

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

Monica Mesa-Perez^{1b}, Phineas T. Hamilton², Alex Miranda², Nicholas Brodie^{1,3}, Connor O'Sullivan¹, Jennifer Christie¹, Bridget C. Ryan⁴, Robert L. Chow⁴, David Goodlett^{1,3}, Christopher J. Nelson¹ and Perry L. Howard^{1,*}

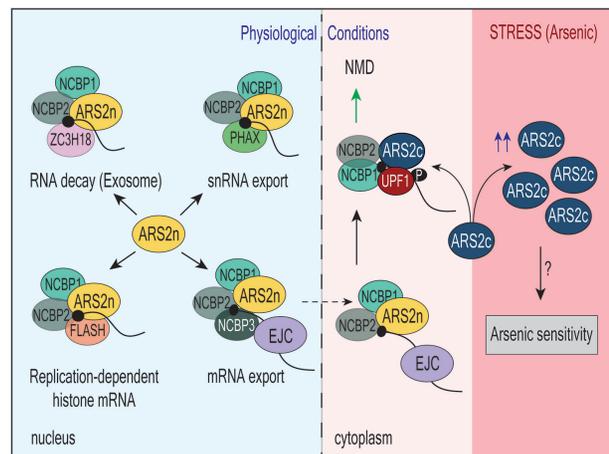
¹Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC V8W 2Y2, Canada, ²Deeley Research Centre, BC Cancer, Victoria, BC V8R 6V5, Canada, ³University of Victoria Genome BC Proteomics Centre, Vancouver Island Technology Park, Victoria, BC V8Z 7X8, Canada and ⁴Department of Biology, University of Victoria, Victoria, BC V8W 3N5, Canada

Received November 16, 2021; Revised December 29, 2021; Editorial Decision January 05, 2022; Accepted January 13, 2022

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), bound to the cap binding complex (CBC). 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. 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 NCBP1 and ERF1, favouring SURF complex formation, SMG7 recruitment and transcript degradation. ARS2 isoform exchange is also relevant during arsenic stress, where cytoplasmic ARS2 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, and stress response.

GRAPHICAL ABSTRACT



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 nuclear cap binding complex (CBC), consisting of NCBP1 and NCBP2, bound to the 7-methyl guanosine cap of RNAPII transcripts (1). ARS2/SRRT plays an important scaffolding role within the CBC by helping to synchronize the dynamic assembly and disassembly of mutually exclusive complexes that regulate mRNA splicing, degradation, and export (1–7). CBC-ARS2 bound mRNAs are exported to the cytoplasm, where NCBP1 is required for the pioneering round of translation and for promoting transcript degradation through nonsense mediated decay (NMD) (8,9). Although ARS2 is translocated to the

*To whom correspondence should be addressed. Tel: +1 260 721 7068; Fax: +1 260 721 8855; Email: phoward@uvic.ca

cytoplasm with the CBC, it is unknown whether ARS2 participates in CBC-related functions in the cytoplasm.

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 (10,11). During nuclear splicing, the multiprotein Exon Junction Complex (EJC) is deposited onto the mRNAs 24 nt upstream of the spliced junctions (12). EJCs are exported to the cytoplasm with the mRNAs and are dissociated by the ribosome during translation. By recruiting several factors involved in splicing, transport, translation and NMD, the EJC establishes a molecular link between nuclear splicing and cytoplasmic NMD (13–15). Mammalian-cell NMD generally occurs when translation terminates more than 50–55 nucleotides upstream of an EJC complex (16). 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 NCBP1. Subsequently, SMG1 kinase joins the complex to generate the SURF complex (SMG1–UPF1–ERF1–ERF3). Next, in a step dependent on DHX34 and NCBP1, 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 (15). Importantly, inhibition of the NMD downstream of the DECID complex results in UPF1 hyperphosphorylation (17). Previous work has shown that CBC complex component NCBP1 promotes NMD in the cytoplasm (9,18). Additional studies have reported the interaction between ARS2 and several components of the NMD pathway (5,19,20). However, the question of whether ARS2 participates in NMD has not been addressed.

Another unanswered question of ARS2 function relates to its known role in the cellular response to arsenic. Arsenic is not only a widespread environmental contaminant, but is also used to treat certain cancers, such as acute promyelocytic leukemia (APL) (21). The initial study of *Ars2* showed that overexpression conferred resistance to arsenic (22). The cDNA used in this study encoded only a small portion of the extreme C-terminus of full-length ARS2 and generated a dominant negