, E). For consistency and to prevent the effect of potential cellular stress in our analysis, I continued to use JetPrime as transfection agent and cell line passes from 15 to 20. Under these conditions, the expression of ARS2n, at protein levels, is about 5 times more than ARS2c.

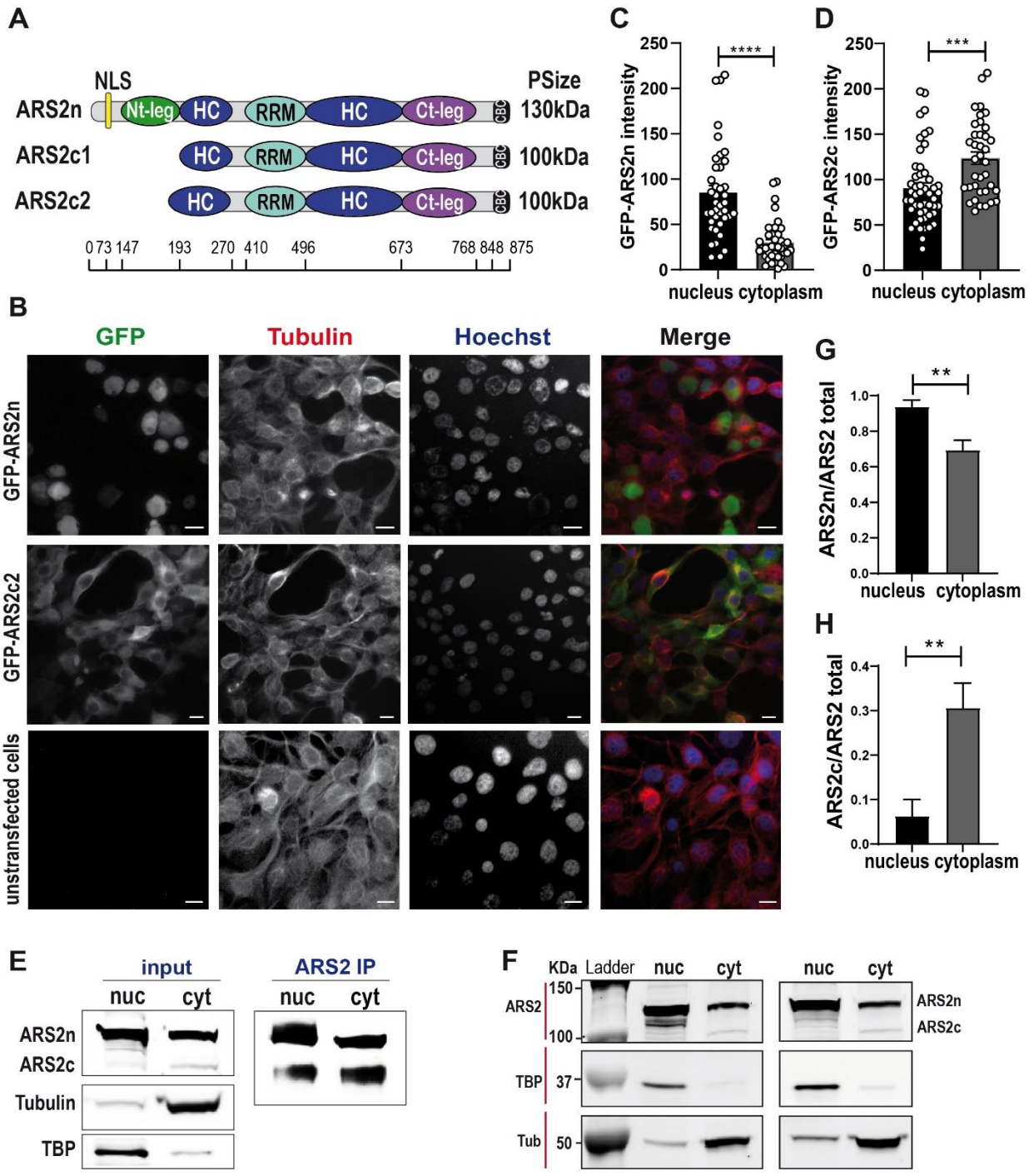


Figure 20: ARS2c isoforms are located in the cytoplasm. A) Protein motif structure of ARS2n and ARS2c1/c2 isoforms, based on the reported structure of ARS2n[21], predicted protein sizes (PSize) are included. **B)** HEK 293T cells were transfected with GFP-ARS2n/GFP-ARS2c. GFP fusion proteins are shown in green, cytoplasms were labelled with anti-tubulin antibodies (red) and nuclei were stained with Hoechst 33342 (blue). Scale bar = 10μM. **C), D)** Quantification of GFP-ARS2n or GFP-ARS2c intensity in nucleus and cytoplasm for fifty individual cells across 5

image panels**. E, F)** Nuclear/cytoplasmic fractionation of C_2C_{12} lysates. Endogenous ARS2 was detected with anti-ARS2 antibodies in the nuclear/cytoplasmic fractions or ARS2 pull downs. TBP and Tubulin detection are used as fractionation quality control. **G, H)** Quantification of ARS2n and ARS2c intensity in nuclear and cytoplasmic fractions. ARS2c or ARS2n intensity was normalized to the total ARS2 intensity (ARS2n + ARS2c) for each fraction, minimizing any potential effects of differential loading. Data are represented as mean ± SEM for n=50 **(C, D)** or n=3 **(G, H)** biologically independent samples. Statistical analysis two-tail unpaired t-test (**p≤0.01; ***p≤0.001, ****p≤0.0001).

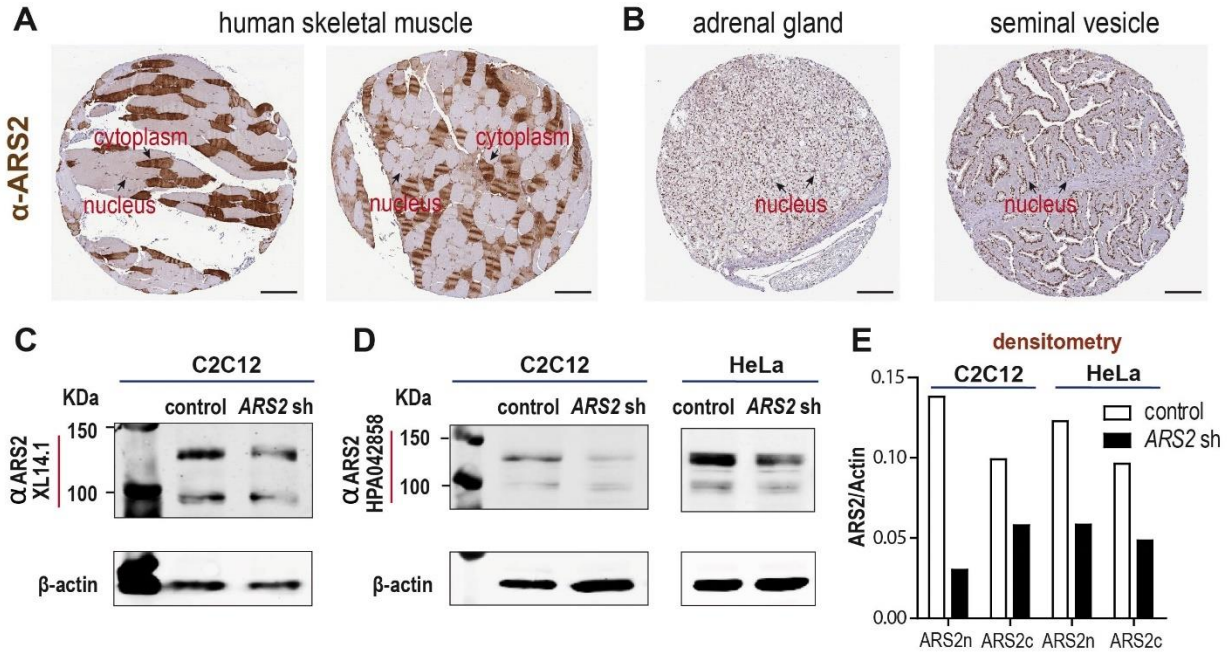


Figure 21: Expression levels of ARS2 isoforms are variable. A), B) Immunohistochemistry images from the Human Protein Atlas. Transverse cut of from skeletal muscle fibers **(A)**, adrenal gland and seminal vesicle **(B)**. ARS2 was detected with an anti-ARS2 antibody (HPA042858) shown in brown. Scale bar = 200μM. Nuclear and cytoplasmic expression of ARS2 are highlighted with arrows. Full image description **(A)** (https://www.proteinatlas.org/ENSG00000087087-SRRT/tissue/skeletal+muscle). **(B)** https://www.proteinatlas.org/ENSG00000087087-SRRT/tissue/adrenal+gland#img, https://www.proteinatlas.org/ENSG00000087087-SRRT/tissue/seminal+vesicle#img. **C)** Western Blot of C_2C_{12} whole cell lysates transfected with *ARS2*(all) sh or control. ARS2 isoforms were detected using anti-ARS2 antibody (XL14.1) and β-actin is shown as loading control. **D)** Western Blot of C_2C_{12} and HeLa whole cell lysates transfected with *ARS2*(all) sh or control. ARS2 isoforms were detected using an anti-ARS2 antibody (HPA042858) validated for Western Blot by the Human Protein Atlas. https://www.proteinatlas.org/ENSG00000087087- SRRT/antibody. β-actin is shown as loading control. **E)** Densitometry analysis of **D)**.

3.3 Discussion

The NCBI database reports the existence of twelve *Ars2* transcripts in both mouse and human that could encode for different proteins isoforms. However, due to the lack of studies on potential ARS2 isoforms, only *ARS2* transcripts coding for canonical, nuclear ARS2 are reported by the NCBI database as validated, while the rest of the detected *ARS2* transcripts are labelled as predicted. Our study constitutes the first characterization of a set of *ARS2* transcripts that retain a portion or the whole intron 5 in both mouse and human cell lines. Confirming previous results of our lab[111], I demonstrated that intron 5 (similar to the exons in the sequence) is highly conserved across vertebrates, suggesting that the intron could be retained during alternative splicing and include functional or regulatory elements. Using multiple sets of primers, I showed that intron 5 is indeed retained in mouse and human cells. Furthermore, full transcripts containing the predicted coding sequence from intron 5 until *Ars2* 3'UTR were amplified and sequenced, showing 99% homology with the reported sequences within the NCBI Gene Database (Appendix).

Retention of the intron 5 generates stop codons in all open reading frames. Thus, protein translation initiated at *ARS2* 5'UTR should be prematurely stopped at the intron 5, leading to transcript degradation by cytoplasmic RNA degradation pathways such as NMD. However, as our study demonstrates, intron 5 containing *ARS2* transcripts are not fully degraded and are translated into isoforms that lack up to 227 amino acids from the N-terminus of ARS2n. Since ARS2n N-terminus contains a structural N-terminal helix-turn-helix, conserved tyrosine phosphorylation sites and the nuclear localization signal, intron 5 containing ARS2 isoforms will likely differ from ARS2n in localization and function.

We demonstrated that intron 5 containing *ARS2* transcripts are translated into proteins of around 100kDa and could be specifically modulated using shRNA/siRNAs that target intron 5. However, since *ARS2n* sequence is shared by all *ARS2* isoforms, I could not design a shRNA/siRNAs that was specific for *ARS2n*. In future studies, a shRNA/siRNA that targets exon5/exon6 junction span could be tried to determine if they can specifically target *ARS2n*.

As predicted by the absence of the nuclear localization signal (NLS), our results confirm that intron 5 containing ARS2 isoforms are enriched in the cytoplasm. However, like the shuttling of ARS2n from nucleus to cytoplasm, the possible translocation of ARS2c isoforms from cytoplasm to nucleus cannot be excluded. As our nuclear and cytoplasmic fractioning studies showed, endogenous ARS2c although enriched in the cytoplasmic fraction, it is also present in the nuclear fraction. Although nuclear localization of ARS2c could be due to the limitations of the cell fractionation technique, it may also reflect the presence of NLS activity in the C-terminal part of the structure (626-875), as reported in the literature[83], which may promote nuclear localization of ARS2c isoforms under specific conditions.

Finally, our results combined with the Human Protein Atlas, suggest that ARS2 isoforms expression varies between tissues, cell line passages and transfection conditions. There are several possible ways in which expression of ARS2c may be regulated. For example, regulation of alternative splicing may determine the inclusion/exclusion of intron 5. Additionally, the cytoplasmic degradation of intron 5 containing transcripts and translation reinitiation at intron 5 or exon 6 could be differentially modulated. As discussed in the next chapter, a clear example of ARS2c expression regulation occurs during the cellular response to arsenic stress, where ARS2c levels are linked to cellular survival. How this regulation takes place is still unknown, but our current hypothesis and some preliminary data will be discussed in the future directions section (**Chapter 6**).

Chapter 4: Differential roles of ARS2 isoforms.

Contributions: LC-MS-MS data acquisition was performed by Nicholas Brodie, University of Victoria Genome BC Proteomics Centre. BioID interactome data analysis was performed by Phineas Hamilton and Alex Miranda Rodriguez, Deeley Research Centre, BC Cancer Agency. Images presented on Figure 23A, were captured by Bridget Ryan, Department of Biology, University of Victoria. All the remaining experiments were performed by MMP.

The data in this chapter contributed to the publication:

Mesa-Perez, M., Hamilton, P. T., Miranda, A., Brodie, N., O'Sullivan, C., Christie, J., Ryan, B., Chow, R., Goodlett, D., Nelson, C. and Howard, P. Cytoplasmic Switch of ARS2 Isoforms Promotes Nonsense-Mediated mRNA Decay and Arsenic Sensitivity, *Nucleic Acids Research, 2022*;, gkac033, https://doi.org/10.1093/nar/gkac033

4.1 Introduction

Ars2 (arsenic resistant gene 2), as discussed in **Chapter 1**, was originally described as a gene that conferred arsenic resistance to Chinese hamster ovary cells[109]. It was later discovered that the cDNA used in this study encoded only a small portion of the extreme C-terminus of full-length ARS2 and it was demonstrated through knockdown studies that full length ARS2 conferred arsenic sensitivity instead[15]. Despite the significant role of ARS2 in arsenic sensitivity, the mechanistic details of this effect are still unknown and haven't been related with ARS2's role in RNA metabolism. Furthermore, arsenic generates endoplasmic reticulum (ER) and mitochondrial stress that leads to apoptosis[206], and this predominantly cytoplasmic response to arsenic could involve our newly described cytoplasmic ARS2 isoforms.

In this chapter, I functionally compare ARS2 isoforms and determine new or unexplained biological roles using BioID and LC-MS/MS. Here I show that ARS2n and ARS2c

interactomes are different and only shared 12% of their interactors. Global network analyses indicate significant differences in the biological process potentially regulated by ARS2 isoforms, where ARS2n interactome was fundamentally involved in RNA metabolism, while ARS2c interactome participated in other processes such as cellular response to stress. Consistent with a role in stress response, expression of ARS2c was specifically induced in response to arsenic stress, and it was responsible for arsenic sensitivity. Indeed, knock down of ARS2c was sufficient to abrogate proteomic remodeling and cellular response to arsenic.

Previous studies implicating the role of ARS2 in arsenic sensitivity examined the effects of *Ars2* knockdown on sensitivity[15]. Since the isoforms were not known, it was assumed the observed effects were due to ARS2n. Our data now show that ARS2c isoforms are more strongly associated with this phenotype and ARS2n is downregulated in response to arsenic. Moreover, the effects of ARS2c on arsenic sensitivity appear to be independent of the CBC complex, which to our knowledge, is the first described function for mammalian ARS2 potentially independent from the CBC complex.

4.2 Results

4.2.1 Comparison of ARS2n and ARS2c interactomes

To understand the biological roles of ARS2c in comparison to nuclear ARS2n, we analyzed their interactomes using BioID and LC-MS/MS. Proximity-dependent biotin identification (BioID) allows the detection of transient and weak protein interactions in living cells[257,269]. In BioID, a promiscuous bacterial biotin ligase (BioID2) is fused to the protein of interest. After the addition of the biotin substrate, proximal proteins are biotinylated and subsequently purified by streptavidin capture (Fig. 22A). Biotinylated proteomes were normalized against the ligase control and enriched proteins were detected by Western Blot or LC-MS/MS. To optimize the assay, I evaluated the effect of the BioID tag position (N-terminus vs C-terminus) in the obtained interactome. Interactome enrichment was higher when the biotin ligase enzyme was placed on the N-terminus of ARS2 versus the C-terminus, which likely reflects the importance of the CBC

binding site within the extreme C-terminus of ARS2 (Fig. 22B). Additionally, I generated Flp-In T-REx 293 cell lines, in which the expression of ARS2 isoforms or biotin ligase control could be induced and regulated with the addition of tetracycline. However, this method returned very few enriched interactors, because I could not adequately control the levels of the biotin ligase control relative to ARS2c1,c2 expression (Fig. 22C). Despite the low recovery, I did confirm ARS2n interactions with known interactors: NCBP3, ZC3H18, ZC3H4, THOC2 and PHAX using this method. To increase the enrichment of ARS2 isoforms interactomes to levels comparable to the literature[21] I overexpressed ARS2 isoforms and control.

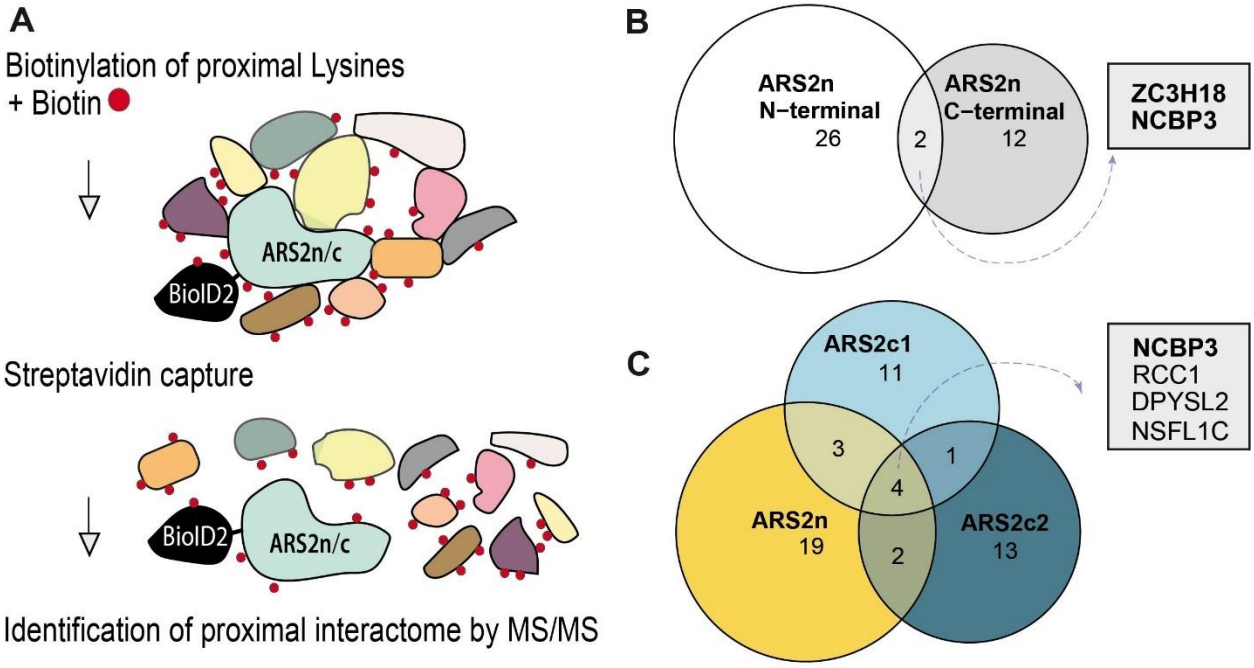


Figure 22: BioID optimization. A) Schematical representation of BioID. Biotynilated proteins are immunoprecipitated by streptavidin capture and interactors are detected by Western Blot or LC/MS-MS. **B), C)** Expression of ARS2n-BioID2 or BioID2-ARS2n/c1/c2 was induced with 10-200ng/ml of Tetracycline for 24h, in Flp-In-TREx cells stably transfected with the plasmids. Cells were treated with biotin for 24h and whole cell lystes were obtained. Biotinylated interactomes of ARS2 forms and BioID2 control were immunoprecipitated with streptavidin beads and detected by LC/MS-MS. ARS2 interactomes were normalized against BioID2 control and analyzed by SAINT. Only proteins with a SAINT score (SP>0.7) were included in the Venn diagrams.

Further optimization of the assay was performed. I confirmed that fusion of ARS2 isoforms to BioID2 ligase did not affect their localization or function: BioID2-ARS2n and BioID2-ARS2c1/c2 localized in the nucleus and cytoplasm respectively, and BioID2-ARS2n maintains its interaction with the CBC complex (Fig. 23A, B). Lastly, the functionality of the BioID tag was confirmed by western blot, where biotinylated interactomes were detected with streptavidin antibodies in whole cell lysates and streptavidin pulldowns (Fig. 23C).

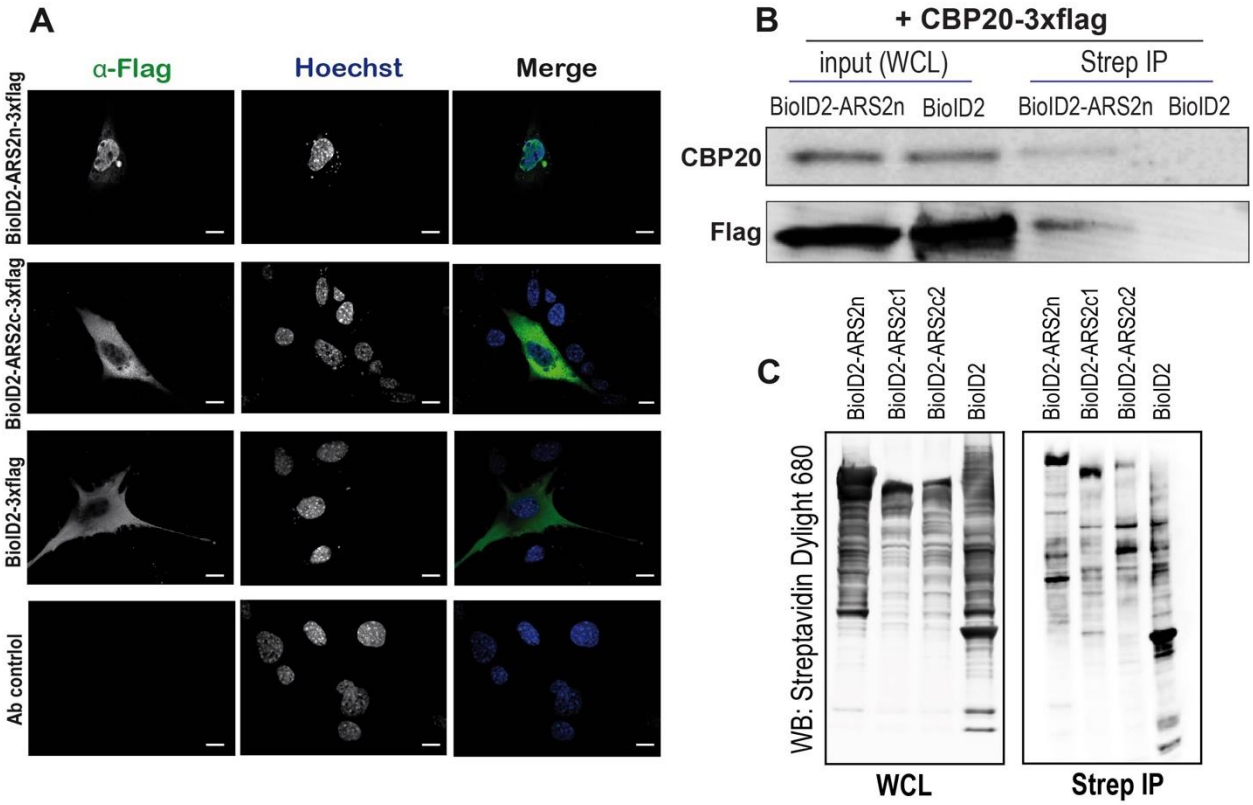


Figure 23: BioID2 tag does not affect localization or function of ARS2 isoforms. A) HEK 293T cells were transfected with BioID2-ARS2n/c-3xflag or control for 48h. ARS2 was detected with anti-flag antibodies (green) and nuclei were stained with Hoechst 33342 (blue). Scale bar indicates 10μm. **B)** HEK 293T cells were co-transfected with BioID2-ARS2n and CBP20-3xflag. After biotin treatment, lysis and streptavidin immunoprecipitation, CBP20 was detected with anti-flag and anti-CBP20 antibodies. **C)** HEK 293T cells were transfected with BioID2-ARS2n/c1/c2-3xflag or control, and treated with biotin for 24h. After lysis, biotinylated interactomes of ARS2 forms and BioID2 control are immunoprecipitated with streptavidin beads and detected by western blot using a streptavidin dylight 680 antibody.

Thus, N-terminus BioID2 tagged ARS2 isoforms and control were expressed in HEK 293T cells for 24h and treated with biotin for additional 24h. For each condition,15 biological replicates were pooled and run in 3 independent experiments (**Chapter 2**). Whole cell lysates were immunoprecipitated with streptavidin beads and 10% of the pulldowns were used for western blot analysis and protein quantification. An equal amount of protein for each condition, as quantified by silver stain, was used for LC-MS/MS. ARS2 isoforms interactomes were normalized against biotin ligase control and analyzed by SAINT (Significance Analysis of INTeractome). As previously reported, only proteins with a saint score (SP>0.7) were included in the subsequent analysis[269]. Dot plot comparison of ARS2 isoforms are shown in Figure 24A.

The proximity interactome of ARS2n confirmed known interactors including NELF, NCBP3, PHAX, XRN2 and RBM7, previously detected by AP-LC-MS/MS (Fig. 24A). As expected, given the different cellular locations, ARS2n and ARS2c1/c2 do not share most of their interactors, while ARSc1 and ARS2c2 do (matches are highlighted in red) (Figure. 24A). ARS2n shared only 12% of its interactome with ARS2c isoforms, while 72% of the ARS2c1 interactome is shared by ARS2c2 (Fig. 24B). Indeed ARS2c isoforms share most of their top 10 interactors (Fig. 24C). Due to the similarities between ARS2c isoforms interactomes, we decided to focus on ARS2c2, as it had a larger number of positive interactors, and refer to it as ARS2c, for simplicity.

Global network analysis of the ARS2n interactome shows an enrichment for proteins related to nucleic acid and macromolecule metabolism, which is consistent with the role of ARS2n in RNA metabolism (Fig. 25A). The ARS2c interactome, on the other hand, is enriched in proteins that participate in quite different biological processes including nitrogen compound metabolic processes, cellular localization, and cellular response to stress (Fig. 25B). This last group is intriguing since the *Ars2* gene was discovered as part of the cellular response to arsenic stress[109], and has been reported to confer sensitivity to this cellular stressor[15].

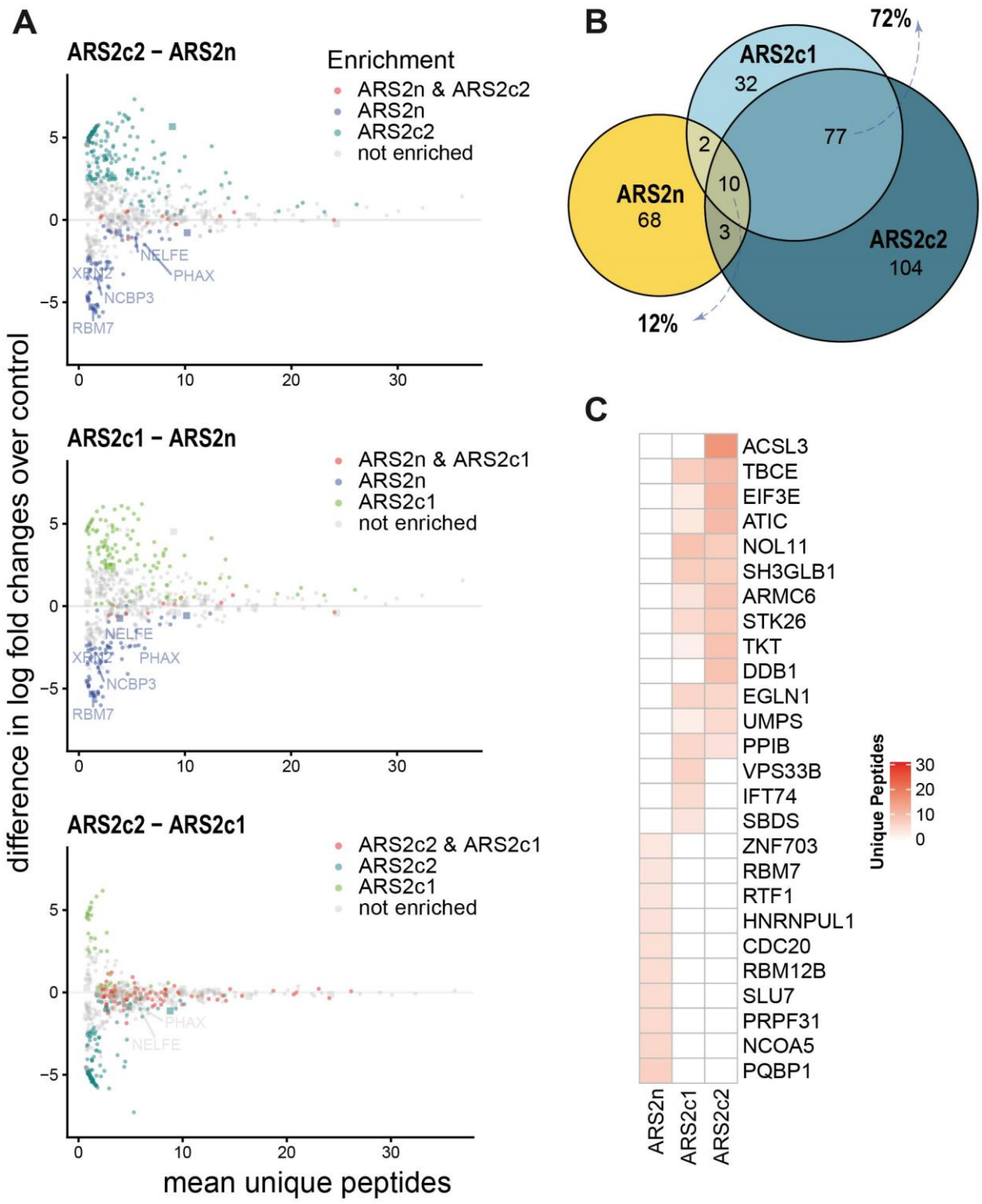


Figure 24: Comparison of ARS2n and ARS2c interactomes. A), B, C) HEK 293T cells were transfected with BioID2-ARS2n/c1/c2-3xflag or control and treated with biotin for 24h. After lysis, biotinylated interactomes of ARS2 isoforms and biotin ligase control were immunoprecipitated with streptavidin beads and detected by LC/MS-MS. 15 independently treated 10cm^2 dishes are

included per condition. ARS2 isoforms interactomes were normalized against biotin ligase control and analyzed by SAINT. Only proteins with a saint score (SP>0.7) were included in the subsequent analysis. **A)** Dot blot comparison of ARS2 isoforms. Previously reported ARS2n interactors like NCBP3, PHAX and RBM7 are highlighted in blue. Shared interactors are shown in red. **B)** Venn diagram comparing ARS2 forms. **C)** Top ten interactors for each isoform.

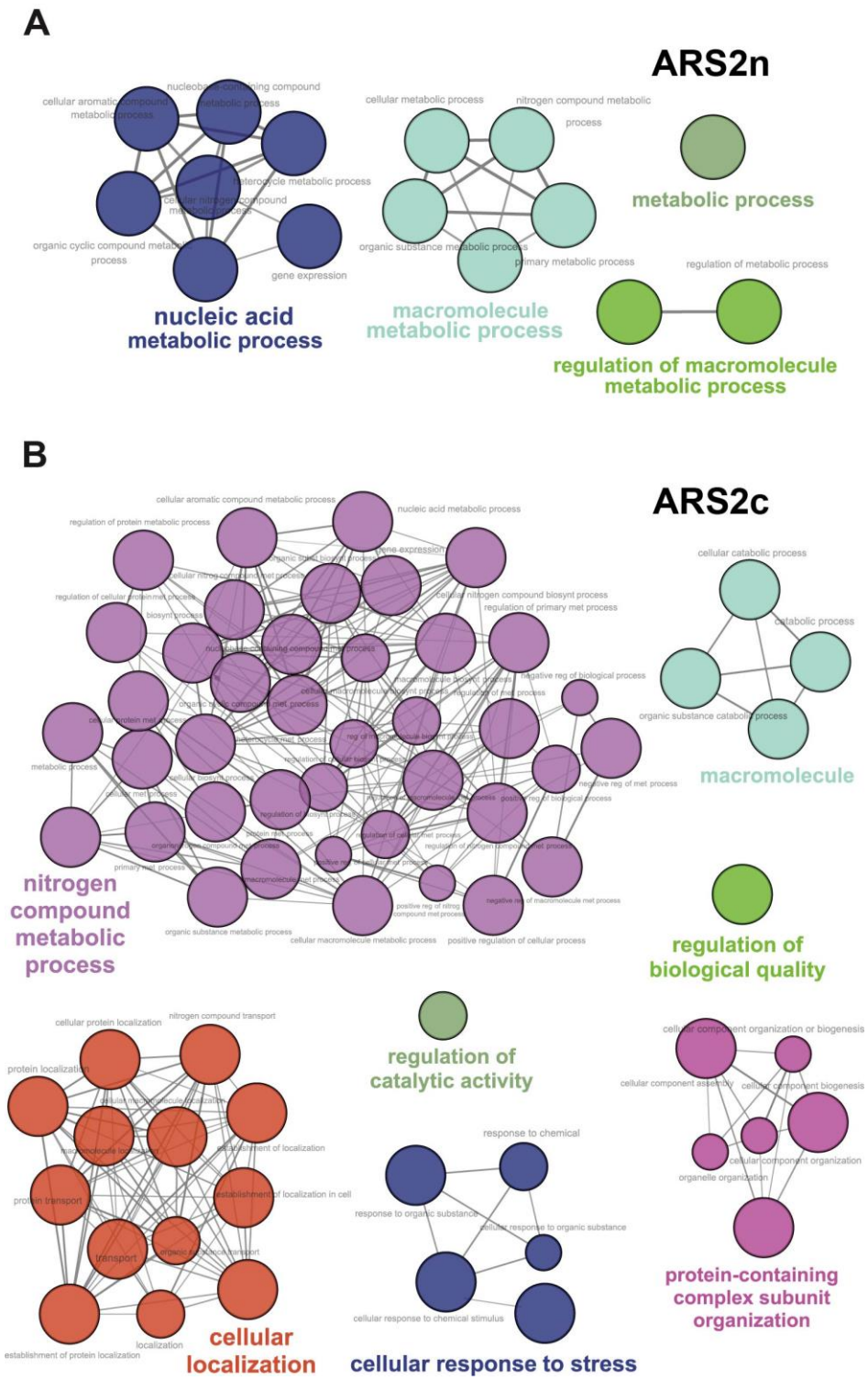


Figure 25: Gene ontology (GO) analysis of ARS2 isoform interactomes. A, B) Analysis was perfomed using the cytoscape plug-in ClueGO[259]. The size of the nodes reflects term significance after Bonferroni correction. Only GO terms with a p.value < 0.01 were considered significant and therefore represented in the network. **A)** ARS2n, **B)** ARS2c.

4.2.2 ARS2c and the cellular response to arsenic stress.

To evaluate the potential differential roles of ARS2 isoforms in the cellular response to stress, I exposed C_2C_{12} and HeLa cells to several types of stress and detected ARS2 isoforms expression by western blot and RT-qPCR. Cellular stressors like serum starvation (Fig. 26A), C_2C_{12} differentiation induction and tunicamycin-induced ER stress (Fig. 26B) did not notably affect expression levels or the ratio of ARS2 isoforms. However, puromycin-induced translation inhibition, increased ARS2c transcripts and protein levels (Fig. 26C-E). This is intriguing because the effect is specific for ARS2c and not for ARS2n or the control protein actin, which implies that ARS2c is specifically translated in a context of global translation inhibition. Similarly, we saw an increase in ARS2c expression when using early passage C_2C_{12} cells and Lipofectamine transfection agent (Fig. 17B). We discontinued the use of Lipofectamine because of toxicity, but this suggests there are likely additional stressors and/or cell states that promote ARS2c expression.

Since ARS2 has been reported to confer sensitivity to the cellular stressor arsenic[15], I evaluated ARS2 expression under arsenic treatment. Arsenic treatment suppresses global transcription and translation, and caused a notable shift in the proteome of cells and a dramatic reduction of protein diversity (Fig. 27A). After arsenic treatment, there is an accumulation of proteins between 50-75KDa (Fig. 27A). Consistent with ARS2c being important for arsenic sensitivity, expression of ARS2c is upregulated in response to arsenic, whereas expression of ARS2n, and known interactors of ARS2: CBP80 and UPF1, are downregulated (Fig. 27B, C). The induction of ARS2c expression appears to be at the transcription or RNA stability level and was inhibited by siRNAs against ARS2c (Fig. 27D).

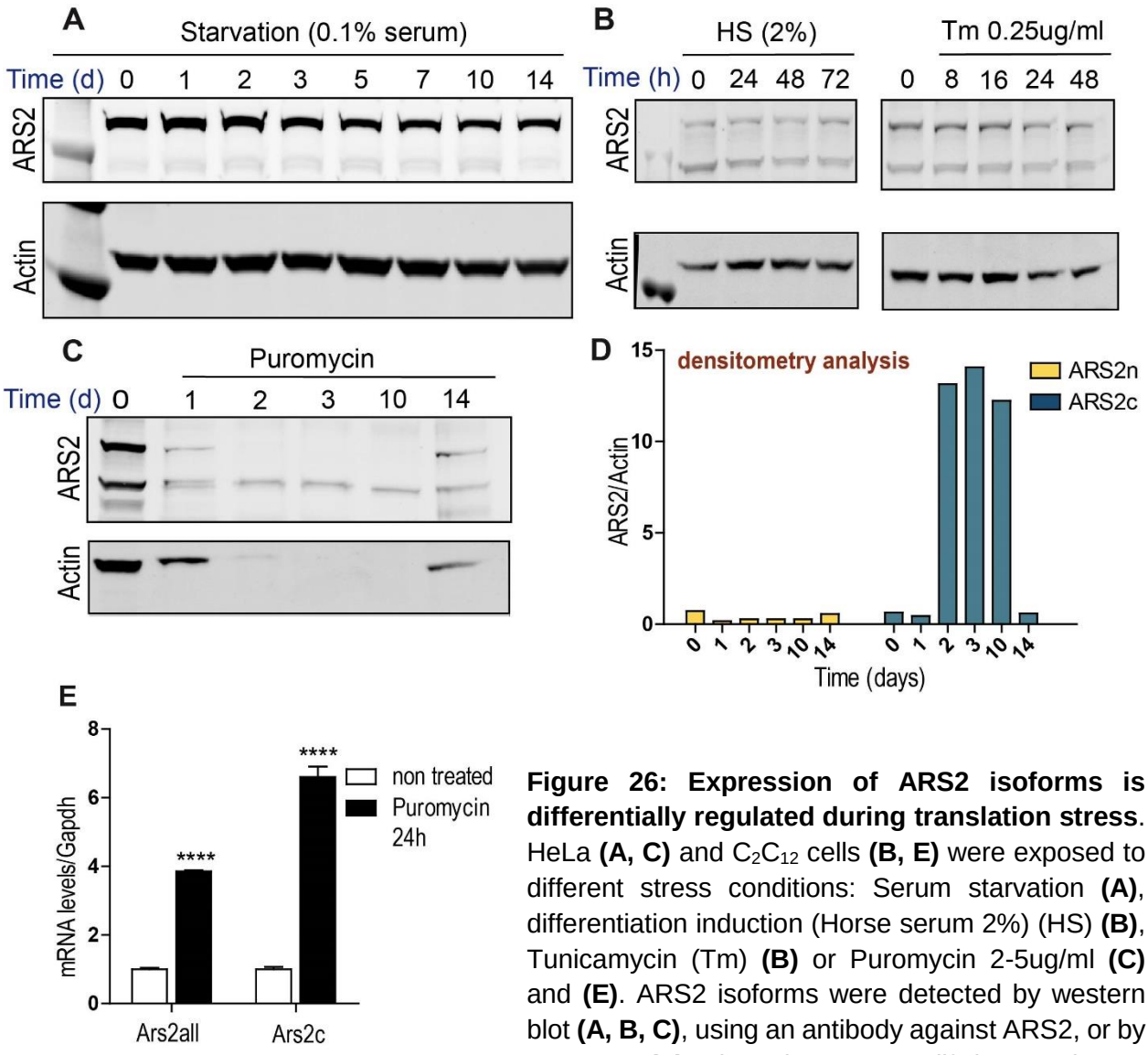


Figure 26: Expression of ARS2 isoforms is differentially regulated during translation stress. HeLa **(A, C)** and C_2C_{12} cells **(B, E)** were exposed to different stress conditions: Serum starvation **(A)**, differentiation induction (Horse serum 2%) (HS) **(B)**, Tunicamycin (Tm) **(B)** or Puromycin 2-5ug/ml **(C)** and **(E)**. ARS2 isoforms were detected by western blot **(A, B, C)**, using an antibody against ARS2, or by RT-qPCR **(E)** using primers to amplify intron 5 (*Ars2c* specific) or all *Ars2* derived transcripts. **C)** Total Protein concentration at time 0 is 30ug, but due to sample limitation, is 15ug from times 1-14 days. **D)** Densitometry analysis of **(C),** where actin was used as a normalizer. **E)** Data are represented as mean ± SEM for n=3 biologically independent samples. Statistical analysis: two-tail unpaired T-test (****p≤0.0001).

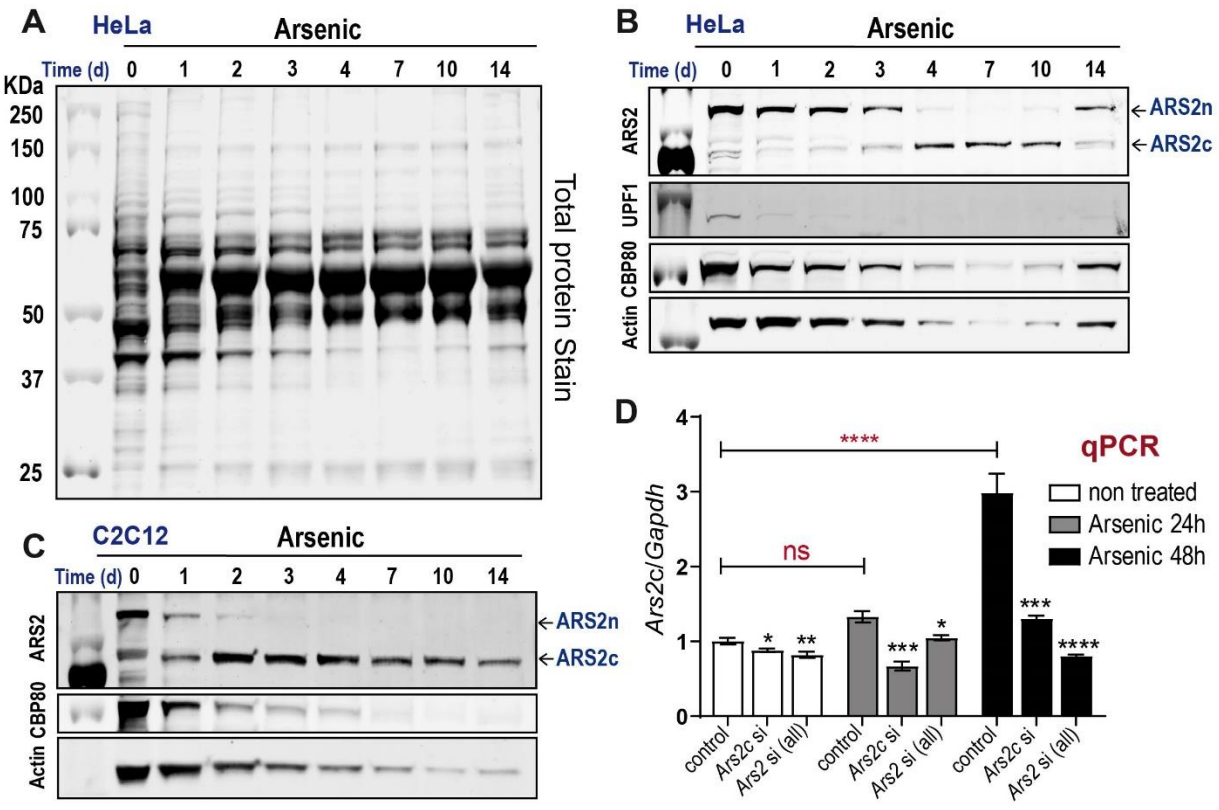


Figure 27: ARS2c expression increases during arsenic stress response. A) B) C) HeLa and C_2C_{12} cells were treated with 40uM arsenic trioxide for 14 days. Total protein expression was detected by western blot, using Revert™ 700 Total Protein Stain. ARS2, UPF1, CBP80 and Actin were detected with specific antibodies, as indicated. **D)** C_2C_{12} cells transfected with *Ars2c*/*Ars2*(all)/control RNAi were treated with arsenic trioxide for 48h. *Ars2c* was detected by RT-qPCR, using a primer against intron 5**.** Data are represented as mean ± SEM for n=3 biologically independent samples**.** Statistical analysis: One way ANOVA, post-test Tukey's (*p≤0.05; **p≤0.01; ***p≤0.001, ****p≤0.0001).

To evaluate the importance of ARS2c upregulation during arsenic treatment, I knocked down ARS2c and examined cellular survival in response to arsenic, using WST-1 and crystal violet assays. The WST-1 assay measures metabolically active cells, while the crystal violet assay evaluates cell viability based on cell attachment and membrane integrity. Depletion of cytoplasmic ARS2 (*Ars2*c si, blue) significantly increases the resistance of C_2C_{12} myoblast cells to arsenic treatment (Fig. 28A, B). Depletion of all ARS2 isoforms (*Ars2*all si, gray), also increases the resistance of C_2C_{12} to arsenic but to a lesser extent than ARS2c knockdown alone (Fig. 28A, B). Since ARS2c is similarly

knocked down by *Ars2* si (all) and *Ars2*c si (Fig. 27D), the significantly different phenotype of cells exposed to these RNAi suggests that ARS2n may be playing an opposite role in the pathway, providing resistance to arsenic. The wholescale changes in the proteome, which occur in response to arsenic, suggest that this change is important for the cellular response to the chemical. I asked whether ARS2c is required for proteomic remodelling in response to arsenic. Consistent with the importance of ARS2c in the arsenic sensitivity, downregulation of ARS2c attenuates the proteomic remodelling (Fig. 28C).

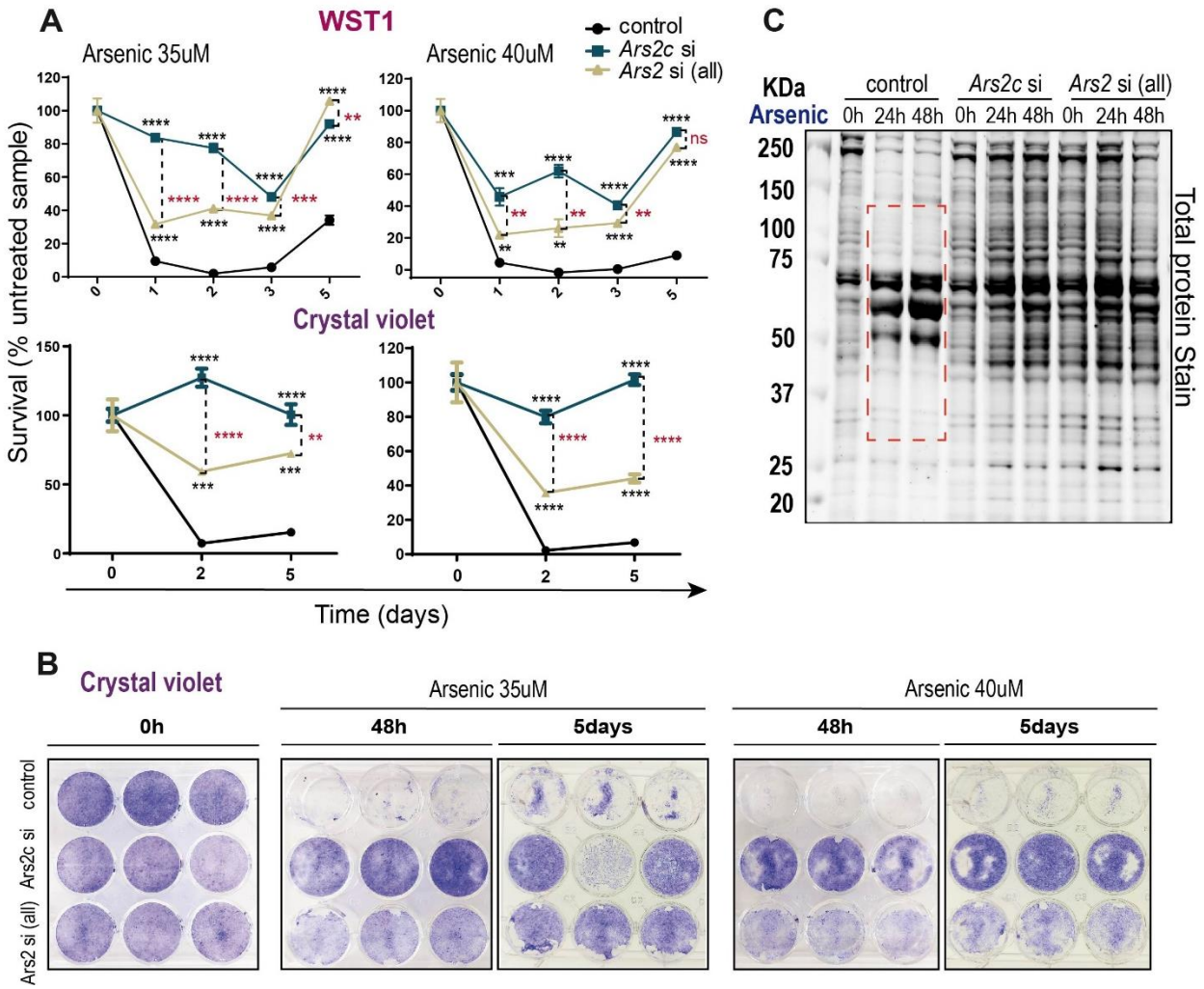


Figure 28: ARS2c is required for arsenic stress response A), B), C) C_2C_{12} cells transfected with *Ars2c*/*Ars2*(all)/control siRNAs were treated with arsenic trioxide (35uM/40uM) for 1-5 days. Survival (% treated/untreated sample) was measured by WST1 **(A)** or crystal violet **(A, B).** Data are represented as mean ± SEM for n=3 biologically independent samples for each time point**.** Statistical analysis: One way ANOVA, post-test Tukey's (*p≤0.05; **p≤0.01; ***p≤0.001, ****p≤0.0001). **C)** C_2C_{12} cells were transfected with *Ars2*c/*Ars2*(all)/control siRNAs and treated

with arsenic trioxide for 48h. Cell lysates were analyzed by Western blot and total protein was detected by Revert™ 700 Total Protein Stain**.** Each lane represents 3 pooled biologically independent samples.

Surprisingly, the ARS2c mediated arsenic response seems to be independent of the CBC complex. Downregulation of CBP80 does not increase cellular survival following arsenic treatment, and co-depletion of ARS2c and CBP80 results in an arsenic resistant phenotype comparable to depletion of ARS2c alone (Fig. 29A-C). Additionally, ARS2c knockdown had no effect on the survival following puromycin treatment (Fig. 29D), indicating that the effect of ARS2c on arsenic sensitivity is not due to the ability of both arsenic and puromycin to block translation.

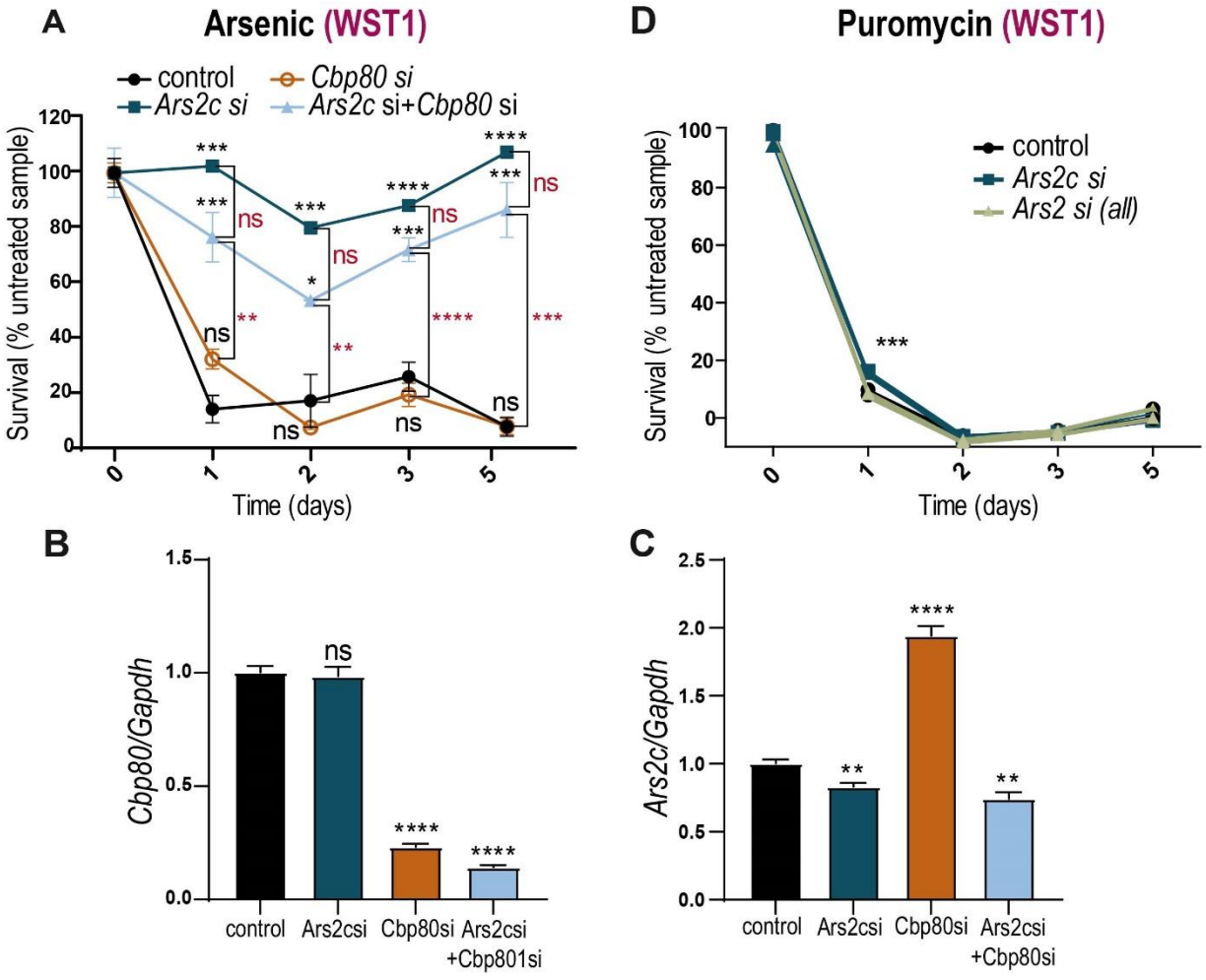


Figure 29: ARS2c mediated arsenic response is independent of the CBC. **A, D)** C_2C_{12} cells transfected with *Ars2(all)*/*Ars2c*/*Cbp80*/*Ars2c*+*Cbp80*/control siRNAs were treated with arsenic trioxide (40uM) **(A)** or puromycin 2ug/ml **(D)** for 1-5 days. Survival (% treated/untreated sample)

was measured by WST1. Data are represented as mean ± SEM for n=3 biologically independent samples for each time point**.** Statistical analysis: One way ANOVA, post-test Tukey's (*p≤0.05; **p≤0.01; ***p≤0.001, ****p≤0.0001). **B, C)** Validation of Cbp80si and Ars2c+Cbp80si combination. C_2C_{12} cells were transfected with *Ars2c*/*Cbp80*/*Ars2c*+*Cbp80*/control siRNAs for 48h. Cbp80 and Ars2c were detected by RT-qPCR. Data are represented as mean ± SEM for n=3 biologically independent samples. Statistical analysis: One-way ANOVA, post-test Dunnett's (**p≤0.01, ****p≤0.0001).

4.3 Discussion

BioID is a widely used method that allows the detection of transient, weak and temporally separated interactions[257,270]. For this reason, it is uniquely suited to detect ARS2 isoforms interactomes, since ARS2 is known to function at multiple stages in RNA metabolism promoting the formation of highly dynamic, transient, and mutually exclusive RNP complexes. ARS2n interactome returned previously known associations with splicing factors, EJC, PHAX and the export machinery, which provided an important validation of our approach. Unexpectedly, we did not find a significant enrichment of the CBC complex (CBP20/80) in our data but detected the enrichment of a newly discovered CBC complex interactor NCBP3. The presence of direct interactors of the CBC complex like NCBP3, PHAX and NELF-E in our dataset, suggests that the inability to detect CBP20 and CBP80 may be due to steric constraints from the placement of the biotin ligase at the N-terminus. I verified that CBP20 was present in BioID2-ARS2n-3flag immunoprecipitations by western blot, confirming that the BioID2 tag did not prevent ARS2n binding to CBP20/80. Thus, consistent with other studies, our data illustrates that BioID is a complementary approach to AP-MS and does not replace it.

Our interactome analyses show that ARS2 isoforms interactomes are substantially different between nuclear and cytoplasmic ARS2 isoforms. Global networks analyses highlighted significant differences in the biological process potentially regulated by ARS2 isoforms, where ARS2n interactome was fundamentally involved in RNA metabolism, while ARS2c interactome participated in other processes like the cellular response to stress. Consistent with this result, expression of ARS2c and not ARS2n was increased at both transcripts and protein levels in response to arsenic stress. Although the

mechanisms of this upregulation are still unknown and will be further discussed (**Chapter 6**), the switch in relative ARS2 isoform expression during the cellular response to stress, is biologically relevant.

Downregulation of ARS2c abrogated the arsenic-induced proteomic remodeling and arsenic sensitivity. Moreover, while specific knockdown of ARS2c strongly increased resistance to arsenic, knockdown of both isoforms decreased this resistance. This suggests ARS2n may confer resistance to arsenic and is consistent with ARS2n downregulation during arsenic treatment. Interestingly, ARS2c mediated arsenic sensitivity appears to be independent of the CBC complex. CBP80 expression, like UPF1 or ARS2n, decreased following arsenic treatment, and depletion of CBP80 had no effect on the survival of arsenic treated cells. Furthermore, co-depletion of ARS2c and CBP80 had a similar phenotype to depletion of ARS2c alone. ARS2c mediated arsenic sensitivity is, to our knowledge, the first described function for mammalian ARS2 that is potentially independent from the CBC complex.

How ARS2c participates in arsenic sensitivity and proteomic remodeling is unclear. The mechanism appears to be unrelated to the effect of arsenic on translation inhibition. Using puromycin to inhibit translation, I found that although puromycin represses expression of ARS2n and induces expression of ARS2c, knockdown of ARS2c had no effect on the survival of puromycin treated cells. Mechanistically, arsenic generates genotoxic reactive oxygen species, ER stress, promotes the induction of the unfolded protein response (UPR) and triggers apoptosis[206]. Interestingly, our BioID data showed interaction between ARS2c and proteins involved in cell redox homeostasis, protein folding, and response to unfolded protein (Fig. 30). How such interactions regulate the cellular response to arsenic will require further study.

The newly described function of cytoplasmic ARS2 as a critical mediator of arsenic sensitivity clarifies a 20-year-old question in the field, demonstrating that ARS2c isoforms are more associated with this phenotype. Furthermore, the involvement of ARS2c in arsenic sensitivity could have implications for cancer treatment, and this subject will be discussed in detail in the Future Directions section (**Chapter 6**). Arsenic trioxide (ATO) is an approved treatment for Acute Promyelocytic Leukemia and it is currently under study

for several other locations, including pancreatic and ovarian cancer. Thus, relative ARS2 isoform expression in tumours may serve as a prognostic factor to stratify those patients most likely to benefit from ATO treatment.

Chapter 5: ARS2 isoforms regulation of shared biological processes.

Contributions: Bioinformatic analysis (Fig. 30) was performed by Phineas Hamilton, Deeley Research Centre, BC Cancer Agency. Tethering assay optimization and evaluation were performed by Jennifer Christie and Connor O'Sullivan (Fig. 32C, D), Perry's Howard lab. Figure 32B diagram was adapted from Jennifer Christie master's thesis[112]. All the remaining experiments and analyses were performed by MMP.

The data in this chapter contributed to the publication:

Mesa-Perez, M., Hamilton, P. T., Miranda, A., Brodie, N., O'Sullivan, C., Christie, J., Ryan, B., Chow, R., Goodlett, D., Nelson, C. and Howard, P. Cytoplasmic Switch of ARS2 Isoforms Promotes Nonsense-Mediated mRNA Decay and Arsenic Sensitivity, *Nucleic Acids Research, 2022*;, gkac033, https://doi.org/10.1093/nar/gkac033

5.1 Introduction

The life of an mRNA is highly coordinated from its beginnings during transcriptional initiation to its inevitable degradation. In the nucleus, the coordination is guided in part by the CBC complex (CBP80/20) and ARS2n, which help to synchronize the dynamic assembly and disassembly of mutually exclusive complexes that regulate mRNA splicing, degradation, and export[11,13,15,23,81,102,271]. CBC-ARS2n bound mRNAs are exported to the cytoplasm, where CBP80 is required for the pioneer round of translation and for promoting transcript degradation through nonsense mediated decay (NMD)[106,108,272]. Although ARS2n is translocated to the cytoplasm with the CBC, it is unknown whether ARS2n participates in CBC-related functions in the cytoplasm.

As discussed in **Chapter 1**, NMD is a translation-dependent surveillance process that promotes the cytoplasmic degradation of mRNA with premature termination codons (PTC), to prevent the synthesis of dysfunctional proteins and maintain mRNA

homeostasis[273,274]. During nuclear splicing, the multiprotein Exon Junction Complex (EJC) is deposited onto the mRNAs 24nt upstream of the spliced junctions[275]. EJCs are exported to the cytoplasm with the mRNAs and are dissociated by the ribosome during translation. Mammalian-cell NMD generally occurs when translation terminates more than 50–55nt upstream of an EJC complex[191]. In this scenario, the termination codon is recognized as a PTC, which triggers RNP recruitment and remodelling. First, UPF1 is recruited to the stalled ribosome by ERF1, ERF3 and CBP80. Subsequently, SMG1 kinase joins the complex to generate the SURF complex (SMG1–UPF1– ERF1–ERF3). Next, in a step dependent on DHX34 and CBP80, UPF1 interacts with the downstream UPF2/3 within the EJC to form the decay-inducing (DECID) complex, triggering SMG1 activation. The phosphorylation of UPF1 by SMG1 leads to translation inhibition and promotes mRNA degradation through the recruitment of SMG6, SMG5–SMG7 and mRNA decay factors[140]. Importantly, inhibition of the NMD downstream of the DECID complex results in UPF1 hyperphosphorylation[197]. Previous work has shown that CBC complex component CBP80 promotes NMD in the cytoplasm[108,272]. Additional studies have reported the interaction between ARS2 and several components of the NMD pathway[21,59,102]. However, the question of whether ARS2 participates in NMD has not been addressed.

In this chapter I evaluate the potential role of ARS2 isoforms in NMD, since components of this pathway were enriched in both ARS2n and ARS2c interactomes. Confirming previous reports in the literature, ARS2n interacts with components of the EJC. ARS2c, on the other hand, associates with cytoplasmic components of the NMD pathway such as UPF1, ERF1 and the translation initiation machinery. Our data demonstrates that both ARS2 isoforms regulate NMD, but strikingly they exert opposite roles: ARS2n inhibits, while ARS2c promotes the NMD pathway. Furthermore, ARS2 isoforms are functionally distinct, and their distinctness is partially explained by the differences in their N-terminal region.

Our results support a model in which ARS2n and ARS2c work in tandem to modulate the NMD pathway. ARS2n primarily functions in the nucleus with the CBC to regulate mRNA splicing, degradation, and export[11,13,15,23,81,102,271]. We propose that once the mRNPs

enter the cytoplasm, an isoform switch occurs in which ARS2c replaces ARS2n within the CBC. ARS2c, in association with the CBC component CBP80, promotes the interaction between UPF1 and ERF1/CBP80, favoring the formation of the SURF complex at the PTC, SMG7 recruitment, and transcript degradation though NMD. Thus, the cytoplasmic switch of ARS2 isoforms tailors the protein to function in a nonredundant manner in nuclear and cytoplasmic CBC-dependent processes.

5.2 Results

5.2.1 ARS2 isoforms interact with different components of NMD pathway

ARS2n is a scaffold protein that joins the CBC to direct RNAPII transcript processing, degradation, and export. To better understand the role of ARS2c isoforms, we asked whether they share functions with ARS2n. GO analysis of the BioID interactome data (**Chapter 4**), showed that both ARS2n and ARS2c isoforms interact with components of the RNA splicing machinery, RNA transport and RNA catabolic process, including nonsense mediated decay (NMD) (Fig. 30, Table 5). Since ARS2n role in RNA splicing and transport has been evaluated in different studies[11,15,81,102], I decided to focus on NMD, where ARS2n's function is yet to be established.

Our proteomic analysis detected proximity interactions of ARS2n with core EJC components MAGOH, MAGOHB and CASC3 (Fig. 31A, left). In contrast, ARS2c isoform interactome was enriched for NMD components UPF1 and ERF1 (Fig. 31A, right). I confirmed the novel interactions between ARS2c and ERF1, UPF1, as well as components of the translation initiation complex EIF3K/E/F, by immunoprecipitation and western blotting (Fig. 31B, C). Previous studies have shown that ARS2n interacts with all four core components of the EJC: CASC3, MAGOH, RBM8A and EIF4A3, the EJC periferal factors: UPF3B, RNPS1, PNN and ACINUS as well as UPF1 and the phosphatases PPP2R1A and PPP2R2A (Fig. 32A)[21,59,102,103,184,276,277]. To evaluate the functionality of these interactions, a tethering assay was performed.

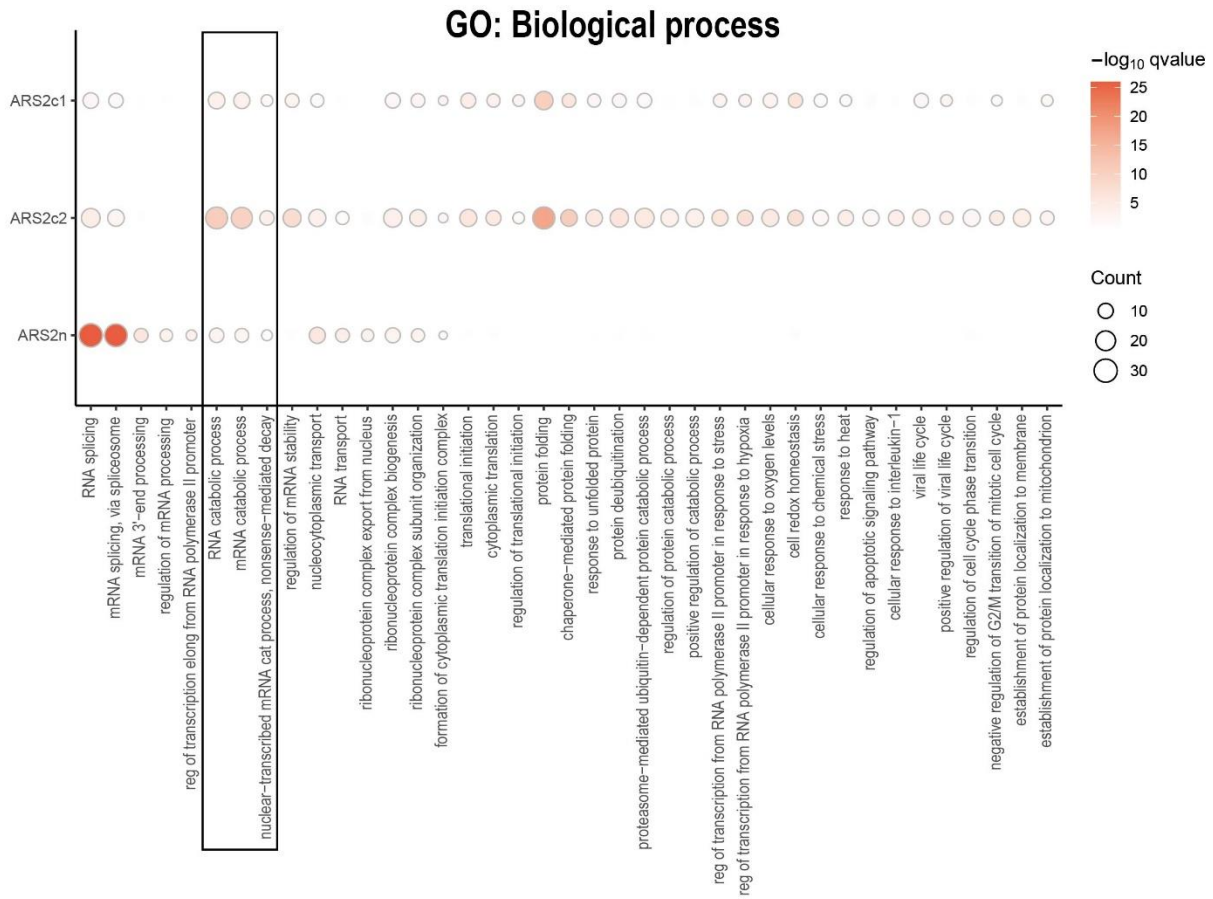


Figure 30: ARS2 isoforms interactomes are enriched in components of the NMD pathway. Gene Ontology (GO) analysis of ARS2 isoforms interactomes. Enriched biological process are shown (p.adjust < 0.05). Relative counts and significance of enrichment are represented for each process. Full list of enriched biological processes is included in Table 5.

In the tethering assay, the protein of interest (ARS2n) is fused with a λN tag and binds to a 5x-B-Box in the 3' UTR of a luciferase reporter (Fig. 32B-D). If ARS2n interacts with EJC components, binding of ARS2n to the 3'UTR will recruit the EJC to that region, triggering the degradation of the reporter through NMD (Fig. 32B). Tethering of ARS2n to the luciferase reporter resulted in reporter degradation, similar to the EJC component RNPS1 (Fig. 32C). To confirm that the observed degradation was dependent on NMD, we knocked down *UPF1*, a key component of the NMD pathway. As shown in Figure 32D, knockdown of *UPF1* rescued the expression of the reporter, indicating that the effect of ARS2n is partially dependent on UPF1. Thus, the tethering assay supports the

functionality of ARS2n-EJC interaction and suggests a potential role of ARS2n in NMD. Since both ARS2n and ARS2c interact with NMD components, they are both likely to be positive with this assay and we did not pursue the testing of ARS2c. Instead, we focussed our efforts on more specific assays that evaluate the direct modulation of NMD by ARS2 isoforms, in a more native context.

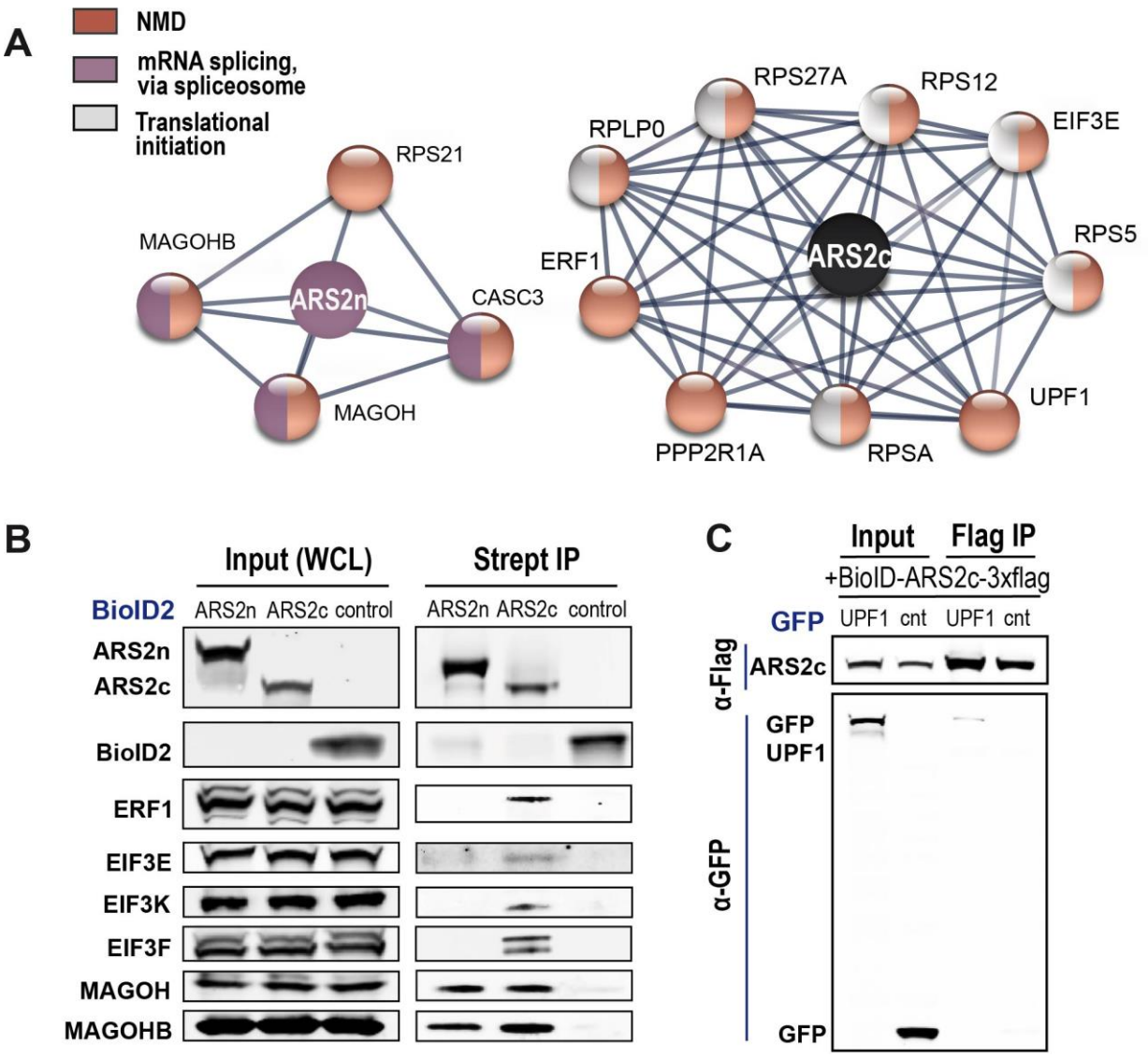


Figure 31: ARS2 isoforms interact with different components of NMD pathway. A) NMD pathway components interacting with the isoforms are represented using STRING[278]. **B) C)** Western Blot confirmation of detected interactors using the specified antibodies. **B)** HEK 293T cells were transfected with BioID2-ARS2n/ARS2c-3xflag or control and treated with biotin for 24h. Biotinylated interactors were pulled down with streptavidin beads and proteins were detected as indicated. **C)** HEK 293T cells were co-transfected with BioID2-ARS2c-3xflag and GFP-

UPF1/GFP. Samples were pulled down using anti-Flag magnetic beads. BioID-ARS2c-3xflag and UPF1 were detected with anti-Flag and anti-GFP antibodies, respectively.

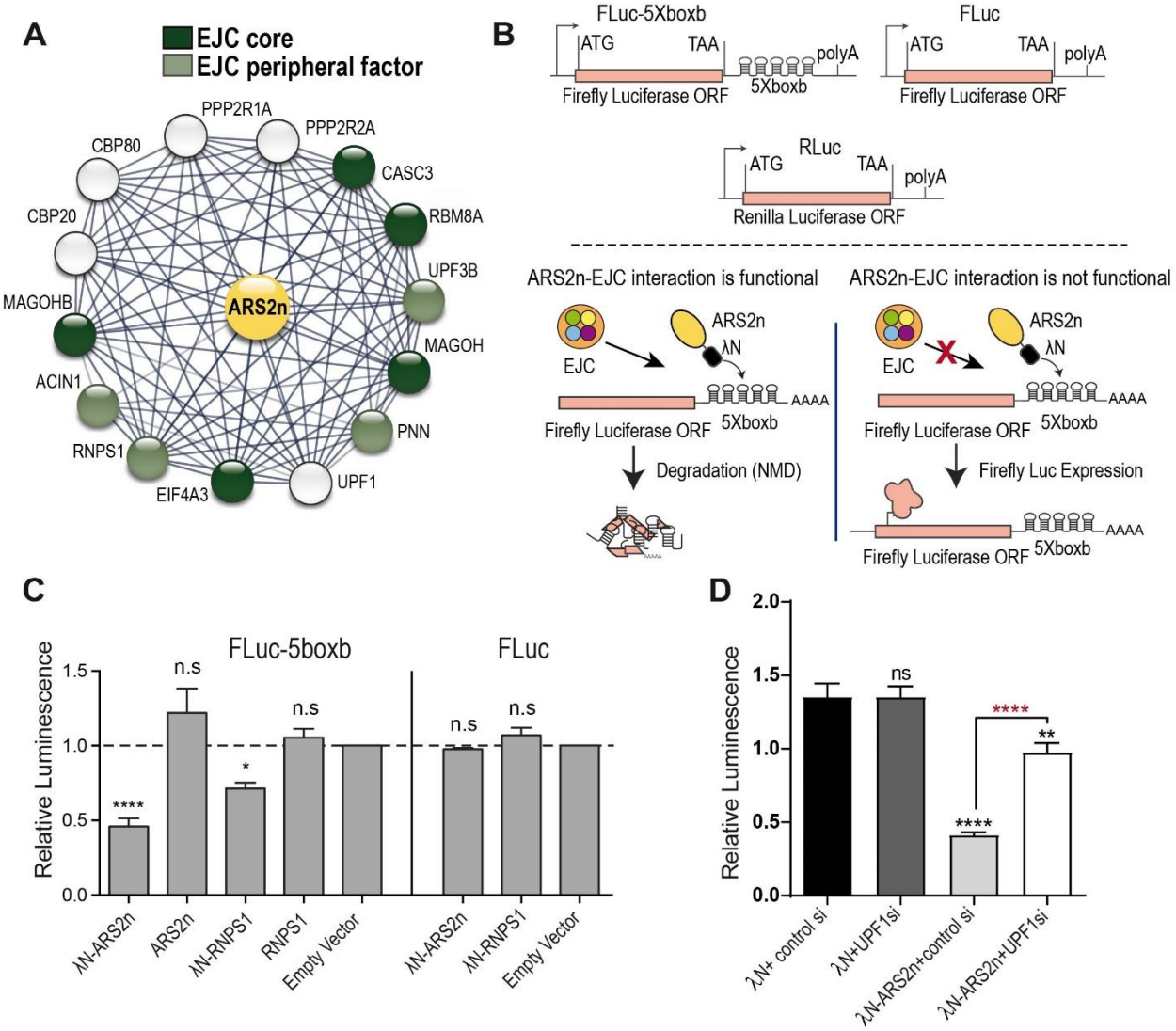


Figure 32: ARS2n has functional interactions with EJC components. A) STRING representation of the interactions between ARS2n and NMD components, reported in the literature. **B)** Schematic representation of the tethering assay. ARS2n-λN fusions are tethered to 5xBoxB sites in the 3'UTR of a Firefly luciferase reporter. If ARS2n recruits the EJC or other NMD components to the reporter, the reporter is degraded through NMD resulting in a loss of Firefly luciferase activity. **C)** HeLa cells were transfected with λN-ARS2n/ARS2n/λN-RNPS1/ RNPS1/control + Fluc-5boxb/Fluc+ Renilla luciferase (normalizer). Normalized luciferase activity was measured at 24h after transfection. **D)** HeLa cells were transfected with λN/λN-ARS2n + control si/*UPF1* si + Firefly luciferase reporter + Renilla luciferase. Normalized luciferase activity was measured at 24h after transfection. Data are represented as mean ± SEM for n=3 **(C)** or n=9

(D) biologically independent samples. Statistical analysis: One way ANOVA, post-test Dunnett's **(C)**, post-test Tukey's **(D)** (*p≤0.05; **p≤0.01; ****p≤0.0001).

5.2.2 ARS2 isoforms have opposite roles in NMD

To gain a better understanding of the role of nuclear and cytoplasmic ARS2 isoforms in NMD, I used a NMD dual fluorescence reporter[250]. The pNMD+ vector contains TagGFP2 cDNA fused to exon 2 and exon 3 of human β-globin, with an intron between them. Splicing of the β-globin intron deposits an exon junction complex (EJC) >50 nucleotides downstream of the termination codon of TagGFP2. Consequently, the TagGFP2 termination codon is recognized as a premature termination codon (PTC) and TagGFP2 mRNA is degraded by NMD. The pNMD- vector is a negative control plasmid where exon 2 is only 35 nucleotides long. After splicing, the distance between the termination codon of TagGFP2 and the EJC will be <50 nucleotides, making TagGFP2 mRNA insensitive to NMD. To normalize for transfection efficiency, the vectors also contain an expression cassette for the far-red protein Katushka (TurboFP635) (Fig. 33A, B). To validate the reporter in our conditions, I transfected HeLa cells with either pNMD+ or pNMD- plasmids. Superposition of the density plots of pNMD+ and pNMD- transfected cells (Fig. 33C), shows that the mean TagGFP2 intensity in pNMD+ transfected cells (~10^2) is decreased relative to the mean TagGFP2 intensity in cells transfected with the pNMD- plasmid (~10^3). As expected, the mean intensity on the Katushka channel is constant for pNMD+ or pNMD- transfected cells. This indicates that the differences observed in the green channel are due to NMD and not due to transfection efficiency differences. Histograms plots for TagGFP2 and Katushka channels at 24h and 48 are shown in Figure 33D. TagGFP2 is progressively degraded in the cells transfected with the pNMD+ plasmid (left), while Katushka expression remains constant in both pNMD+/pNMD- transfected cells (right).

In our experiments, HeLa cells were co-transfected with (pNMD+ and *UPF1/ARS2c/ARS2*/control si) or (pNMD- and *UPF1/ARS2c/ARS2*/control si) (Fig. 33E-K). For each sample, TagGFP2 is normalized to Katushka and then expressed as the ratio of pNMD-/pNMD+. The pNMD-/pNMD+ ratio represents NMD activity, where higher

ratio corresponds to higher NMD activity. By normalizing pNMD- vector to pNMD+ vector, the NMD activity calculation focusses on NMD and excludes from the analysis any potential effects of ARS2n or ARS2c on mRNA transcription, splicing, processing, transport and NMD- independent degradation.

Using this reporter system, I found that ARS2c and ARS2n have opposing impacts on NMD. Specific knockdown of *ARS2*c, promotes a decrease in the NMD activity, as measured by flow cytometry and RT-qPCR at 48h (Fig. 33E, I). A similar effect was observed with downregulation of UPF1 (Fig. 33G, K), suggesting that ARS2c isoforms, like UPF1, are required for NMD. I confirmed the role of ARS2c in NMD is conserved in mouse cells by testing this reporter assay in C_2C_{12} cells. A similar NMD activity decrease was seen in these cells (Fig. 33H). There are no sequences unique to *ARS2*n to specifically target this isoform by RNAi. Thus, to test the impact of ARS2n on NMD I used RNAi to exon 4, which disrupts expression of both *ARS2*n and *ARS2*c (Fig. 18A). Surprisingly, this RNAi had the opposite effect to the specific knockdown of *ARS2c*, and increased NMD activity as shown at both protein and transcript levels (Fig. 33F, J). Collectively, these experiments suggest that ARS2n inhibits the NMD pathway, while ARS2c promotes it.

To confirm the opposing roles of ARS2n and ARS2c in NMD, I next examined the effect of isoform overexpression and downregulation on endogenous genes that are naturally regulated by NMD. Consistent with a promoting role in NMD, ARS2c overexpression increased the activity of the NMD pathway and promoted degradation of endogenous NMD targets (Fig. 34A), an effect that was counteracted by downregulation of UPF1 (Fig. 34B). Conversely, both UPF1 and ARS2c depletion inhibited the activity of the NMD pathway and favored the accumulation of endogenous NMD targets (Fig. 34C, D). These results confirm that ARS2c, similar to UPF1, promotes the NMD pathway. In contrast, overexpression of ARS2n decreased the activity of the NMD pathway and induced the accumulation of NMD regulated transcripts (Fig. 34E). Consistent with an inhibitory role, downregulation of ARS2n increased NMD activity and promoted the degradation of endogenous NMD targets (Fig. 34F). These results were observed in both human HeLa

and mouse C_2C_{12} cells, confirming the effects of the isoforms on endogenous targets are conserved between the species.

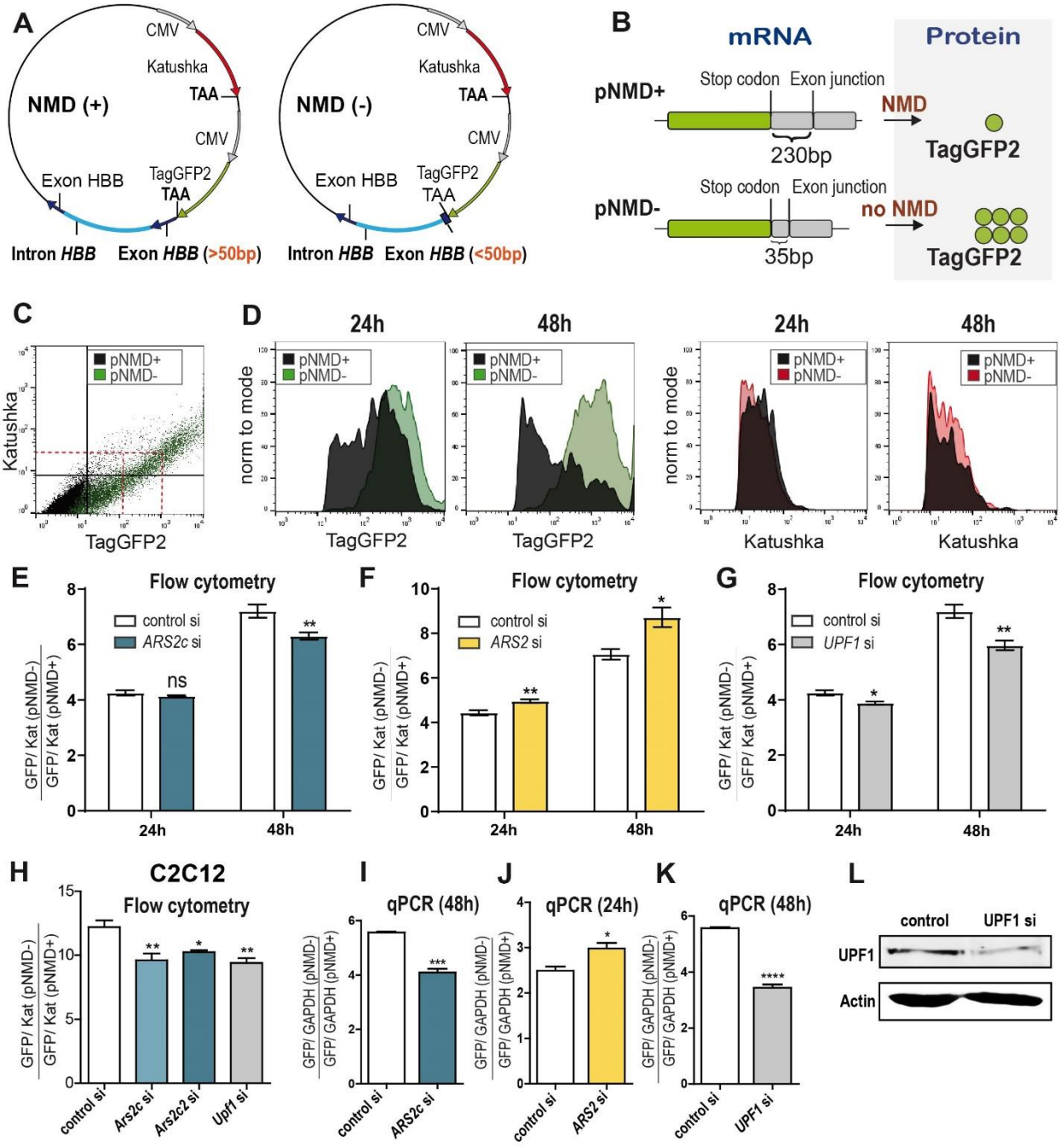


Figure 33: ARS2 isoforms have opposite roles in the regulation of NMD. A), B) Schematic representation of pNMD+/pNMD- reporters. pNMD+ has an EJC deposited >50 nucleotides from the TagGFP2 stop codon, which leads to TagGFP2 degradation by NMD. pNMD- has an EJC placed <50 nucleotides from TagGFP2 stop codon, making TagGFP2 insensitive to NMD

degradation. Both reporters also contain the protein Katushka, used as a transfection normalizer. **C), D)** Flow cytometry showing the functionality of the reporter. **C)** Superposition of density plots (TagGFP2 and Katushka channels) for pNMD+ or pNMD- transfected cells. Mean intensities in both channels are represented with red discontinuous lines to facilitate visualization. **D)** Histogram plots of the TagGFP2+/Kat+ cell population, for TagGFP2 and Katushka channels at 24 and 48h. **E-K)** HeLa **(E, F, G, I, J, K)** and C_2C1_2 **(H)** cells were transfected with pNMD+/pNMD- and *ARS2c*/*Ars2c2*/*ARS2*/*UPF1* or control RNAi. Expression levels were detected by Flow cytometry **(E-H)** or RT-qPCR **(I-K)**. TagGFP2 levels were normalized to Katushka or *GAPDH*, shown are the pNMD-/pNMD+ ratios. Data are represented as mean ± SEM for n=3 or n=6 biologically independent samples. Statistical analysis: two-tail unpaired t-test **(E, F, G, I, J, K)** or One-way ANOVA, post-test Dunnett's **(H)** (*p≤0.05; **p≤0.01; ***p≤0.001; ****p≤0.0001). **L)** Western blot validation of UPF1 siRNA. Endogenous UPF1 is detected using anti-UPF1 antibodies. Actin is used as a normalizer.

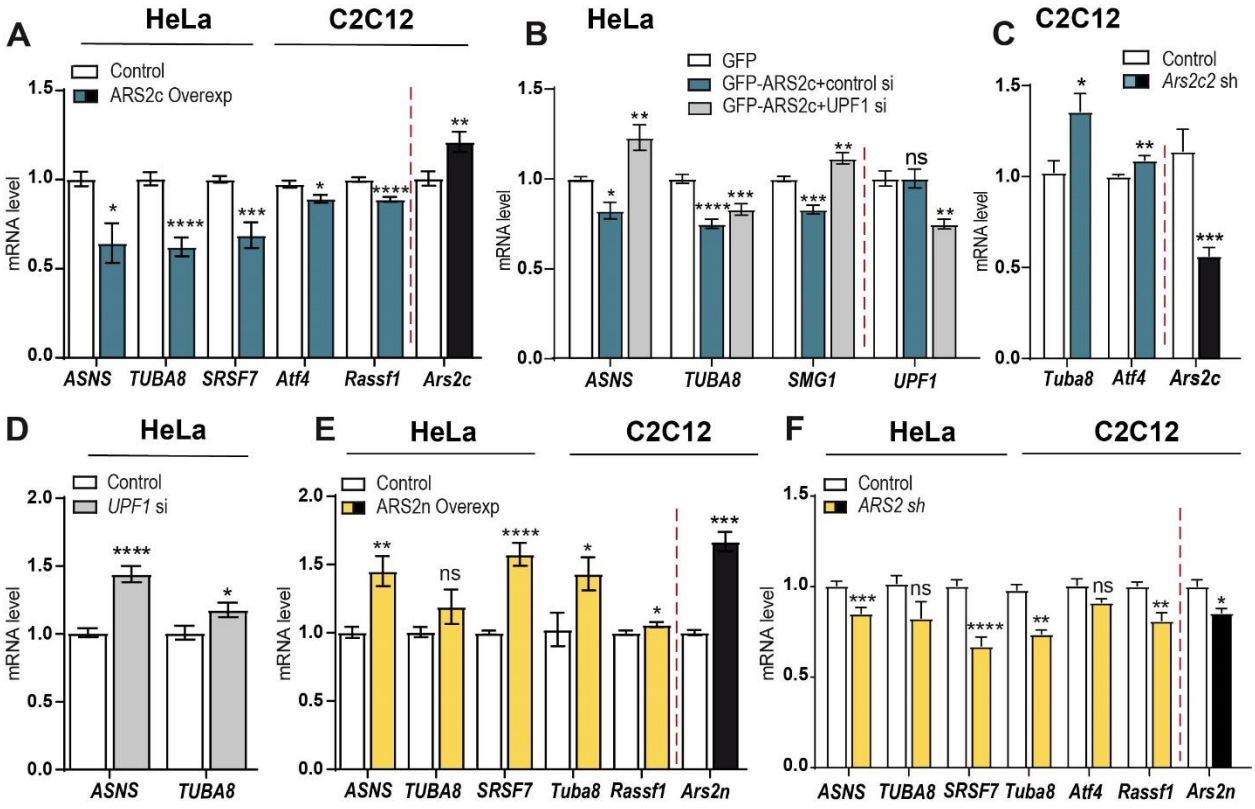


Figure 34: ARS2c promotes, while ARS2n inhibits the degradation of endogenous transcripts targeted by NMD. A-F) HeLa or C_2C_{12} cells were transfected with GFP-ARS2c/GFP-ARS2n/GFP/GFP-ARS2c+UPF1si/GFP-ARS2c+controlsi or *ARS2*sh/*Ars2c*2sh/*UPF1*si/negative controls for 48h. Expression levels were detected by RT-qPCR and each gene was normalized against *GAPDH*. Data are represented as mean ± SEM for n=3 to n=6 biologically independent samples (see Materials and Methods section). Statistical analysis: Two-tail unpaired t-test (**A, C, D, E, F)**, One-way ANOVA, post-test Dunnett's **(B)** (*p≤0.05; **p≤0.01; ***p≤0.001, ****p≤0.0001).

I next demonstrated that the effect of ARS2 isoforms on the pathway was NMD dependent and not due to off-targeting. I used a set of primers to amplify alternatively spliced transcripts of the endogenous *HNRNPA2AB1* gene[279] in HeLa or in a Flp-In T-REx (ARS2n) inducible cell line, in which ARS2n-3xFlag has been knocked into an *FRT* locus within the cell line. The *HNRNPA2AB1* gene generates two alternative spliced transcripts represented here as NMD- and NMD+ (Fig. 35A). Similar to the pNMD-/+ reporter (Fig. 33), the NMD- transcript is insensitive to NMD, while the NMD+ product is degraded by NMD. The total transcript levels are measured by an amplicon common to all the splicing variants. As expected, downregulation of UPF1 results in an accumulation of the NMD+ transcript (Fig. 35B). Supporting its NMD promoting role, overexpression of ARS2c preferentially induces the degradation of the NMD+ transcript (Fig. 35C). In contrast, overexpression of ARS2n promotes the accumulation of the NMD+ transcript, while downregulation of ARS2n preferentially induces NMD+ degradation (Fig. 35D, E). To confirm that the effects of *ARS2n* knockdown on the pathway were specific to *ARS2n*, I repeated the experiment but induced expression of ARS2n-3xFlag with the addition of tetracycline. Restoring *ARS2*n levels rescued the expression of the *HNRNPA2AB1* NMD+ transcript, showing that the inhibitory effect of *ARS2n* is specific. (Fig. 35F). In summary, these results show ARS2n inhibits while ARS2c promotes the NMD pathway.

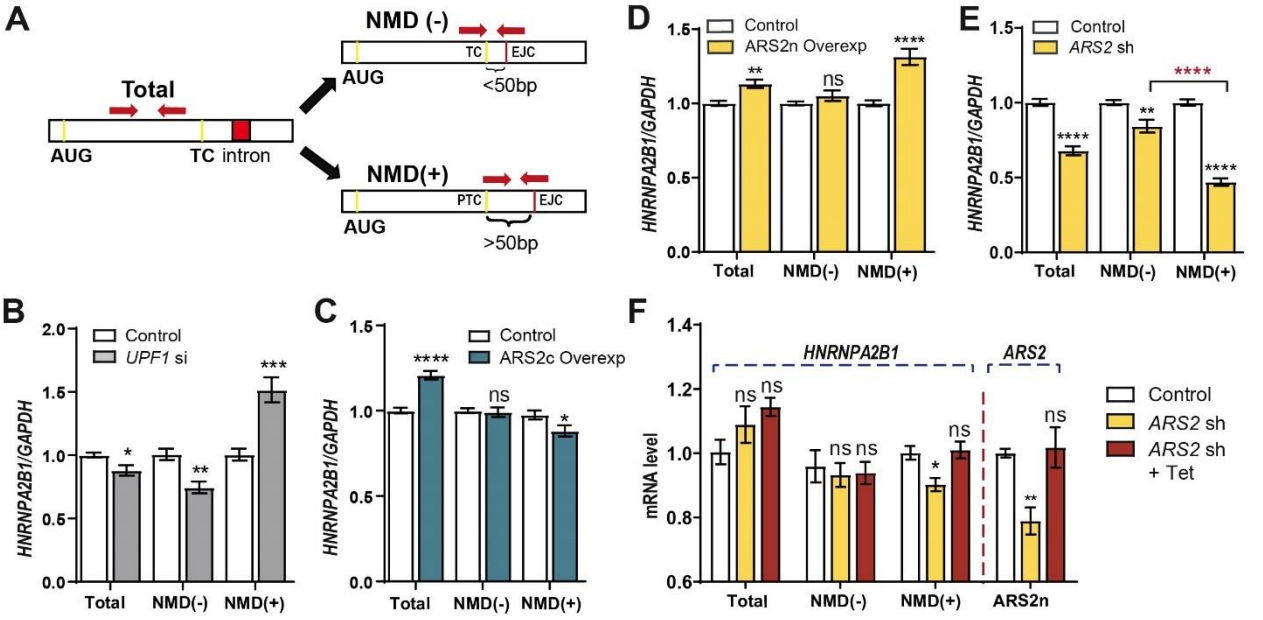


Figure 35: ARS2 isoforms differentially regulate the degradation of PTC containing transcripts. A) Schematic representation of primers used **(B-F)**. HeLa cells were transfected with

GFP-ARS2c/GFP-ARS2n/GFP or *ARS2*sh/*UPF1*si/ negative controls for 48h. **F)** Flp-In T-REx 293 cells inducibly expressing *ARS2*n were transfected with *ARS2* sh or control for 48h and treated with 2.27ng/ml of tetracycline for 36h to induce *ARS2n* to endogenous levels. Expression levels were detected by RT-qPCR and each gene was normalized against *GAPDH*. Data are represented as mean ± SEM for n=3 biologically independent samples. Statistical analysis: Two-tail unpaired t-test **(B-E)**, One way ANOVA, post-test Dunnett's **(F)** (*p≤0.05; **p≤0.01; ***p≤0.001, ****p≤0.0001).

5.2.3 ARS2 isoforms are functionally distinct and work in tandem to regulate NMD.

To understand how ARS2 isoforms differentially regulate NMD, we first asked whether the inhibitory role of ARS2n on the NMD pathway was dependent on its nuclear localization. I narrowed the nuclear localization signal region from 105 amino acids[83] to 12, and generated an ARS2n mutant which lacks amino acids (73-LSPPQKRMRRDW-84) (Fig. 36A). As shown in Fig. 36B, deletion of these 12 amino acids localizes ARS2n to the cytoplasm. Interestingly, deletion of the nuclear localization signal (NLS) was sufficient to abrogate the inhibition of NMD caused by ARS2n overexpression, suggesting that the inhibitory effects of ARS2n on the pathway originate in the nucleus (Fig. 36C).

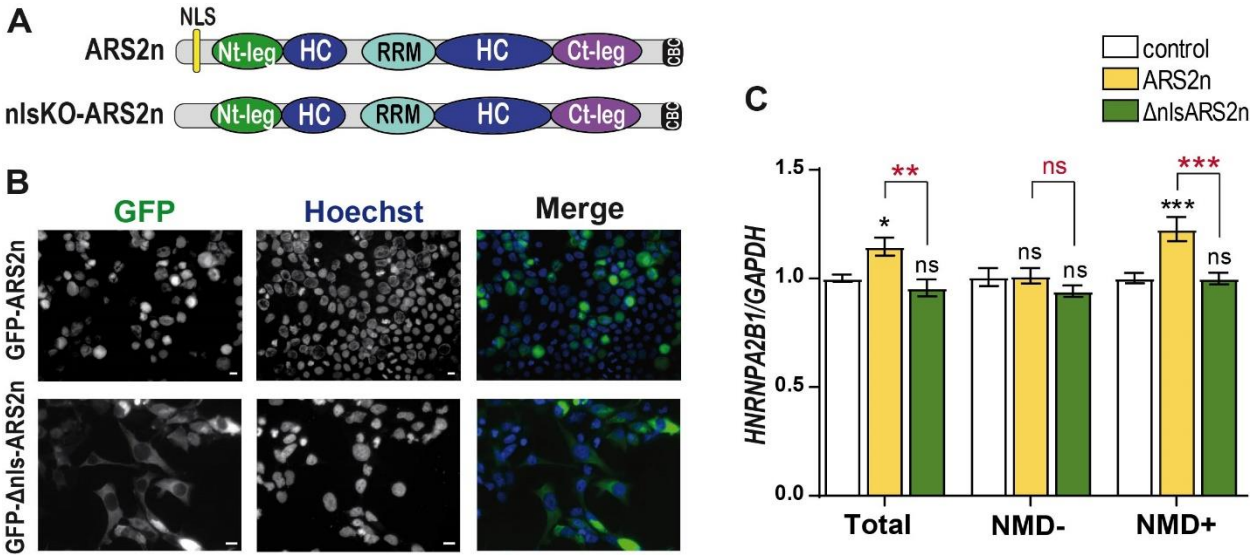


Figure 36: Inhibitory effect of ARS2n on NMD pathway, originates in nucleus. A) Schematic representation of nlsKO-ARS2n mutant. **B)** HEK 293T cells were transfected with GFP-ARS2n/ GFP-Δnls-ARS2n. GFP fusion proteins are shown in green and nuclei were stained with Hoechst 33342 (blue). Scale bar = 10μM. **C)** HeLa cells were transfected with GFP-ARS2n/ GFP-Δnls-

ARS2n or GFP control for 48h. Expression levels were detected by RT-qPCR. *HNRNPA2AB1* gene isoforms were detected using Fig. 35A primers and each gene was normalized against *GAPDH*. Data are represented as mean ± SEM for n=6 biologically independent samples. Statistical analysis: One way ANOVA, post-test Tukey's (*p≤0.05; **p≤0.01; ***p≤0.001).

To evaluate the mechanism of the isoforms and mutant Δnls-ARS2n on NMD, I looked at their ability to interact with UPF1 from whole cell lysates. All three ARS2 proteins interacted with UPF1. However, ARS2c was more enriched in UPF1 pull-downs than ARS2n or the ΔNLS mutant (Fig. 37A, 39A). To test whether the interaction with UPF1 takes place in the cytoplasm, I performed endogenous UPF1 pull-downs on cytoplasmic fractions of cell lysates (Fig. 37B-H). Both ARS2 isoforms and the mutant were detected in cytoplasmic UPF1 pull-downs (Fig. 37B). I adjusted protein concentrations (Fig. 37C-F) and compared ARS2 isoforms/mutant expression in the input and the supernatant (SN), obtained after UPF1 pull-down (Fig. 37G). Consistent with Figure 37A results, ARS2c is about 2 times more depleted in the SN than ARS2n or the mutant, suggesting that ARS2c interacts with UPF1 in the cytoplasm, more than ARS2n or the mutant (Fig. 37G). Expression of ARS2 isoforms and mutant in nuclear and cytoplasmic fractions are shown in Figure 37H.

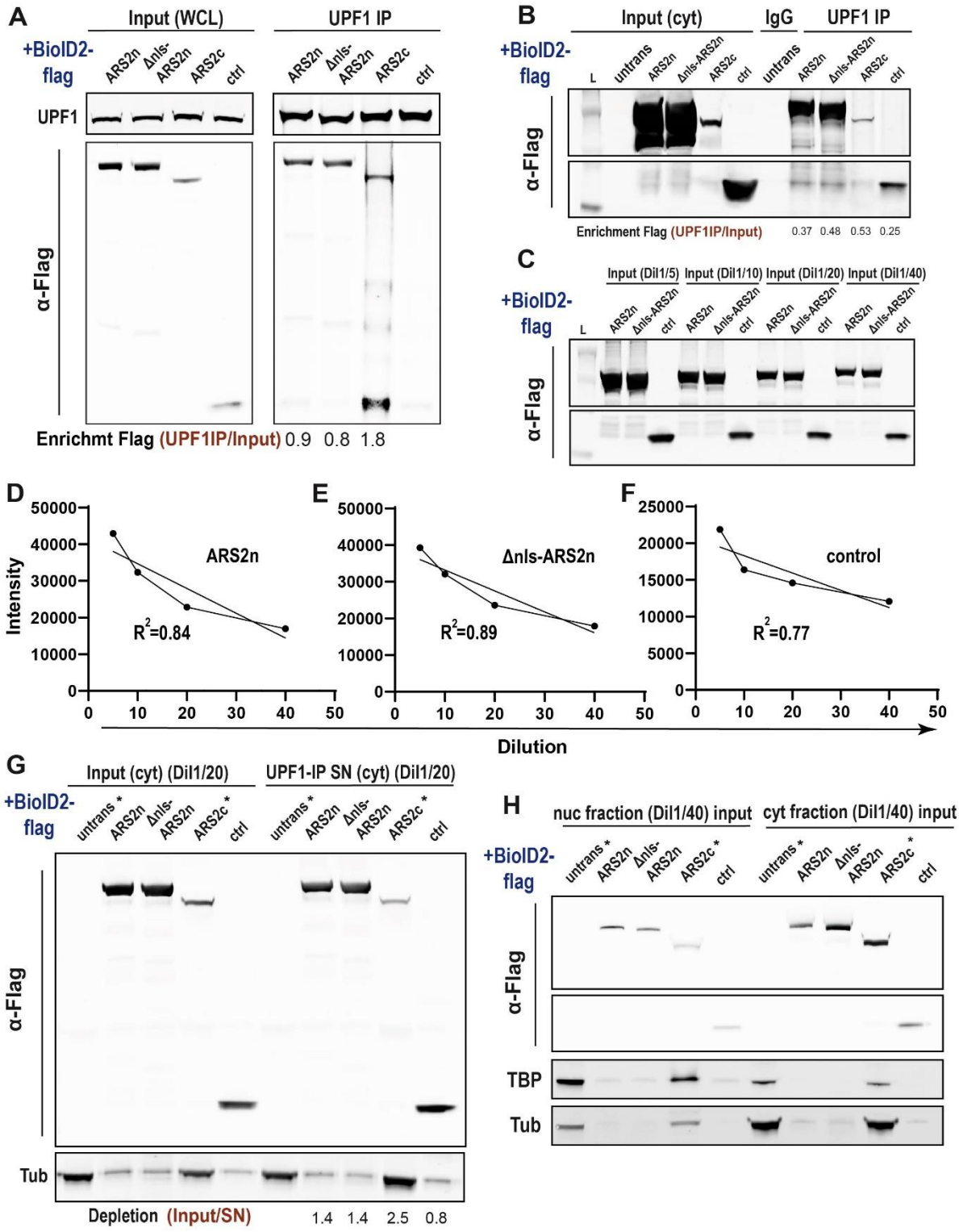


Figure 37: ARS2c interacts with UPF1 more than ARS2n or the ΔNLS-mutant. HEK 293T cells were transfected with BioID2-ARS2n/Δnls-ARS2n/ARS2c/control-3xflag for 24h **(A)** or 48h **(B-H)**. Endogenous UPF1 was pulled down with anti-UPF1 antibodies from WCL **(A)** or

cytoplasmic fractions **(B-G)**. Untransfected cells control was also included. ARS2n/Δnls-ARS2n/ARS2c/control and UPF1 were detected using anti-flag and anti-UPF1 antibodies respectively. Enrichment is defined as the ratio between ARS2n/Δnls-ARS2n/ARS2c/control in the pull-down, versus input. **(C-F)** Dilution curve of cytoplasmic fraction lysates to ensure that all samples are detected in the linear range of the Flag antibody. We chose Dil1/20 (1ug of total protein levels) as the concentration to compare ARS2n/mutant/control expression with ARS2c/untransfected control, where 20ug of total proteins were added. **G)** Depletion of ARS2 isoforms from the supernatant (SN), obtained after UPF1 pull down. The ratio input/SN measures the depletion of the analyzed proteins from the SN, due to interaction with UPF1. For each sample the same amount of total proteins either 1ug or 20ug were added in the input and the SN, as shown by the tubulin loading control **(bottom). H)** Nuclear/cytoplasmic fractionation of HEK293 lysates. ARS2n/Δnls-ARS2n/ARS2c/control were detected using an anti-flag antibody. TBP and Tubulin detection are used as fractionation quality control. **(D-F)** Statistical analysis non-linear fit. L: Ladder, * (20ug of total proteins were loaded in these samples).

Next, to explore the mechanism of ARS2c promotion of NMD, I examined whether ARS2c preferentially interacts with other components of the NMD pathway: the CBC and SURF complexes. ARS2 interacts directly with the CBC through a CBC binding motif within the C-terminus of ARS2, which is present in both ARS2n and ARS2c[22]. As expected, ARS2c (like ARS2n) interacts with the CBC component CBP80 (Fig. 38A). Strikingly, ARS2c is more strongly associated with CBP80, ERF1 and SMG1 than either ARS2n or the mutant (Fig. 38B). Since UPF1, ERF1 and SMG1 are components of the NMD-SURF complex, which interact with the CBC complex (CBP20/80) to promote NMD, our data suggests that ARS2c promotes NMD through the CBC and the SURF complexes.

To test this idea, I first evaluated the effect of ARS2 isoforms on UPF1 phosphorylation. SMG1-mediated UPF1 phosphorylation, promotes mRNA degradation through the recruitment of SMG6, SMG5-7 and mRNA decay factors [140]. Importantly, inhibition of the NMD downstream of UPF1 mRNA assembly results in increased UPF1 phosphorylation[197]. ARS2 isoforms differentially affected the phosphorylation of UPF1: overexpression of ARS2n increased UPF1 phosphorylation, while ARS2c and the ΔNLS mutant overexpression decreased UPF1 phosphorylation (Fig. 38C, D). I next examined the impact of ARS2 isoform expression on UPF1 interactions with SURF and DECID complex members. Consistent with a promoting role in NMD, overexpression of ARS2c increased the binding of UPF1 to CBP80, ERF1 and SMG7, effect that was not observed

during overexpression of either ARS2n or the ΔNLS mutant (Fig. 39A-E). The promoting role of ARS2c was confirmed with endogenous UPF1, were expression of ARS2c but not ARS2n or the mutant, promoted the cytoplasmic binding of endogenous UPF1 to CBP80 (Fig. 39F). Interestingly, despite its cytoplasmic localization the ΔNLS mutant behaved like ARS2n with respect to UPF1/ERF1/CBP80/SMG1 binding and the inability to promote the association of UPF1 with NMD components (Fig. 37A, 37G, 38B, 39A-F). These results suggest the ΔNLS mutant is not able to replace ARS2c, implying that the loss of the N-terminal leg in ARS2c is functionally important for its role in the promotion of NMD.

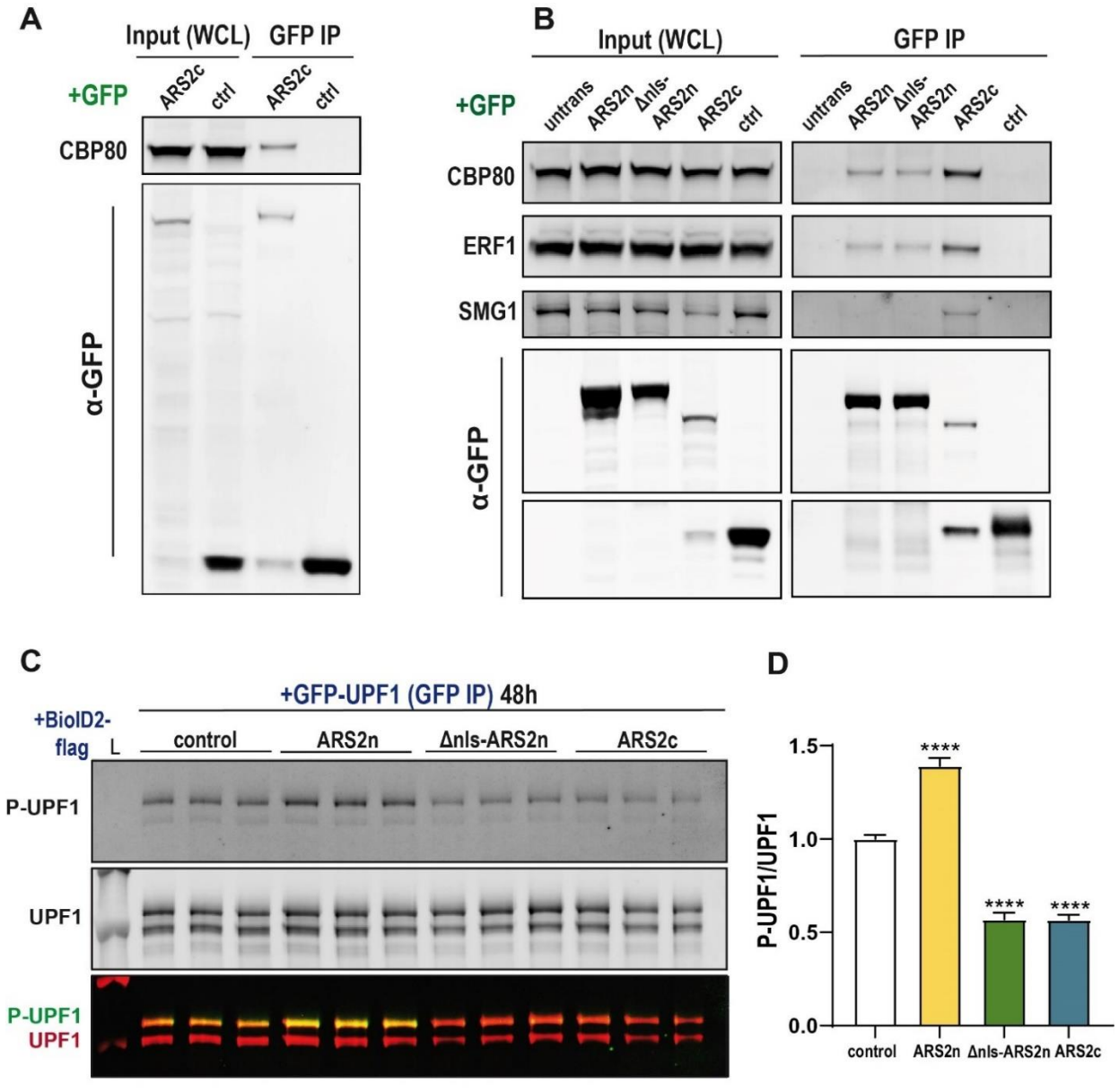


Figure 38: ARS2c preferentially interacts with CBC and SURF complex components, affecting UPF1 phosphorylation. A) HEK 293T cells were transfected with GFP-ARS2c/ GFP control. Samples were pulled down using anti-GFP beads and GFP-ARS2c/ GFP and CBP80 were detected with anti-GFP and anti-CBP80 antibodies, respectively. **B)** HEK 293T cells were transfected with GFP-ARS2n/Δnls-ARS2n/ARS2c or GFP control and samples were pulled down using anti-GFP beads. GFP-ARS2n/Δnls-ARS2n/ARS2c/control and their interactors were detected with specific antibodies, as indicated. Untransfected cells, immunoprecipitated under the same conditions, were included as negative control. **C)** HEK 293T cells were co-transfected with BioID2-ARS2n/Δnls-ARS2n/ARS2c/control-3xflag and GFP-UPF1. GFP-UPF1 was pulled down using anti-GFP beads and specified proteins were detected in GFP-UPF1 pull-downs. P-GFP-UPF1 and GFP-UPF1 were detected using Phospho-(Ser/Thr) ATM/ATR substrate antibodies (green) and anti-UPF1 antibodies (red), respectively. Bottom gel image shows the superposition of both antibodies in the green (DyLight 800) and red channels (DyLight 680). P-GFP-UPF1 was normalized to immunoprecipitated GFP-UPF1 (both bands) and represented as P-UPF1/UPF1 ratio. The ratio P-UPF1/UPF1 was additionally normalized against the BioID2-3xflag control sample and represented in **(D)** for n=3 biologically independent samples, repeated in 3 independent western blots. Statistical analysis: One way ANOVA, post-test Dunnett's (****p≤0.0001).

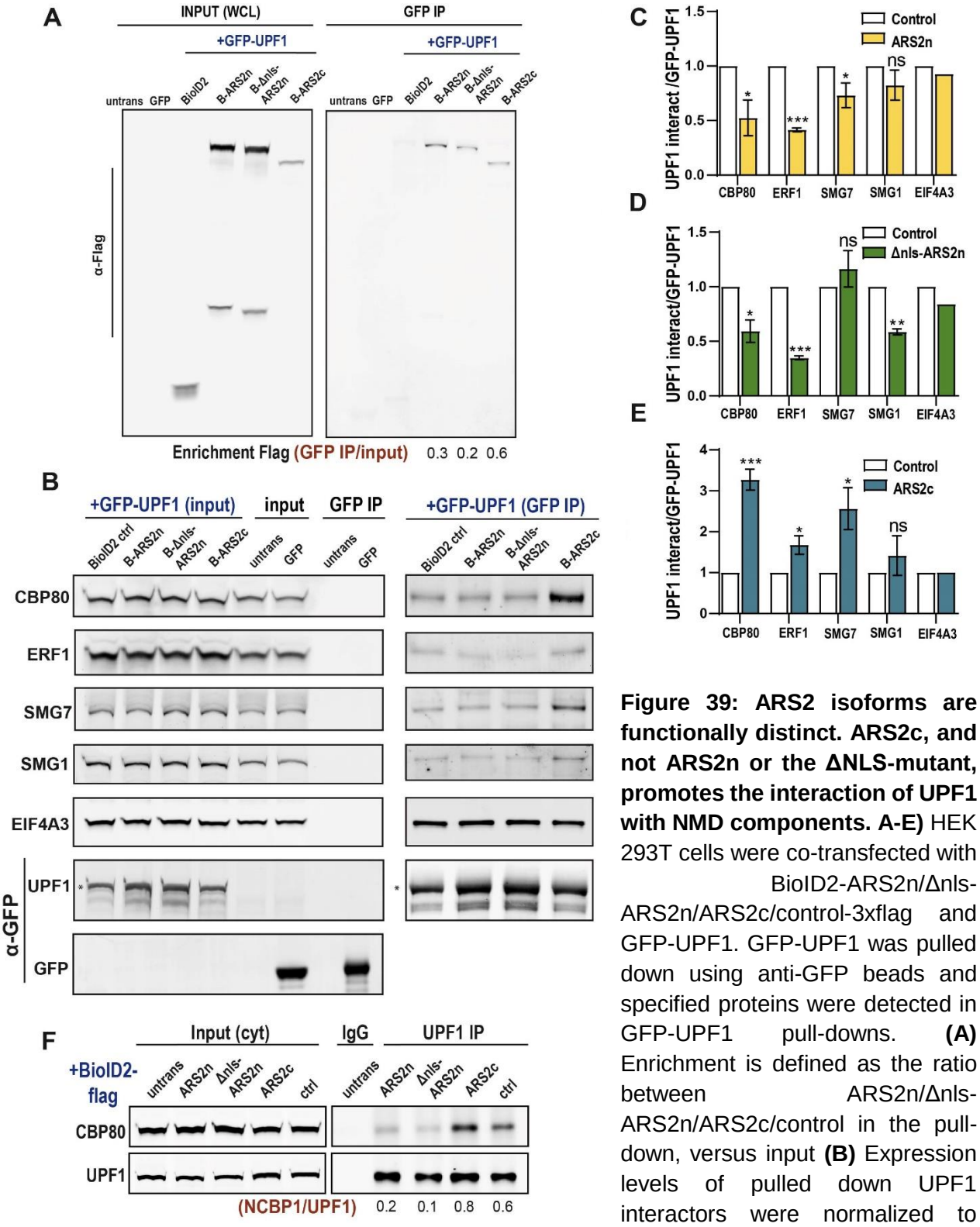


Figure 39: ARS2 isoforms are functionally distinct. ARS2c, and not ARS2n or the ΔNLS-mutant, promotes the interaction of UPF1 with NMD components. A-E) HEK 293T cells were co-transfected with BioID2-ARS2n/Δnls-ARS2n/ARS2c/control-3xflag and GFP-UPF1. GFP-UPF1 was pulled down using anti-GFP beads and specified proteins were detected in GFP-UPF1 pull-downs. **(A)** Enrichment is defined as the ratio between ARS2n/Δnls-ARS2n/ARS2c/control in the pull-down, versus input **(B)** Expression levels of pulled down UPF1 interactors were normalized to immunoprecipitated GFP-UPF1, marked with an asterisk. The ratio UPF1 interactors/GFP-UPF1 was additionally normalized against the BioID2-3xflag control sample for n=2 or n=3 biologically independent samples **(C-E).** Untransfected and single GFP transfected cells, immunoprecipitated

with GFP beads under the same experimental conditions, were included as a negative control. Statistical analysis two-tail unpaired t-test (*p≤0.05; **p≤0.01; ***p≤0.001). **F)** HEK 293T cells were transfected with BioID2-ARS2n/Δnls-ARS2n/ARS2c/control-3xflag for 48h. Cytoplasmic fractions were extracted and endogenous UPF1 was pulled down with an anti-UPF1 antibody. Cytoplasmic fraction of untransfected cells, pulled down with IgG mouse control antibody were used as a control. CBP80 and UPF1 were detected using anti-CBP80 and anti-UPF1 antibodies respectively. Expression levels of the prey (CBP80) in the IP were normalized to immunoprecipitated UPF1 and represented as CBP80/UPF1 ratio.

5.3 Discussion

The BioID interactome data identified the NMD pathway as one of the enriched biological processes that is shared between ARS2n and ARS2c interactomes. Previous studies have shown that ARS2n interacts with EJC core components and peripheral factors, as well as other NMD components including UPF1[21,59,102,103,184,276,277]. Furthermore, the CBC complex stabilizes the binding of UPF1-ERF1/ERF3 to the EJC, promoting NMD[108,272]. However, to our knowledge, no role for ARS2n in NMD has been described.

To test the functionality of the interaction between ARS2n and NMD components we performed a tethering assay. Tethering of ARS2n to the 3'UTR of the reporter, triggered its degradation in a UPF1 dependent manner, suggesting a role for ARS2n in NMD. While this experiment does not rule out ARS2n participation in other degradation pathways like the nuclear exosome, the rescue observed with *UPF1* knockdown shows that in addition to its role in exosomal degradation[14,83], ARS2n also participates in NMD. An additional limitation of the tethering assay is that it forces the interaction of ARS2n with the 3'UTR of the reporter, where this interaction doesn't necessarily take place. Due to these limitations, we did not pursue the testing of ARS2c with this assay and focussed instead on other approaches more specific for NMD.

Using a NMD dual fluorescence reporter, RT-qPCR of endogenous targets and both gain-of-function and loss-of-function experiments, I found that ARS2n inhibits while ARS2c promotes NMD. This effect is specific for NMD targets (PTC containing transcripts) and is not a secondary effect from ARS2 role on RNA metabolism, as shown by our NMD specific assays like the NMD reporter, *HNRNPA2B1* isoforms amplification and recovery

experiments. The role of ARS2n appears to be primarily in the nucleus through its interactions with the EJC and export machinery, whereas ARS2c preferentially interacts with CBC and SURF complex components and promotes the binding of UPF1 to CBP80/ERF1/SMG7 in the cytoplasm (Fig. 31A-C, 37A, 37G, 38B, 39A-F).

To directly test whether the isoforms are functionally the same and differ only in cellular localization, or whether the isoforms are functionally distinct and differ in location and function, I created a cytoplasmic version of ARS2n that lacked the NLS but maintained the N-terminal leg and remaining unstructured N-terminus. Deletion of the nuclear localization signal was sufficient to confer cytoplasmic localization to the full-length isoform and abrogated the nuclear isoform's ability to inhibit NMD (Fig. 36B, C). Despite the cytoplasmic localization, the ΔNLS mutant behaved like ARS2n with respect to UPF1/ERF1/CBP80/SMG1 binding and the inability to promote the association of UPF1 with ERF1 and CBP80 (Fig. 37A, 37G, 38B, 39A-F). Unlike ARS2n however, expression of ΔNLS mutant did not increase UPF1 phosphorylation (Fig. 38C, D). Consistent with this finding, the ΔNLS mutant decreased the binding of UPF1 to SMG1, the kinase responsible for UPF1 phosphorylation (Figure 39D). Although more mechanistic evaluations will need to be performed, our results suggest that cytoplasmic expression of ARS2n is not sufficient to promote NMD, and that the isoforms are functionally distinct.

The functional distinctness of the isoforms must be related to the differences in their N-termini. The N-terminal leg of ARS2n physically associates with the C-terminus of the protein in the crystal structure and is thus poised to affect binding events at the C-terminus[21]. Consistent with this prediction, we have shown previously that mutations within the N-terminal leg alter the RNA binding ability at the C-terminal leg. Specifically, mutation of three conserved tyrosine sites within the N-terminal leg reduces the association of ARS2 with RNA[18]. Interestingly, two of these tyrosine residues are also phosphorylated within the phosphoproteome data set, suggesting interactions to this region of the protein may be regulated[280,281]. It is possible that the presence of the N-terminal leg sterically hinders ARS2n from promoting NMD and that an isoform switch is necessary for the pathway. However, understanding how these isoforms precisely work to coordinate mRNA fate during NMD will require further study.

Based on our results and the literature in NMD, we propose a model in which ARS2n and ARS2c work in tandem to regulate NMD (Fig. 40). As previously described, ARS2n primarily functions in the nucleus in conjunction with the CBC to promote mRNA splicing, degradation and export[11,13,15,23,81,102,271]. We propose that once in the cytoplasm, an isoform switch occurs in which ARS2c replaces ARS2n within the CBC. This ARS2 "isoform switching" has dramatic functional consequences changing ARS2 from a NMD inhibitor to a NMD promoter. ARS2c, in association with the CBC component CBP80, promotes the interaction between UPF1 and ERF1/CBP80, favoring the formation of the SURF complex at the PTC, SMG7 recruitment, and transcript degradation though NMD. Our model is consistent with previous literature showing that CBP80 promotes both UPF1 interaction with ERF1 and ERF3 and the association of UPF1 with the EJC[108,272] and extends this work to include an ARS2 isoform switch as an important step in this process. We suggest that failure to switch ARS2 isoforms in the cytoplasm creates a CBC that is inefficient at promoting NMD. Thus, ARS2 isoform switching tailors the protein to function in a nonredundant manner in nuclear and cytoplasmic CBC-dependent processes.

Although our model is supported by our results and the literature, other alternative models are also possible. For example, ARS2n could act as an inhibitor of the NMD pathway in nucleus and cytoplasm. ARS2n could compete with UPF1 for binding to the CBC complex, thus disfavoring NMD target recognition by UPF1, and their degradation through NMD. In nucleus, where ARS2n is predominant, this inhibitory mechanism would be reinforced. However, once the PTC containing mRNA is exported to the cytoplasm, the scenario changes. ARS2c, enriched in the cytoplasm and with higher affinity for the CBC complex than ARS2n, outcompetes and substitutes ARS2n in the CBCA complex. The lack of the N-terminal leg in ARS2c then favors the binding of UPF1 to the CBCA complex, PTC transcript recognition and degradation through NMD. Thus, cytoplasmic switch of ARS2 isoforms could release ARS2n-mediated inhibition of NMD and promote transcript degradation. Furthermore, this cytoplasmic interchange of ARS2 isoforms could be regulated by additional factors like importins, the SURF complex, and the pioneer round of translation. ARS2n interaction with importin-α[13], should promote ARS2n protein translocation to nucleus, resulting in a decreased ARS2n concentration in the cytoplasm. Secondly, as our data suggest, ARS2c interacts and favors SURF complex formation,

which could “lock” a conformation of the CBCA complex that, bound to the SURF complex, impedes further ARS2 isoforms exchange. Finally, our interactome data showed that ARS2n link to NMD was fundamentally mediated by ARS2n interaction with EJC components. The removal of EJC complexes from the mRNA, during the pioneer round of translation, should severed a significant part of the connection between ARS2n and the mRNA, thus favoring a transition to ARS2c, which binds to translation initiation and termination proteins.

To conclude, we propose a third model that contemplates the existence of independent mRNA pools that bind to ARS2n or ARS2c. In this model a set of mRNAs, bound to the CBC but not to ARS2n, would be processed and exported independently of ARS2n, to be recognized by ARS2c in the cytoplasm. However, due to the key role of ARS2n in nuclear RNA metabolism, ARS2n functions as a CBC effector should be substituted by another protein(s) in nucleus, to ensure the proper processing of these mRNAs. This alternative protein, similar to ARS2n could be exchanged in the cytoplasm by ARS2c or could interact with ARS2c and regulate its functionality. If true, this model could open many interesting questions in the field related the specific structure, protein interactors and functionality of these mRNAs. Furthermore, the concept of independent mRNA pools, differentially recognized by ARS2 isoforms could be extended to other RNA families.

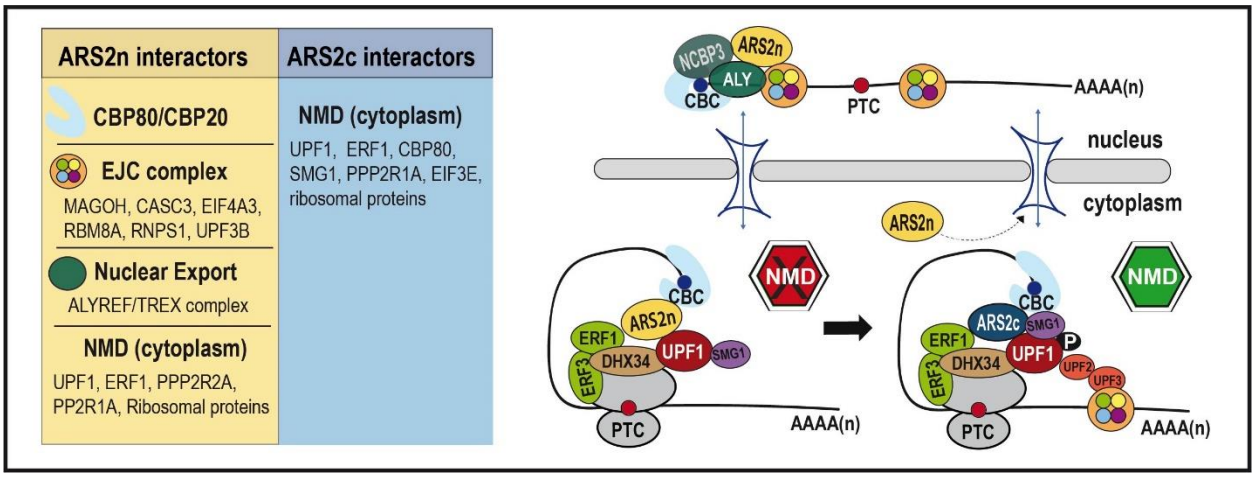


Figure 40: ARS2 isoforms work in tandem to regulate NMD. ARS2n, bound to CBC and NCBP3, interacts with the EJC and regulates the translocation to the cytoplasm of NMD-sensitive mRNAs. Once in the cytoplasm, ARS2n is recycled back to the nucleus and is substituted by

ARS2c. ARS2c interacts with CBP80, UPF1, ERF1 and SMG1, enhances the binding of UPF1 to CBP80/ERF1, promoting the formation of the SURF complex. The presence of an EJC more than 50nt downstream of the termination codon favors the formation of the DECID complex and transcript degradation though NMD. Left panel: ARS2n and ARS2c interactors reported in our study and/or the literature.

Chapter 6: Conclusions and Future Directions

6.1 Concluding remarks

The life of an mRNA is highly regulated from biogenesis to degradation. This regulation starts in nucleus with the addition of the 7-methyl guanosine cap at the 5' end, followed by the binding of the CBC complex and ARS2. Through the formation of dynamic, mutually exclusive complexes, CBC-ARS2 escort the mRNA from nucleus to cytoplasm, where it's translated to a functional protein. As a critical partner to the CBC, ARS2 plays an important role in nuclear CBC dependent mRNA processing events, including splicing, 3'end processing, polyadenylation, and export[11,13,15,23,81,102,271]. However, at the time of my thesis no roles for ARS2 in cytoplasmic mRNA metabolism had been described. My thesis work focuses on this fundamental gap in the literature and addresses another poorly explained function of ARS2: its role in arsenic sensitivity.

In Chapter 3, I identified the existence of *ARS2* isoforms that contain a highly conserved intron, retained during alternative splicing and absent from the canonical *ARS2n* sequence. These alternatively spliced transcripts produce cytoplasmic ARS2 proteins that share the helical core, C-terminal domain, RNA and CBC binding regions with ARS2n, but lack 270 amino acids from the N-terminus of the nuclear isoform. I developed the tools required for the specific targeting and detection of these cytoplasmic isoforms in mouse and human cells, which were used to evaluate and compare ARS2 isoforms functions in the following chapters.

As part of the CBC, ARS2 interacts with highly dynamic and transient RNP complexes. The diversity of RNAPII transcripts recognized by ARS2 combined with the weak nature of ARS2's interaction with RNP complexes, has meant that traditional mass spectrometry based proteomic approaches often fail to detect less abundant but equally important interactions. Therefore, I performed the first BioID proximity ligation characterization of both ARS2n and ARS2c, which allows the detection of transient, weak and temporally separated interactions. This analysis showed that interactomes of nuclear and

cytoplasmic ARS2 isoforms were strikingly different (**Chapter 4**). Notably, ARS2c interactome was enriched in proteins involved in the cellular response to stress. In **Chapter 4,** I showed that ARS2c is upregulated during arsenic stress, and is responsible for arsenic sensitivity. Whereas ARS2n is downregulated and may mediate arsenic resistance. I further demonstrated that ARS2c is required for proteomic remodeling during arsenic stress and that this cellular response is independent of the CBC complex. These results demonstrate that ARS2c isoforms and not ARS2n are associated with arsenic sensitivity. Furthermore, ARS2c mediated arsenic sensitivity is the first described function for mammalian ARS2 that is potentially independent from the CBC complex.

In **Chapter 5**, I turned my attention to biological processes that are shared between ARS2 isoforms. Remarkably, most of the shared processes were related to RNA metabolism. This suggests that *ARS2*, has evolved to express non-redundant protein isoforms that regulate the life of an mRNA from beginning to end, nucleus to cytoplasm. Consistent with this crucial role in RNA metabolism, I demonstrated that ARS2 isoforms work in tandem to modulate the RNA catabolic pathway, NMD (**Chapter 5**). ARS2n primarily functions in the nucleus with the CBC to regulate mRNA splicing, degradation, and export[11,13,15,23,81,102,271]. We propose that once the mRNPs enter the cytoplasm, ARS2n "hands off" the CBC bound mRNA to ARS2c, which promotes the formation of the SURF complex at the stalled ribosome and transcript degradation though NMD. Thus, ARS2 isoforms function in a non-redundant manner escorting the mRNA from nuclear processing to cytoplasmic degradation.

Alterations in RNA metabolism are often found in diseases such as neurodegeneration, diabetes, and cancer. Therefore, understanding how RNA metabolism is regulated during physiological conditions and disease, is important for addressing issues related to human health and ageing. Moreover, misregulation of ARS2 expression have been related to cancer[282], and the two novel cytoplasmic functions of ARS2 described in our study, could have implications for cancer treatment. NMD has a well-documented role in tumorigenesis[283], and drugs that modulate NMD activity are being considered in clinical trials to increase patients response to immunotherapy[284]. Due to the opposing roles of ARS2 isoforms in the regulation of NMD, the relative expression levels of each ARS2

isoform could impact NMD-driven tumorigenesis and the tumour response to these therapies. Similarly, although arsenic trioxide (ATO) is an approved treatment for Acute Promyelocytic Leukemia[110,246,247], the use of arsenic as an anticancer agent is limited by its toxicity and the lack of mechanistic studies in different tumour scenarios. Since ARS2c downregulation significantly abrogates cell sensitivity to high concentrations of arsenic, expression levels of ARS2c within tumours could serve as a prognostic factor to stratify those patients most likely to benefit from ATO treatment. Understanding the mechanism of ARS2c-mediated arsenic sensitivity could lead to new treatment options that promote higher arsenic toxicity within tumours, with minor secondary effects on patients' health.

6.2 Future directions

6.2.1 ARS2c expression regulation under physiological conditions

My work indicates that ARS2c expression is critical for a global proteomic response to arsenic stress, and consequently, ARS2c levels must be regulated under physiological conditions and stress. However, the precise mechanisms of this regulation need to be addressed. Retention of the intron 5 generates transcripts with upstream stop codons in all open reading frames. Thus, protein translation initiated at 5' end of the gene should be prematurely stopped at intron 5, leading to transcript degradation by NMD. One interesting aspect of NMD is that it is self regulated. Under physiological conditions, components of the NMD pathway often induce their own transcript degradation through NMD[122], as a negative feedback loop that controls the pathway efficiency. Based on ARS2c's role in NMD, expression patterns, and long 5'UTR, we predict that ARS2c could be regulating itself through NMD. Thus, high expression of ARS2c will increase NMD efficiency, which induces the degradation of the NMD sensitive *ARS2c* transcript, leading to the contraction of ARS2c pools and NMD activity.

ARS2c expression is also likely regulated by alternative splicing, controlling the inclusion or exclusion of intron 5. Under physiological conditions, exclusion of the intron 5 seems to be favored, since intron 5 containing transcripts are around 17 times less expressed than transcripts that exclude this intron (Fig. 41A). However, the potential role of NMD in

the degradation of intron 5 containing transcripts, will also contribute to the observed difference. Interestingly, downregulation of the CBC component CBP80, induces the expression of all *Ars2* isoforms (Fig. 41B-C) which implies there are feedback mechanisms controlling expression of CBCA complex components.

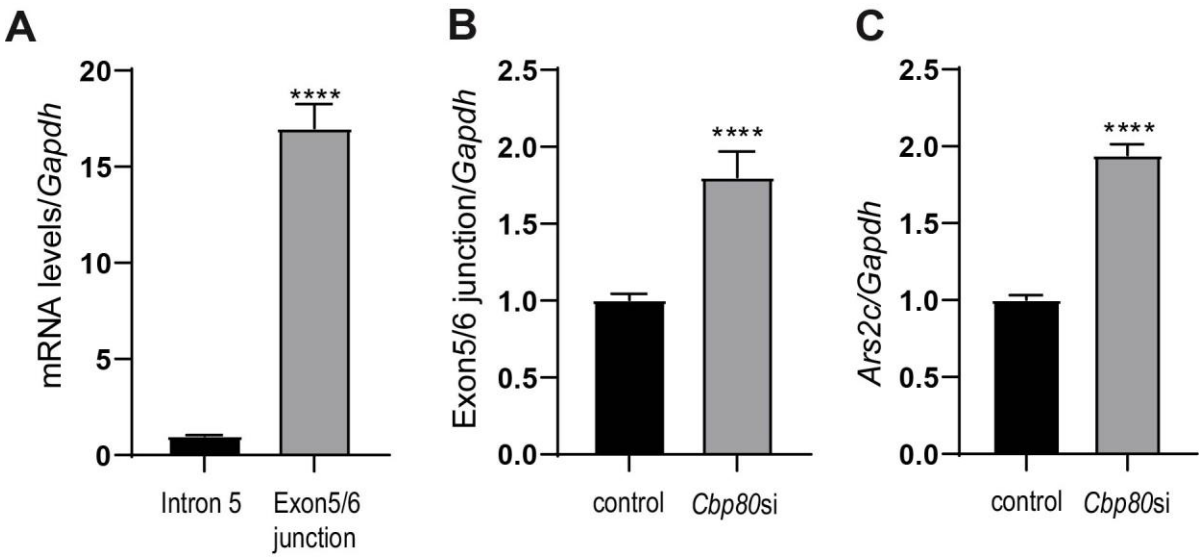


Figure 41: *Ars2* isoforms expression is affected by CBP80 downregulation. **A)-C)** Expression levels of *Ars2* isoforms were detected by RT-qPCR and normalized against *Gapdh*, in C_2C_{12} cells. *Ars2c* was detected using a primer that amplifies intron 5, which is specifically retained in *Ars2c* isoforms. *Ars2n*, was detected with a primer that amplifies Exon5/6 junction. **B)-C)** C_2C_{12} cells were transfected with *Cbp80*/control RNAi. Data are represented as mean ± SEM for n=3 biologically independent samples**.** Statistical analysis two-tail unpaired t-test (****p≤0.0001).

6.2.2 ARS2c expression regulation under stress

Under puromycin and arsenic stress, our data shows that *Ars2c* expression increases at transcript levels. Both puromycin and arsenic stress, block translation which leads to inhibition of NMD. Thus, if *Ars2c* transcript is indeed degraded by NMD under physiological conditions, inhibition of NMD due to translation blockade could explain the increase of *Ars2c* transcripts levels. However, puromycin and arsenic stress also result in an increase of ARS2c protein. In contrast, although both *Ars2n* and *Cbp80* transcripts levels increase after translation inhibition, their protein levels decrease (Fig. 42A-C).

Therefore, protein induction of ARS2c is specific and takes place in a context of global translation inhibition. A critical question for the future is therefore: how does ARS2c escape translation inhibition?

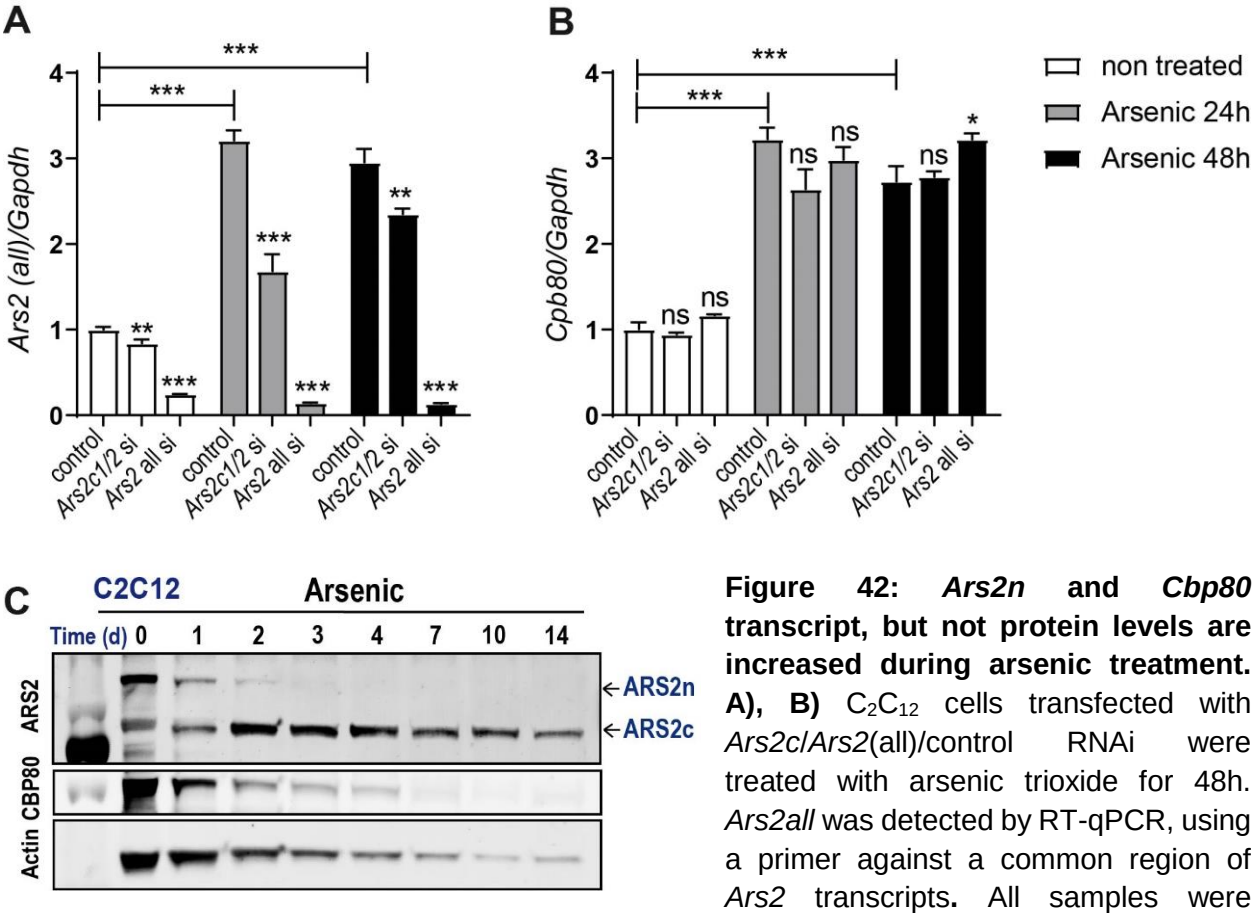


Figure 42: *Ars2n* and *Cbp80* transcript, but not protein levels are increased during arsenic treatment. A), B) C_2C_{12} cells transfected with *Ars2c*/*Ars2*(all)/control RNAi were treated with arsenic trioxide for 48h. *Ars2all* was detected by RT-qPCR, using a primer against a common region of *Ars2* transcripts**.** All samples were normalized against *Gapdh*. Data are represented as mean ± SEM for n=3 biologically independent samples. Statistical analysis: One way ANOVA, post-test Tukey's (*p≤0.05; **p≤0.01; ***p≤0.001). **C)** C_2C_{12} cells were treated with 40uM arsenic trioxide for 14 days. Western blots of whole cell lysates are shown with the indicated antibodies**.** The image was previously presented in Figure 27C and repeated here for illustrative purposes.

The unfolded protein response (UPR) might shed some light to how ARS2c is being translated. During the UPR, translation is drastically attenuated to reduce the burden on the cellular protein folding system and conserve limited amino acids pools[285]. In this situation translation is inhibited using two main mechanisms, which are successfully avoided by mRNAs that code for proteins involved in the resolution of the stress response. The first translation attenuation mechanism is through phosphorylation of eIF2α, the

initiator factor that delivers initiator methionyl tRNA (Met-tRNAi^{Met}) to the 40S subunit of the ribosome to initiate translation[285,286]. P-eIF2α engages in a non-productive protein complex with its guanine nucleotide exchange factor eIF2B, reducing the ability of cells to recycle eIF2 and deliver Met-tRNAi^{Met} to ribosomes, resulting in translation inhibition[286]. Transcripts that code for proteins involved in the stress response contain uORFs in their 5' UTR, which allows them to express functional proteins and bypass P-eIF2α mediated translation attenuation. An excellent example of this strategy is described for ATF4[223] and was previously discussed in **Chapter 1**.

The second mechanism to decrease translation under ER stress is mediated by 4E-BP. 4E-BP binds and inhibits eIF4E association to the 5'-cap, which affects the loading of the 43S ribosome complex to initiate translation. Thus, 4E-BP serves as an inhibitor of cap-dependent translation[285,287]. External stimuli such as growth factors, mitogens and hormones activate the PI3K/AKT/mTOR signalling pathway, resulting in mTOR-mediated phosphorylation of 4E-BP. P-4E-BP dissociates from eIF4E, which is now free to engage in cap-dependent translation[288]. In contrast, as part of the UPR during ER stress, the transcription factor ATF4 induces the expression of 4E-BP. 4E-BP sequesters eIF4E molecules, leading to inhibition of cap-dependent translation[289]. Notably, this translation inhibition mechanism is avoided by stress protein-mRNAs, through the presence of Internal Ribosome Entry Site (IRES) in their 5'UTRs[285]. IRES elements can directly recruit the ribosomes to the start codon, allowing the mRNA to undergo cap-independent translation[285,290]. Thus, the presence of "escaping" features in the 5'UTR of stress proteins-mRNAs, ensures their specific translation in a context of global translation inhibition.

Interestingly, *Ars2c* isoforms have significantly long 5'UTRs. While the median length of 5'UTRs is 175nt in mouse[290], *Ars2c* isoforms 5'UTRs are greater than 1500nt in length. Long 5'UTR often contain structures that regulate translation, including IRES-like structures. As IRES elements can directly recruit the ribosome and promote cap-independent protein translation under global translation inhibition, we evaluated if *Ars2c* 5'UTR contained an IRES sequence. Using the IRESpy software to predict the presence of IRES sequences[291], I analyzed *Ars2c* 5'UTR and compared it to *Ars2n* 5'UTR and

Ars2n/c full sequences (Table 6). The software showed that *Ars2c* intron 5 contained a potential IRES sequence between nucleotides 1950-2200, which is absent from *Ars2n* transcripts. This result was confirmed by Michael Amiss, a previous honour student in the lab, who contrasted the results with a second software IRESFinder[292]. Michael modeled the potential *Ars2c* IRES secondary and tertiary structure and demonstrated that they resemble to previously described cellular IRES[292]. Furthermore, he found two potential binding sites for PTB, an IRES-Trans activating factor (ITAF) that has been shown to perform initial remodelling of cellular IRES elements, and is required for the recruitment of the 40S ribosomal subunit[293,294]. Although more experiments are required to confirm the existence of an IRES in the *Ars2c* 5'UTR, our current data suggest that *Ars2c* isoforms could use IRES elements to bypass translation inhibition and ensure ARS2c expression during the cellular response to stress.

Based on its domain architecture, ARS2c should retain the RNA binding capacity of ARS2n and could potentially modulate translation. ARS2n interacts with ssRNA, miRNA and RDH mRNAs, through its RRM domain (ssRNA) and the C-terminal leg (miRNA/RDH mRNAs)[18,21]. Since both domains are also present in ARS2c, ARS2n and ARS2c could potentially interact with RNAs in the cytoplasm, including double stranded RNA sequences, like miRNA or IRES. Therefore, under stress conditions ARS2c may control its own translation. In support of this, I identified several components of the 43S complex, often used to bypass the cap-dependent translation[290,295], in ARS2c interactome. eIF3F/E/H/K were detected in our interactome and confirmed by western blot (Fig. 31B, Table 3), while eIF2D and eIF2B2, important for the recruitment of Met-tRNAi^{Met} to the initiation codon, were also found. Moreover, using its RNA binding capacity and the association to 43S complex components, ARS2c could also regulate the translation of proteins involved in the arsenic stress response, thus explaining why ARS2c expression is necessary for proteome remodelling during arsenic treatment (Fig. 28C).

6.2.3 NMD

Although our study addressed the mechanism of ARS2c regulation of NMD in the cytoplasm, a few questions remain. One of them is how ARS2n inhibits NMD. Our results

suggest that ARS2n inhibition of NMD originates in the nucleus (Fig. 36C), however both ARS2n and the ΔNLS mutant bind to UPF1/ERF1/CBP80 and decrease the interaction of UPF1 with NMD components in the cytoplasm (Fig. 37A, 37B, 37G, 38B, 39A-D, F). Therefore, although ARS2n role in NMD is at least partially nuclear, ARS2n could compete with ARS2c for CBC-bound mRNAs in the cytoplasm, affecting the efficiency of the NMD.

Another idea to consider is the potential formation of mutually exclusive ARS2 containing complexes in the cytoplasm. ARS2n regulates RNA metabolism in nucleus through the formation of dynamic, mutually exclusive complexes that ensure the correct processing of the different families of RNAPII transcripts. ARS2c could exert a similar role in the cytoplasm. The CBC complex, besides NMD, is also involved in the pioneer round of translation[106]. CBC bound ARS2c could participate in this process as well, and consistent with this idea, our interactome data shows that ARS2c interactome is enriched in components of the translation initiation machinery and ribosomal proteins (Table 3). Consequently, ARS2c could form mutually exclusive complexes with the CBC complex that promote translation or NMD, in the cytoplasm. Understanding how the decisions between translation and NMD are regulated will be important in future research.

6.2.4 Pathological consequences of ARS2 isoforms misregulation

6.2.4.1 NMD pathway is enriched in high *ARS2*-expressing tumours

Misregulation of *ARS2* has pathological consequences[282]. Phineas Hamilton, a collaborator of our lab, analyzed gene expression and survival data for solid cancers publicly available through TCGA. These analyses revealed that total *ARS2* expression is significantly associated with poorer survival in several cancers, including liver hepatocellular carcinoma, glioma, and kidney renal cell cancer (i.e. LIHC, LGG and KIRC; Cox proportional hazards models, $P_{adj} < 0.05$), but predicted better outcome in pancreatic adenocarcinoma (PAAD; Fig. 43A, B).

To evaluate which biological processes were correlated with expression of *ARS2*, P. Hamilton compared cancers in the first and fourth quartiles of *ARS2* expression for each

evaluable cancer type and performed gene set enrichment analysis based on the ranked log-fold changes of gene expression between quartiles. Strikingly, this unbiased approach revealed that the NMD pathway is highly enriched in most of the evaluated tumour types. The top 20 GO terms associated with *ARS2* expression (biological processes; based on the average absolute normalized enrichment score across cancers) are shown in Fig. 44A. Moreover, analysis of the specific NMD components that contribute to this score across tumour types, shows that approximately half are confirmed interactors of ARS2 (Fig. 44B). The correlation analysis was performed with total *ARS2* expression, and further studies will be required to determine the independent contribution of each *ARS2* isoform to the enrichment of each biological process. Nevertheless, the enrichment of NMD in high *ARS2* expressing tumours, and my data showing the ability of ARS2 isoforms to regulate NMD, suggests that my findings could be clinically important. NMD has both promoting and inhibitor roles in tumorigenesis, depending on tumour type. ARS2 isoforms expression ratio and their opposing regulation of NMD, could influence tumorigenesis and tumour response to therapies that modulate NMD activity.

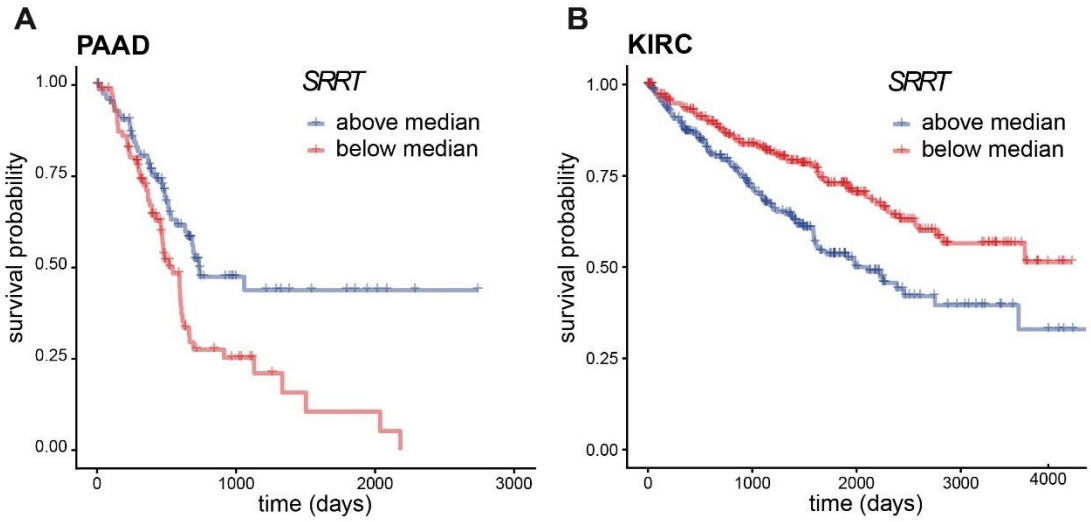
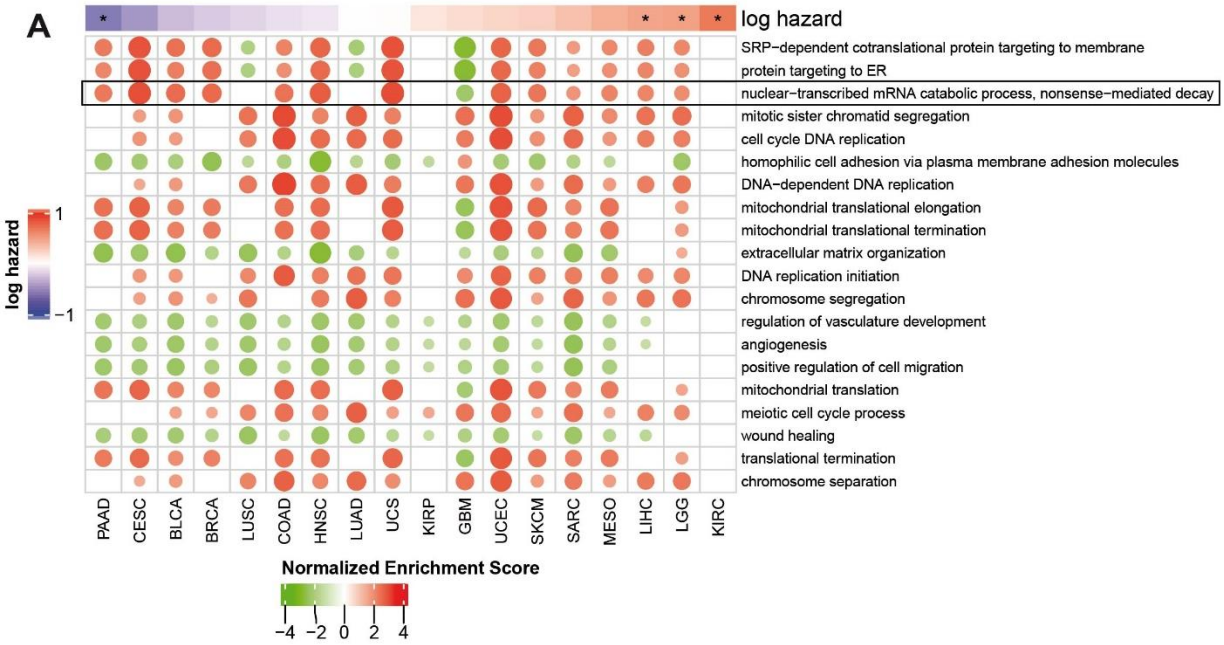


Figure 43: Expression of *ARS2* is significantly associated with patient survival KIRC and PAAD tumours. A), B) Kaplan-Meier survival plots for kidney renal clear carcinoma (KIRC) and pancreatic adenocarcinoma (PAAD) patients with high or low *ARS2* expression. Data was obtained from the TCGA database and analyses were performed by Phineas Hamilton, Deeley Research Centre, BC Cancer Agency.



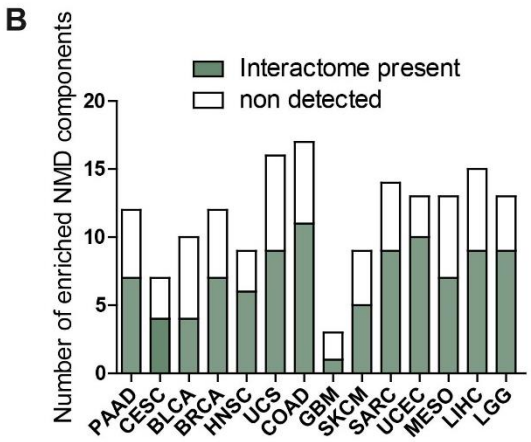


Figure 44: NMD pathway is enriched in high *ARS2*-expressing tumours. A) *ARS2* survival data for solid cancers evaluated across tumours in the TCGA database. (Cox proportional hazards models, Padj<0.05 represented by asterisk). GSEA was performed using genes ranked-fold change of expression difference between patients with high (first quartile) and low *ARS2* expression (fourth quartile) for each represented tumour. Top 20 GO terms associated with *ARS2* expression (biological processes; based on average absolute normalized enrichment score (NES) across cancers) are shown; circle sizes are proportional to NES. Analysis was performed by Phineas Hamilton, Deeley Research Centre, BC Cancer Agency. **B)** NMD components enriched in the GSEA were compared against ARS2n interactome, reported in the literature. The number of enriched NMD components that are reported ARS2n interactors is represented in green. Total enriched NMD components are represented in white.

6.2.4.2 *ARS2c* is downregulated in tumours

Using the intron 5 sequence in transcriptome databases as a proxy for *ARS2c* expression, Michael Amiss analyzed *ARS2c* expression across malignant and healthy human tissue samples. Comparison between healthy tissue samples with their matched tumour type,

showed that *ARS2c*, and not total *ARS2* expression, was strikingly decreased in tumours samples (Fig 45A, B, Fig 46).

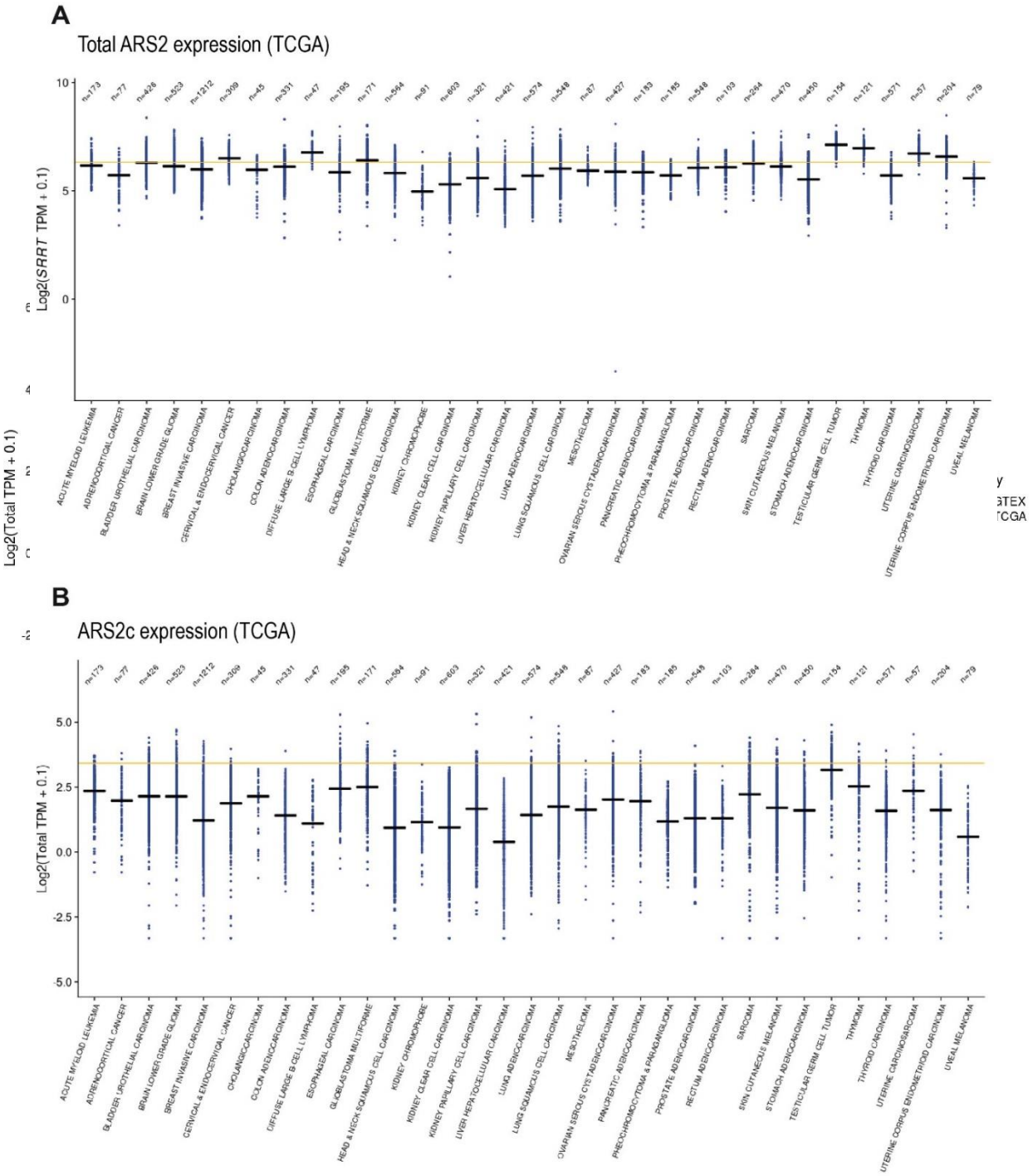


Figure 45: Expression levels of total *ARS2* or *ARS2c* in tumour samples. The data was obtained from TCGA (tumour samples) and GTEx (healthy tissues) in the Xenabrowser database. **A)** Total *ARS2* expression across TCGA cancers with the median level of total *ARS2* expression in healthy tissues shown in yellow. **B)** *ARS2c* expression across TCGA cancers with the median

level of *ARS2c* expression in healthy tissues shown in yellow. Performed by Michael Amiss, P. Howard lab and Phineas Hamilton, Deeley Research Centre, BC Cancer Agency.

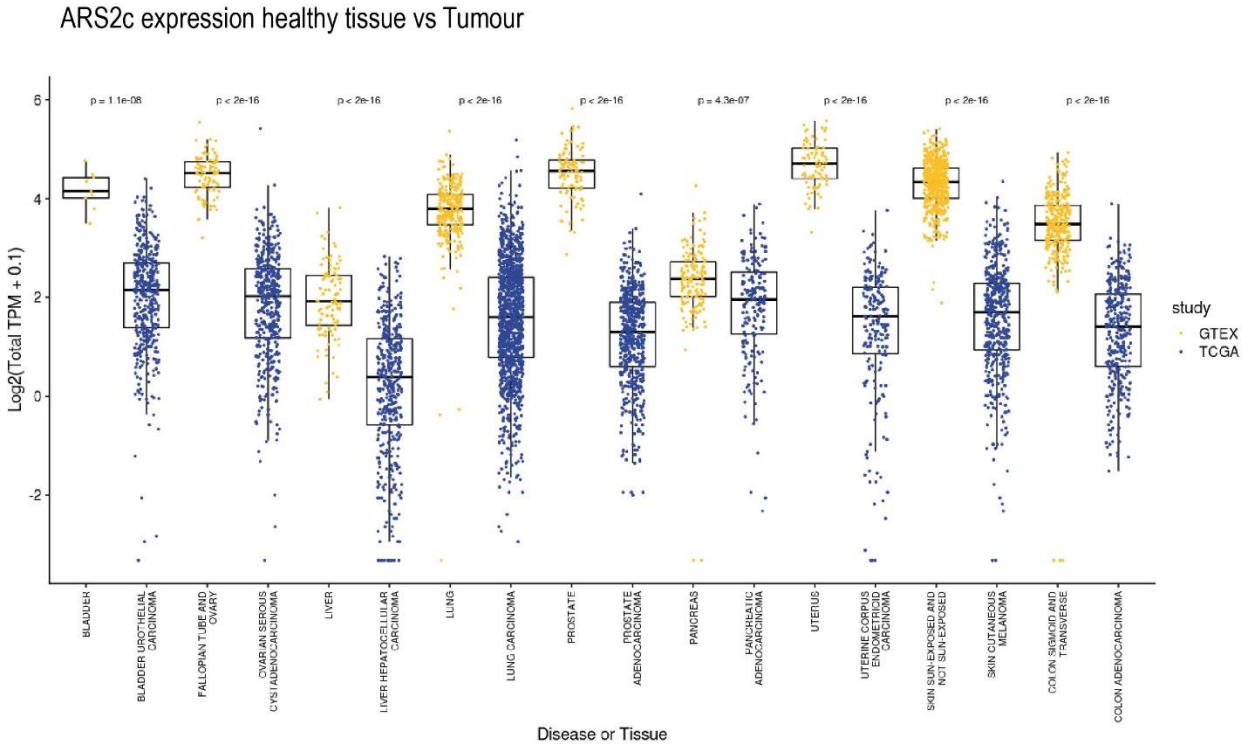


Figure 46: *ARS2c* expression across matched healthy and cancer tissue pairings. Tumour samples are represented in blue, and healthy tissues are represented in yellow. The data was obtained from TCGA and GTEx in Xenabrowser. Performed by Michael Amiss, P. Howard lab and Phineas Hamilton, Deeley Research Centre, BC Cancer Agency.

Moreover, our BioID data showed that ARS2c interactome is enriched in proteins involved in the cellular response to stress, hypoxia, chemical stress, and apoptosis (**Chapter 5,** Table 5). Thus, downregulation of ARS2c could be relevant for tumour survival under several stress conditions, often found in the tumour microenvironment or during chemotherapy treatment. Indeed, as our data shows, downregulation of ARS2c increases cells' resistance to arsenic stress and prevents apoptosis, even when exposed to high concentrations of the chemical (**Chapter 4**). Since arsenic it's currently under study for several tumour locations, including pancreatic and ovarian cancer, studies that evaluate ARS2c expression during tumour progression and the correlation of ARS2c expression with patient survival are needed. Based on the marked arsenic resistant phenotype

observed during ARS2c downregulation, we predict that low ARS2c expression in tumours could significantly increase their resistance to arsenic-based therapies.

6.2.5 ARS2 isoforms in global RNA metabolism

This document has focussed on the role of ARS2 isoforms in RNA metabolism, based on the classical maturation model for RNAPII transcripts. In this model the CBC and ARS2 bind to capped RNAs early during transcription and chaperone the RNAs through nuclear processing, export, and cytoplasmic translation or degradation through NMD. After the pioneer round of translation, the RNAs are handed off to eIF4E for steady-state translation. Thus, in the classical model there is a clear separation of functions based on the cellular localization of CBC or eIF4E, where CBC is fundamentally involved in nuclear processes, while eIF4E dominates cytoplasmic translation[296]. However, accumulated evidence in the literature suggest that eIF4E can also function in splicing, cleavage and polyadenylation, and nuclear export of selected transcripts [296–302]. Furthermore, both CBC and eIF4E localize and bind to capped RNA in nucleus and cytoplasm, positioning them to act in all of these functions simultaneously[8,58,296,300,303–306].

Consequently, a new model for RNA maturation has been proposed recently[296]. The parallel, non-linear, interleaved RNA maturation model proposes that multiple, parallel RNA maturation paths coexist, and the cap-chaperones CBC or eIF4E don't have a monopoly on nuclear or cytoplasmic events, respectively[296]. In this model, RNAs can transit through the maturation steps in parallel, directed by either CBC or eIF4E. Furthermore, the RNA does not have to be restricted to a maturation path for life. Referred to as interleaving, the authors of this model suggest that some RNAs can undergo cap-chaperone switching during their maturation and translation[296].

How ARS2 isoforms fit in this new model? Proteomic data from our study and others have detected the interaction between ARS2n or ARS2c with several 43S complex components, including eIF3D/H/F/E/H/L/C/A, eIF2D, eIF2B2 and ribosomal proteins[21] (Table 2-4). However, interactions with eIF4 complex components are notably absent and the enrichment of eIF4E or eIF4G, has not been detected. While this may be a limitation

of BioID and AP-LC-MS/MS, the exclusion of eIF4E or eIF4G may also be functionally relevant. In this regard, it is intriguing that our data shows that ARS2n interacts with eIF3D, which has been recently shown to drive cap-dependent, eIF4E-independent translation[307,308]. Although the current literature suggests ARS2 isoforms function in RNA metabolism, through the interaction with the CBC, the new RNA maturation model should be considered in future functional analysis of ARS2 isoforms.

References

1. Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: Control of transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* vol. 19 464–478 (2018).

2. Eaton, J. D. & West, S. Termination of Transcription by RNA Polymerase II : BOOM ! *Trends Genet.* **36**, 664–675 (2020).

3. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).

4. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics* vol. 13 720–731 (2012).

5. Ghosh, A., Shuman, S. & Lima, C. D. Structural Insights to How Mammalian Capping Enzyme Reads the CTD Code. *Mol. Cell* **43**, 299–310 (2011).

6. Furuichi, Y., Lafiandra, A. & Shatkin, A. J. 5’ -Terminal structure and mRNA stability. *Nature* **266**, 235–239 (1977).

7. Izaurralde, E., Lewis, J., Mcguigan, C. & Darzynkiewicz, E. A Nuclear Cap Binding Protein Complex Involved in Pre-mRNA Splicing. *Cell* **78**, 657–668 (1994).

8. Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B. & Mattaj, I. W. A Nuclear Cap-binding Complex Binds Balbiani Ring Pre-mRNA Cotranscriptionally and Accompanies the Ribonucleoprotein Particle during Nuclear Export. *J. Cell Biol.* **133**, 5–14 (1996).

9. Flaherty, M, Fortes, P, Izaurralde, E, Mattaj, I and Gilmartin, G. Participation of the nuclear cap binding complex in pre-mRNA 3 ˜processing. *PNAS* **94**, 11893–11898 (1997).

10. Narita, T. *et al.* NELF Interacts with CBC and Participates in 3′ End Processing of Replication-Dependent Histone mRNAs. *Mol. Cell* **26**, 349–365 (2007).

11. Laubinger, S. *et al.* Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *PNAS* **105**, 8795–8800 (2008).

12. Gruber, J. J. *et al.* Ars2 Promotes proper replication-dependent histone mRNA 3' end formation. *Mol. Cell* **45**, 87–98 (2012).

13. Hallais, M. *et al.* CBC-ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* **20**, 1358–1366 (2013).

14. Andersen, P. R. *et al.* The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Publ. Gr.* **20**, 1367–1376 (2013).

15. Gruber, J. J. *et al.* Ars2 Links the Nuclear Cap-Binding Complex to RNA Interference and Cell Proliferation. *Cell* **138**, 328–339 (2009).

16. Kiriyama, M., Kobayashi, Y., Saito, M., Ishikawa, F. & Yonehara, S. Interaction of FLASH with Arsenite Resistance Protein 2 Is Involved in Cell Cycle Progression at S Phase. *Mol. Cell. Biol.* **29**, 4729–4741 (2009).

17. Sullivan, C. O. & Howard, P. L. The Diverse Requirements of ARS2 in nuclear cap-binding complex-dependent RNA Processing. *Rna Dis.* **3**, 1–16 (2016).

18. O'Sullivan, C. *et al.* Mutagenesis of ARS2 Domains To Assess Possible Roles in Cell Cycle Progression and MicroRNA and Replication-Dependent Histone mRNA Biogenesis. *Mol. Cell. Biol.* **35**, 3753–3767 (2015).

19. Machida, S., Chen, H. & Yuan, Y. A. Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. *Nucleic Acids Res.* **39**, 7828–7836 (2011).

20. Iwata, Y., Takahashi, M., Fedoroff, N. V & Hamdan, S. M. Dissecting the interactions of SERRATE with RNA and DICER-LIKE 1 in Arabidopsis microRNA precursor processing. *Nucleic Acids Res.* **41**, 9129–9140 (2013).

21. Schulze, W. M., Stein, F., Rettel, M., Nanao, M. & Cusack, S. Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting. *Nat. Commun.* **9**, 1–15 (2018).

22. Schulze, W. M. & Cusack, S. Structural basis for mutually exclusive co-transcriptional nuclear cap-binding complexes with either NELF-E or ARS2. *Nat. Commun.* **8**, (2017).

23. Giacometti, S. *et al.* Mutually Exclusive CBC-Containing Complexes Contribute to RNA Fate. *Cell Rep.* **18**, 2635–2650 (2017).

24. Brien, J. O., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front. Endocrinol* **9**, 1–12 (2018).

25. Denli, A. M., Tops, B. B. J. & Plasterk, R. H. A. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).

26. Okada, C. *et al.* A High-Resolution Structure of the Pre-microRNA Nuclear Export Machinery. *Science (80-.).* **455**, 1275–1280 (2009).

27. Kawamata, T. & Tomari, Y. Making RISC. *Trends Biochem. Sci.* **35**, 368–376 (2010).

28. Lee, Y. S. *et al.* Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. *Cell* **117**, 69–81 (2004).

29. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H. & Tomari, Y. Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat. Struct. Mol. Biol.* **18**, 1153–1159 (2011).

30. Okamura, K., Ishizuka, A., Siomi, H. & Siomi, M. C. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *GENES Dev.* **18**, 1655–1666 (2004).

31. Sabin, L. R. *et al.* Ars2 regulates both miRNA- and siRNA-dependent silencing and suppresses RNA virus infection in Drosophila Leah. *Cell* **138**, 340–351 (2009).

32. Ding, S. & Voinnet, O. Antiviral Immunity Directed by Small RNAs. *Cell* **130**, 413–426 (2007).

33. Kim, K., Lee, Y. S., Harris, D., Nakahara, K. & Carthew, R. W. The RNAi Pathway

Initiated by Dicer-2 in Drosophila. *Cold Spring Harb. Symp. Quant. Biol.* **LXXI**, 39–44 (2006).

34. Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**, 843–854 (2009).

35. Mandel, C. R. *et al.* Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**, 953–956 (2006).

36. Dominski, Z., Yang, X. & Marzluff, W. F. The Polyadenylation Factor CPSF-73 Is Involved in Histone-Pre-mRNA Processing. *Cell* **123**, 37–48 (2005).

37. Ryan, K., Calvo, O. & Manley, J. L. Evidence that polyadenylation factor CPSF-73 is the mRNA 3 ˜processing endonuclease. *RNA* **10**, 565–573 (2004).

38. Sabath, I. *et al.* 3 *′* -End processing of histone pre-mRNAs in Drosophila : U7 snRNP is associated with FLASH and polyadenylation factors. *RNA* **19**, 1726–1744 (2013).

39. Wang, Z., Whitfield, M. L., Ingledue, C., Dominski, Z. & Marzluff, W. F. The protein that binds the 3 ' end of histone mRNA : a novel RNA-binding protein required for histone pre-mRNA processing. *GENES Dev.* **10**, 3028–3040 (1996).

40. Dominski, Z., Yang, X., Kaygun, H., Dadlez, M. & Marzluff, W. F. A 3 ˜ Exonuclease that Specifically Interacts with the 3 ˜End of Histone mRNA. *Mol. Cell* **12**, 295–305 (2003).

41. Scharl, E. C. & Steitz, J. A. The site of 3 ' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. *EMBO J.* **13**, 2432–2440 (1994).

42. Yang, X., Torres, M. P., Marzluff, W. F. & Dominski, Z. Three Proteins of the U7-Specific Sm Ring Function as the Molecular Ruler To Determine the Site of 3' -End Processing in Mammalian Histone Pre-mRNA. *Mol. Cell. Biol.* **29**, 4045–4056 (2009).

43. Sullivan, E. *et al.* Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *GENES Dev.* **15**, 173–187 (2001).

44. Yang, X., Burch, B. D., Yan, Y., Marzluff, W. F. & Dominski, Z. FLASH , a Proapoptotic Protein Involved in Activation of Caspase-8 , Is Essential for 3' End Processing of Histone Pre-mRNAs. *Mol. Cell* **36**, 267–278 (2009).

45. Yang, X. *et al.* FLASH Is Required for the Endonucleolytic Cleavage of Histone Pre-mRNAs but Is Dispensable for the 5' Exonucleolytic Degradation of the Downstream Cleavage Product. *Mol. Cell. Biol.* **31**, 1492–1502 (2011).

46. Yang, X. *et al.* A Complex Containing the CPSF73 Endonuclease and Other Polyadenylation Factors Associates with U7 snRNP and Is Recruited to Histone Pre-mRNA for 3'-End Processing. *Mol. Cell. Biol.* **33**, 28–37 (2013).

47. Burch, B. D. *et al.* Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila. *RNA* **17**, 1132–1147 (2011).

48. Barcaroli, D. *et al.* FLASH is required for histone transcription and S-phase progression. *PNAS* **103**, 14808–14812 (2006).

49. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).

50. Will, C. L. & Luhrman, R. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **3**, 1–23 (2011).

51. Kornblihtt, A. R. *et al.* Alternative splicing : a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* **14**, 153–166 (2013).

52. Raczynska, K. D. *et al.* Involvement of the nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Res.* **38**, 265–278 (2010).

53. Raczynska, K. D. *et al.* The SERRATE protein is involved in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Res.* **42**, 1224–1244 (2014).

54. Speth, C. *et al.* Arabidopsis RNA processing factor SERRATE regulates the transcription of intronless genes. *Elife* **7**, 1–21 (2018).

55. Lewis, J. D. & Izaurfllde, E. The role of the cap structure in RNA processing and nuclear export. *Eur. J. Biochem.* **469**, 461–469 (1998).

56. Mullane, L. O. & Eperon, I. A. N. C. The Pre-mRNA 5’ Cap Determines Whether U6 Small Nuclear RNA Succeeds U1 Small Nuclear Ribonucleoprotein Particle at 5’ Splice Sites. *Mol. Cell. Biol.* **18**, 7510–7520 (1998).

57. Lenasi, T., Peterlin, B. M. & Barboric, M. Cap-binding Protein Complex Links Pre-mRNA Capping to Transcription Elongation and Alternative Splicing through Positive Transcription Elongation Factor b (P-TEFb). *J. Biol. Chem.* **286**, 22758–22768 (2011).

58. Pabis, M. *et al.* The nuclear cap-binding complex interacts with the U4 / U6 · U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA* **19**, 1054–1063 (2013).

59. Thillainadesan, G. *et al.* Conserved protein Pir2ARS2 mediates gene repression through cryptic introns in lncRNAs. *Nat. Commun.* **11**, (2020).

60. Shi, Y. *et al.* Molecular architecture of the human pre-mRNA 3′ processing complex. *Mol. Cell* **33**, 365–376 (2009).

61. Hollerer, I., Grund, K., Hentze, M. W. & Kulozik, A. E. mRNA 3 ′ end processing : A tale of the tail reaches the clinic. *EMBO Mol. Med.* **6**, 16–26 (2014).

62. Millevoi, S. & Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3’ end processing regulation. *Nucleic Acids Res.* **38**, 2757–2774 (2010).

63. Sun, Y. *et al.* Molecular basis for the recognition of the human AAUAAA polyadenylation signal. *PNAS* **115**, 1419–1428 (2018).

64. Soles, L. V & Shi, Y. Crosstalk Between mRNA 3 ’-End Processing and Epigenetics. *Front. Genet.* **12**, 1–8 (2021).

65. Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA

polyadenylation. *GENES Dev.* **11**, 2755–2766 (1997).

66. Shi, Y. Alternative polyadenylation : New insights from global analyses. *RNA* **18**, 2105–2117 (2012).

67. Tian, B. & Manley, J. L. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18**, 18–30 (2017).

68. Serena Chan, Eun-A Choi, and Y. S. Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip Rev RNA* **2**, 321–335 (2011).

69. Vries, H. De *et al.* Human pre-mRNA cleavage factor II m contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.* **19**, 5895–5904 (2000).

70. Ntini, E. *et al.* Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat. Struct. Mol. Biol.* **20**, 923–928 (2013).

71. Baillat, D. *et al.* Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. *Cell* **123**, 265–276 (2005).

72. Schmid, M. & Jensen, T. H. Controlling nuclear RNA levels. *Nature Reviews Genetics* vol. 19 518–529 (2018).

73. Kilchert, C., Wittmann, S. & Vasiljeva, L. The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17**, 227–239 (2016).

74. Chlebowski, A., Lubas, M., Heick, T. & Dziembowski, A. RNA decay machines : The exosome. *Biochim. Biophys. Acta* **1829**, 552–560 (2013).

75. Schneider, C. & Tollervey, D. Looking into the barrel of the RNA exosome. *Nat. Publ. Gr.* **21**, 17–18 (2014).

76. Meola, N. Targeting the nuclear RNA exosome : Poly (A) binding proteins enter the stage. *RNA Biol.* **14**, 820–826 (2017).

77. Lubas, M. *et al.* The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. *Cell Rep.* **10**, 178–192

(2015).

78. Meola, N. *et al.* Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol. Cell* **64**, 520–533 (2016).

79. Lubas, M. *et al.* Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. *Mol. Cell* **43**, 624–637 (2011).

80. Hrossova, D. *et al.* RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3 -end extended forms of snRNAs. *Nucleic Acids Res.* **43**, 4236–4248 (2015).

81. Gromadzka, A. M., Steckelberg, A., Singh, K. K., Hofmann, K. & Gehring, N. H. A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res.* **44**, 2348–2361 (2016).

82. Fan, J. *et al.* Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J.* **36**, (2017).

83. Melko, M. *et al.* Mapping domains of ARS2 critical for its RNA decay. *Nucleic Acids Res.* **48**, 6943–6953 (2019).

84. Joe, D., Izaurralde, E., Jarmolowski, A., Mcguigan, C. & Mattaj, I. W. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5 ' splice site. *GENES Dev.* **10**, 1683–1698 (1996).

85. Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I. W. PHAX , a Mediator of U snRNA Nuclear Export Whose Activity Is Regulated by Phosphorylation. *Cell* **101**, 187–198 (2000).

86. McCloskey, A., Taniguchi, I., Shinmyozu, K. & Ohno, M. hnRNP C Tetramer Measures RNA Length to Classify RNA Polymerase II Transcripts for Export. *Science (80-.).* 1643–1647 (2012).

87. Masuyama, K., Taniguchi, I., Kataoka, N. & Ohno, M. RNA length defines RNA export pathway. *GENES Dev.* **18**, 2074–2085 (2004).

88. Chi, B. *et al.* Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res.* **41**, 1294–1306 (2013).

89. Dufu, K. *et al.* ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *GENES Dev.* **24**, 2043–2053 (2010).

90. Masuda, S. *et al.* Recruitment of the human TREX complex to mRNA during splicing. *GENES Dev.* **19**, 1512–1517 (2005).

91. Zhou, Z. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401–405 (2000).

92. Luo, M. *et al.* Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**, 644–647 (2001).

93. Hir, Â. Le, Gat, D., Izaurralde, E. & Moore, M. J. The exon ± exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–4997 (2001).

94. Viphakone, N. *et al.* TREX exposes the RNA binding domain of Nxf1 to enable mRNA export. *Nat. Commun.* **3**, 1–24 (2012).

95. Hautbergue, G. M., Hung, M., Golovanov, A. P., Lian, L. & Wilson, S. A. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *PNAS* **105**, 5154–5159 (2008).

96. Braun, I. C., Herold, A., Rode, M., Conti, E. & Izaurralde, E. Overexpression of TAP / p15 Heterodimers Bypasses Nuclear Retention and Stimulates Nuclear mRNA Export *. *J. Biol. Chem.* **276**, 20536–20543 (2001).

97. Braun, I. C., Izaurralde, E. & Conti, E. Structural Basis for the Recognition of a Nucleoporin FG Repeat by the NTF2-like Domain of the TAP / p15 mRNA Nuclear Export Factor. *Mol. Cell* **8**, 645–656 (2001).

98. Nojima, T., Hirose, T., Kimura, H. & Hagiwara, M. The Interaction between Cap-

binding Complex and RNA Export Factor Is Required for Intronless mRNA Export * □. *J. Biol. Chem.* **282**, 15645–15651 (2007).

99. Chi, B. *et al.* A Sub-Element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. *Nucleic Acids Res.* **42**, 7305–7318 (2014).

100. McCloskey, A., Taniguchi, I., Shinmyozu, K. & Ohno, M. Length to Classify RNA Polymerase II Transcripts for Export. *Science (80-.).* 1643–1647 (2012).

101. Cheng, H. *et al.* Human mRNA Export Machinery Recruited to the 5′ End of mRNA. *Cell* **127**, 1389–1400 (2006).

102. Dou, Y. *et al.* NCBP3 positively impacts mRNA biogenesis. *Nucleic Acids Res.* **48**, 10413–10427 (2020).

103. Dou, Y. *et al.* Affinity proteomic dissection of the human nuclear cap-binding complex interactome. *Nucleic Acids Res.* **48**, 10456–10469 (2020).

104. Calero, G. *et al.* Structural basis of mtGpppG binding to the nuclear cap-binding protein complex. *Nat Struct Biol* **9**, 912–917 (2002).

105. Mazza, C., Segref, A., Mattaj, I. W. & Cusack, S. Large-scale induced fit recognition of an m 7 GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.* **21**, 5548–5557 (2002).

106. Choe, J. *et al.* Translation Initiation on mRNAs Bound by Nuclear Cap-binding Protein Complex CBP80 / 20 Requires Interaction between CBP80 / 20-dependent Translation Initiation Factor and Eukaryotic Translation Initiation Factor 3g. *J. Biol. Chem.* **287**, 18500–18509 (2012).

107. Isken, O. *et al.* Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay. *Cell* **133**, 314–327 (2008).

108. Hwang, J., Sato, H., Tang, Y., Matsuda, D. & Maquat, L. E. UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. *Mol. Cell* **39**, 396–409 (2010).

109. Rossman, T. G. & Wang, Z. Expression cloning for arsenite-resistance resulted in isolation of tumor-suppressor fau cDNA: Possible involvement of the ubiquitin system in arsenic carcinogenesis. *Carcinogenesis* **20**, 311–316 (1999).

110. Iland, H. J. & Seymour, J. F. Role of Arsenic Trioxide in Acute Promyelocytic Leukemia. *Curr. Treat. Options Oncol.* **14**, 170–184 (2013).

111. Wilson, M. D. *et al.* ARS2 Is a Conserved Eukaryotic Gene Essential for Early Mammalian Development. *Mol. Cell. Biol.* **28**, 1503–1514 (2008).

112. Christie, J. Characterizing ARS2 localization and function in differentiating myoblasts. *Unpubl. Univ. Victoria* (2015).

113. Andjus, S., Antonin Morillon & Wery, M. From Yeast to Mammals, the Nonsense-Mediated mRNA Decay as a Master Regulator of Long Non-Coding RNAs Functional Trajectory. *Noncoding RNA* **7**, (2021).

114. Muhlrad, D. & Parker, R. Premature translational termination triggers mRNA decapping. *Nature* **370**, 5–8 (1994).

115. Yi, Z., Sanjeev, M. & Singh, G. The Branched Nature of the Nonsense-Mediated mRNA Decay Pathway. *Trends Genet.* **37**, 143–159 (2021).

116. He, F. & Jacobson, A. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annu Rev Genet* **49**, 339–366 (2016).

117. Karousis, E. D., Mühlemann, O., Nasif, S. & Mühlemann, O. Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip. Rev. RNA* **7**, 661–682 (2016).

118. Lykke-andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nat. Publ. Gr.* **16**, 665–677 (2015).

119. Kishor, A., Fritz, S. E. & Hogg, J. R. Nonsense-mediated mRNA decay: the challenge of telling right from wrong in a complex transcriptome. *Wiley Interdiscip*

Rev RNA **10**, 1–35 (2020).

120. Kurosaki, T., Popp, M. W., Maquat, L. E. & Popp, M. W. Quality and quantity control of gene expression by nonsense- mediated mRNA decay. *Nat Rev Mol Cell Biol* **20**, 406–420 (2019).

121. Kebaara, B. W. & Atkin, A. L. Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. *Nucleic Acids Res.* **37**, 2771–2778 (2009).

122. Yepiskoposyan, H. & Aeschimann, F. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17**, 2108–2118 (2011).

123. Huang, L. *et al.* Article RNA Homeostasis Governed by Cell Type-Specific and Branched Feedback Loops Acting on NMD. *Mol. Cell* **43**, 950–961 (2011).

124. Amrani, N. *et al.* A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432**, 112–118 (2004).

125. Karousis, E. D. & Mühlemann, O. Nonsense-mediated mRNA decay begins where translation ends. *Cold Spring Harb. Perspect. Biol.* **11**, 1–18 (2019).

126. Lavysh, D. & Neu-yilik, G. UPF1-Mediated RNA Decay — Danse Macabre in a Cloud. *Biomol. Rev.* **4**, (2020).

127. Karam, R. *et al.* The unfolded protein response is shaped by the NMD pathway. *EMBO Rep.* **16**, 599–609 (2015).

128. Lou, C., Shum, E. Y. & Wilkinson, M. F. RNA degradation drives stem cell differentiation. *EMBO J.* **34**, 2013–2015 (2015).

129. Jaffrey, S. R. & Wilkinson, M. F. Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. *Nat. Rev. Neurosci.* **19**, 715–728 (2018).

130. Weischenfeldt, J. *et al.* NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.* **22**, 1381–1396 (2008).

131. Medghalchi, S. M. *et al.* Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* **10**, 99–105 (2001).

132. Goetz, A. E. & Wilkinson, M. *Stress and the nonsense-mediated RNA decay pathway*. *Cell. Mol. Life Sci.* vol. 74 (2017).

133. Rodriguez-Gabriel, M. A., Watt, S., Bahler, J. & Russell, P. Upf1, an RNA Helicase Required for Nonsense-Mediated mRNA Decay, Modulates the Transcriptional Response to Oxidative Stress in Fission Yeast. *Mol. Cell. Biol.* **26**, 6347–6356 (2006).

134. Thoren, L. A. *et al.* UPF2 is a critical regulator of liver development, function and regeneration. *PLoS One* **5**, (2010).

135. Rock Pulak & Philip Anderson Department. mRNA surveillance by the Caenorhabditis elegans stag genes. *GENES Dev.* 1885–1897 (1993).

136. He, F., Brown, A. H. & Jacobson, A. Upf1p , Nmd2p , and Upf3p Are Interacting Components of the Yeast Nonsense-Mediated mRNA Decay Pathway. *Mol. Cell. Biol.* **17**, 1580–1594 (1997).

137. Wittkopp, N. *et al.* Nonsense-Mediated mRNA Decay Effectors Are Essential for Zebrafish Embryonic Development and Survival. *Mol. Cell. Biol.* **29**, 3517–3528 (2009).

138. Woodward, L. A., Mabin, J. W., Gangras, P. & Singh, G. The exon junction complex: a lifelong guardian of mRNA fate. *Wiley Interdiscip. Rev. RNA* **8**, (2017).

139. Boehm, V. & Gehring, N. H. Exon Junction Complexes: Supervising the Gene Expression Assembly Line. *Trends in Genetics* vol. 32 724–735 (2016).

140. Hir, H. Le, Saulière, J. & Wang, Z. The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* **17**, 41–54 (2016).

141. Tange, T., Nott, A. & Moore, M. J. The ever-increasing complexities of the exon junction complex. *Current Opinion in Cell Biology* vol. 16 279–284 (2004).

142. Buchwald, G. *et al.* Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *PNAS* **107**, 10050–10055 (2010).

143. Chamieh, H., Ballut, L., Bonneau, F. & Le Hir, H. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* **15**, 85–93 (2008).

144. Gehring, N. H., Neu-Yilik, G., Schell, T., Hentze, M. W. & Kulozik, A. E. Y14 and hUpf3b form an NMD-activating complex. *Mol. Cell* **11**, 939–949 (2003).

145. Kashima, I. *et al.* Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.* **20**, 355–367 (2006).

146. Ohnishi, T. *et al.* Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol. Cell* **12**, 1187–1200 (2003).

147. Nott, A., Hir, H. Le & Moore, M. J. Splicing enhances translation in mammalian cells : an additional function of the exon junction complex. *Genes Dev.* **18**, 210–222 (2004).

148. Causier, B. *et al.* Conservation of Nonsense-Mediated mRNA Decay Complex Components Throughout Eukaryotic Evolution. *Sci. Rep.* **7**, 1–12 (2017).

149. Kim, Y. K. I. & Maquat, L. E. UPFront and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *RNA* **25**, 407–422 (2019).

150. Fiorini, F., Bagchi, D., Hir, L. & Croquette, V. Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat. Commun.* **6**, (2015).

151. Lee, S. R., Pratt, G. A., Martinez, F. J., Yeo, G. W. & Lykke-Andersen, J. Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity. *Mol. Cell* **59**, 413–425 (2015).

152. Franks, T. M., Singh, G. & Lykke-andersen, J. Upf1 ATPase-Dependent mRNP

Disassembly Is Required for Completion of Nonsense- Mediated mRNA Decay. *Cell* **143**, 938–950 (2010).

153. Kurosaki, T. & Maquat, L. E. Rules that govern UPF1 binding to mRNA 3′ UTRs. *Proc. Natl. Acad. Sci.* **110**, 3357–3362 (2013).

154. Zünd, D., Gruber, A. R., Zavolan, M. & Mühlemann, O. Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3′ UTRs. *Nat. Struct. Mol. Biol.* **20**, 936–943 (2013).

155. Hogg, J. R. & Goff, S. P. Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* **143**, 379–389 (2010).

156. Ivanov, P. V, Gehring, N. H., Kunz, J. B., Hentze, M. W. & Kulozik, A. E. Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. *EMBO J.* **27**, 736–747 (2008).

157. Singh, G., Rebbapragada, I. & Lykke-Andersen, J. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.* **6**, 860–871 (2008).

158. Eberle, A. B., Stalder, L., Mathys, H., Orozco, R. Z. & Mühlemann, O. Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. *PLoS Biol.* **6**, 849–859 (2008).

159. Behm-ansmant, I., Gatfield, D., Rehwinkel, J. & Izaurralde, E. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J.* **26**, 1591–1601 (2007).

160. Peixeiro, I. *et al.* Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. *Nucleic Acids Res.* **40**, 1160–1173 (2012).

161. Grimson, A., O'Connor, S., Newman, C. L. & Anderson, P. SMG-1 Is a Phosphatidylinositol Kinase-Related Protein Kinase Required for Nonsense-Mediated mRNA Decay in Caenorhabditis elegans. *Mol. Cell. Biol.* **24**, 7483–7490

(2004).

162. Denning, G., Jamieson, L., Maquat, L. E., Thompson, E. A. & Fields, A. P. Cloning of a Novel Phosphatidylinositol Kinase-related Kinase. *J. Biol. Chem.* **276**, 22709–22714 (2001).

163. Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y. & Ohno, S. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.* **15**, 2215–2228 (2001).

164. Chakrabarti, S. *et al.* Molecular Mechanisms for the RNA-Dependent ATPase Activity of Upf1 and Its Regulation by Upf2. *Mol. Cell* **41**, 693–703 (2011).

165. Lopez-Perrote, A. *et al.* Human nonsense-mediated mRNA decay factor UPF2 interacts directly with eRF3 and the SURF complex Andr es. *Nucleic Acids Res.* **44**, 1909–1923 (2016).

166. Shum, E. Y. *et al.* The Antagonistic Gene Paralogs Upf3a and Upf3b Article The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay. *Cell* **165**, 382–395 (2016).

167. Kunz, J. B., Neu-yilik, G., Hentze, M. W., Kulozik, A. E. & Gehring, N. H. Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. *RNA* **12**, 1015–1022 (2006).

168. Neu- Yilik, G. *et al.* Dual function of UPF3B in early and late translation termination. *EMBO J.* **36**, 2968–2986 (2017).

169. Yamashita, A. *et al.* SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. *Genes Dev.* **23**, 1091–1105 (2009).

170. Arias-palomo, E. *et al.* The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8. *GENES Dev.* **10**, 153–164 (2011).

171. Deniaud, A. A. *et al.* A network of SMG-8, SMG-9 and SMG-1 C-terminal insertion domain regulates UPF1 substrate recruitment and phosphorylation. *Nucleic Acids Res.* **43**, 7600–7611 (2015).

172. Fernandez, I. *et al.* Characterization of SMG-9, an essential component of the nonsense-mediated mRNA decay SMG1C complex. *Nucleic Acids Res.* **39**, 347–358 (2011).

173. Melero, R. *et al.* Structures of SMG1-UPFs Complexes: SMG1 Contributes to Regulate UPF2-Dependent Activation of UPF1 in NMD. *Structure* **3**, 1105–1119 (2014).

174. Okada-Katsuhata, Y. *et al.* N-and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res.* **40**, 1251–1266 (2012).

175. Chakrabarti, S., Bonneau, F., Sch, S. & Eppinger, E. Phospho-dependent and phospho-independent interactions of the helicase UPF1 with the NMD factors SMG5-SMG7 and SMG6. *Nucleic Acids Res.* **42**, 9447–9460 (2014).

176. Jonas, S., Weichenrieder, O. & Izaurralde, E. An unusual arrangement of two 14-3-3- like domains in the SMG5 – SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *GENES Dev.* **27**, 211–225 (2013).

177. Nicholson, P., Josi, C., Kurosawa, H., Yamashita, A. & Mühlemann, O. A novel phosphorylation-independent interaction between SMG6 and UPF1 is essential for human NMD. *Nucleic Acids Res.* **42**, 9217–9235 (2014).

178. Lykke-Andersen, S. *et al.* Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* **28**, 2498–2517 (2014).

179. Boehm, V. *et al.* 3′ UTR Length and Messenger Ribonucleoprotein Composition Determine Endocleavage Efficiencies at Termination Codons. *Cell Rep.* **9**, 555–568 (2014).

180. Loh, B., Jonas, S. & Izaurralde, E. The SMG5 – SMG7 heterodimer directly

recruits the CCR4 – NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *GENES Dev.* **27**, 2125–2138 (2013).

181. Nicholson, P., Gkratsou, A., Josi, C. & Colombo, M. Dissecting the functions of SMG5 , SMG7 and PNRC2 in nonsense- mediated mRNA decay of human cells. *RNA* **24**, 557–573 (2018).

182. Colombo, M., Karousis, E. D., Bourquin, J., Bruggmann, R. & Mühlemann, O. Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA* **23**, 189–201 (2017).

183. Singh, G. *et al.* The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus. *Cell* **151**, 750–764 (2012).

184. Mabin, J. W. *et al.* The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. *Cell Rep.* **25**, 2431-2446.e7 (2018).

185. Aznarez, I. *et al.* Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1. *Cell Rep.* **23**, 2186–2198 (2018).

186. Sato, H., Hosoda, N. & Maquat, L. E. Efficiency of the Pioneer Round of Translation Affects the Cellular Site of Nonsense-Mediated mRNA Decay. *Mol. Cell* **29**, 255–262 (2008).

187. Ge, Z., Quek, B. L., Beemon, K. L. & Hogg, J. R. Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *Elife* **5**, 1–25 (2016).

188. Kishor, A., Ge, Z. & Hogg, J. R. hnRNP L- dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. *EMBO J.* **38**, 1–19 (2019).

189. Fritz, S. E., Ranganathan, S., Wang, C. D. & Hogg, J. R. The RNA-binding protein PTBP1 promotes ATPase- dependent dissociation of the RNA helicase UPF1 to protect transcripts from nonsense-mediated mRNA decay. *J. Biol. Chem.* **295**,

11613–11625 (2020).

190. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family : insights into their role in health and disease. *Hum. Genet.* **135**, 851–867 (2016).

191. Isken, O. & Maquat, L. E. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.* **9**, 699–712 (2008).

192. Gehring, N. H. *et al.* Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol. Cell* **20**, 65–75 (2005).

193. Chan, W. *et al.* An alternative branch of the nonsense-mediated decay pathway. *EMBO J.* **26**, 1820–1830 (2007).

194. Bao, J. *et al.* UPF2-Dependent Nonsense-Mediated mRNA Decay Pathway Is Essential for Spermatogenesis by Selectively Eliminating Longer 3'UTR Transcripts. *PLoS Genet.* **12**, 1–24 (2016).

195. Rufener, S. C. & Mühlemann, O. EIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.* **20**, 710–717 (2013).

196. Ishigaki, Y. *et al.* Evidence for a Pioneer Round of mRNA Translation : mRNAs Subject to Nonsense-Mediated Decay in Mammalian Cells Are Bound by CBP80 and CBP20. *Cell* **106**, 607–617 (2001).

197. Durand, S., Franks, T. M. & Lykke-Andersen, J. Hyperphosphorylation amplifies UPF1 activity to resolve stalls in nonsense-mediated mRNA decay. *Nat. Commun.* **7**, (2016).

198. Hoek, T. A. *et al.* Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. *Mol. Cell* **75**, 324-339.e11 (2019).

199. Gardner, L. B. Hypoxic Inhibition of Nonsense-Mediated RNA Decay Regulates Gene Expression and the Integrated Stress Response. *Mol. Cell. Biol.* **28**, 3729–3741 (2008).

200. Brown, J. A. L. *et al.* A Novel Role for hSMG-1 in Stress Granule Formation. *Mol. Cell. Biol.* **31**, 4417–4429 (2011).

201. Wang, D. *et al.* Inhibition of Nonsense-Mediated RNA Decay by the Tumor Microenvironment Promotes Tumorigenesis. *Mol. Cell. Biol.* **31**, 3670–3680 (2011).

202. Hetz, C. & Kaufman, R. J. Mechanisms, regulation and functions of the unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **21**, 421–438 (2020).

203. Fulda, S., Gorman, A. M., Hori, O. & Samali, A. Cellular Stress Responses: Cell Survival and Cell Death. *Int J Cell Biol* **2010**, (2010).

204. Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science (80-.).* **334**, 1081–1086 (2005).

205. Tsai, Y. C. & Weissman, A. M. The unfolded protein response, degradation from the endoplasmic reticulum, and cancer. *Genes and Cancer* **1**, 764–778 (2010).

206. Srivastava, R. K. *et al.* ATF4 regulates arsenic trioxide-mediated NADPH oxidase, ER- mitochondrial crosstalk and apoptosis. *Arch. Biochem. Biophys.* **609**, 39–50 (2016).

207. Zhou, J. *et al.* The crystal structure of human IRE1 luminal domain reveals a conserved dimerization interface required for activation of the unfolded protein response. *PNAS* **103**, 14343–14348 (2006).

208. Credle, J. J., Finer-Moore, J. S., Papa, F. R., Stroud, R. M. & Walter, P. On the mechanism of sensing unfolded protein in the endoplasmic reticulum. *PNAS* **102**, 18773–18784 (2005).

209. Shen, X. *et al.* Complementary signaling pathways regulate the unfolded protein response and are required for C. elegans development. *Cell* **107**, 893–903 (2001).

210. Calfon, M. *et al.* IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA. *Nature* **415**, 1–6 (2002).

211. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA Is Induced by ATF6 and Spliced by IRE1 in Response to ER Stress to Produce a Highly Active Transcription Factor. *Cell* **107**, 881–891 (2001).

212. John-Paul Upton *et al.* IRE1a Cleaves Select microRNAs During ER Stress to Derepress Translation of Proapoptotic Caspase-2. *Science (80-.).* **338**, 818–823 (2012).

213. Hollien, J. & Jonathan S. Weissman. Decay of Endoplasmic Reticulum-Localized mRNAs During the Unfolded Protein Response. *Science (80-.).* **313**, 104–108 (2006).

214. Hollien, J. *et al.* Regulated Ire1-dependent decay of messenger RNAs in mammalian cells. *JCB:Report* **186**, 323–331 (2009).

215. Hetz, C. & Papa, F. R. The Unfolded Protein Response and Cell Fate Control. *Mol. Cell* **69**, 169–181 (2018).

216. Ye, J. *et al.* ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Mol. Cell* **6**, 1355–1364 (2000).

217. Shaffer, A. L. *et al.* XBP1, Downstream of Blimp-1, Expands the Secretory Apparatus and Other Organelles, and Increases Protein Synthesis in Plasma Cell Differentiation. *Immunity* **21**, 81–93 (2004).

218. Sriburi, R., Jackowski, S., Mori, K. & Brewer, J. W. XBP1: a link between the unfolded protein response, lipid biosynthesis, and biogenesis of the endoplasmic reticulum. *JCB:Report* **167**, 35–41 (2004).

219. Bommiasamy, H. *et al.* ATF6 α induces XBP1-independent expansion of the endoplasmic reticulum. *J. Cell Sci.* **1**, 1626–1636 (2007).

220. Kaufman, R. J. Orchestrating the unfolded protein response in health and disease. *J. Clin. Invest.* **110**, 1389–1398 (2002).

221. Ron, D. & Walter, P. Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **8**, 519–529 (2007).

222. Lu, P. D., Harding, H. P. & Ron, D. Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *JCB:Report* **167**, 27–33 (2004).

223. Vattem, K. M. & Wek, R. C. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *PNAS* **101**, 11269–11274 (2004).

224. Dever, T. E. *et al.* Phosphorylation of Initiation Factor 2a by Protein Kinase GCN2 Mediates Gene-Specific Translational Control of GCN4 in Yeast. *Cell* **68**, 585–596 (1992).

225. Chan, C. P., Kok, K. H., Tang, H. M. V., Wong, C. M. & Jin, D. Y. Internal ribosome entry site-mediated translational regulation of ATF4 splice variant in mammalian unfolded protein response. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2165–2175 (2013).

226. Harding, H. P. *et al.* An Integrated Stress Response Regulates Amino Acid Metabolism and Resistance to Oxidative Stress. *Mol. Cell* **11**, 619–633 (2003).

227. Han, J. *et al.* ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. *Nat. Cell Biol.* **15**, 481–490 (2013).

228. Muñoz-Pinedo, C. & López-Rivas, A. A role for caspase-8 and TRAIL-R2/DR5 in ER-stress-induced apoptosis. *Cell Death Differ.* **25**, 226 (2018).

229. Lam, M., Lawrence, D. A., Ashkenazi, A. & Walter, P. Confirming a critical role for death receptor 5 and caspase-8 in apoptosis induction by endoplasmic reticulum stress. *Cell Death and Differentiation* vol. 25 1530–1531 (2018).

230. Chang, T. K. *et al.* Coordination between Two Branches of the Unfolded Protein Response Determines Apoptotic Cell Fate. *Mol. Cell* **71**, 629-636.e5 (2018).

231. Lu, M. *et al.* Opposing unfolded-protein-response signals converge on death receptor 5 to control apoptosis. *Science (80-.).* **345**, 98–101 (2015).

232. Chae, H. *et al.* BI-1 Regulates an Apoptosis Pathway Linked to Endoplasmic Reticulum Stress. *Mol. Cell* **15**, 355–366 (2004).

233. Colombo, A., Martı, G., Kiviluoto, S., Rodrı, D. & Patron, M. TMBIM3 / GRINA is a novel unfolded protein response (UPR) target gene that controls apoptosis through the modulation of ER calcium homeostasis. *Cell Death Differ.* **19**, 1013–1026 (2012).

234. Trachootham, D., Lu, W., Ogasawara, M. A., Valle, N. R. & Huang, P. Redox regulation of cell survival. *Antioxid Redox Signal* **10**, (2008).

235. Orrenius, S., Gogvadze, V. & Zhivotovsky, B. Mitochondrial Oxidative Stress : Implications for Cell Death. *Annu Rev Pharmacol Toxicol* **47**, 143–183 (2007).

236. Kuniyasu Niizuma, Hidenori Endo & Pak H. Chan. Oxidative stress and mitochondrial dysfunction as determinants of ischemic neuronal death and survival. *J Neurochem* **109**, 133–138 (2010).

237. John D. Robertson, Chandra, J., Gogvadze, V. & Sten Orrenius. Biological reactive intermediates and mechanisms of cell death. *Adv Exp Med Biol* **500**, 1–10 (2001).

238. Vizcaya-ruiz, A. De, Barbier, O., Ruiz-ramos, R. & Cebrian, M. E. Mutation Research/Genetic Toxicology and Environmental Mutagenesis Biomarkers of oxidative stress and damage in human populations exposed to arsenic. *Mutat. Res.* **674**, 85–92 (2009).

239. Jomova, K. *et al.* Arsenic: toxicity, oxidative stress and human disease. *J. Appl. Toxicol.* **31**, 95–107 (2011).

240. Thakur, M., Rachamalla, M., Niyogi, S., Datusalia, A. K. & Flora, S. J. S. Molecular mechanism of arsenic-induced neurotoxicity including neuronal dysfunctions. *Int J Mol Sci* **22**, (2021).

241. Kligerman, A. D. & Tennant, A. H. Insights into the carcinogenic mode of action of arsenic. *Toxicol. Appl. Pharmacol.* **222**, 281–288 (2007).

242. Díaz-villaseñor, A., Burns, A. L., Hiriart, M., Cebrián, M. E. & Ostrosky-wegman, P. Arsenic-induced alteration in the expression of genes related to type 2 diabetes mellitus. *Toxicol. Appl. Pharmacol.* **225**, 123–133 (2007).

243. Navas-Acien, A. *et al.* Arsenic exposure and cardiovascular disease: A systematic review of the epidemiologic evidence. *American Journal of Epidemiology* vol. 162 1037–1049 (2005).

244. Prozialeck, W. C. *et al.* The Vascular System as a Target of Metal Toxicity. *Toxicol. Sci.* **102**, 207–218 (2008).

245. Schmuck, E. M. *et al.* Characterization of the monomethylarsonate reductase and dehydroascorbate reductase activities of Omega class glutathione transferase variants: Implications for arsenic metabolism and the age-at-onset of Alzheimer's and Parkinson's diseases. *Pharmacogenet. Genomics* **15**, 493–501 (2005).

246. Xia, Y., Fang, H., Zhang, J. & Du, Y. Endoplasmic reticulum stress-mediated apoptosis in imatinib-resistant leukemic K562-r cells triggered by AMN107 combined with arsenic trioxide. *Exp. Biol. Med.* **238**, 932–942 (2013).

247. Du, Y. *et al.* Coordination of intrinsic, extrinsic, and endoplasmic reticulum – mediated apoptosis by imatinib mesylate combined with arsenic trioxide in chronic myeloid leukemia. *Blood* **107**, 1582–1590 (2006).

248. Pourahmad, J., Rabiei, M., Jokar, F. & Brien, P. J. O. A comparison of hepatocyte cytotoxic mechanisms for chromate and arsenite. *Toxicology* **206**, 449–460 (2005).

249. Dilworth, D. *et al.* The basic tilted helical bundle domain of the prolyl isomerase FKBP25 is a novel double-stranded RNA binding module. *Nucleic Acids Res.* **45**, 1–16 (2017).

250. Pereverzev, A. P. *et al.* Method for quantitative analysis of nonsense-mediated mRNA decay at the single cell level. *Sci. Rep.* **5**, 1–10 (2015).

251. Mendell, J. T., Rhys, C. M. J. & Dietz, H. C. Separable Roles for rent1/hUpf1 in Altered Splicing and Decay of Nonsense Transcripts. *Science (80-.).* **298**, 419–422 (2002).

252. Rehwinkel, J. A. N., Behm-ansmant, I., Gatfield, D. & Izaurralde, E. A crucial role for GW182 and the DCP1: DCP2 decapping complex in miRNA-mediated gene

silencing. *RNA* **11**, 1640–1647 (2005).

253. Domanski, M. *et al.* Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. *Biotechniques* **0**, 1–6 (2012).

254. Roux, K. J., Kim, D. I. & Burke, B. BioID: A screen for protein-protein interactions. *Curr Protoc Protein Sci* **74**, 19.23.1-19.23.14 (2013).

255. Senko, M. W. *et al.* Novel Parallelized Quadrupole/Linear Ion Trap/Orbitrap Tribrid Mass Spectrometer Improving Proteome Coverage and Peptide Identification Rates. *Anal Chem.* **85**, 11710−11714 (2013).

256. Ma, B. *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **17**, 2337–2342 (2003).

257. Lambert, J. P., Tucholska, M., Go, C., Knight, J. D. R. R. & Gingras, A. C. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J. Proteomics* **118**, 81–94 (2015).

258. Youn, J. *et al.* High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies Resource High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* **69**, 517–532 (2018).

259. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).

260. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **13**, 2498–2504 (2003).

261. Jensen, L. J. *et al.* STRING 8 — a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, 412–416 (2009).

262. Gurskaya, N. G., Pereverzev, A. P., Staroverov, D. B., Markina, N. M. & Lukyanov, K. A. Analysis of Nonsense-Mediated mRNA Decay at the Single-Cell Level Using Two Fluorescent Proteins. *Methods Enzymol.* **572**, 291–314 (2016).

263. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347**, 1260419-1-1260419–9 (2015).

264. Lobbes, D., Rallapalli, G., Schmidt, D. D., Martin, C. & Clarke, J. SERRATE: A new player on the plant microRNA scene. *EMBO Rep.* **7**, 1052–1058 (2006).

265. Oh, S. W. *et al.* A P-element insertion screen identified mutations in 455 novel essential genes in Drosophila. *Genetics* **163**, 195–201 (2003).

266. Golling, G. *et al.* Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat. Genet.* **31**, 135–140 (2002).

267. Amsterdam, A. *et al.* Identification of 315 genes essential for early zebrafish development. *PNAS* **101**, 12792–12797 (2004).

268. Kim, D. D. U. *et al.* Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. *Nat. Biotechnol.* **28**, 617–623 (2010).

269. Gingras, A., Abe, K. T. & Raught, B. Getting to know the neighborhood: using proximity- dependent biotinylation to characterize protein complexes and map organelles. *Curr. Opin. Chem. Biol.* **48**, 44–54 (2019).

270. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification – mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).

271. Müller-mcnicoll, M. & Neugebauer, K. M. Good cap/bad cap: how the cap-binding complex determines RNA fate. *Nat. Publ. Gr.* **21**, 9–12 (2014).

272. Hosoda, N., Yoon, K. K., Lejeune, F. & Maquat, L. E. CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.* **12**, 893–901 (2005).

273. Popp, M. & Maquat, L. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.* **47**, 139–165 (2013).

274. Nickless, A., Bailis, J. M. & You, Z. Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci.* **7**, 1–12 (2017).

275. Hir, Â. Le, Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20 ± 24 nucleotides upstream of mRNA exon ± exon junctions. *EMBO J.* **19**, 6860–6869 (2000).

276. Tange, T. Ø., Shibuya, T., Jurica, M. S. & Moore, M. J. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* **11**, 1869–1883 (2005).

277. Ewing, R. M. *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3**, 1–17 (2007).

278. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

279. McGlincy, N. J. *et al.* Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. *BMC Genomics* **11**, 1–19 (2010).

280. Hornbeck, P. V *et al.* PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **40**, 261–270 (2012).

281. Amanchy, R. *et al.* Identification of c-Src Tyrosine Kinase Substrates Using Mass Spectrometry and Peptide Microarrays research articles. *J. Proteome Res.* **7**, 3900–3910 (2008).

282. Yin, J. *et al.* ARS2/MAGL signaling in glioblastoma stem cells promotes self-renewal and M2-like polarization of tumor-associated macrophages. *Nat. Commun.* **11**, (2020).

283. Popp, M. W. & Maquat, L. E. Nonsense-mediated mRNA Decay and Cancer. *Curr. Opin. Genet. Dev.* **48**, 44–50 (2018).

284. Supek, F., Lehner, B. & Lindeboom, R. G. H. To NMD or Not To NMD: Nonsense-Mediated mRNA Decay in Cancer and Other Genetic Diseases. *Trends Genet.* **37**, 657–668 (2021).

285. Ryoo, H. D. & Vasudevan, D. Two distinct nodes of translational inhibition in the Integrated Stress Response. *BMB Reports* vol. 50 539–545 (2017).

286. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* vol. 136 731–745 (2009).

287. Preston, A. M. & Hendershot, L. M. Examination of a second node of translational control in the unfolded protein response. *J. Cell Sci.* **126**, 4253–4261 (2013).

288. Mamane, Y., Petroulakis, E., LeBacquer, O. & Sonenberg, N. mTOR, translation initiation and cancer. *Oncogene* **25**, 6416–6422 (2006).

289. Yamaguchi, S. *et al.* ATF4-Mediated Induction of 4E-BP1 Contributes to Pancreatic β Cell Survival under Endoplasmic Reticulum Stress. *Cell Metab.* **7**, 269–276 (2008).

290. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* **19**, 158–174 (2018).

291. Wang, J. & Gribskov, M. IRESpy: An XGBoost model for prediction of internal ribosome entry sites. *BMC Bioinformatics* **20**, 1–15 (2019).

292. Amiss, J. M. The Bioinformatic Identification and Characterization of an Internal Ribosomal Entry Site within the 5 ’ UTR of an Ars2 Transcript Variant. *Unpubl. Univ. Victoria* (2021).

293. Spriggs, K. A., Bushell, M., Mitchell, S. A. & Willis, A. E. Internal ribosome entry segment-mediated translation during apoptosis: The role of IRES-trans-acting factors. *Cell Death Differ.* **12**, 585–591 (2005).

294. Gritsenko, A. A. *et al.* Sequence features of viral and human Internal Ribosome Entry Sites predictive of their activity. *PLoS Comput. Biol.* **13**, 1–23 (2017).

295. Godet, A. C. *et al.* IRES trans-acting factors, key actors of the stress response. *Int. J. Mol. Sci.* **20**, (2019).

296. Mars, J.-C., Ghram, M., Culjkovic-Kraljaic, B. & Borden, K. The Cap-Binding Complex CBC and the Eukaryotic Translation Factor eIF4E: Co-Conspirators in Cap-Dependent RNA Maturation and Translation. *Cancers (Basel).* **13**, (2021).

297. Culjkovic-Kraljacic, B. *et al.* The eukaryotic translation initiation factor eIF4E elevates steady-state m7G capping of coding and noncoding transcripts. *PNAS* **117**, 26773–26783 (2020).

298. Culjkovic-Kraljacic, B., Baguet, A., Volpon, L., Amri, A. & Borden, K. L. B. The Oncogene eIF4E Reprograms the Nuclear Pore Complex to Promote mRNA Export and Oncogenic Transformation. *Cell Rep.* **2**, 207–215 (2012).

299. Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M. & Borden, K. L. B. eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3′UTR. *J. Cell Biol.* **169**, 245–256 (2005).

300. Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M. & Borden, K. L. B. eIF4E is a central node of an RNA regulon that governs cellular proliferation. *J. Cell Biol.* **175**, 415–426 (2006).

301. Davis, M. R., Delaleau, M. & Borden, K. L. B. Nuclear eIF4E Stimulates 3′-End Cleavage of Target RNAs. *Cell Rep.* **27**, 1397-1408.e4 (2019).

302. Volpon, L. *et al.* A biochemical framework for eIF4E-dependent mRNA export and nuclear recycling of the export machinery. *RNA* **23**, 927–937 (2017).

303. Rambout, X. & Maquat, L. E. The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. *Genes Dev.* **34**, 1113–1127 (2020).

304. Cohen, N. *et al.* PML RING suppresses oncogenic transformation by reducing the affinity of eIF4E for mRNA. *EMBO J.* **20**, 4547–4559 (2001).

305. Volpon, L. *et al.* Importin 8 mediates m7G cap-sensitive nuclear import of the

eukaryotic translation initiation factor eIF4E. *PNAS* **113**, 5263–5268 (2016).

306. Park, Y. *et al.* Translation mediated by the nuclear cap-binding complex is confined to the perinuclear region via a CTIF-DDX19B interaction. *Nucleic Acids Res.* **49**, 8261–8276 (2021).

307. Fan, Y. & Guo, Y. Knockdown of eIF3D inhibits breast cancer cell proliferation and invasion through suppressing the Wnt/β-catenin signaling pathway. *Int. J. Clin. Exp. Pathol.* **8**, 10420–10427 (2015).

308. Lee, A. S. Y., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. D. EIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* **536**, 96–99 (2016).

309. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

Appendix A: Supplementary Figures

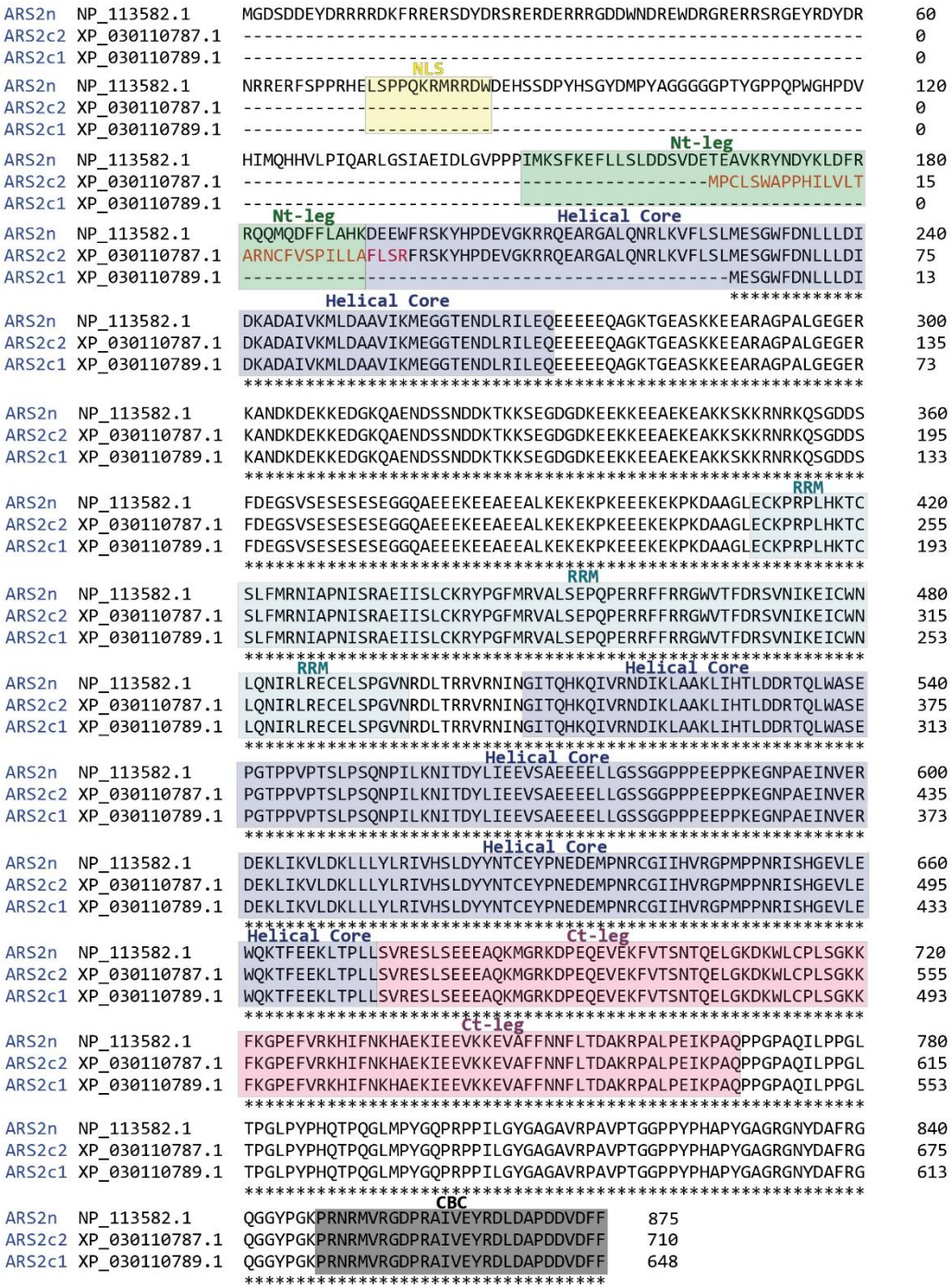


Figure 47: Bioinformatic alignment of ARS2 isoforms. The alignments were generated using Clustal Omega [309]. Additional amino acids corresponding to intron 5 coding sequence were manually added in ARS2c2 (labelled in red). Regions that correspond to structural motifs or domains are outlined with boxes and labeled above the alignment. Nt-leg domain is highlighted in green, Helical Core domain in blue, RRM domain in teal, Ct-leg domain in purple and CBC binding region in black. Asterisks indicate amino acids conservation.

Appendix B: Supplementary Tables

Table 1: List of primers and RNAi used in the study.

SRRT isoforms *Mus musculus*		
Target name	**Forward**	**Reverse**
Ars2n (AF/3UTR)	ATGGGTGACAGTGATGATGAAT	GAGGGGACATGGATGGCTCAA
Ars2n (AF/AR)	ATGGGTGACAGTGATGATGAAT	GACACACTGCCCTCATCGAA
Full sequence amplification (XF/3UTR)	ATGCCCTGCCTTTCTTGGG	GAGGGGACATGGATGGCTCAA
Intron 5 (P1)	AGTGCCCCTGCTTGCCT	AAACGACCCTGAGATCTTGGC
Intron 5 (P5)	CTGTGTGTCTTGGGATCGCT	GACACACTGCCCTCATCGAA
Intron 5 (P3)	CTGTGTGTCTTGGGATCGCT	CCCAAGAAAGGCAGGGCATA
Intron 5 (P4)	GGAGTTGGTGGGAACCTCAT	GACACACTGCCCTCATCGAA
Intron 5 (P2)	GTCTCTCCCATCCTGCTTGC	GACACACTGCCCTCATCGAA
SRRT isoforms *Homo sapiens*	**Forward**	**Reverse**
Intron 5 (P1)	CACCATACCCAGCCCATACAT	TAGCTGCCTGAGCAATCACAT
Intron 5 (P2)	GGGAACTACACCCGCTCATC	AGCTAAACCGCTGCATTGTG
Full sequence amplification (XF/3UTR)	CACCATACCCAGCCCATACAT	CTGTTGCGAGGTTTCCCAGG
NMD regulated genes	**Forward**	**Reverse**
HNRNPA2B1 total human	TGAGCCAAAACGTGCTGTAG	TTTAATTCCGCCAACAAACAG
*HNRNPA2B1*_a human	CTATTTGCCATGGGCTTCAC	AATGGCTCTCTGCATCTGCT
*HNRNPA2B1*_b human	TCTGCTGCCACAAAGACTGTA	GCAGCAAGACACCTTCCATT
Tuba 8 mouse	CACTCAGGCCAGCAAGATCA	GCCCGCACCTCATCTACTAC
Atf4 mouse	ATGGCCGGCTATGGATGAT	CGAAGTCAAACTCTTTCAGATC CATT
Rassf1 mouse	CTGTAGAGCGGGAGACACC	GCACTGAAACAGGACGCACT
GAPDH hum/mouse	ATGACCACAGTCCATGCCATC	CCAGTGAGCTTCCCGTTCA
GFP1	ATGGACGAGCTGTACAGGTGAA	GCACACAGACCAGCACGTTG
GFP2	AACGTGTACATCATGCCCGA	CGTTGGTCTGGTAGTGGTCG

TUBA 8 human	TCAGATTGGCAATGCCTGCT	CATCAAAAGTGCCGTCTGCC
SRSF7 human	TCGAGGTCAAGATCAAGATCCA	CTGCGAGGACTTCCTGATGG
ASNS human	CCTCTCCAGACATTTGCAATT	ACTTCATCCAGAGCCTGAATG
Cbp80 mouse	TCCTTTGTACAGTTGCACGTCT	TTCCTGGCATTCAGCAGTCC
List of sh/RNAi used in the study		
Ars2 si1 mouse	CCCGTCGTGTCCGCAACATAA	
Ars2 si2 mouse	CCGAAGAAGCACTTAAAGAAA	
Ars2 si3 mouse	CCCAGCTTTGCCTGAGATCAA	
ARS2 sh/si (all) human/mouse	ATGCAGCACCATGTCCTGCCTATCCAGGC	
ARS2c sh/si (c1, c2) human/mouse	ATCTTGAGCTCTGTTTGACCAAGCGGCCA	
Ars2c2 sh/si (c2 only) mouse	CATTCTGGTCTTGACTGCTCGTAACTGTT	
UPF1 si human	GUGACGAGUUUAAAUCACAAAUCGA	

Table 2: SAINT analysis of LC-MS/MS identified proteins in BioID2-ARS2n precipitates. Proteins identified by LC-MS/MS from streptavidin affinity purifications with BioID2-ARS2n constructs in HEK 293 cells. Three independent experiments with 5 biological replicates each were analyzed by LC-MS/MS. The table contains the results for protein enrichment in BioID2-ARS2n vs BioID2 control. Additional negative controls were included: HEK293 cell background control and NLS-BirA control. SAINT software was used on the analysis. SP score > 0.7 was defined as cut-off for significantly enriched proteins.

Bait	Prey	Spec	Num Replicates	Ctrl Counts	Saint Score	Fold Change
ARS2n	KPNA3	6|4|6	3	0|0|2|0|0|0|0|0|0|0|0|0	1	34.67
ARS2n	EIF3D	8|6|11	3	1|0|2|0|0|0|2|2|0|0|0|0	1	15.48
ARS2n	PQBP1	7|7|7	3	2|0|0|0|0|0|0|0|0|0|0|0	1	45.5
ARS2n	SLU7	7|5|4	3	0|0|2|0|0|0|0|0|0|0|0|0	1	34.67
ARS2n	HNRNPH3	9|8|13	3	2|1|2|0|0|0|1|1|0|1|0|0|0	1	16.25
ARS2n	KPNA2	24|22|31	3	8|4|9|0|0|0|2|0|0|0|6|3|2	1	9.81
ARS2n	CDC20	5|4|5	3	0|0|0|0|0|0|0|0|0|0|0|0	1	46.67
ARS2n	PRPF31	5|5|7	3	1|0|0|0|0|0|0|0|1|0|0|0	1	36.83
ARS2n	RTF1	5|3|4	3	0|0|0|0|0|0|0|0|0|0|0|0	1	40
ARS2n	HNRNPUL1	3|4|6	3	0|0|0|0|0|0|0|0|0|0|0|0	1	43.33
ARS2n	ZNF703	3|3|5	3	0|0|1|0|0|0|0|0|0|0|0|0	1	36.67
ARS2n	NCOA5	6|6|6	3	1|0|0|0|0|0|0|0|0|0|0|0	1	60
ARS2n	RALY	9|10|9	3	2|1|0|0|0|4|3|2|2|2|1|1|0	1	6.74
ARS2n	RBM7	5|4|3	3	0|0|0|0|0|0|0|0|0|0|0|0	1	40

ARS2n	RBM6	3|4|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	30
ARS2n	NCBP3	6|5|5	3	2|1|1|0|0|0|0|0|0|1|1|0	0.99	11.56
ARS2n	RBM12B	5|2|8	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	50
ARS2n	DAZAP1	3|2|3	3	0|0|0|0|1|0|0|0|0|0|0|0	0.99	26.67
ARS2n	PSMA4	5|2|6	3	0|0|1|0|0|0|0|0|0|0|1|0	0.98	28.17
ARS2n	RPS21	2|3|2	3	0|0|1|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2n	CSNK2B	3|2|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2n	DHX38	2|2|4	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	26.67
ARS2n	UFD1	6|4|5	3	3|0|0|0|0|0|0|0|0|1|0|0	0.98	16.25
ARS2n	ZNF503	2|2|5	3	0|0|1|0|0|0|0|0|0|0|0|0	0.98	30
ARS2n	YJU2	2|2|4	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	26.67
ARS2n	RTF2	2|2|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2n	DHX16	2|3|5	3	0|0|2|0|0|0|0|0|0|0|0|0	0.97	21.67
ARS2n	XRN2	4|4|6	3	0|0|0|0|1|2|2|0|1|0|0|0|0	0.97	10.11
ARS2n	SUGP1	5|5|6	3	3|0|2|0|0|0|0|0|0|1|0|2	0.96	8.67
ARS2n	ESS2	12|13|12	3	7|3|8|0|0|0|0|0|0|0|1|0	0.96	8.44
ARS2n	USP22	2|2|5	3	2|0|0|0|0|0|0|0|0|0|0|0	0.96	19.5
ARS2n	ELAVL1	13|10|13	3	4|4|5|0|2|0|5|0|2|3|6|4|5	0.95	3.9
ARS2n	CDC73	3|4|11	3	0|0|5|0|0|0|0|0|0|0|0|0	0.95	15.6
ARS2n	RBM42	6|4|8	3	2|0|4|0|0|0|0|0|0|1|0|0	0.95	11.14
ARS2n	HNRNPDL	10|10|10	3	6|5|4|0|0|1|0|0|0|2|4|3|3	0.94	4.64
ARS2n	TCEAL4	6|6|5	3	0|3|5|0|0|0|0|0|0|0|0|0	0.94	9.21
ARS2n	TXNL1	11|11|12	3	7|6|6|0|0|0|0|0|0|3|4|0	0.93	5.67
ARS2n	MFAP1	10|9|7	3	3|4|4|0|0|0|0|0|0|4|4|3	0.92	5.12
ARS2n	RAVER1	6|5|6	3	3|1|4|0|0|0|0|0|0|2|1|1	0.92	6.14
ARS2n	ATF1	3|3|4	3	2|0|2|0|0|0|0|0|0|1|1|0	0.91	7.22
ARS2n	MPP1	7|7|10	3	6|5|2|0|0|0|0|0|0|2|1|2	0.91	5.78
ARS2n	PPIL4	10|11|14	3	8|6|7|3|0|0|0|0|0|3|2|1	0.91	5.06
ARS2n	POLDIP3	13|13|15	3	7|8|8|0|0|0|0|0|0|5|6|5	0.91	4.56
ARS2n	PHAX	14|15|17	3	7|7|12|0|0|0|0|0|0|5|6|6	0.91	4.64
ARS2n	WDR70	6|7|9	3	4|5|6|0|0|0|0|0|0|0|0|0	0.91	6.36
ARS2n	TPX2	6|5|8	3	3|6|2|0|0|0|0|0|0|0|0|1	0.91	6.86
ARS2n	EIF3H	8|6|10	3	4|4|9|0|0|0|0|0|0|0|0|0	0.9	6.12
ARS2n	GOLPH3	4|4|7	3	3|0|4|0|0|0|0|0|0|1|0|0	0.9	8.12
ARS2n	PIN1	3|2|2	3	2|1|0|0|0|0|0|0|0|0|0|0	0.89	10.11
ARS2n	RCC2	17|16|17	3	13|16|14|0|0|0|0|1|2|0|0|2|2	0.89	4.33

ARS2n	ZNF281	2|3|2	3	0|1|2|0|0|0|0|0|0|0|0|0	0.89	10.11
ARS2n	GTF2F1	5|4|5	3	3|2|4|0|0|0|0|0|0|1|0|0	0.88	6.07
ARS2n	SMNDC1	3|2|4	3	1|2|2|0|0|0|0|0|0|0|0|0	0.87	7.8
ARS2n	NELFE	10|11|9	3	10|5|5|0|0|0|0|0|0|2|4|3	0.87	4.48
ARS2n	FTO	2|2|5	3	2|0|2|0|0|0|0|0|0|0|0|0	0.87	9.75
ARS2n	MAGOH	5|5|6	3	3|2|4|0|0|0|1|0|0|0|3|1|1	0.86	4.62
ARS2n	DDX42	6|9|10	3	6|4|5|0|0|0|0|0|0|0|5|0|2	0.86	4.92
ARS2n	PRCC	6|6|11	3	3|3|8|0|0|0|0|0|0|0|3|2|0	0.86	5.25
ARS2n	MAGOHB	5|5|6	3	3|2|4|0|0|0|1|0|0|0|3|1|1	0.86	4.62
ARS2n	DNAJC8	9|7|13	3	6|4|5|0|0|0|0|0|0|0|4|4|4	0.85	4.65
ARS2n	ZRANB2	9|11|14	3	7|8|14|0|0|0|0|0|0|0|2|1|1	0.85	4.46
ARS2n	NACA	8|7|12	3	7|5|0|0|0|0|0|0|0|0|4|5|5	0.83	4.5
ARS2n	H2AX	3|3|3	3	1|0|0|0|0|0|0|0|1|0|1|3|1	0.83	5.57
ARS2n	TMPO	13|14|14	3	7|8|8|0|5|1|4|3|2|3|5|6|6	0.83	3.06
ARS2n	ZMAT2	5|4|4	3	4|0|3|0|0|0|0|0|0|0|2|1|1	0.83	5.12
ARS2n	SMARCE1	8|5|5	3	4|4|4|0|0|0|0|0|0|0|2|1|2	0.82	4.59
ARS2n	PDIA6	7|6|7	3	4|6|6|0|0|0|0|0|0|0|1|2|2	0.81	4.13
ARS2n	RBM22	10|9|14	3	7|4|9|0|0|0|0|0|0|0|5|5|5	0.81	4.09
ARS2n	DNAJC17	3|4|3	3	3|1|2|0|0|0|0|0|0|0|1|1|1	0.8	4.81
ARS2n	CASC3	2|4|5	3	1|2|3|0|0|0|0|0|0|0|1|1|1	0.79	5.3
ARS2n	TRIP6	5|3|4	3	4|5|2|0|0|0|0|0|0|0|0|0	0.79	4.73
ARS2n	TXNDC5	10|7|13	3	10|10|10|0|0|0|0|0|0|0|0|1|0	0.79	4.19
ARS2n	CSNK1A1	5|5|6	3	4|2|6|0|0|0|0|0|0|0|3|2|0	0.78	4.08
ARS2n	IK	16|15|17	3	13|9|13|0|0|0|0|0|0|0|8|7|7	0.78	3.65
ARS2n	H2AC8	3|3|3	3	0|0|0|3|0|0|0|0|0|0|1|3|1	0.77	4.88
ARS2n	WDR18	3|3|3	3	5|3|0|0|0|0|0|1|0|0|0|0|0	0.77	4.33
ARS2n	RNF113A	9|9|9	3	6|8|6|0|0|0|0|0|0|0|5|4|4	0.76	3.55
ARS2n	CD2BP2	10|10|12	3	8|5|10|0|0|0|0|0|0|0|7|4|5	0.74	3.56
ARS2n	PPP1CC	15|12|17	3	15|9|3|0|0|1|1|0|1|0|8|7|7	0.74	3.67
ARS2n	ARFGAP1	6|5|5	3	5|4|5|0|0|0|0|0|0|0|3|1|1	0.73	3.65
ARS2n	LUC7L	5|7|5	3	4|2|4|0|0|0|0|3|0|1|3|2|2	0.73	3.51
ARS2n	SAP30BP	6|7|6	3	4|4|7|0|0|0|0|0|0|0|4|2|2	0.73	3.58
ARS2n	ANP32A	2|4|2	3	3|1|2|0|0|0|0|0|0|0|1|0|0	0.72	4.95
ARS2n	ZNF207	2|2|3	3	2|2|3|0|0|0|0|0|0|0|0|0|0	0.7	4.33
ARS2n	BTF3	5|7|9	3	10|5|5|0|0|0|0|0|0|0|4|0|1	0.7	3.64
ARS2n	CSRP2	6|3|2	3	2|1|0|0|0|3|2|0|0|0|1|0|1	0.7	4.77
ARS2n	C1orf52	4|5|3	3	4|5|4|0|0|0|0|0|0|0|0|0|1	0.7	3.71
ARS2n	H2AC1	3|3|3	3	1|0|0|0|0|2|2|0|0|0|1|3|1	0.7	3.9

Table 3: SAINT analysis of LC-MS/MS identified proteins in BioID2-ARS2c2 precipitates. Proteins identified by LC-MS/MS from streptavidin affinity purifications with BioID2-ARS2c2 constructs in HEK 293 cells. Three independent experiments with 5 biological replicates each were analyzed by LC-MS/MS. Table contains the results for protein enrichment in BioID2-ARS2c2 vs BioID2 control. Additional negative controls were included: HEK293 cell background control and BirA-GFP control. SAINT software was used on the analysis. SP score > 0.7 was defined as cut-off for significantly enriched proteins.

Bait	Prey	Spec	Num Replic	Ctrl Counts	Saint Score	Fold Change
ARS2c2	AIP	4|4|5	3	0|0|1|0|0|0|0|0|0|0|0|0|0	1	43.33
ARS2c2	EIF3F	8|9|12	3	0|0|0|0|2|0|0|0|1|1|0|0|0	1	31.42
ARS2c2	PSMD14	4|3|5	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	40
ARS2c2	KPNA6	6|5|5	3	1|0|0|0|0|0|0|0|0|0|0|1|0	1	34.67
ARS2c2	PRKRA	5|5|5	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	50
ARS2c2	TIPRL	4|4|5	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	43.33
ARS2c2	ZWINT	3|4|4	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	36.67
ARS2c2	ACSL3	17|17|13	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	156.67
ARS2c2	HPRT1	4|4|6	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	46.67
ARS2c2	ANXA6	6|4|7	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	56.67
ARS2c2	RPSA	7|6|6	3	0|0|0|0|1|0|2|1|0|2|0|0|0	1	13.72
ARS2c2	UMPS	6|6|5	3	0|0|1|0|0|0|0|0|0|0|0|0|0	1	56.67
ARS2c2	HSPA6	5|6|4	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	50
ARS2c2	PPIB	4|4|5	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	43.33
ARS2c2	PSMA4	7|6|4	3	0|0|1|0|0|0|0|0|0|0|0|1|0	1	36.83
ARS2c2	TKT	7|6|14	3	0|0|1|0|0|0|0|0|0|0|0|0|0	1	90
ARS2c2	DNAJA1	12|13|16	3	1|0|2|0|1|2|4|2|0|0|1|1|0	1	12.69
ARS2c2	ATIC	9|8|14	3	0|0|2|0|0|0|0|0|0|0|0|0|0	1	67.17
ARS2c2	RFC2	3|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	30
ARS2c2	PSMC4	5|5|5	3	0|0|2|0|0|0|0|0|0|0|0|0|0	1	32.5
ARS2c2	MAP2K4	3|4|4	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	36.67
ARS2c2	PSMB2	5|4|4	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	43.33
ARS2c2	PEX5	3|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	30
ARS2c2	KPNA2	26|25|30	3	8|4|9|0|0|0|2|0|0|0|6|3|2	1	10.32
ARS2c2	AK2	3|4|3	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	33.33
ARS2c2	EIF3E	12|9|12	3	0|0|0|0|0|0|0|0|1|1|0|0|0	1	71.5
ARS2c2	NUTF2	3|3|5	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	36.67
ARS2c2	GNAS	4|3|4	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	36.67
ARS2c2	CLTC	5|6|8	3	0|0|0|0|0|0|1|0|2|1|0|0|0	1	20.58
ARS2c2	PPP2R5C	3|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	30

ARS2c2	SF3B3	15|20|21	3	0|0|4|0|0|0|0|0|0|1|0|0|0	1	48.53
ARS2c2	TBCE	10|10|11	3	0|0|1|0|0|0|0|0|0|0|0|0|0	1	103.33
ARS2c2	DDB1	8|6|13	3	0|0|2|0|0|0|0|0|0|0|0|0|0	1	58.5
ARS2c2	ARMC6	9|8|9	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	86.67
ARS2c2	SND1	16|14|11	3	0|0|0|0|0|0|0|0|0|0|3|3|1	1	25.38
ARS2c2	PJA1	4|4|4	3	1|0|0|0|0|0|0|0|0|0|0|0|0	1	40
ARS2c2	DCUN1D4	3|4|5	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	40
ARS2c2	OTUB1	3|4|5	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	40
ARS2c2	CFAP36	3|3|4	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	33.33
ARS2c2	PUS7	11|9|12	3	2|1|1|0|0|0|0|0|0|1|2|1	1	17.33
ARS2c2	DNAJC7	13|15|16	3	4|1|5|0|0|0|0|0|0|1|1|0	1	15.89
ARS2c2	PDCD2L	4|6|5	3	1|0|0|0|0|0|0|0|0|0|0|0|0	1	50
ARS2c2	EGLN1	9|5|4	3	0|0|1|0|0|0|0|0|0|0|0|0|0	1	60
ARS2c2	NOL11	7|8|8	3	0|0|2|0|0|0|0|0|0|1|0|0	1	33.22
ARS2c2	PDLIM7	3|3|3	3	0|0|1|0|0|0|0|0|0|0|0|0|0	1	30
ARS2c2	STK26	9|6|9	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	80
ARS2c2	TRMT112	4|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0|0	1	33.33
ARS2c2	STUB1	8|11|10	3	0|2|3|0|0|0|0|0|0|0|0|2	1	17.95
ARS2c2	DRG1	9|8|4	3	0|0|2|0|0|0|0|0|0|0|0|0	1	45.5
ARS2c2	SH3GLB1	7|7|9	3	0|0|1|0|0|0|0|0|0|1|0|0	1	49.83
ARS2c2	MAGED1	3|4|4	3	0|0|0|0|0|0|0|0|0|0|0|0	1	36.67
ARS2c2	NFKBIE	2|3|3	3	1|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c2	UBE2C	2|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c2	GIPC1	3|2|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c2	CTNND1	6|5|4	3	0|0|2|0|0|0|0|0|0|0|0|0	0.99	32.5
ARS2c2	TTC4	22|20|25	3	8|8|6|0|0|0|0|0|0|4|4|4	0.99	8.54
ARS2c2	BAG3	21|21|23	3	0|0|11|0|0|0|0|0|0|0|0|2	0.99	21.67
ARS2c2	UCHL3	3|2|4	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	30
ARS2c2	GSTM3	6|2|5	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	43.33
ARS2c2	VARS1	14|12|18	3	6|2|6|0|0|0|0|0|0|0|0|0	0.99	13.62
ARS2c2	PSMA5	5|2|5	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	40
ARS2c2	COPB2	19|17|21	3	4|3|12|0|0|0|0|0|0|1|0|0|0	0.99	12.35
ARS2c2	TALDO1	9|9|16	3	3|3|4|0|0|0|0|0|0|1|0|1	0.99	12.28
ARS2c2	ATP6V1A	4|4|5	3	1|0|2|0|0|0|0|0|0|0|0|0	0.99	18.78
ARS2c2	PEX19	14|14|11	3	0|1|6|0|0|0|0|0|0|0|0|1	0.99	21.12

ARS2c2	EIF2D	2|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c2	GDI2	2|4|4	3	0|0|0|0|0|0|0|0|1|0|0|0	0.99	33.33
ARS2c2	CLNS1A	5|5|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	40
ARS2c2	DCAF7	3|2|4	3	0|0|0|0|1|0|0|0|0|0|0|0	0.99	30
ARS2c2	RBM6	2|5|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	33.33
ARS2c2	CCT2	22|22|29	3	8|3|12|0|2|2|3|0|1|2|4|2|0	0.99	8.11
ARS2c2	FKBP4	3|4|5	3	0|0|2|0|0|0|0|0|0|0|0|0	0.99	26
ARS2c2	KPNB1	14|14|15	3	0|0|3|0|4|3|2|2|5|6|0|0|0	0.99	7.45
ARS2c2	GIGYF2	3|3|2	3	0|1|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c2	THOC3	3|3|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c2	VPS25	8|7|8	3	0|0|4|0|0|0|0|0|0|0|0|0	0.99	24.92
ARS2c2	CDK1	13|14|13	3	0|0|1|0|4|0|0|3|3|7|0|0|0	0.98	9.63
ARS2c2	XRCC5	5|3|5	3	0|0|1|0|0|0|0|0|0|0|1|2|0	0.98	14.08
ARS2c2	TPT1	11|10|13	3	4|3|6|0|0|0|0|0|0|0|2|2|1	0.98	8.19
ARS2c2	TCP1	21|21|18	3	0|0|3|0|7|0|7|9|5|7|1|2|0	0.98	6.34
ARS2c2	ABCE1	2|2|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2c2	PPP2R5D	4|6|4	3	0|0|3|0|0|0|0|0|0|0|0|0	0.98	20.22
ARS2c2	PTGES3	2|2|4	3	0|0|0|0|0|1|0|0|0|0|0|0	0.98	26.67
ARS2c2	CHAC2	2|2|6	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	33.33
ARS2c2	UBXN4	2|2|3	3	0|0|1|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2c2	DCUN1D5	2|3|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2c2	KNSTRN	2|3|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2c2	TXNL1	13|13|17	3	7|6|6|0|0|0|0|0|0|3|4|0	0.97	7.17
ARS2c2	DHFR	2|2|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.97	20
ARS2c2	RPLP0	9|8|11	3	0|1|3|0|0|3|2|4|5|0|2|0|1	0.97	5.78
ARS2c2	JUN	2|2|2	3	0|1|0|0|0|0|0|0|0|0|0|0	0.97	20
ARS2c2	ADSS2	3|3|5	3	2|1|0|0|0|0|0|0|0|0|0|0	0.97	15.89
ARS2c2	DSTN	6|6|7	3	0|0|0|0|0|0|0|5|1|0|0|0	0.97	13.72
ARS2c2	PSMC1	3|3|5	3	1|0|2|0|0|0|0|0|0|0|0|0	0.97	15.89
ARS2c2	UBXN1	2|2|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.97	20
ARS2c2	UBR7	6|2|3	3	0|0|2|0|0|0|0|0|0|0|0|0	0.97	23.83
ARS2c2	TXNDC17	2|2|2	3	1|0|0|0|0|0|0|0|0|0|0|0	0.97	20
ARS2c2	CACYBP	12|12|13	3	6|4|4|0|0|0|0|0|0|5|5|2	0.97	6.17
ARS2c2	COMMD9	3|3|3	3	1|1|1|0|0|0|0|0|0|0|0|0	0.97	13
ARS2c2	PLAA	5|3|11	3	0|0|4|0|0|0|0|0|0|0|0|0	0.97	20.58
ARS2c2	SNX9	8|6|8	3	0|0|7|0|0|0|0|0|0|0|0|0	0.97	13.62

ARS2c2	HSPB1	7|5|3	3	0|0|0|0|2|2|1|0|0|0|0|0|0	0.96	13
ARS2c2	PDIA4	30|31|32	3	16|19|18|0|0|0|0|0|0|0|10|8|4	0.96	5.37
ARS2c2	PPP2R1A	3|3|6	3	0|0|0|0|0|0|0|0|0|3|0|0|0	0.96	17.33
ARS2c2	CSK	3|7|8	3	0|0|5|0|0|0|0|0|0|0|0|0|0	0.96	15.6
ARS2c2	ETF1	2|2|3	3	0|0|0|0|0|0|0|1|0|1|0|0|0	0.96	15.17
ARS2c2	PDIA6	12|12|10	3	4|6|6|0|0|0|0|0|0|0|1|2|2	0.96	7.02
ARS2c2	SNRNP27	6|6|3	3	1|1|0|2|0|0|0|0|0|0|1|0|1	0.96	10.83
ARS2c2	HSPH1	24|24|17	3	13|5|6|0|0|0|0|0|0|0|8|9|2	0.96	6.55
ARS2c2	HAUS8	3|3|5	3	2|0|1|0|0|0|0|0|0|0|0|0|1	0.96	11.92
ARS2c2	UNC45A	7|6|3	3	0|0|2|0|2|0|0|0|1|0|0|1|0	0.96	11.56
ARS2c2	P4HB	24|25|26	3	14|15|13|0|0|0|1|0|0|0|7|6|6	0.95	5.24
ARS2c2	USP5	6|4|5	3	2|0|4|0|0|0|0|0|0|0|0|0|0	0.95	10.83
ARS2c2	VCP	13|12|10	3	5|0|8|0|0|0|0|0|0|0|5|3|0	0.95	7.22
ARS2c2	FMR1	22|20|25	3	11|8|13|0|0|0|0|0|0|0|10|7|6	0.95	5.28
ARS2c2	BZW1	14|17|20	3	7|5|4|0|0|0|0|0|0|0|8|10|3	0.95	5.97
ARS2c2	AARSD1	4|6|6	3	0|0|6|0|0|0|0|0|0|0|0|0|0	0.95	11.56
ARS2c2	VTA1	3|3|6	3	0|0|4|0|0|0|0|0|0|0|0|0|0	0.95	13
ARS2c2	PRMT7	5|6|4	3	3|2|2|0|0|0|0|0|0|0|0|0|0	0.95	9.29
ARS2c2	RTCB	4|5|7	3	0|0|0|0|0|3|0|2|0|2|0|0|0	0.95	9.9
ARS2c2	PDIA3	20|21|21	3	11|12|12|0|0|0|0|0|0|0|5|8|4	0.94	5.17
ARS2c2	NMT1	12|13|16	3	9|8|1|0|0|0|0|0|0|0|3|3|5	0.94	6.13
ARS2c2	MPP1	8|9|10	3	6|5|2|0|0|0|0|0|0|0|2|1|2	0.94	6.5
ARS2c2	LARP4	2|2|2	3	0|0|2|0|0|0|0|0|0|0|0|0|0	0.94	13
ARS2c2	TXNDC5	16|15|15	3	10|10|10|0|0|0|0|0|0|0|0|1|0	0.94	6.43
ARS2c2	FTO	3|3|4	3	2|0|2|0|0|0|0|0|0|0|0|0|0	0.94	10.83
ARS2c2	TOMM22	2|2|2	3	0|0|0|0|0|0|0|0|0|2|0|0|0	0.94	13
ARS2c2	PIH1D1	10|9|10	3	5|6|0|0|0|0|0|0|0|0|4|2|3	0.94	6.28
ARS2c2	EZR	13|13|24	3	8|6|9|0|0|0|0|0|2|0|5|5|5	0.93	5.42
ARS2c2	PSMC3	6|5|5	3	3|3|2|0|0|0|0|0|0|0|2|0|0	0.93	6.93
ARS2c2	FXR2	12|11|11	3	8|4|2|0|0|0|0|0|0|0|6|4|2	0.93	5.67
ARS2c2	TXN2	4|4|2	3	1|1|2|0|0|0|0|0|0|0|0|0|0	0.93	10.83
ARS2c2	GPN1	4|4|5	3	2|2|2|0|0|0|0|0|0|0|2|1|1	0.93	5.63
ARS2c2	HSPA4	11|10|9	3	5|0|8|0|0|0|0|0|0|2|4|3|1	0.92	5.65

ARS2c2	PRPF38A	2|2|6	3	0|0|0|3|0|0|0|0|0|0|0|0|0	0.92	14.44
ARS2c2	BTF3L4	3|3|6	3	1|0|4|0|0|0|0|0|0|0|0|0	0.92	10.4
ARS2c2	TJP2	3|2|3	3	0|0|2|0|0|0|0|0|0|1|0|0	0.92	11.56
ARS2c2	GARS1	5|4|7	3	3|0|6|0|0|0|0|0|0|0|0|0	0.91	7.7
ARS2c2	CCT3	28|27|20	3	17|13|23|0|5|0|1|0|0|2|4|2|0	0.91	4.85
ARS2c2	PPID	7|7|10	3	6|7|3|0|0|0|0|0|0|0|0|0	0.91	6.5
ARS2c2	TRIP6	7|6|5	3	4|5|2|0|0|0|0|0|0|0|0|0	0.91	7.09
ARS2c2	GOLPH3L	5|2|9	3	0|1|3|0|0|0|0|0|0|1|0|0	0.91	13.87
ARS2c2	NOP58	2|2|4	3	0|0|0|0|0|0|0|3|0|0|0|0	0.91	11.56
ARS2c2	U2AF1L5	10|9|8	3	4|5|5|0|3|3|1|1|0|2|4|4|2	0.9	3.44
ARS2c2	HNRNPH3	4|3|5	3	2|1|2|0|0|0|1|1|0|1|0|0|0	0.9	6.5
ARS2c2	EIF2B2	5|6|5	3	2|2|4|0|0|0|0|0|0|3|0|0	0.9	6.3
ARS2c2	AFDN	3|2|3	3	0|0|4|0|0|0|0|0|0|0|0|0	0.9	8.67
ARS2c2	U2AF1	10|9|8	3	4|5|5|0|3|3|1|1|0|2|4|4|2	0.9	3.44
ARS2c2	EDC3	14|13|15	3	10|7|8|0|0|0|0|0|0|8|5|3	0.9	4.44
ARS2c2	ZPR1	9|8|8	3	4|4|6|0|0|0|0|0|0|4|3|2	0.89	4.71
ARS2c2	RPS12	6|7|6	3	0|0|0|0|4|3|4|0|1|3|0|0|0	0.89	5.49
ARS2c2	EEF1G	16|17|21	3	7|6|19|0|2|0|0|2|0|0|6|8|3	0.89	4.42
ARS2c2	TXNRD1	6|7|11	3	4|3|7|0|0|0|0|0|0|0|2|1|1	0.89	5.78
ARS2c2	PSMD1	12|10|13	3	9|6|15|0|0|0|0|0|0|0|1|0|1	0.89	4.74
ARS2c2	DPH5	3|3|2	3	2|0|2|0|0|0|0|0|0|0|0|0	0.89	8.67
ARS2c2	ARPC3	5|5|5	3	0|1|1|0|0|0|0|0|7|2|0|0|0	0.88	5.91
ARS2c2	GLRX3	6|5|8	3	3|3|4|0|0|0|0|0|0|0|2|3|1	0.88	5.15
ARS2c2	ST13	8|7|10	3	4|2|5|0|0|0|0|0|0|0|5|4|3	0.88	4.71
ARS2c2	RPS5	3|3|3	3	1|0|1|0|2|2|0|0|0|0|1|0|0	0.87	5.57
ARS2c2	RPS27A	7|7|6	3	6|0|0|4|0|0|2|0|0|0|0|5|0	0.87	5.1
ARS2c2	EIF3H	6|6|7	3	4|4|9|0|0|0|0|0|0|0|0|0	0.86	4.84
ARS2c2	YWHAZ	5|5|7	3	1|0|3|0|2|0|0|0|0|7|0|0|0	0.86	5.67
ARS2c2	ILF2	6|6|5	3	0|0|0|0|0|5|5|2|2|0|0|0|0	0.86	5.26
ARS2c2	RP9	9|8|10	3	3|4|2|12|0|0|0|0|0|0|0|2|3	0.86	4.5
ARS2c2	LIMD1	12|10|11	3	7|5|7|0|0|0|0|0|0|0|5|6|5	0.86	4.09
ARS2c2	BCAS2	4|6|6	3	3|2|4|0|0|0|0|0|0|0|2|2|2	0.84	4.62
ARS2c2	PXN	6|6|6	3	5|4|4|0|0|0|0|0|0|0|2|2|0	0.84	4.59
ARS2c2	PICALM	5|3|5	3	3|2|4|0|0|0|0|0|0|0|1|0|0	0.84	5.63
ARS2c2	NKAP	6|5|12	3	1|2|2|11|0|0|0|0|0|0|0|2|1	0.84	5.25
ARS2c2	SNX3	6|5|6	3	4|4|3|0|0|0|0|0|0|0|2|1|3	0.83	4.33
ARS2c2	STX5	2|3|6	3	0|4|3|0|0|0|0|0|0|0|0|0|0	0.83	6.81

ARS2c2	RANGAP1	26|26|25	3	18|14|14|0|0|0|0|0|0|0|17|13|12	0.82	3.79
ARS2c2	HAT1	10|10|8	3	4|4|9|0|0|0|0|0|0|0|6|5|3	0.81	3.91
ARS2c2	DNAJA2	3|5|2	3	0|0|0|0|0|3|3|0|0|0|0|0|1	0.81	6.19
ARS2c2	TXN	9|10|12	3	5|5|8|0|0|3|3|1|0|0|4|5|3	0.81	3.63
ARS2c2	XRCC6	12|12|15	3	6|0|11|0|2|0|0|0|5|4|8|5|4	0.81	3.76
ARS2c2	BTF3	7|7|10	3	10|5|5|0|0|0|0|0|0|0|4|0|1	0.81	4.16
ARS2c2	COPS5	4|4|5	3	3|2|4|0|0|0|0|0|0|0|1|1|1	0.81	4.69
ARS2c2	TOP1	3|6|9	3	0|0|3|14|0|0|0|0|0|0|0|0	0.79	4.59
ARS2c2	SCRN1	3|4|2	3	2|3|2|0|0|0|0|0|0|0|0|0	0.79	5.57
ARS2c2	UBAP2	8|9|10	3	6|6|6|0|0|0|0|0|0|6|5|2	0.79	3.77
ARS2c2	SKA3	4|3|3	3	2|2|3|0|0|0|0|0|0|2|0|0	0.78	4.81
ARS2c2	SREK1IP1	2|2|2	3	0|0|0|7|0|0|0|0|0|0|0|0	0.78	3.71
ARS2c2	H2AC8	3|3|3	3	0|0|0|3|0|0|0|0|0|0|1|3|1	0.77	4.88
ARS2c2	C12orf45	2|2|2	3	0|2|3|0|0|0|0|0|0|0|0|0	0.77	5.2
ARS2c2	DDX18	5|3|2	3	0|0|0|0|0|0|2|0|3|3|0|0|0	0.77	5.42
ARS2c2	MAGED2	5|8|3	3	4|1|4|0|0|0|0|0|0|0|2|3|0	0.77	4.95
ARS2c2	VBP1	9|8|8	3	6|7|5|0|0|0|0|0|0|0|4|5|3	0.76	3.61
ARS2c2	SH3GL1	10|11|9	3	10|10|9|0|0|0|0|0|0|0|2|1|4	0.76	3.61
ARS2c2	ACTR2	2|4|6	3	0|0|0|0|0|0|0|0|8|4|0|0|0	0.75	4.33
ARS2c2	TFCP2	4|3|2	3	2|4|1|0|0|0|0|0|0|0|1|0|0	0.75	4.88
ARS2c2	FGFR1OP	5|3|7	3	4|3|5|0|0|0|0|0|0|0|3|0|0	0.74	4.33
ARS2c2	CSTF2	6|7|8	3	3|3|4|0|0|0|0|0|0|0|7|5|3	0.74	3.64
ARS2c2	NHLRC2	4|3|5	3	4|2|6|0|0|0|0|0|0|0|0|1	0.74	4
ARS2c2	GOLPH3	3|2|3	3	3|0|4|0|0|0|0|0|0|0|1|0|0	0.73	4.33
ARS2c2	HSPD1	20|19|22	3	9|4|16|0|8|8|8|9|3|10|1|1|1	0.72	3.39
ARS2c2	SMN1	2|4|2	3	1|0|3|0|1|0|1|0|0|0|1|0|0	0.72	4.95
ARS2c2	UPF1	5|4|3	3	3|0|1|0|0|0|0|0|0|0|4|3|2	0.72	4
ARS2c2	PTPN1	2|5|3	3	1|1|5|0|0|0|0|0|0|0|1|2|0	0.71	4.33
ARS2c2	SRSF4	2|3|3	3	3|0|1|1|0|0|0|0|0|0|1|1|1	0.71	4.33
ARS2c2	CEP170	5|4|5	3	3|1|7|0|0|0|0|0|0|0|2|3|1	0.71	3.57
ARS2c2	OBI1	3|2|2	3	3|0|3|0|0|0|0|0|0|0|1|0|0	0.71	4.33
ARS2c2	NACA	6|6|11	3	7|5|0|0|0|0|0|0|0|0|4|5|5	0.7	3.83
ARS2c2	CD2AP	13|15|12	3	12|8|11|0|0|0|0|0|0|0|8|7|4	0.7	3.47

Table 4: SAINT analysis of LC-MS/MS identified proteins in BioID2-ARS2c1 precipitates. Proteins identified by LC-MS/MS from streptavidin affinity purifications with BioID2-ARS2c1 constructs in HEK 293 cells. Three independent experiments with 5 biological replicates each were analyzed by LC-MS/MS. Table contains the results for protein enrichment in BioID2-ARS2c1 vs BioID2 control. Additional negative controls were included: HEK293 cell background control and BirA-GFP control. SAINT software was used on the analysis. SP score > 0.7 was defined as cut-off for significantly enriched proteins.

Bait	Prey	Spec	Num Replic	Ctrl Counts	Saint Score	Fold Change
ARS2c1	PPIB	5|6|7	3	0|0|0|0|0|0|0|0|0|0|0|0	1	60
ARS2c1	TBCE	6|8|8	3	0|0|1|0|0|0|0|0|0|0|0|0	1	73.33
ARS2c1	ARMC6	5|4|3	3	0|0|0|0|0|0|0|0|0|0|0|0	1	40
ARS2c1	GIGYF2	4|3|4	3	0|1|0|0|0|0|0|0|0|0|0|0	1	36.67
ARS2c1	PALLD	7|5|7	3	0|2|1|0|0|0|0|0|0|0|0|0	1	27.44
ARS2c1	DCUN1D4	4|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0	1	33.33
ARS2c1	IFT74	4|7|5	3	0|0|0|0|0|0|0|0|0|1|0|0	1	53.33
ARS2c1	EGLN1	9|4|6	3	0|0|1|0|0|0|0|0|0|0|0|0	1	63.33
ARS2c1	VPS33B	9|6|5	3	0|0|0|0|0|0|0|0|0|0|0|0	1	66.67
ARS2c1	UNC45A	11|13|6	3	0|0|2|0|2|0|0|0|1|0|0|1|0	1	21.67
ARS2c1	NOL11	12|7|8	3	0|0|2|0|0|0|0|0|0|1|0|0	1	39
ARS2c1	STK26	5|6|6	3	0|0|0|0|0|0|0|0|0|0|0|0	1	56.67
ARS2c1	TJP2	9|5|6	3	0|0|2|0|0|0|0|0|0|1|0|0	1	28.89
ARS2c1	SH3GLB1	6|9|8	3	0|0|1|0|0|0|0|0|0|1|0|0	1	49.83
ARS2c1	SBDS	5|3|4	3	0|0|0|0|0|0|0|0|0|0|0|0	1	40
ARS2c1	EIF3F	4|5|6	3	0|0|0|0|2|0|0|0|1|1|0|0|0	0.99	16.25
ARS2c1	DYNC1LI2	2|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c1	BAG3	15|15|17	3	0|0|11|0|0|0|0|0|0|0|0|2	0.99	15.67
ARS2c1	APRT	2|3|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c1	PSMA4	4|3|3	3	0|0|1|0|0|0|0|0|0|0|1|0	0.99	21.67
ARS2c1	DNAJA1	9|6|9	3	1|0|2|0|1|2|4|2|0|0|1|1|0	0.99	7.43
ARS2c1	RPS5	6|5|4	3	1|0|1|0|2|2|0|0|0|1|0|0	0.99	9.29
ARS2c1	SELENOH	3|3|5	3	1|0|1|0|0|0|0|0|0|0|0|0	0.99	23.83
ARS2c1	SNRNP27	6|6|4	3	1|1|0|2|0|0|0|0|0|1|0|1	0.99	11.56
ARS2c1	SELENOM	2|3|5	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	33.33
ARS2c1	VCPIP1	3|2|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	26.67
ARS2c1	RBSN	4|4|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.99	33.33
ARS2c1	DRG1	4|3|5	3	0|0|2|0|0|0|0|0|0|0|0|0	0.99	26
ARS2c1	KPNA2	20|16|22	3	8|4|9|0|0|0|2|0|0|0|6|3|2	0.98	7.39
ARS2c1	EIF3E	3|2|4	3	0|0|0|0|0|0|0|0|1|1|0|0|0	0.98	19.5
ARS2c1	DHX29	4|3|2	3	0|0|1|0|0|0|0|0|0|0|1|0|0	0.98	19.5

ARS2c1	NUDCD3	2|2|3	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2c1	UBXN4	2|2|3	3	0|0|1|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2c1	DNAJC7	10|9|10	3	4|1|5|0|0|0|0|0|0|1|1|0	0.98	10.47
ARS2c1	HAUS8	5|3|4	3	2|0|1|0|0|0|0|0|0|0|0|1	0.98	13
ARS2c1	VPS51	3|2|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.98	23.33
ARS2c1	STUB1	6|6|9	3	0|2|3|0|0|0|0|0|0|0|0|2	0.98	13
ARS2c1	TXNL1	14|15|14	3	7|6|6|0|0|0|0|0|0|3|4|0	0.97	7.17
ARS2c1	CTNND1	3|4|2	3	0|0|2|0|0|0|0|0|0|0|0|0	0.97	19.5
ARS2c1	P4HB	26|26|31	3	14|15|13|0|0|0|1|0|0|0|7|6|6	0.97	5.8
ARS2c1	PDIA4	33|30|38	3	16|19|18|0|0|0|0|0|0|0|10|8|4	0.97	5.84
ARS2c1	GDI2	2|2|2	3	0|0|0|0|0|0|0|0|1|0|0|0	0.97	20
ARS2c1	OBI1	5|7|6	3	3|0|3|0|0|0|0|0|0|1|0|0	0.97	11.14
ARS2c1	GPRASP2	2|2|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.97	20
ARS2c1	TXN2	4|4|3	3	1|1|2|0|0|0|0|0|0|0|0|0	0.97	11.92
ARS2c1	TMA7	2|2|2	3	0|0|0|0|0|0|0|0|0|0|0|0	0.97	20
ARS2c1	CDK1	11|11|9	3	0|0|1|0|4|0|0|3|3|7|0|0|0	0.96	7.46
ARS2c1	TPT1	8|10|12	3	4|3|6|0|0|0|0|0|0|0|2|2|1	0.96	7.22
ARS2c1	PDIA3	23|21|26	3	11|12|12|0|0|0|0|0|0|0|5|8|4	0.96	5.83
ARS2c1	PPP2R5D	3|3|9	3	0|0|3|0|0|0|0|0|0|0|0|0	0.96	21.67
ARS2c1	SND1	5|5|12	3	0|0|0|0|0|0|0|0|0|3|3|1	0.96	13.62
ARS2c1	TXNDC5	15|16|24	3	10|10|10|0|0|0|0|0|0|0|1|0	0.96	7.69
ARS2c1	HSPH1	20|19|20	3	13|5|6|0|0|0|0|0|0|8|9|2	0.96	5.95
ARS2c1	TTC4	15|17|12	3	8|8|6|0|0|0|0|0|0|4|4|4	0.95	5.61
ARS2c1	CCT2	14|15|18	3	8|3|12|0|2|2|3|0|1|2|4|2|0	0.95	5.22
ARS2c1	PDIA6	11|13|9	3	4|6|6|0|0|0|0|0|0|1|2|2	0.95	6.81
ARS2c1	TOP1	8|8|14	3	0|0|3|14|0|0|0|0|0|0|0|0	0.94	7.65
ARS2c1	COPB2	10|9|14	3	4|3|12|0|0|0|0|0|0|1|0|0|0	0.94	7.15
ARS2c1	ATP6V1A	2|3|5	3	1|0|2|0|0|0|0|0|0|0|0|0	0.94	14.44
ARS2c1	PRPF38A	2|3|5	3	0|0|0|3|0|0|0|0|0|0|0|0	0.94	14.44
ARS2c1	TXLNG	7|4|4	3	0|0|1|0|0|0|0|0|0|3|2|1	0.94	9.29
ARS2c1	COMMD9	3|2|3	3	1|1|1|0|0|0|0|0|0|0|0|0	0.94	11.56
ARS2c1	SNX9	8|3|6	3	0|0|7|0|0|0|0|0|0|0|0|0	0.94	10.52
ARS2c1	U2AF1L5	8|11|11	3	4|5|5|0|3|3|1|1|0|2|4|4|2	0.93	3.82
ARS2c1	U2AF1	8|11|11	3	4|5|5|0|3|3|1|1|0|2|4|4|2	0.93	3.82
ARS2c1	UFD1	3|4|3	3	3|0|0|0|0|0|0|0|0|1|0|0	0.93	10.83
ARS2c1	VTA1	4|2|4	3	0|0|4|0|0|0|0|0|0|0|0|0	0.93	10.83

ARS2c1	VARS1	9|5|13	3	6|2|6|0|0|0|0|0|0|0|0|0	0.92	8.36
ARS2c1	FXR2	9|11|13	3	8|4|2|0|0|0|0|0|0|6|4|2	0.92	5.5
ARS2c1	DSTN	4|3|6	3	0|0|0|0|0|0|0|5|1|0|0|0	0.92	9.39
ARS2c1	CLTC	2|3|5	3	0|0|0|0|0|1|0|2|1|0|0|0	0.92	10.83
ARS2c1	TAF9	3|3|2	3	2|0|1|0|0|0|0|0|0|0|0|0	0.92	11.56
ARS2c1	FMR1	18|18|21	3	11|8|13|0|0|0|0|0|0|0|10|7|6	0.91	4.49
ARS2c1	PRPF38B	7|6|9	3	0|0|0|17|0|0|0|0|0|0|1|0|0	0.91	5.3
ARS2c1	HSPB1	3|3|3	3	0|0|0|0|2|2|1|0|0|0|0|0	0.9	7.8
ARS2c1	CCT3	26|20|26	3	17|13|23|0|5|0|1|0|0|2|4|2|0	0.9	4.66
ARS2c1	MPP1	9|6|9	3	6|5|2|0|0|0|0|0|0|0|2|1|2	0.9	5.78
ARS2c1	BZW1	12|12|17	3	7|5|4|0|0|0|0|0|0|0|8|10|3	0.9	4.8
ARS2c1	CCDC85C	4|4|2	3	2|0|2|0|0|0|0|0|0|0|1|0	0.89	8.67
ARS2c1	NACA	9|8|15	3	7|5|0|0|0|0|0|0|0|4|5|5	0.89	5.33
ARS2c1	NMT1	10|9|16	3	9|8|1|0|0|0|0|0|0|3|3|5	0.89	5.23
ARS2c1	PICALM	4|5|6	3	3|2|4|0|0|0|0|0|0|1|0|0	0.89	6.5
ARS2c1	RANGAP1	26|29|28	3	18|14|14|0|0|0|0|0|0|17|13|12	0.88	4.09
ARS2c1	RP9	9|10|10	3	3|4|2|12|0|0|0|0|0|0|2|3	0.88	4.83
ARS2c1	BTF3L4	5|2|3	3	1|0|4|0|0|0|0|0|0|0|0|0	0.88	8.67
ARS2c1	PRIM2	3|4|2	3	2|0|2|0|0|0|0|0|0|1|0|0	0.87	7.8
ARS2c1	TXNRD1	7|5|13	3	4|3|7|0|0|0|0|0|0|2|1|1	0.86	6.02
ARS2c1	NKAP	7|6|8	3	1|2|2|11|0|0|0|0|0|0|2|1	0.86	4.79
ARS2c1	RSRC1	4|4|6	3	0|0|0|14|0|0|0|0|0|0|0|0	0.86	4.33
ARS2c1	PPID	5|6|7	3	6|7|3|0|0|0|0|0|0|0|0|0	0.85	4.88
ARS2c1	PRMT7	5|2|5	3	3|2|2|0|0|0|0|0|0|0|0|0	0.85	7.43
ARS2c1	BTF3	7|8|11	3	10|5|5|0|0|0|0|0|0|4|0|1	0.84	4.51
ARS2c1	FAM114A2	7|9|8	3	5|3|8|0|0|0|0|0|0|5|3|0	0.84	4.33
ARS2c1	BCAS2	5|5|5	3	3|2|4|0|0|0|0|0|0|2|2|2	0.83	4.33
ARS2c1	AARSD1	2|2|3	3	0|0|6|0|0|0|0|0|0|0|0|0	0.83	5.06
ARS2c1	LIMD1	13|11|9	3	7|5|7|0|0|0|0|0|0|5|6|5	0.83	4.09
ARS2c1	TXN	9|10|15	3	5|5|8|0|0|3|3|1|0|0|4|5|3	0.82	3.98
ARS2c1	SMAP1	3|4|4	3	1|3|1|0|0|0|0|0|0|0|3|1	0.82	5.3
ARS2c1	RPS27A	6|5|6	3	6|0|0|4|0|0|2|0|0|0|0|5|0	0.81	4.33
ARS2c1	GPN1	3|3|5	3	2|2|2|0|0|0|0|0|0|0|2|1|1	0.81	4.77
ARS2c1	H2AC8	3|3|4	3	0|0|0|3|0|0|0|0|0|0|1|3|1	0.8	5.42
ARS2c1	CPSF6	3|3|2	3	0|0|0|0|0|0|1|0|0|0|2|2|1	0.8	5.78
ARS2c1	NECAP2	2|2|2	3	1|0|2|0|0|0|0|0|0|0|1|0|0	0.8	6.5
ARS2c1	TRIP6	5|4|3	3	4|5|2|0|0|0|0|0|0|0|0|0|0	0.79	4.73
ARS2c1	SPART	7|6|3	3	3|2|4|0|0|0|0|0|0|0|2|2|1	0.79	4.95

ARS2c1	HDLBP	19|18|11	3	10|10|8|0|0|1|0|0|0|0|6|9|5	0.78	4.24
ARS2c1	AMOT	12|11|3	3	9|5|4|0|0|0|0|0|0|0|0|0|0	0.78	6.26
ARS2c1	NKAPD1	2|2|4	3	0|0|3|4|0|0|0|0|0|0|0|0|0	0.77	4.95
ARS2c1	CCT7	10|9|13	3	10|3|7|0|0|0|0|0|1|2|5|5|4	0.77	3.75
ARS2c1	ANXA5	4|5|6	3	1|4|4|0|0|0|0|0|0|4|0|3	0.76	4.06
ARS2c1	ARL2	4|4|3	3	1|0|4|0|2|0|0|0|0|2|2|0	0.76	4.33
ARS2c1	SH3GL1	11|10|9	3	10|10|9|0|0|0|0|0|0|0|2|1|4	0.76	3.61
ARS2c1	GOLPH3	4|3|2	3	3|0|4|0|0|0|0|0|0|1|0|0	0.76	4.88
ARS2c1	CACYBP	7|6|13	3	6|4|4|0|0|0|0|0|0|5|5|2	0.76	4.33
ARS2c1	SKA3	7|3|2	3	2|2|3|0|0|0|0|0|0|2|0|0	0.75	5.78
ARS2c1	UBAP2	10|9|7	3	6|6|6|0|0|0|0|0|0|6|5|2	0.74	3.63
ARS2c1	RPSA	3|2|2	3	0|0|0|0|1|0|2|1|0|2|0|0|0	0.73	5.06
ARS2c1	EEF1G	13|14|17	3	7|6|19|0|2|0|0|2|0|0|6|8|3	0.73	3.6
ARS2c1	RPL36	4|4|3	3	3|1|2|0|1|0|0|3|2|0|0|0|0	0.73	3.97
ARS2c1	EIF3H	8|5|3	3	4|4|9|0|0|0|0|0|0|0|0|0|0	0.72	4.08
ARS2c1	PSMD12	4|4|6	3	4|0|11|0|0|0|0|0|0|0|2|0|1	0.71	3.37
ARS2c1	CSTF2	7|11|5	3	3|3|4|0|0|0|0|0|0|0|7|5|3	0.71	3.99
ARS2c1	TOMM34	3|2|4	3	2|2|3|0|0|0|0|0|0|0|0|0|2	0.71	4.33
ARS2c1	PSMD1	7|8|10	3	9|6|15|0|0|0|0|0|0|0|1|0|1	0.71	3.39
ARS2c1	MAGED2	6|4|3	3	4|1|4|0|0|0|0|0|0|0|2|3|0	0.71	4.02

Table 5: GO analysis (biological process) of LC-MS/MS identified proteins in BioID2-ARS2n and BioID2-ARSc1/2 precipitates. P. adjusted value <0.05 was used as cut off. Biological processes represented in Figure 30 are highlighted in yellow.

Cluster	Description	GeneRatio	pvalue	p.adjust	qvalue
ARS2n	RNA splicing	31/79	2.72E-29	1.20E-26	1.02E-26
ARS2n	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	29/79	3.74E-29	1.20E-26	1.02E-26
ARS2n	mRNA splicing, via spliceosome	29/79	3.74E-29	1.20E-26	1.02E-26
ARS2n	RNA splicing, via transesterification reactions	29/79	4.71E-29	1.20E-26	1.02E-26
ARS2n	mRNA 3'-end processing	8/79	8.28E-09	1.68E-06	1.44E-06
ARS2n	RNA 3'-end processing	9/79	1.40E-08	2.37E-06	2.02E-06
ARS2n	nucleocytoplasmic transport	12/79	2.07E-08	2.90E-06	2.48E-06
ARS2n	nuclear transport	12/79	2.28E-08	2.90E-06	2.48E-06
ARS2n	regulation of mRNA metabolic process	10/79	1.08E-06	0.00012	0.0001

ARS2n	ribonucleoprotein complex localization	7/79	1.16E-06	0.00012	0.0001
ARS2n	nucleic acid transport	8/79	1.57E-06	0.00012	0.0001
ARS2n	RNA transport	8/79	1.57E-06	0.00012	0.0001
ARS2n	nuclear export	8/79	1.63E-06	0.00012	0.0001
ARS2n	RNA export from nucleus	7/79	1.66E-06	0.00012	0.0001
ARS2n	establishment of RNA localization	8/79	1.76E-06	0.00012	0.0001
ARS2n	mRNA transport	7/79	3.66E-06	0.00023	0.0002
ARS2n	RNA localization	8/79	5.78E-06	0.00035	0.0003
ARS2n	regulation of transcription elongation from RNA polymerase II promoter	4/79	7.48E-06	0.00038	0.00033
ARS2n	mRNA export from nucleus	6/79	7.53E-06	0.00038	0.00033
ARS2n	mRNA-containing ribonucleoprotein complex export from nucleus	6/79	7.53E-06	0.00038	0.00033
ARS2n	nucleobase-containing compound transport	8/79	8.13E-06	0.00039	0.00034
ARS2n	protein export from nucleus	7/79	1.07E-05	0.0005	0.00042
ARS2n	ribonucleoprotein complex export from nucleus	6/79	1.63E-05	0.00072	0.00062
ARS2n	regulation of RNA splicing	6/79	2.31E-05	0.00098	0.00083
ARS2n	regulation of mRNA processing	6/79	2.51E-05	0.00102	0.00087
ARS2n	ribonucleoprotein complex biogenesis	10/79	2.73E-05	0.00107	0.00091
ARS2n	RNA catabolic process	9/79	4.49E-05	0.00169	0.00144
ARS2n	ribonucleoprotein complex assembly	7/79	5.22E-05	0.00189	0.00162
ARS2n	positive regulation of transcription elongation from RNA polymerase II promoter	3/79	5.68E-05	0.00199	0.0017
ARS2n	regulation of DNA-templated transcription, elongation	4/79	5.90E-05	0.002	0.00171
ARS2n	ribonucleoprotein complex subunit organization	7/79	6.31E-05	0.00207	0.00177
ARS2n	mRNA catabolic process	8/79	0.00015	0.0048	0.0041
ARS2n	protein-containing complex localization	7/79	0.00019	0.00575	0.00491
ARS2n	positive regulation of DNA-templated transcription, elongation	3/79	0.0003	0.00878	0.0075
ARS2n	alternative mRNA splicing, via spliceosome	4/79	0.0003	0.00878	0.0075
ARS2n	transcription elongation from RNA polymerase II promoter	4/79	0.00044	0.01251	0.01069
ARS2n	regulation of mRNA splicing, via spliceosome	4/79	0.00079	0.02179	0.01862
ARS2n	regulation of cellular amide metabolic process	8/79	0.0009	0.02372	0.02027
ARS2n	histone ubiquitination	3/79	0.00091	0.02372	0.02027
ARS2n	histone H2B ubiquitination	2/79	0.00114	0.02883	0.02464

ARS2n	DNA-templated transcription, elongation	4/79	0.0013	0.03233	0.02763
ARS2n	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	4/79	0.00168	0.04069	0.03478
ARS2n	endodermal cell fate commitment	2/79	0.00179	0.04232	0.03617
ARS2n	formation of cytoplasmic translation initiation complex	2/79	0.00204	0.04609	0.0394
ARS2n	positive regulation of histone H3-K4 methylation	2/79	0.00204	0.04609	0.0394
ARS2n	regulation of translation	7/79	0.00211	0.04656	0.03979
ARS2n	protein localization to kinetochore	2/79	0.00231	0.04988	0.04263
ARS2c2	protein folding	27/186	1.81E-21	4.59E-18	3.73E-18
ARS2c2	chaperone-mediated protein folding	13/186	1.69E-14	2.14E-11	1.74E-11
ARS2c2	RNA catabolic process	26/186	2.64E-14	2.23E-11	1.81E-11
ARS2c2	mRNA catabolic process	24/186	2.43E-13	1.54E-10	1.25E-10
ARS2c2	regulation of mRNA stability	16/186	2.68E-11	1.36E-08	1.10E-08
ARS2c2	regulation of RNA stability	16/186	4.45E-11	1.88E-08	1.52E-08
ARS2c2	regulation of mRNA catabolic process	16/186	1.57E-10	5.66E-08	4.60E-08
ARS2c2	cell redox homeostasis	11/186	2.33E-10	7.36E-08	5.98E-08
ARS2c2	regulation of transcription from RNA polymerase II promoter in response to hypoxia	11/186	2.69E-10	7.56E-08	6.15E-08
ARS2c2	anaphase-promoting complex-dependent catabolic process	11/186	4.72E-10	1.19E-07	9.70E-08
ARS2c2	regulation of transcription from RNA polymerase II promoter in response to stress	12/186	3.05E-09	7.01E-07	5.70E-07
ARS2c2	protein deubiquitination	17/186	3.71E-09	7.81E-07	6.35E-07
ARS2c2	regulation of DNA-templated transcription in response to stress	12/186	5.31E-09	1.03E-06	8.41E-07
ARS2c2	protein modification by small protein removal	17/186	8.44E-09	1.46E-06	1.19E-06
ARS2c2	translational initiation	14/186	8.65E-09	1.46E-06	1.19E-06
ARS2c2	hematopoietic stem cell differentiation	10/186	9.80E-09	1.55E-06	1.26E-06
ARS2c2	response to unfolded protein	13/186	2.41E-08	3.59E-06	2.92E-06
ARS2c2	regulation of mRNA metabolic process	17/186	2.76E-08	3.88E-06	3.16E-06
ARS2c2	regulation of hematopoietic stem cell differentiation	9/186	3.95E-08	5.26E-06	4.28E-06
ARS2c2	proteasome-mediated ubiquitin-dependent protein catabolic process	19/186	4.22E-08	5.33E-06	4.34E-06
ARS2c2	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	9/186	5.68E-08	6.84E-06	5.56E-06
ARS2c2	cytoplasmic translation	10/186	6.02E-08	6.93E-06	5.63E-06
ARS2c2	proteasomal protein catabolic process	20/186	6.60E-08	7.26E-06	5.90E-06

ARS2c2	antigen processing and presentation of exogenous peptide antigen via MHC class I	9/186	8.99E-08	9.48E-06	7.71E-06
ARS2c2	cellular response to oxygen levels	14/186	9.80E-08	9.92E-06	8.07E-06
ARS2c2	response to topologically incorrect protein	13/186	1.03E-07	1.00E-05	8.14E-06
ARS2c2	regulation of cellular amine metabolic process	9/186	1.12E-07	1.05E-05	8.54E-06
ARS2c2	'de novo' protein folding	7/186	1.47E-07	1.33E-05	1.08E-05
ARS2c2	cellular response to hypoxia	13/186	1.63E-07	1.42E-05	1.16E-05
ARS2c2	regulation of hematopoietic progenitor cell differentiation	9/186	1.71E-07	1.44E-05	1.17E-05
ARS2c2	regulation of cellular amino acid metabolic process	8/186	1.78E-07	1.45E-05	1.18E-05
ARS2c2	cellular response to decreased oxygen levels	13/186	2.81E-07	2.22E-05	1.81E-05
ARS2c2	negative regulation of G2/M transition of mitotic cell cycle	9/186	3.73E-07	2.86E-05	2.32E-05
ARS2c2	antigen processing and presentation of peptide antigen via MHC class I	9/186	4.90E-07	3.65E-05	2.96E-05
ARS2c2	ribonucleoside monophosphate metabolic process	7/186	6.05E-07	4.37E-05	3.56E-05
ARS2c2	interleukin-1-mediated signaling pathway	9/186	6.95E-07	4.88E-05	3.97E-05
ARS2c2	ribonucleoprotein complex subunit organization	13/186	7.31E-07	5.00E-05	4.07E-05
ARS2c2	hematopoietic progenitor cell differentiation	11/186	8.81E-07	5.79E-05	4.71E-05
ARS2c2	interaction with host	12/186	8.93E-07	5.79E-05	4.71E-05
ARS2c2	chaperone-mediated protein complex assembly	5/186	9.66E-07	6.11E-05	4.96E-05
ARS2c2	negative regulation of cell cycle G2/M phase transition	9/186	1.05E-06	6.49E-05	5.28E-05
ARS2c2	RNA splicing	18/186	1.12E-06	6.75E-05	5.48E-05
ARS2c2	establishment of protein localization to membrane	15/186	1.25E-06	7.35E-05	5.98E-05
ARS2c2	Wnt signaling pathway, planar cell polarity pathway	9/186	1.33E-06	7.67E-05	6.24E-05
ARS2c2	purine ribonucleoside monophosphate metabolic process	6/186	1.37E-06	7.70E-05	6.26E-05
ARS2c2	regulation of establishment of planar polarity	9/186	1.56E-06	8.21E-05	6.67E-05
ARS2c2	response to heat	11/186	1.57E-06	8.21E-05	6.67E-05
ARS2c2	activation of innate immune response	10/186	1.61E-06	8.21E-05	6.67E-05
ARS2c2	purine nucleoside monophosphate metabolic process	6/186	1.62E-06	8.21E-05	6.67E-05
ARS2c2	'de novo' posttranslational protein folding	6/186	1.62E-06	8.21E-05	6.67E-05

ARS2c2	cellular response to interleukin-1	11/186	1.85E-06	9.18E-05	7.46E-05
ARS2c2	positive regulation of viral life cycle	7/186	2.41E-06	0.00012	9.43E-05
ARS2c2	regulation of stem cell differentiation	9/186	2.43E-06	0.00012	9.43E-05
ARS2c2	protein refolding	6/186	2.61E-06	0.00012	9.95E-05
ARS2c2	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	9/186	3.22E-06	0.00015	0.00012
ARS2c2	ribonucleoprotein complex assembly	12/186	3.32E-06	0.00015	0.00012
ARS2c2	RNA localization	12/186	3.47E-06	0.00015	0.00013
ARS2c2	establishment of planar polarity	9/186	4.22E-06	0.00018	0.00015
ARS2c2	establishment of tissue polarity	9/186	4.22E-06	0.00018	0.00015
ARS2c2	regulation of G2/M transition of mitotic cell cycle	11/186	4.46E-06	0.00019	0.00015
ARS2c2	ribonucleoprotein complex biogenesis	17/186	4.71E-06	0.0002	0.00016
ARS2c2	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process	8/186	4.84E-06	0.0002	0.00016
ARS2c2	viral life cycle	14/186	5.51E-06	0.00022	0.00018
ARS2c2	positive regulation of catabolic process	16/186	5.79E-06	0.00023	0.00018
ARS2c2	cellular amine metabolic process	9/186	5.84E-06	0.00023	0.00018
ARS2c2	T cell receptor signaling pathway	11/186	5.95E-06	0.00023	0.00019
ARS2c2	response to temperature stimulus	12/186	6.10E-06	0.00023	0.00019
ARS2c2	regulation of protein catabolic process	15/186	6.72E-06	0.00025	0.0002
ARS2c2	nucleoside monophosphate metabolic process	7/186	7.42E-06	0.00027	0.00022
ARS2c2	response to interleukin-1	11/186	7.51E-06	0.00027	0.00022
ARS2c2	positive regulation of protein localization to nucleus	7/186	8.92E-06	0.00032	0.00026
ARS2c2	nucleocytoplasmic transport	14/186	9.17E-06	0.00032	0.00026
ARS2c2	regulation of cell cycle G2/M phase transition	11/186	9.85E-06	0.00034	0.00027
ARS2c2	response to oxygen levels	15/186	1.00E-05	0.00034	0.00027
ARS2c2	nuclear transport	14/186	1.01E-05	0.00034	0.00027
ARS2c2	amine metabolic process	9/186	1.01E-05	0.00034	0.00027
ARS2c2	antigen processing and presentation of exogenous peptide antigen	10/186	1.04E-05	0.00034	0.00028
ARS2c2	morphogenesis of a polarized epithelium	9/186	1.20E-05	0.00039	0.00032
ARS2c2	negative regulation of canonical Wnt signaling pathway	10/186	1.21E-05	0.00039	0.00032
ARS2c2	protein localization to nucleus	12/186	1.30E-05	0.00041	0.00034
ARS2c2	regulation of morphogenesis of an epithelium	10/186	1.34E-05	0.00042	0.00034
ARS2c2	protein polyubiquitination	13/186	1.45E-05	0.00045	0.00036

ARS2c2	antigen processing and presentation of exogenous antigen	10/186	1.47E-05	0.00045	0.00036
ARS2c2	chaperone cofactor-dependent protein refolding	5/186	1.51E-05	0.00045	0.00037
ARS2c2	response to hypoxia	14/186	1.53E-05	0.00046	0.00037
ARS2c2	post-translational protein modification	14/186	1.63E-05	0.00048	0.00039
ARS2c2	positive regulation of cellular catabolic process	14/186	1.68E-05	0.00048	0.00039
ARS2c2	positive regulation of canonical Wnt signaling pathway	9/186	1.68E-05	0.00048	0.00039
ARS2c2	stimulatory C-type lectin receptor signaling pathway	8/186	1.75E-05	0.0005	0.0004
ARS2c2	non-canonical Wnt signaling pathway	9/186	1.98E-05	0.00056	0.00045
ARS2c2	antigen processing and presentation of peptide antigen	10/186	2.04E-05	0.00057	0.00046
ARS2c2	innate immune response activating cell surface receptor signaling pathway	8/186	2.12E-05	0.00058	0.00047
ARS2c2	response to decreased oxygen levels	14/186	2.14E-05	0.00058	0.00047
ARS2c2	peptidyl-proline modification	6/186	2.15E-05	0.00058	0.00047
ARS2c2	innate immune response-activating signal transduction	8/186	2.26E-05	0.0006	0.00049
ARS2c2	protein import	10/186	2.34E-05	0.00062	0.0005
ARS2c2	regulation of protein stability	12/186	2.89E-05	0.00076	0.00061
ARS2c2	Fc receptor signaling pathway	11/186	3.11E-05	0.0008	0.00065
ARS2c2	toxin transport	5/186	3.13E-05	0.0008	0.00065
ARS2c2	positive regulation of protein-containing complex assembly	11/186	3.48E-05	0.00088	0.00072
ARS2c2	G2/M transition of mitotic cell cycle	11/186	3.89E-05	0.00097	0.00079
ARS2c2	positive regulation of response to biotic stimulus	11/186	4.18E-05	0.00104	0.00084
ARS2c2	nuclear-transcribed mRNA catabolic process	10/186	4.44E-05	0.00109	0.00089
ARS2c2	cellular amino acid metabolic process	13/186	4.83E-05	0.00118	0.00096
ARS2c2	negative regulation of Wnt signaling pathway	10/186	5.01E-05	0.00121	0.00098
ARS2c2	Fc-epsilon receptor signaling pathway	9/186	5.06E-05	0.00121	0.00098
ARS2c2	spindle organization	9/186	5.30E-05	0.00125	0.00102
ARS2c2	positive regulation of DNA biosynthetic process	6/186	5.44E-05	0.00127	0.00104
ARS2c2	stem cell differentiation	11/186	5.57E-05	0.00129	0.00105
ARS2c2	regulation of innate immune response	12/186	5.78E-05	0.00133	0.00108
ARS2c2	positive regulation of innate immune response	10/186	5.87E-05	0.00133	0.00108
ARS2c2	positive regulation of protein catabolic process	10/186	5.87E-05	0.00133	0.00108

ARS2c2	cellular response to heat	8/186	6.98E-05	0.00155	0.00126
ARS2c2	establishment of protein localization to mitochondrion	8/186	6.98E-05	0.00155	0.00126
ARS2c2	regulation of Wnt signaling pathway	13/186	7.42E-05	0.00163	0.00133
ARS2c2	protein stabilization	9/186	7.56E-05	0.00164	0.00133
ARS2c2	cell cycle G2/M phase transition	11/186	7.59E-05	0.00164	0.00133
ARS2c2	protein export from nucleus	9/186	7.90E-05	0.00168	0.00136
ARS2c2	positive regulation of Wnt signaling pathway	9/186	7.90E-05	0.00168	0.00136
ARS2c2	protein localization to mitochondrion	8/186	8.55E-05	0.0018	0.00147
ARS2c2	antigen processing and presentation	10/186	9.25E-05	0.00193	0.00157
ARS2c2	NIK/NF-kappaB signaling	9/186	9.36E-05	0.00194	0.00158
ARS2c2	positive regulation of viral process	7/186	9.80E-05	0.00202	0.00164
ARS2c2	interleukin-12-mediated signaling pathway	5/186	0.0001	0.00207	0.00169
ARS2c2	positive regulation of cellular protein localization	12/186	0.0001	0.00208	0.00169
ARS2c2	neutrophil degranulation	15/186	0.00011	0.00217	0.00177
ARS2c2	protein localization to nuclear body	3/186	0.00011	0.00217	0.00177
ARS2c2	positive regulation of establishment of protein localization to telomere	3/186	0.00011	0.00217	0.00177
ARS2c2	protein localization to Cajal body	3/186	0.00011	0.00217	0.00177
ARS2c2	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	13/186	0.00011	0.00221	0.00179
ARS2c2	mRNA splicing, via spliceosome	13/186	0.00011	0.00221	0.00179
ARS2c2	neutrophil activation involved in immune response	15/186	0.00012	0.00222	0.00181
ARS2c2	protein targeting	14/186	0.00012	0.00232	0.00189
ARS2c2	RNA splicing, via transesterification reactions	13/186	0.00012	0.00233	0.00189
ARS2c2	cellular response to interleukin-12	5/186	0.00012	0.00233	0.00189
ARS2c2	regulation of viral life cycle	8/186	0.00013	0.00234	0.0019
ARS2c2	negative regulation of cellular protein localization	7/186	0.00013	0.00241	0.00196
ARS2c2	response to interleukin-12	5/186	0.00014	0.00251	0.00204
ARS2c2	establishment of protein localization to chromosome	4/186	0.00014	0.00255	0.00207
ARS2c2	regulation of canonical Wnt signaling pathway	11/186	0.00014	0.0026	0.00211
ARS2c2	neutrophil activation	15/186	0.00014	0.0026	0.00211
ARS2c2	nuclear export	9/186	0.00015	0.0026	0.00211
ARS2c2	neutrophil mediated immunity	15/186	0.00015	0.00262	0.00213
ARS2c2	regulation of mitotic cell cycle phase transition	14/186	0.00015	0.00264	0.00215

ARS2c2	regulation of establishment of protein localization to telomere	3/186	0.00015	0.00264	0.00215
ARS2c2	regulation of protein localization to nucleus	7/186	0.00016	0.00282	0.00229
ARS2c2	protein localization to chromosome	6/186	0.00018	0.0031	0.00252
ARS2c2	protein localization to chromosome, telomeric region	4/186	0.00019	0.00319	0.00259
ARS2c2	regulation of response to biotic stimulus	13/186	0.00019	0.00329	0.00268
ARS2c2	negative regulation of mitotic cell cycle phase transition	10/186	0.0002	0.00329	0.00268
ARS2c2	positive regulation of establishment of protein localization	14/186	0.0002	0.00329	0.00268
ARS2c2	regulation of establishment of protein localization to chromosome	3/186	0.0002	0.00329	0.00268
ARS2c2	positive regulation of protein localization to chromosome, telomeric region	3/186	0.0002	0.00329	0.00268
ARS2c2	protein localization to nucleoplasm	3/186	0.0002	0.00329	0.00268
ARS2c2	telomere maintenance	8/186	0.00022	0.00365	0.00297
ARS2c2	regulation of animal organ morphogenesis	10/186	0.00023	0.00377	0.00306
ARS2c2	regulation of viral process	9/186	0.00025	0.00396	0.00322
ARS2c2	intracellular transport of virus	5/186	0.00026	0.0041	0.00333
ARS2c2	IMP metabolic process	3/186	0.00026	0.00411	0.00334
ARS2c2	tumor necrosis factor-mediated signaling pathway	8/186	0.00027	0.00434	0.00353
ARS2c2	establishment of protein localization to plasma membrane	5/186	0.00028	0.00437	0.00355
ARS2c2	ribonucleoside monophosphate biosynthetic process	4/186	0.00031	0.00483	0.00393
ARS2c2	positive regulation of ubiquitin-protein transferase activity	4/186	0.00031	0.00483	0.00393
ARS2c2	transport of virus	5/186	0.00033	0.00492	0.004
ARS2c2	GMP metabolic process	3/186	0.00033	0.00492	0.004
ARS2c2	organelle inheritance	3/186	0.00033	0.00492	0.004
ARS2c2	Golgi inheritance	3/186	0.00033	0.00492	0.004
ARS2c2	regulation of protein localization to chromosome, telomeric region	3/186	0.00033	0.00492	0.004
ARS2c2	regulation of cell cycle phase transition	14/186	0.00033	0.00501	0.00408
ARS2c2	antigen receptor-mediated signaling pathway	11/186	0.00034	0.00506	0.00411
ARS2c2	cytoplasmic translational initiation	4/186	0.00035	0.00512	0.00416
ARS2c2	positive regulation of telomere maintenance via telomerase	4/186	0.00035	0.00512	0.00416
ARS2c2	multi-organism transport	5/186	0.00035	0.00512	0.00416

ARS2c2	multi-organism localization	5/186	0.00035	0.00512	0.00416
ARS2c2	negative regulation of cell cycle phase transition	10/186	0.00036	0.00515	0.00419
ARS2c2	telomere organization	8/186	0.00038	0.00541	0.0044
ARS2c2	regulation of protein-containing complex assembly	13/186	0.00038	0.00545	0.00443
ARS2c2	regulation of interspecies interactions between organisms	9/186	0.0004	0.00563	0.00458
ARS2c2	positive regulation of telomerase RNA localization to Cajal body	3/186	0.00041	0.00573	0.00466
ARS2c2	positive regulation of telomerase activity	4/186	0.00044	0.00615	0.005
ARS2c2	regulation of cellular ketone metabolic process	8/186	0.00045	0.00635	0.00516
ARS2c2	modulation of process of other organism involved in symbiotic interaction	6/186	0.00047	0.00652	0.0053
ARS2c2	anatomical structure homeostasis	13/186	0.00047	0.00655	0.00532
ARS2c2	positive regulation of telomere maintenance via telomere lengthening	4/186	0.00049	0.0067	0.00545
ARS2c2	formation of cytoplasmic translation initiation complex	3/186	0.0005	0.00677	0.00551
ARS2c2	regulation of proteasomal protein catabolic process	8/186	0.00051	0.00689	0.0056
ARS2c2	cellular response to unfolded protein	7/186	0.00051	0.0069	0.00561
ARS2c2	positive regulation of cellular protein catabolic process	7/186	0.00051	0.0069	0.00561
ARS2c2	canonical Wnt signaling pathway	11/186	0.00055	0.00742	0.00603
ARS2c2	negative regulation of mitotic cell cycle	11/186	0.0006	0.00787	0.0064
ARS2c2	DNA damage response, detection of DNA damage	4/186	0.0006	0.00787	0.0064
ARS2c2	Golgi to plasma membrane protein transport	4/186	0.0006	0.00787	0.0064
ARS2c2	positive regulation of proteasomal protein catabolic process	6/186	0.00061	0.008	0.0065
ARS2c2	response to interleukin-7	4/186	0.00066	0.00854	0.00695
ARS2c2	cellular response to interleukin-7	4/186	0.00066	0.00854	0.00695
ARS2c2	telomere maintenance via telomerase	5/186	0.00067	0.0086	0.00699
ARS2c2	viral gene expression	8/186	0.00067	0.00863	0.00701
ARS2c2	establishment of protein localization to telomere	3/186	0.00071	0.00904	0.00735
ARS2c2	regulation of telomerase RNA localization to Cajal body	3/186	0.00071	0.00904	0.00735
ARS2c2	regulation of DNA biosynthetic process	6/186	0.00074	0.00942	0.00766

ARS2c2	regulation of establishment of protein localization to mitochondrion	5/186	0.00076	0.00954	0.00776
ARS2c2	regulation of apoptotic signaling pathway	12/186	0.00079	0.00986	0.00802
ARS2c2	DNA biosynthetic process	8/186	0.0008	0.00986	0.00802
ARS2c2	cellular response to chemical stress	11/186	0.0008	0.00986	0.00802
ARS2c2	cellular ketone metabolic process	9/186	0.00083	0.01005	0.00817
ARS2c2	purine nucleoside monophosphate biosynthetic process	3/186	0.00084	0.01005	0.00817
ARS2c2	purine ribonucleoside monophosphate biosynthetic process	3/186	0.00084	0.01005	0.00817
ARS2c2	RNA localization to Cajal body	3/186	0.00084	0.01005	0.00817
ARS2c2	telomerase RNA localization to Cajal body	3/186	0.00084	0.01005	0.00817
ARS2c2	telomerase RNA localization	3/186	0.00084	0.01005	0.00817
ARS2c2	RNA localization to nucleus	3/186	0.00084	0.01005	0.00817
ARS2c2	regulation of cellular protein catabolic process	9/186	0.00085	0.01017	0.00826
ARS2c2	protein peptidyl-prolyl isomerization	4/186	0.00087	0.01032	0.00839
ARS2c2	regulation of intracellular protein transport	9/186	0.00093	0.01096	0.00891
ARS2c2	modulation of process of other organism	6/186	0.00094	0.01101	0.00895
ARS2c2	immune response-activating cell surface receptor signaling pathway	13/186	0.00094	0.01101	0.00895
ARS2c2	immune response-activating signal transduction	13/186	0.00094	0.01101	0.00895
ARS2c2	nucleoside monophosphate biosynthetic process	4/186	0.00095	0.01101	0.00895
ARS2c2	error-free translesion synthesis	3/186	0.00098	0.01128	0.00917
ARS2c2	ncRNA metabolic process	13/186	0.00098	0.01128	0.00917
ARS2c2	regulation of cellular amide metabolic process	13/186	0.001	0.01145	0.00931
ARS2c2	RNA-dependent DNA biosynthetic process	5/186	0.00103	0.01167	0.00949
ARS2c2	interaction with symbiont	5/186	0.00103	0.01167	0.00949
ARS2c2	movement in host environment	7/186	0.00105	0.01184	0.00963
ARS2c2	regulation of translational fidelity	3/186	0.00113	0.01275	0.01036
ARS2c2	regulation of translation	12/186	0.00115	0.01282	0.01042
ARS2c2	regulation of translational initiation	5/186	0.00115	0.01282	0.01042
ARS2c2	regulation of cellular response to heat	5/186	0.00115	0.01282	0.01042
ARS2c2	cellular response to topologically incorrect protein	7/186	0.00117	0.01292	0.0105
ARS2c2	regulation of intracellular transport	11/186	0.00125	0.01371	0.01114
ARS2c2	import into nucleus	7/186	0.00126	0.01376	0.01119

ARS2c2	regulation of small molecule metabolic process	12/186	0.00127	0.01388	0.01128
ARS2c2	positive regulation of proteolysis involved in cellular protein catabolic process	6/186	0.00129	0.01397	0.01136
ARS2c2	regulation of telomere maintenance	5/186	0.00129	0.01397	0.01136
ARS2c2	negative regulation of protein localization to plasma membrane	3/186	0.0013	0.01403	0.01141
ARS2c2	regulation of telomerase activity	4/186	0.00132	0.01413	0.01149
ARS2c2	telomere maintenance via telomere lengthening	5/186	0.00137	0.01457	0.01185
ARS2c2	regulation of proteolysis involved in cellular protein catabolic process	8/186	0.0014	0.01486	0.01208
ARS2c2	nucleotide-excision repair, DNA damage recognition	3/186	0.00149	0.01568	0.01275
ARS2c2	proteasomal ubiquitin-independent protein catabolic process	3/186	0.00149	0.01568	0.01275
ARS2c2	positive regulation of telomere maintenance	4/186	0.00165	0.01736	0.01412
ARS2c2	positive regulation of defense response	11/186	0.00167	0.01749	0.01422
ARS2c2	negative regulation of protein localization to cell periphery	3/186	0.00169	0.01757	0.01429
ARS2c2	endomembrane system organization	12/186	0.00173	0.01795	0.01459
ARS2c2	regulation of ubiquitin-protein transferase activity	4/186	0.00178	0.01836	0.01493
ARS2c2	endosomal transport	8/186	0.00197	0.02026	0.01647
ARS2c2	regulation of telomere maintenance via telomerase	4/186	0.00205	0.02095	0.01703
ARS2c2	regulation of protein modification by small protein conjugation or removal	8/186	0.00226	0.02303	0.01872
ARS2c2	response to endoplasmic reticulum stress	9/186	0.00229	0.02322	0.01888
ARS2c2	tRNA metabolic process	7/186	0.00258	0.02601	0.02115
ARS2c2	ribonucleotide biosynthetic process	7/186	0.00258	0.02601	0.02115
ARS2c2	regulation of protein localization to plasma membrane	5/186	0.00261	0.02624	0.02133
ARS2c2	regulation of protein localization to membrane	7/186	0.00274	0.0274	0.02228
ARS2c2	regulation of DNA metabolic process	10/186	0.0028	0.02791	0.02269
ARS2c2	endothelial cell development	4/186	0.00283	0.0281	0.02284
ARS2c2	protein targeting to mitochondrion	5/186	0.00299	0.02956	0.02403
ARS2c2	ribosome biogenesis	9/186	0.00301	0.02962	0.02408
ARS2c2	positive regulation of DNA metabolic process	7/186	0.00308	0.03022	0.02457
ARS2c2	protein import into nucleus	6/186	0.00312	0.03051	0.0248

ARS2c2	ribose phosphate biosynthetic process	7/186	0.00317	0.03087	0.02509
ARS2c2	positive regulation of proteolysis	10/186	0.00323	0.03105	0.02524
ARS2c2	modulation by host of viral process	3/186	0.00324	0.03105	0.02524
ARS2c2	mRNA destabilization	3/186	0.00324	0.03105	0.02524
ARS2c2	negative regulation of protein localization to membrane	3/186	0.00324	0.03105	0.02524
ARS2c2	nucleic acid transport	7/186	0.00326	0.03105	0.02524
ARS2c2	RNA transport	7/186	0.00326	0.03105	0.02524
ARS2c2	negative regulation of intracellular transport	4/186	0.00339	0.03202	0.02603
ARS2c2	regulation of telomere maintenance via telomere lengthening	4/186	0.00339	0.03202	0.02603
ARS2c2	negative regulation of cell cycle process	10/186	0.00342	0.03218	0.02617
ARS2c2	vesicle budding from membrane	5/186	0.00355	0.03313	0.02694
ARS2c2	establishment of RNA localization	7/186	0.00355	0.03313	0.02694
ARS2c2	modulation by virus of host process	3/186	0.00356	0.03313	0.02694
ARS2c2	Golgi to plasma membrane transport	4/186	0.00359	0.03331	0.02708
ARS2c2	negative regulation of establishment of protein localization	7/186	0.00365	0.03374	0.02743
ARS2c2	regulation of nucleocytoplasmic transport	5/186	0.00386	0.03552	0.02888
ARS2c2	RNA destabilization	3/186	0.0039	0.03564	0.02897
ARS2c2	protein-lipid complex assembly	3/186	0.0039	0.03564	0.02897
ARS2c2	SRP-dependent cotranslational protein targeting to membrane	5/186	0.00402	0.03661	0.02976
ARS2c2	mitotic spindle organization	5/186	0.00419	0.0369	0.03
ARS2c2	nucleoside metabolic process	5/186	0.00419	0.0369	0.03
ARS2c2	protein targeting to membrane	7/186	0.00419	0.0369	0.03
ARS2c2	mRNA transport	6/186	0.00421	0.0369	0.03
ARS2c2	IMP biosynthetic process	2/186	0.00421	0.0369	0.03
ARS2c2	GDP metabolic process	2/186	0.00421	0.0369	0.03
ARS2c2	IRES-dependent viral translational initiation	2/186	0.00421	0.0369	0.03
ARS2c2	establishment of endothelial intestinal barrier	2/186	0.00421	0.0369	0.03
ARS2c2	positive regulation of transcription of nucleolar large rRNA by RNA polymerase I	2/186	0.00421	0.0369	0.03
ARS2c2	negative regulation of protein polyubiquitination	2/186	0.00421	0.0369	0.03
ARS2c2	positive regulation of oxidative phosphorylation	2/186	0.00421	0.0369	0.03
ARS2c2	multivesicular body sorting pathway	3/186	0.00426	0.03717	0.03022
ARS2c2	regulation of protein ubiquitination	7/186	0.00442	0.03843	0.03125

ARS2c2	regulation of DNA replication	5/186	0.00453	0.03929	0.03194
ARS2c2	protein localization to plasma membrane	8/186	0.00463	0.03995	0.03248
ARS2c2	cotranslational protein targeting to membrane	5/186	0.00471	0.04053	0.03295
ARS2c2	positive regulation of protein transport	11/186	0.00473	0.04053	0.03295
ARS2c2	DNA duplex unwinding	5/186	0.0049	0.04188	0.03405
ARS2c2	receptor catabolic process	3/186	0.00504	0.04259	0.03463
ARS2c2	positive regulation of viral genome replication	3/186	0.00504	0.04259	0.03463
ARS2c2	negative regulation of translation	7/186	0.00504	0.04259	0.03463
ARS2c2	viral latency	2/186	0.00512	0.04259	0.03463
ARS2c2	very-low-density lipoprotein particle assembly	2/186	0.00512	0.04259	0.03463
ARS2c2	positive regulation of RNA polymerase II transcription preinitiation complex assembly	2/186	0.00512	0.04259	0.03463
ARS2c2	chaperone-mediated protein transport	2/186	0.00512	0.04259	0.03463
ARS2c2	regulation of RNA binding	2/186	0.00512	0.04259	0.03463
ARS2c2	positive regulation of DNA replication	3/186	0.00546	0.04526	0.03679
ARS2c2	nucleotide-excision repair, DNA incision, 5'-to lesion	3/186	0.00589	0.04835	0.03931
ARS2c2	binding of sperm to zona pellucida	3/186	0.00589	0.04835	0.03931
ARS2c2	positive regulation of protein import into nucleus	3/186	0.00589	0.04835	0.03931
ARS2c2	regulation of protein localization to cell periphery	5/186	0.00591	0.04835	0.03931
ARS2c2	cellular response to abiotic stimulus	9/186	0.00607	0.0487	0.03959
ARS2c2	cellular response to environmental stimulus	9/186	0.00607	0.0487	0.03959
ARS2c2	transcription-coupled nucleotide-excision repair	4/186	0.00608	0.0487	0.03959
ARS2c2	glycerol ether metabolic process	2/186	0.0061	0.0487	0.03959
ARS2c2	cellular response to salt stress	2/186	0.0061	0.0487	0.03959
ARS2c2	cellular response to X-ray	2/186	0.0061	0.0487	0.03959
ARS2c2	regulation of protein K63-linked ubiquitination	2/186	0.0061	0.0487	0.03959
ARS2c2	regulation of IRE1-mediated unfolded protein response	2/186	0.0061	0.0487	0.03959
ARS2c2	DNA replication	8/186	0.00617	0.04906	0.03988
ARS2c2	regulation of intrinsic apoptotic signaling pathway	6/186	0.00625	0.0496	0.04033
ARS2c1	protein folding	17/112	2.94E-14	4.96E-11	4.13E-11
ARS2c1	cell redox homeostasis	9/112	7.40E-10	6.24E-07	5.20E-07
ARS2c1	chaperone-mediated protein folding	8/112	2.24E-09	1.26E-06	1.05E-06
ARS2c1	translational initiation	10/112	2.51E-07	0.00011	8.80E-05

ARS2c1	RNA catabolic process	13/112	7.57E-07	0.00026	0.00021
ARS2c1	mRNA catabolic process	12/112	1.93E-06	0.00054	0.00045
ARS2c1	cytoplasmic translation	7/112	2.34E-06	0.00056	0.00047
ARS2c1	toxin transport	5/112	2.66E-06	0.00056	0.00047
ARS2c1	regulation of transcription from RNA polymerase II promoter in response to hypoxia	6/112	6.83E-06	0.00128	0.00107
ARS2c1	regulation of transcription from RNA polymerase II promoter in response to stress	7/112	8.33E-06	0.0014	0.00117
ARS2c1	regulation of DNA-templated transcription in response to stress	7/112	1.14E-05	0.00152	0.00127
ARS2c1	regulation of mRNA stability	8/112	1.14E-05	0.00152	0.00127
ARS2c1	cellular response to oxygen levels	9/112	1.17E-05	0.00152	0.00127
ARS2c1	regulation of RNA stability	8/112	1.46E-05	0.00176	0.00146
ARS2c1	protein localization to nuclear body	3/112	2.45E-05	0.00237	0.00197
ARS2c1	positive regulation of establishment of protein localization to telomere	3/112	2.45E-05	0.00237	0.00197
ARS2c1	protein localization to Cajal body	3/112	2.45E-05	0.00237	0.00197
ARS2c1	response to topologically incorrect protein	8/112	2.67E-05	0.00237	0.00197
ARS2c1	regulation of mRNA catabolic process	8/112	2.67E-05	0.00237	0.00197
ARS2c1	protein localization to nucleus	9/112	2.87E-05	0.00242	0.00202
ARS2c1	positive regulation of viral life cycle	5/112	3.23E-05	0.00257	0.00214
ARS2c1	regulation of establishment of protein localization to telomere	3/112	3.35E-05	0.00257	0.00214
ARS2c1	cellular response to hypoxia	8/112	3.54E-05	0.00259	0.00216
ARS2c1	regulation of establishment of protein localization to chromosome	3/112	4.44E-05	0.00288	0.0024
ARS2c1	positive regulation of protein localization to chromosome, telomeric region	3/112	4.44E-05	0.00288	0.0024
ARS2c1	protein localization to nucleoplasm	3/112	4.44E-05	0.00288	0.0024
ARS2c1	cellular response to decreased oxygen levels	8/112	4.94E-05	0.00309	0.00257
ARS2c1	negative regulation of cellular protein localization	6/112	5.81E-05	0.0035	0.00292
ARS2c1	ribonucleoprotein complex assembly	8/112	7.21E-05	0.0041	0.00341
ARS2c1	regulation of protein localization to chromosome, telomeric region	3/112	7.29E-05	0.0041	0.00341
ARS2c1	positive regulation of protein localization to nucleus	5/112	8.23E-05	0.00448	0.00373
ARS2c1	ribonucleoprotein complex subunit organization	8/112	8.90E-05	0.00459	0.00383
ARS2c1	positive regulation of telomerase RNA localization to Cajal body	3/112	9.07E-05	0.00459	0.00383
ARS2c1	response to unfolded protein	7/112	9.26E-05	0.00459	0.00383

ARS2c1	formation of cytoplasmic translation initiation complex	3/112	0.00011	0.00527	0.00439
ARS2c1	regulation of translational initiation	5/112	0.00011	0.00527	0.00439
ARS2c1	RNA splicing	11/112	0.00012	0.00546	0.00455
ARS2c1	anaphase-promoting complex-dependent catabolic process	5/112	0.00013	0.00562	0.00468
ARS2c1	response to oxygen levels	10/112	0.00013	0.00565	0.00471
ARS2c1	regulation of mRNA metabolic process	9/112	0.00015	0.00618	0.00515
ARS2c1	establishment of protein localization to telomere	3/112	0.00016	0.00631	0.00525
ARS2c1	regulation of telomerase RNA localization to Cajal body	3/112	0.00016	0.00631	0.00525
ARS2c1	viral life cycle	9/112	0.00016	0.00631	0.00525
ARS2c1	RNA localization to Cajal body	3/112	0.00019	0.00682	0.00568
ARS2c1	telomerase RNA localization to Cajal body	3/112	0.00019	0.00682	0.00568
ARS2c1	telomerase RNA localization	3/112	0.00019	0.00682	0.00568
ARS2c1	RNA localization to nucleus	3/112	0.00019	0.00682	0.00568
ARS2c1	regulation of translational fidelity	3/112	0.00026	0.00908	0.00756
ARS2c1	maintenance of protein location in nucleus	3/112	0.0003	0.01022	0.00851
ARS2c1	protein deubiquitination	8/112	0.00031	0.01022	0.00851
ARS2c1	response to hypoxia	9/112	0.00031	0.01022	0.00851
ARS2c1	regulation of protein stability	8/112	0.00032	0.01022	0.00851
ARS2c1	peptidyl-proline modification	4/112	0.00038	0.01199	0.00999
ARS2c1	protein hydroxylation	3/112	0.00039	0.01199	0.00999
ARS2c1	response to decreased oxygen levels	9/112	0.00039	0.01199	0.00999
ARS2c1	endosomal transport	7/112	0.00043	0.01287	0.01072
ARS2c1	protein modification by small protein removal	8/112	0.00044	0.01287	0.01072
ARS2c1	maintenance of protein location in cell	4/112	0.00046	0.01287	0.01072
ARS2c1	positive regulation of viral process	5/112	0.00046	0.01287	0.01072
ARS2c1	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	9/112	0.00047	0.01287	0.01072
ARS2c1	mRNA splicing, via spliceosome	9/112	0.00047	0.01287	0.01072
ARS2c1	RNA splicing, via transesterification reactions	9/112	0.00049	0.0134	0.01116
ARS2c1	ribonucleoprotein complex biogenesis	10/112	0.00051	0.01378	0.01148
ARS2c1	establishment of protein localization to chromosome	3/112	0.00055	0.01459	0.01215
ARS2c1	proteasomal protein catabolic process	10/112	0.0006	0.01548	0.0129
ARS2c1	antigen processing and presentation of exogenous peptide antigen	6/112	0.00066	0.01676	0.01396

ARS2c1	regulation of protein localization to nucleus	5/112	0.00067	0.0169	0.01408
ARS2c1	protein localization to chromosome, telomeric region	3/112	0.00069	0.017	0.01416
ARS2c1	protein stabilization	6/112	0.00072	0.01752	0.01459
ARS2c1	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	5/112	0.00078	0.01886	0.01571
ARS2c1	antigen processing and presentation of exogenous antigen	6/112	0.0008	0.01912	0.01593
ARS2c1	regulation of establishment of protein localization to mitochondrion	4/112	0.00092	0.02161	0.018
ARS2c1	proteasome-mediated ubiquitin-dependent protein catabolic process	9/112	0.00095	0.02195	0.01829
ARS2c1	antigen processing and presentation of peptide antigen	6/112	0.00098	0.02231	0.01859
ARS2c1	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	4/112	0.00107	0.02375	0.01979
ARS2c1	regulation of oxidative stress-induced cell death	4/112	0.00107	0.02375	0.01979
ARS2c1	cytoplasmic translational initiation	3/112	0.0011	0.02375	0.01979
ARS2c1	positive regulation of telomere maintenance via telomerase	3/112	0.0011	0.02375	0.01979
ARS2c1	purine ribonucleoside monophosphate metabolic process	3/112	0.0013	0.02751	0.02292
ARS2c1	antigen processing and presentation of exogenous peptide antigen via MHC class I	4/112	0.0013	0.02751	0.02292
ARS2c1	interaction with host	6/112	0.00138	0.02777	0.02313
ARS2c1	binding of sperm to zona pellucida	3/112	0.00141	0.02777	0.02313
ARS2c1	purine nucleoside monophosphate metabolic process	3/112	0.00141	0.02777	0.02313
ARS2c1	'de novo' posttranslational protein folding	3/112	0.00141	0.02777	0.02313
ARS2c1	positive regulation of telomere maintenance via telomere lengthening	3/112	0.00141	0.02777	0.02313
ARS2c1	establishment of protein localization to mitochondrion	5/112	0.00142	0.02777	0.02313
ARS2c1	endomembrane system organization	9/112	0.00144	0.02798	0.02331
ARS2c1	hippo signaling	3/112	0.00152	0.02884	0.02402
ARS2c1	regulation of protein export from nucleus	3/112	0.00152	0.02884	0.02402
ARS2c1	GDP metabolic process	2/112	0.00156	0.02904	0.02419
ARS2c1	protein localization to chromosome	4/112	0.00157	0.02904	0.02419
ARS2c1	protein localization to mitochondrion	5/112	0.00161	0.02947	0.02455
ARS2c1	maintenance of protein localization in organelle	3/112	0.00164	0.02951	0.02458

ARS2c1	Golgi organization	5/112	0.00166	0.02951	0.02458
ARS2c1	response to endoplasmic reticulum stress	7/112	0.00166	0.02951	0.02458
ARS2c1	clathrin-dependent endocytosis	3/112	0.00177	0.03104	0.02586
ARS2c1	regulation of RNA binding	2/112	0.00189	0.03234	0.02694
ARS2c1	'de novo' protein folding	3/112	0.0019	0.03234	0.02694
ARS2c1	negative regulation of intracellular protein transport	3/112	0.0019	0.03234	0.02694
ARS2c1	regulation of cellular response to oxidative stress	4/112	0.00194	0.03278	0.0273
ARS2c1	regulation of viral life cycle	5/112	0.00205	0.0342	0.02849
ARS2c1	maintenance of protein location	4/112	0.00211	0.03489	0.02907
ARS2c1	positive regulation of cellular amide metabolic process	5/112	0.00223	0.03602	0.03001
ARS2c1	keratinocyte development	2/112	0.00226	0.03602	0.03001
ARS2c1	glycerol ether metabolic process	2/112	0.00226	0.03602	0.03001
ARS2c1	pre-mRNA cleavage required for polyadenylation	2/112	0.00226	0.03602	0.03001
ARS2c1	negative regulation of G2/M transition of mitotic cell cycle	4/112	0.00238	0.03714	0.03094
ARS2c1	cell death in response to oxidative stress	4/112	0.00238	0.03714	0.03094
ARS2c1	antigen processing and presentation	6/112	0.00243	0.03755	0.03128
ARS2c1	movement in host environment	5/112	0.00264	0.03915	0.03261
ARS2c1	RNA localization	6/112	0.00265	0.03915	0.03261
ARS2c1	peptidyl-proline hydroxylation	2/112	0.00266	0.03915	0.03261
ARS2c1	mRNA cleavage involved in mRNA processing	2/112	0.00266	0.03915	0.03261
ARS2c1	aminoacyl-tRNA metabolism involved in translational fidelity	2/112	0.00266	0.03915	0.03261
ARS2c1	antigen processing and presentation of peptide antigen via MHC class I	4/112	0.00267	0.03915	0.03261
ARS2c1	regulation of response to oxidative stress	4/112	0.00277	0.04029	0.03356
ARS2c1	negative regulation of intrinsic apoptotic signaling pathway	4/112	0.00288	0.04146	0.03453
ARS2c1	interleukin-1-mediated signaling pathway	4/112	0.00309	0.04317	0.03596
ARS2c1	GMP metabolic process	2/112	0.0031	0.04317	0.03596
ARS2c1	organelle inheritance	2/112	0.0031	0.04317	0.03596
ARS2c1	Golgi inheritance	2/112	0.0031	0.04317	0.03596
ARS2c1	regulation of intrinsic apoptotic signaling pathway	5/112	0.00318	0.04402	0.03667
ARS2c1	ribonucleoside monophosphate metabolic process	3/112	0.00336	0.04602	0.03834
ARS2c1	positive regulation of cellular protein localization	7/112	0.0034	0.04631	0.03857

ARS2c1	positive regulation of telomere maintenance	3/112	0.00355	0.04768	0.03972
ARS2c1	regulation of nucleocytoplasmic transport	4/112	0.00356	0.04768	0.03972
ARS2c1	spindle organization	5/112	0.00362	0.04804	0.04001
ARS2c1	SRP-dependent cotranslational protein targeting to membrane	4/112	0.00369	0.04819	0.04015
ARS2c1	negative regulation of cell cycle G2/M phase transition	4/112	0.00369	0.04819	0.04015
ARS2c1	mitotic spindle organization	4/112	0.00381	0.04947	0.04121

Table 6: Intron 5 IRES prediction. *Ars2n* and *Ars2c2* sequences were analyzed using IRESpy[291]. Sequence region, probability score and prediction are shown.

Name	Sequence	Prob IRES	Prediction label
ARS2n	full seq	0.05	Non-IRES
ARS2n	CDS	0.047	Non-IRES
ARS2n	5'UTR (1-270)	0.2	Non-IRES
ARS2c2	full seq (1-4781)	0.1188	Non-IRES
ARS2c2	CDS (2649-4781)	0.0476	Non-IRES
ARS2c2	5'UTR (1-2649)	0.2678	Non-IRES
ARS2c2	intron 5 (1030-2740)	0.4723	Non-IRES
ARS2c2	intron 5 (1030-1885)	0.186	Non-IRES
ARS2c2	intron 5 (1885-2740)	0.6856	Potential IRES
ARS2c2	intron 5 (1506-1838) in common X3, X7 and X6	0.04	Non-IRES
ARS2c2	intron 5 (1219-1838) in common X3 and X5	0.1848	Non-IRES
ARS2c2	intron 5 (1030-1838) in common X3 and X4	0.1766	Non-IRES
ARS2c2	intron 5 (1506-2647) in common X3 and X2	0.5146	Potential IRES
ARS2c2	intron 5 (1506-2740)	0.5316	Potential IRES
ARS2c2	intron 5 (1600-2740)	0.571	Potential IRES
ARS2c2	intron 5 (1700-2740)	0.6279	Potential IRES
ARS2c2	intron 5 (1800-2740)	0.649	Potential IRES
ARS2c2	intron 5 (1900-2740)	0.673	Potential IRES
ARS2c2	intron 5 (2000-2740)	0.669	Potential IRES
ARS2c2	intron 5 (2100-2740)	0.5468	Potential IRES
ARS2c2	intron 5 (1950-2740)	0.699	Potential IRES
ARS2c2	intron 5 (1950-2700)	0.669	Potential IRES
ARS2c2	intron 5 (1950-2500)	0.694	Potential IRES
ARS2c2	intron 5 (1950-2200)	0.71	Potential IRES
ARS2c2	intron 5 (1950-2100)	0.66	Potential IRES
ARS2c2	intron 5 (1950-2150)	0.62	Potential IRES

Appendix C: Sequence data

1) **ARS2n (AF/AR)**

AGAGCGCAGCGATTATGACCGTTTCCCCGGGAAAGGGATGAAAGACGGCGAGGG
GACGATTGGAATGACCGAGAGTGGGACCGTGGCCGGGAGCGCCGCAGTCGGGGT
GAATATCGAGACTACGACAGGAACCGAAGGGAACGCTTCTCTCCCCCTCGACACG
AACTCAGCCCCCCCCAGAAGCGCATGCGGAGAGACTGGGATGAGCACAGCTCTGA
CCCATACCACAGTGGCTATGACATGCCCTATGCTGGGGGGGGTGGGGGACCAACT
TACGGCCCCCCTCAGCCCTGGGGCCACCCAGACGTCCACATCATGCAGCACCATG
TCCTGCCCATCCAGGCCAGGCTGGGCAGCATCGCAGAGATTGACTTGGGGGTGCC
ACCGCCCATAATGAAGTCCTTCAAAGAGTTCCTCCTGTCTCTGGATGACTCTGTGG
ATGAGACGGAGGCAGTTAAACGCTACCATGACTACAAGCTGGACTTCCGAAGGCA
GCAGATGCAGGACTTTTTCCTGGCTCACAAAGACGAGGAGTGGTTCCGATCTAAGT
ACCACCCTGATGAGGTGGGAAAGCGTCGGCAGGAGGCCCGGGGGGCCCTGCAGA
ACCGGCTGAAGGTGTTCCTGTCCCTCATGGAGAGTGGCTGGTTTGATAACCTTCTC
TTGGACATAGACAAAGCTGATGCCATTGTCAAGATGCTAGATGCAGCTGTCATTAA
GATGGAAGGTGGCACAGAGAACGATCTCCGAATTTTGGAGCAGGAGGAGGAGGAG
GAACAGGCAGGCAAGACTGGGGAGGCCAGCAAGAAAGAGGAGGCCCGTGCTGGA
CCAGCCCTGGGGGAAGGAGAGCGCAAAGCCAATGATAAGGATGAGAAGAAAGAAG
ATGGAAAACAGGCTGAGAATGACAGTTCCAACGATGACAAAACTAAAAAATCTGAG
GGTGATGGGGACAAGGAGGAGAAGAAAGAAGAGGCTGAGAAAGGAAGCCAAAAA
GAGCAAGAAGCGGAACAGGAAGCAGAGTGGCGATGACAGCTTCGATGAGGNAGNT
GTGTCA

2) **Intron 5 region (1618-3328) based on transcript XM_030254927.2**

2.1) Primer P1:

P1F:

NNNNNNNNNNNTNNNTGGTTTNNNTGGAGGAGCTGCAGCCCTGTCCTCTGGTTC
TCCCCACCGTGGCACTTCCTGGGAGTGGGGAAGGGGAGGGAAACACTGTTGGCT
GGTGGGATCACACCCACCTCCTTCTCTGCCTTCCTCACTGGGACACCTCCAGGCA
GCTGACCAGAGCCACTTAGCCTCTTGGTACAGCACACAGACTGCGATCCTCTTGTT
GGTGTGGGTCATTGTCATCCCCAGTCAACAGGACCCAGTGTGGCCGTGGTCTCCC
AAAGGCAACGAGTGAAGCCAGACAAGAACAGCCAAGATNTANAGGGGANCCGTTT
A

P1R:

NNNNNNNNNNNTNGTTGCCTTATGGGANANCACGGCCACACTGGGTCCTGTTGA
CTGGGGATGACAATGACCCACACCAACAAGAGGATCGCAGTCTGTGTGCTGTACC
AAGAGGCTAAGTGGCTCTGGTCAGCTGCCTGGAGGTGTCCCAGTGAGGAAGGCAG
AGAAGGAGGTGGGTGTGATCCCACCAGCCAACAGTGTTTCCCTCCCCTTCCCCACT
CCCAGGAAGTGCCACGGTGGGGAGAACCAGAGGACAGGGCTGCAGCTCCTCCAG

CCAAACCAGAGAAATGGGCCAGGGCCAAGGGGGCAGGCANNAGGGGGGCACTAA
A

Sequence: 100% homology transcript variants XM_030254927.2, XM_030254928.2 and XM_036165632.1 (1672-1959 from XM_030254927.2)
TGGAGGAGCTGCAGCCCTGTCCTCTGGTTCTCCCCACCGTGGCACTTCCTGGGAG
TGGGGAAGGGGAGGGAAACACTGTTGGCTGGTGGGATCACACCCACCTCCTTCTC
TGCCTTCCTCACTGGGACACCTCCAGGCAGCTGACCAGAGCCACTTAGCCTCTTG
GTACAGCACACAGACTGCGATCCTCTTGTTGGTGTGGGTCATTGTCATCCCCAGTC
AACAGGACCCAGTGTGGCCGTGGTCTCCCAAAGGCAACGAGTGAAGCCAGACAAG
AACAGCCAAGAT

2.2) Primer P5:
P5F:
NNNNNNNNNNNNNNNNNNNNTGTTCTTGCTCTCTGCTACACACTTCTTGTTTCT
GTTTTCTCAACCTGTCTGTCTTTGTTTTAGATTTCGGGGCCCTCACATTTTCCCCATT
TCTCCTTTTCTCCTACCCCTAATTTCTTCTGCCTTTGTTTTTCTCCCCTGCCCTCAGA
GGCCTATTAAAATGTTTTCTTTTTCATGCTCTGCCCAGATCCTAACCCTCTAGCCCT
CTTCTATCCCATGGCCTACCTGAGGCTGTGCAGGCACTTGGCCTCTTTCTTTGCTC
TTTCCTGCCCCTCTGCCTTTTCTTCCTTCGAGGGGAGAAAGGTGTCCTGCCTTTTG
GCCTTTTCTATCTGGTACACTTGAGGTTTCCTTCTGGCTCCACCTGGGGTGGAGGG
AGTTGGTGGGAACCTCATCTGGGCTTGGCCAGGAGAGGGAGGTTCTCTGGGGCTC
CTGCTCCCTCAGCTTTTCTGAAGCTATGGGAAAGGGCACCTTCTCAGCCTAGTGCT
TCCCAGAGGCTCACAGCAGTAGAGCTGCCTCATAACCCTGGCCTTCCCCAGAGCC
ATGGTCCTCCTAACCCAGAGCTCTCAATTCCTGGTTCTCCCCAGATCTTTCCAGAGT
ACGACCTCTGGTTCCTTTTGGGGTGTTGGGTGGTTTTAGTGCACCCCTTTCTTGATT
TTAAGACCTTCTGTGGGTATGAGAGGCTATGCCCTGCCTTTCTTGGGCCCCTCCCC
ACATTCTGGTCTTGACTGCTCGTAACTGTTTTGTCTCTCCCATCCTGCTTGCTTTCTT
GTCCAGGTTCCGATCTAAGTACCACCCTGATGAGGTGGGAAAGCGTCGGCAGGAG
GCCCGGGGGGCCCTGCAGAACCGGCTGAAGGTGTTCCTGTCCCTCATGGAGAGT
GGCTGGTTTGATACCTTCTCTGGACTAGANAAGCTGATGCCTTGTCAGATGCTAAA
GCAGCTGTCTTAGNGGAAGGNGGCCAAAAACGATNTCCATTTGGAACNGGGGAGG
NGANAANNGCGGCAAATGGGAGCNACAAAAANNNCCGGCTGAACCCNGGGGAGG
AAGCCACCCNNANGGAAAAAAAAAGGAAGGGGNGNNTCCCGGAAANTTTTTGGGG
GGGNNNAAAAAAAAAAANCANGNAAANNCCACTCCCCCTTTTTNNTNNNNNNNNNN
NTCTCGNGAGGGGNNNNN

P5R:
NNNNNNNNNNNNTTCTGTTCCGCTTCTTGCTCTTTTTGGCTTCCTTCTCAGCCTCTT
CTTTCTTCTCCTCCTTGTCCCCATCACCCTCAGATTTTTTAGTTTTGTCATCGTTGGA
ACTGTCATTCTCAGCCTGTTTTCCATCTTCTTTCTTCTCATCCTTATCATTGGCTTTG
CGCTCTCCTTCCCCCAGGGCTGGTCCAGCACGGGCCTCCTCTTTCTTGCTGGCCT
CCCCAGTCTTGCCTGCCTGTTCCTCCTCCTCCTCCTGCTCCAAAATTCGGAGATCG
TTCTCTGTGCCACCTTCCATCTTAATGACAGCTGCATCTAGCATCTTGACAATGGCA
TCAGCTTTGTCTATGTCCAAGAGAAGGTTATCAAACCAGCCACTCTCCATGAGGGA
CAGGAACACCTTCAGCCGGTTCTGCAGGGCCCCCCGGGCCTCCTGCCGACGCTTT
CCCACCTCATCAGGGTGGTACTTAGATCGGAACCTGGACAAGAAAGCAAGCAGGA

TGGGAGAGACAAAACAGTTACGAGCAGTCAAGACCAGAATGTGGGGAGGGGCCCA
AGAAAGGCAGGGCATAGCCTCTCATACCCACAGAAGGTCTTAAAATCAAGAAAGGG
GTGCACTAAAACCACCCAACACCCCAAAAGGAACCAGAGGTCGTACTCTGGAAAGA
TCTGGGGAGAACCAGGAATTGAGAGCTCTGGGTTAGGAGGACCATGGCTCTGGAG
AAGGCCAGGGTTATGAGGCAGCTCTACTGCTGTGAGCCTCTGGGAAGCACTAGGC
TGAGAAGGTGCCCTTTCGCATAGCTTCAGAAAAGCTGAGGGAGCAGGAGCCCCAG
AGAACCTCCCTCTCCTGGGCAAGCCCAGATGAGGTTCCACCAACTNCCTCCACCCA
GGTGGNNNCGAAGGAAACCTCAGTGTACCAANNNNAAAGGCAAAAGGNNGAAACT
TTCTCCCTNGAAGAAAAAAGGNANGGGCAGAAAAACCAAAA

Sequence: 99% homology transcript variants XM_030254927.2, XM_006504631.3 and XM_036165632.1 (2554- 3799 from XM_030254927.2)
TGTTCTTGCTCTCTGCTACACACTTCTTGTTTCTGTTTTCTCAACCTGTCTGTCTTTG
TTTTAGATTTCGGGGCCCTCACATTTTCCCCATTTCTCCTTTTCTCCTACCCCTAATT
TCTTCTGCCTTTGTTTTTCTCCCCTGCCCTCAGAGGCCTATTAAAATGTTTTCTTTTT
CATGCTCTGCCCAGATCCTAACCCTCTAGCCCTCTTCTATCCCATGGCCTACCTGA
GGCTGTGCAGGCACTTGGCCTCTTTCTTTGCTCTTTCCTGCCCCTCTGCCTTTTCTT
CCTTCGAGGGGAGAAAGGTGTCCTGCCTTTTGGCCTTTTCTATCTGGTACACTTGA
GGTTTCCTTCTGGCTCCACCTGGGGTGGAGGGAGTTGGTGGGAACCTCATCTGGG
CTTGGCCAGGAGAGGGAGGTTCTCTGGGGCTCCTGCTCCCTCAGCTTTTCTGAAG
CTATGGGAAAGGGCACCTTCTCAGCCTAGTGCTTCCCAGAGGCTCACAGCAGTAG
AGCTGCCTCATAACCCTGGCCTTCCCCAGAGCCATGGTCCTCCTAACCCAGAGCTC
TCAATTCCTGGTTCTCCCCAGATCTTTCCAGAGTACGACCTCTGGTTCCTTTTGGGG
TGTTGGGTGGTTTTAGTGCACCCCTTTCTTGATTTTAAGACCTTCTGTGGGTATGAG
AGGCTATGCCCTGCCTTTCTTGGGCCCCTCCCCACATTCTGGTCTTGACTGCTCGT
AACTGTTTTGTCTCTCCCATCCTGCTTGCTTTCTTGTCCAGGTTCCGATCTAAGTAC
CACCCTGATGAGGTGGGAAAGCGTCGGCAGGAGGCCCGGGGGGCCCTGCAGAAC
CGGCTGAAGGTGTTCCTGTCCCTCATGGAGAGTGGCTGGTTTGATAACCTTCTCTT
GGACATAGACAAAGCTGATGCCATTGTCAAGATGCTAGATGCAGCTGTCATTAAGA
TGGAAGGTGGCACAGAGAACGATCTCCGAATTTTGGAGCAGGAGGAGGAGGAGGA
ACAGGCAGGCAAGACTGGGGAGGCCAGCAAGAAAGAGGAGGCCCGTGCTGGACC
AGCCCTGGGGGAAGGAGAGCGCAAAGCCAATGATAAGGATGAGAAGAAAGAAGAT
GGAAAACAGGCTGAGAATGACAGTTCCAACGATGACAAAACTAAAAAATCTGAGGG
TGATGGGGACAAGGAGGAGAAGAAAGAAGAGGCTGAGAAGGAAGCCAAAAAGAGC
AAGAAGCGGAACAG

2.3) Primer P3
P3F:
NNNNNNNNTNNTTNTCTNTNNTGTTCTTGCTCTCTGCTACACACTTCTTGTTTCTGTT
TTCTCAACCTGTCTGTCTTTGTTTTAGATTTCGGGGCCCTCACATTTTCCCCATTTCT
CCTTTTCTCCTACCCCTAATTTCTTCTGCCTTTGTTTTTCTCCCCTGCCCTCAGAGGC
CTATTAAAATGTTTTCTTTTTCATGCTCTGCCCAGATCCTAACCCTCTAGCCCTCTTC
TATCCCATGGCCTACCTGAGGCTGTGCAGGCACTTGGCCTCTTTCTTTGCTCTTTC
CTGCCCCTCTGCCTTTTCTTCCTTCGAGGGGAGAAAGGTGTCCTGCCTTTTGGCCT
TTTCTATCTGGTACACTTGAGGTTTCCTTCTGGCTCCACCTGGGGTGGAGGGAGTT
GGTGGGAACCTCATCTGGGCTTGGCCAGGAGAGGGAGGTTCTCTGGGGCTCCTG
CTCCCTCAGCTTTTCTGAAGCTATGGGAAAGGGCACCTTCTCAGCCTAGTGCTTCC

CAGAGGCTCACAGCAGTAGAGCTGCCTCATAACCCTGGCCTTCCCCAGAGCCATG
GTCCTCCTAACCCAGAGCTCTCAATTCCTGGTTCTCCCCAGATCTTTCCAGAGTAC
GACCTCTGGTTCCTTTTGGGGTGTTGGGTGGTTTTAGTGCACCCCTTTCTTGATTTT
AAGACCTTCTGTGGGTATGAGAGGCTATGCTNTGNNTTTTTTTGGGGGNGAAAAGC

P3R:
NNNNNNNNNNNNNTTNNNTNNAGAAAGGGGTGCACTAAAACCACCCAACACCCC
AAAAGGAACCAGAGGTCGTACTCTGGAAAGATCTGGGGAGAACCAGGAATTGAGA
GCTCTGGGTTAGGAGGACCATGGCTCTGGGGAAGGCCAGGGTTATGAGGCAGCTC
TACTGCTGTGAGCCTCTGGGAAGCACTAGGCTGAGAAGGTGCCCTTTCCCATAGCT
TCNGAAAAGCTGAGGGAGCAGGAGCCCCNGANAACCTCCCTCTCCTGGNCNAGCC
CNGATGANGTTNCCNCCNACTNCCTNCNCCCCNGGTGG

Sequence: 99% homology transcript variants XM_030254927.2, XM_006504631.3 and XM_036165632.1 (2554- 3240 from XM_030254927.2)
TGTTCTTGCTCTCTGCTACACACTTCTTGTTTCTGTTTTCTCAACCTGTCTGTCTTTG
TTTTAGATTTCGGGGCCCTCACATTTTCCCCATTTCTCCTTTTCTCCTACCCCTAATT
TCTTCTGCCTTTGTTTTTCTCCCCTGCCCTCAGAGGCCTATTAAAATGTTTTCTTTTT
CATGCTCTGCCCAGATCCTAACCCTCTAGCCCTCTTCTATCCCATGGCCTACCTGA
GGCTGTGCAGGCACTTGGCCTCTTTCTTTGCTCTTTCCTGCCCCTCTGCCTTTTCTT
CCTTCGAGGGGAGAAAGGTGTCCTGCCTTTTGGCCTTTTCTATCTGGTACACTTGA
GGTTTCCTTCTGGCTCCACCTGGGGTGGAGGGAGTTGGTGGGAACCTCATCTGGG
CTTGGCCAGGAGAGGGAGGTTCTCTGGGGCTCCTGCTCCCTCAGCTTTTCTGAAG
CTATGGGAAAGGGCACCTTCTCAGCCTAGTGCTTCCCAGAGGCTCACAGCAGTAG
AGCTGCCTCATAACCCTGGCCTTCCCCAGAGCCATGGTCCTCCTAACCCAGAGCTC
TCAATTCCTGGTTCTCCCCAGATCTTTCCAGAGTACGACCTCTGGTTCCTTTTGGGG
TGTTGGGTGGTTTTAGTGCACCCCTTTCTTGATTTTAAGACCTTCTGTGGGTATGAG
AGGCTATGCT

2.4) Primer P4
P4F:
NNNNNTNNNNNNANNANGTTCTCTGGGGCTCCTGCTCCCTCAGCTTTTCTGAAGCT
ATGGGAAAGGGCACCTTCTCAGCCTAGTGCTTCCCAGAGGCTCACAGCAGTAGAG
CTGCCTCATAACCCTGGCCTTCCCCAGAGCCATGGTCCTCCTAACCCAGAGCTCTC
AATTCCTGGTTCTCCCCAGATCTTTCCAGAGTACGACCTCTGGTTCCTTTTGGGGTG
TTGGGTGGTTTTAGTGCACCCCTTTCTTGATTTTAAGACCTTCTGTGGGTATGAGAG
GCTATGCCCTGCCTTTCTTGGGCCCCTCCCCACATTCTGGTCTTGACTGCTCGTAA
CTGTTTTGTCTCTCCCATCCTGCTTGCTTTCTTGTCCAGGTTCCGATCTAAGTACCA
CCCTGATGAGGTGGGAAAGCGTCGGCAGGAGGCCCGGGGGGCCCTGCAGAACCG
GCTGAAGGTGTTCCTGTCCCTCATGGAGAGTGGCTGGTTTGATAACCTTCTCTTGG
ACATAGACAAAGCTGATGCCATTGTCAAGATGCTAGATGCAGCTGTCATTAAGATG
GAAGGTGGCACAGAGAACGATCTCCGAATTTTGGAGCAGGAGGAGGAGGAGGAAC
AGGCAGGCAAGACTGGGGAGGCCAGCAAGAAAGAGGAGGCCCGTGCTGGACCAG
CCCTGGGGGAAGGAGAGCGCAAAGCCCATGATAAGGATGAGAAGAAAGAAGATGG
AAAACAGGCTGAGAATGACAGTTCCCACGATGACAAAACTAAAAATCTGAGGGTGA

TGGGGACAGGAGGAGAAAAAAGANNGGCTGAGAGGAGCCCCAAGAGCAGAACCG
GANAGGAGCCAAATGCGAGNCNNT
P4R:
NNNNNNNNTCTGNTTNNTGTTCCGCTTCTTGCTCTTTTTGGCTTCCTTCTCAGCCTC
TTCTTTCTTCTCCTCCTTGTCCCCATCACCCTCAGATTTTTTAGTTTTGTCATCGTTG
GAACTGTCATTCTCAGCCTGTTTTCCATCTTCTTTCTTCTCATCCTTATCATTGGCTT
TGCGCTCTCCTTCCCCCAGGGCTGGTCCAGCACGGGCCTCCTCTTTCTTGCTGGC
CTCCCCAGTCTTGCCTGCCTGTTCCTCCTCCTCCTCCTGCTCCAAAATTCGGAGAT
CGTTCTCTGTGCCACCTTCCATCTTAATGACAGCTGCATCTAGCATCTTGACAATGG
CATCAGCTTTGTCTATGTCCAAGAGAAGGTTATCAAACCAGCCACTCTCCATGAGG
GACAGGAACACCTTCAGCCGGTTCTGCAGGGCCCCCCGGGCCTCCTGCCGACGC
TTTCCCACCTCATCAGGGTGGTACTTAGATCGGAACCTGGACAAGAAAGCAAGCAG
GATGGGAGAGACAAAACAGTTACGAGCAGTCAAGACCAGAATGTGGGGAGGGGCC
CAAGAAAGGCAGGGCATAGCCTCTCATACCCACAGAAGGTCTTAAAATCAAGAAAG
GGGTGCACTAAAACCACCCAACACCCCAAAAGGAACCAGAGGTCGTACTCTGGAA
AGATCTGGGGAGAACCAGGAATTGAGAGCTCTGGGTTAGGAGGACCATGGCTCTG
GGGAAGGCCAGGGTTATGAGGCAGCTCTACTGCTGTGAGCCTCTGGGAAGCACTA
GGCTGAGAAGGTGCCCTTTCCCATAGCTTCAGAAAGCGAGGGGAGCAGGAGCCCC
ACAGAACTCCCTCTCTGGCAGCCATGAGTTTCCCCCAAATCACAAAAAAAAAAACAC
NAAAAGGGGGAGGGGCCCTTTCCCTCCCNTCAAAAAGCTGAGGGANNNAGGAGCC
CCAGAAAACTCCCNCTCCTGGCCAACCCAAAGAAAGTTCCCCCACCCNACTACCCC
AANNNAGGGAAGGGANANCCCAAAACCCCCCNCCCGG

Sequence: 99% homology transcript variants XM_030254927.2, XM_006504631.3 and XM_036165632.1 (2970- 3798 from XM_030254927.2)
GTTCTCTGGGGCTCCTGCTCCCTCAGCTTTTCTGAAGCTATGGGAAAGGGCACCTT
CTCAGCCTAGTGCTTCCCAGAGGCTCACAGCAGTAGAGCTGCCTCATAACCCTGG
CCTTCCCCAGAGCCATGGTCCTCCTAACCCAGAGCTCTCAATTCCTGGTTCTCCCC
AGATCTTTCCAGAGTACGACCTCTGGTTCCTTTTGGGGTGTTGGGTGGTTTTAGTG
CACCCCTTTCTTGATTTTAAGACCTTCTGTGGGTATGAGAGGCTATGCCCTGCCTTT
CTTGGGCCCCTCCCCACATTCTGGTCTTGACTGCTCGTAACTGTTTTGTCTCTCCCA
TCCTGCTTGCTTTCTTGTCCAGGTTCCGATCTAAGTACCACCCTGATGAGGTGGGA
AAGCGTCGGCAGGAGGCCCGGGGGGCCCTGCAGAACCGGCTGAAGGTGTTCCTG
TCCCTCATGGAGAGTGGCTGGTTTGATAACCTTCTCTTGGACATAGACAAAGCTGA
TGCCATTGTCAAGATGCTAGATGCAGCTGTCATTAAGATGGAAGGTGGCACAGAGA
ACGATCTCCGAATTTTGGAGCAGGAGGAGGAGGAGGAACAGGCAGGCAAGACTGG
GGAGGCCAGCAAGAAAGAGGAGGCCCGTGCTGGACCAGCCCTGGGGGAAGGAGA
GCGCAAAGCCAATGATAAGGATGAGAAGAAAGAAGATGGAAAACAGGCTGAGAAT
GACAGTTCCAACGATGACAAAACTAAAAAATCTGAGGGTGATGGGGACAAGGAGGA
GAAGAAAGAAGAGGCTGAGAAGGAAGCCAAAAAGAGCAAGAAGCGGAACA

3) Full predicted sequence amplification (XF/3UTR)
P3 F:
NNNGNNNNNNNNNNNNNNNNNNGCTCGTACTGTTTTGTCTCTCCCATCCTGCTT
GCTTTCTTGTCCAGGTTCCGATCTAAGTACCACCCTGATGAGGTGGGAAAGCGTCG
GCAGGAGGCCCGGGGGGCCCTGCAGAACCGGCTGAAGGTGTTCCTGTCCCTCAT
GGAGAGTGGCTGGTTTGATAACCTTCTCTTGGACATAGACAAAGCTGATGCCATTG

TCAAGATGCTAGATGCAGCTGTCATTAAGATGGACAGGTGGCACAGAGAACGATCT
CCGAATTTTGGAGCAGGAGGCCAGGAGGAGGAACAGGCAGGCAAGACTGGGGAG
GCCAGCAAGATAGGGAGGCCCGTGCTGGACCAGCCCTGGGGGAAGGAGAGCGCA
CAGCCAATGATAAGGATGAGAAGAAAGAAGATGGAAAACAGGCTGAGAATGACAGT
TCCAACGATGACGAACTAAAAAATCTGAGGGTGATGGGGACAAGGAGGAGAAGAA
AGAAGAGGCTGAGAAGGAAGCCAAAAAGAGCAAGAAGCGGAACAGGAAGCAGAG
GGCGATGACAGCTTCGATGAGGGCAGTGTGTCCGAGTCTGAGTCCGAGTCTGAGG
ATGGCCAGGCCGAGGAGGAGAAGGAGGAGGCCGAAGACTCTTNGGNTGNANTGT
AAGCCCCNGNNCNTGNATAAGGANTTGCTCTCTCGATGCTGCCCNGGTTCGANNN
NAACCTTTCGNNGTTGCATATCATTTGCTCTCTCATNNTGCACNNCATCTTTACGCN
ANTNTCACTGTCAGAGATCCTTTCTCTTTGTGANGTATNCTCACGCTGTCTGCGAGA
GTTTTTGATCNCANCGTCANCCTTANAGGAAAGNTNTTNCGACCCTGCCTGAGNGT
CTNTTTCCNGGNAGTGTGNAACTGTNNTCGNGNTCTGATCGGNNNCTGNGANANN
NTCTNNGNNTCNNNCAGTGANTGNNNGTTNNNCNGNNNNNTGACANAAAGATGCT
GNNNCTGN

P4 F:
GNNNNNNNNNNNNNNNNNNGAGGCCGAGAGCACTTAAAGAAAAGGAGAAGCCC
AAAGAGGAGGAGAAGGAGAAGCCTAAGGATGCTGCAGGGTTGGAGTGTAAGCCCC
GGCCCTTGCATAAGACTTGCTCTCTCTTCATGCGCAACATCGCACCCAACATTTCAA
GGGCAGAGATCATTTCTCTTTGTAAACGATACCCAGGCTTTATGCGAGTGGCACTG
TCAGAGCCCCAGCCAGAGAGGAGGTTTTTTCGCCGTGGCTGGGTGACTTTTGACC
GCAGTGTTAACATTAAGGAGATCTGTTGGAACCTGCAGAACATTCGGCTCCGGGAG
TGTGAACTGAGTCCCGGTGTGAACAGAGACCTGACCCGTCGTGTCCGCAACATAAA
TGGCATTACACAGCACAAGCAGATAGTGCGCAATGACATCAAGTTGGCAGCCAAGC
TAATCCACACACTGGATGACAGGACCCAGCTCTGGGCCTCTGAGCCTGGGACGCC
TCCTGTGCCCACAAGCCTCCCCTCGCAAAACCCCATCCTGAAGAACATCACTGACT
ACCTGATTGAGGAAGTGAGTGCGGAGGAGGAGGAGCTTCTGGGGAGCAGTGGGG
GACCCCCTCCTGAGGAGCCTCCCAAGGAGGGCAACCCAGCCGAGATCAATGTGGA
GAGGGATGAGAAGCTGATCAAGGTCTTGGATAAACTTCTTCTCTATTTGCGTATTGT
GCATTCTCTGGATTATTATAACACCTGTGAGTACCCTAATGAAGACGAGATGCCCAA
CCGCTGTGGCATAATCCACGTTCGGGGGCCCATGCCTCCCAACCGAATTAGTCAC
GGAGAAGTGCTGGAGTGGCAGAAGACATTTGAGGAGAAACTGACTCCACTGTTGA
GTGTGCGTGAATCCCTTTCTGAGGAAGAGGCCCANANATGGGTCGAAAAGACCCA
GAGCAGAAGTNNNAGTTGTCNCCTCCANACGCNGACTGGGCNNNTANNTGNTATG
N

P5 R:
NNGNNNNNNNNNNNNNNNNNNNNNNNNNGTCCCGATACTCCACTATGGCCCTTG
GGTCTCCTCGAACCATCCTGTTCCGAGGTTTCCCAGGATAACCGCCTTGGCCTCGA
AAAGCATCATAGTTCCCACGGCCGGCACCATATGGAGCATGGGGGTATGGAGGCC
CTCCTGTTGGGACTGCAGGGCGGACAGCACCAGCTCCATAGCCCAAGATGGGAGG
CCGGGGCTGACCATACGGCATCAAGCCCTGTGGCGTCTGATGTGGGTAGGGAAGT
CCTGGGGTCAGGCCTGGGGGGAGTATCTGGGCAGGGCCAGGTGGCTGAGCTGGC
TTGATCTCAGGCAAAGCTGGGCGCTTGGCGTCTGTGAGAAAGTTATTGAAGAACGC
CACCTCCTTCTTCACCTCCTCGATCTTCTCGGCATGCTTATTGAAGATATGCTTGCG
CACAAACTCCGGGCCCTTGAATTTCTTGACAATGACAGGACATATGCACTTATCCTT

GCCCAGTTCCTGCGTGTTGGACGTGACAAACTTCTCCACTTCTTGCTCTGGGTCTT
TTCGACCCATCTTCTGGGCCTCTTCCTTCAAAGGGATTCACGCACACTCATCAGTG
GAGTCAGTTTCTCCTCAAATGTCTTCTGCCACTCCTGCACTTCTCCGTGACTAATTC
NGTTGGCATGCANNTCCCCNTCAACGTGAATTATGTTACAGCGGTTGTGCATCTCG
TCTTCTTCANGGTACTCACAGGTGTTATACTAATCCAGAGAATGTGTCATACTGCAA
ATACAGCAGATNTTTATCCACGACTNTGATCAGCATCTCATCCCTCACCAAAGTGAA
CTCGGACTGGGNTGCACGCCCTTGGANGCTCCTCAGGAGTGNNGNTCNCNCGCT
GCTCCCANNAAGNATCATCCTCCTCAGCANTCACNTNNACAATCATGNAGTNNATG
NTGATCNTCNNCATGNGNCTNCTGCNNAGTGNANGNNNNGCNGNGNATCNN

Sequence: 99% homology transcript variants XM_030254927.2, XM_006504631.3 and XM_036165632.1 (3282- 5336 from XM_030254927.2)

GCTCGTACTGTTTTGTCTCTCCCATCCTGCTTGCTTTCTTGTCCAGGTTCCGATCTA
AGTACCACCCTGATGAGGTGGGAAAGCGTCGGCAGGAGGCCCGGGGGGCCCTGC
AGAACCGGCTGAAGGTGTTCCTGTCCCTCATGGAGAGTGGCTGGTTTGATAACCTT
CTCTTGGACATAGACAAAGCTGATGCCATTGTCAAGATGCTAGATGCAGCTGTCATT
AAGATGGACAGGTGGCACAGAGAACGATCTCCGAATTTTGGAGCAGGAGGCCAGG
AGGAGGAACAGGCAGGCAAGACTGGGGAGGCCAGCAAGATAGGGAGGCCCGTGC
TGGACCAGCCCTGGGGGAAGGAGAGCGCACAGCCAATGATAAGGATGAGAAGAAA
GAAGATGGAAAACAGGCTGAGAATGACAGTTCCAACGATGACGAACTAAAAAATCT
GAGGGTGATGGGGACAAGGAGGAGAAGAAAGAAGAGGCTGAGAAGGAAGCCAAA
AAGAGCAAGAAGCGGAACAGGAAGCAGAGGGCGATGACAGCTTCGATGAGGGCA
GTGTGTCCGAGTCTGAGTCCGAGTCTGAGGATGGCCAGGCCGAGGAGGAGAAGG
AGGAGGCCGAAGAAGCACTTAAAGAAAAGGAGAAGCCCAAAGAGGAGGAGAAGGA
GAAGCCTAAGGATGCTGCAGGGTTGGAGTGTAAGCCCCGGCCCTTGCATAAGACT
TGCTCTCTCTTCATGCGCAACATCGCACCCAACATTTCAAGGGCAGAGATCATTTCT
CTTTGTAAACGATACCCAGGCTTTATGCGAGTGGCACTGTCAGAGCCCCAGCCAGA
GAGGAGGTTTTTTCGCCGTGGCTGGGTGACTTTTGACCGCAGTGTTAACATTAAGG
AGATCTGTTGGAACCTGCAGAACATTCGGCTCCGGGAGTGTGAACTGAGTCCCGG
TGTGAACAGAGACCTGACCCGTCGTGTCCGCAACATAAATGGCATTACACAGCACA
AGCAGATAGTGCGCAATGACATCAAGTTGGCAGCCAAGCTAATCCACACACTGGAT
GACAGGACCCAGCTCTGGGCCTCTGAGCCTGGGACGCCTCCTGTGCCCACAAGCC
TCCCCTCGCAAAACCCCATCCTGAAGAACATCACTGACTACCTGATTGAGGAAGTG
AGTGCGGAGGAGGAGGAGCTTCTGGGGAGCAGTGGGGGACCCCCTCCTGAGGAG
CCTCCCAAGGAGGGCAACCCAGCCGAGATCAATGTGGAGAGGGATGAGAAGCTGA
TCAAGGTCTTGGATAAACTTCTTCTCTATTTGCGTATTGTGCATTCTCTGGATTATTA
TAACACCTGTGAGTACCCTAATGAAGACGAGATGCCCAACCGCTGTGGCATAATCC
ACGTTCGGGGGCCCATGCCTCCCAACCGAATTAGTCACGGAGAAGTGCTGGAGTG
GCAGAAGACATTTGAGGAGAAACTGACTCCACTGTTGAGTGTGCGTGAATCCCTTT
CTGAGGAAGAGGCCCANANATGGGTCGAAAAGACCCAGAGCAGGAAGTGGAGAA
GTTTGTCACGTCCAACACGCAGGAACTGGGCAAGGATAAGTGCATATGTCCTGTCA
TTGTCAAGAAATTCAAGGGCCCGGAGTTTGTGCGCAAGCATATCTTCAATAAGCAT
GCCGAGAAGATCGAGGAGGTGAAGAAGGAGGTGGCGTTCTTCAATAACTTTCTCAC
AGACGCCAAGCGCCCAGCTTTGCCTGAGATCAAGCCAGCTCAGCCACCTGGCCCT
GCCCAGATACTCCCCCCAGGCCTGACCCCAGGACTTCCCTACCCACATCAGACGC
CACAGGGCTTGATGCCGTATGGTCAGCCCCGGCCTCCCATCTTGGGCTATGGAGC
TGGTGCTGTCCGCCCTGCAGTCCCAACAGGAGGGCCTCCATACCCCCATGCTCCA

TATGGTGCCGGCCGTGGGAACTATGATGCTTTTCGAGGCCAAGGCGGTTATCCTG
GGAAACCTCGGAACAGGATGGTTCGAGGAGACCCAAGGGCCATAGTGGAGTATCG
GGAC

4) Human: Intron 5 region (963-1289) based on transcript XM_024446794.1

4.1) Primer P1

P1F:
CACCCGGGCTGAACGGCGGGGCCGGCCGGCAGACAGGCAGGGCACCGCTGTGC
ACAATGCAGTGGTCTAGCTGAATGTGATTGCTCAGGCAGCTAGACCACTGCATTGT
GCACAGCGGTGCCCTGCCTGTCTGCCTGGCCGCCCCCGCCGCCCAGCCCGCAGG
TATGGGCTGGGTATGGTNA

P1R:
TCACTCTGCATGAGCAATCTACCATACCCAGCCCATACATGAGCCCCCGGGCTGAG
CGGCGGGGCCGGCCAGGCAGACAGGCAGGGCACCGCTGGCACAAGNNGTCGTT

Sequence: 93% homology transcript variant XM_024446794.1 (1027- 1148 from XM_024446794.1)
ACCATACCCAGCCCATACATGAGCCCCCGGGCTGAGCGGCGCGGGGCCGGCCAG
GCAGACAGGCAGGGCACCGCTGTGCACAATGCAGTGGTCTAGCTGAATGTGATTG
CTCAGGCAGCTA

4.2) Primer P2

P2F:
GCTGACACCACCTCCACCCCCACCATACCCAGCCCATACATGAGCCCCCGGGCTG
AGCGGCGGGGCCGGCCAGGCAGACAGGCAGGGCACCGCTGTGCACAATGCAGC
GGTTTAGCTAGTGA

P2R
NNNNNNNNNTNNTGTCTGCCTGGCCGGCCCCGCCGCTCAGCCCGGGGGCTCATG
TATGGGCTGGGTATGGTGGGGGTGGAGGTGGGTTTGCTCCGAGCTCCGTCGATGA
GCGGGTGTAGTTCCCAACC

Sequence: 95% homology transcript variant XM_024446794.1 (974- 1125 from XM_024446794.1)
GGGAACTACACCCGCTCATCGACGGAGCTCGGAGCAAACCCACCTCCACCCCCAC
CATACCCAGCCCATACATGAGCCCCCGGGCTGAGCGGCGGGGCCGGCCAGGCAG
ACAGGCAGGGCACCGCTGTGCACAATGCAGCGGTTTAGCTAGTGA

Appendix D: Additional Publications

1. O'Sullivan C, Nickerson P, Krupke O, Christie J, Chen L, **Mesa Perez M**, Zhu M, Ryan B, Chow P, Howard P. (2019). ARS2 is required for retinal progenitor cell S-phase progression and Müller glial cell fate specification. Biochemistry and Cell Biology. 11: 1-11.
   Published.
   Refereed?: No, Open Access?: Yes

Other Publications

1. Ocean Han, Kyle Bromma, Nicholas Palmerley, Ariadne T Bido, **Mesa Monica**, Abdulaziz Alhussan, Perry L Howard, Alexandre G Brolo, Wayne Beckham, Abraham S Alexander, Devika B Chithrani. (2021). Nanotechnology driven cancer chemoradiation: Exploiting the full potential of radiotherapy with a unique combination of gold nanoparticles and Bleomycin. Pharmaceutics.
   Under Review
   Refereed?: No, Open Access?: Yes

2. Alhussan A, Bromma K, **Perez M.M**, Beckham W, Alexander A, Howard PL, Chithrani, DB. (2021). Docetaxel Mediated Uptake and Retention of Gold Nanoparticles in Tumor Cells and in Cancer Associated Fibroblasts. Cancers. 13
   Published.
   Refereed?: No, Open Access?: Yes

3. Bromma K, Alhussan A, **Perez M.M**, Howard P, Beckham W, Chithrani DB. (2021). Three-Dimensional Tumor Spheroids as a Tool for Reliable Investigation of Combined Gold Nano particle and Docetaxel Treatment. Cancers. 13(1465)
   Published.
   Refereed?: No, Open Access?: Yes

4. Aaron Henry Bannister, Kyle Bromma, Wonmo Sung, **Mesa Monica**, Leah Cicon, Perry Howard, Robert L Chow, Jan Schuemann and Devika Basnagge Chithrani. (2019). Modulation of nanoparticle uptake, intracellular distribution, and retention with docetaxel to enhance radiotherapy. BJR. 99(1106)
   Published.
   Refereed?: Yes, Open Access?: No

5. Cabellero I, Aira LE, Lavastida A, Popa X, Rivero J, Gonzalez J, **Mesa M**, Gonzalez N, Coba K, Lorenzo Luaces P, Wilkinson B, Santiesteban Y, Troche M, Suarez E, Crombet T, Sanchez B, Casaco A, Macias A, Mazorra Z. (2017). Safety and immunogenicity of a Human Epidermal Growth Factor Receptor 1 (HER1)- based vaccine in Prostate Castration-Resistant carcinoma patients: A dose-escalation Phase I Study Trial. Frontiers in Pharmacology. 8(263) Published.
   Refereed?: No, Open Access?: Yes

6. Miranda A, Funes JM, Sanchez N, Limia CM, **Mesa M**, Quezada S, Perez R, de Leon J. (2015). Oncogenic transformation can orchestrate immune evasion and inflammation in human mesenchymal stem cells independently of extrinsic immune-selective pressure. Cancer Research. 75(15): 3032-3042. Published.
   Refereed?: Yes, Open Access?: Yes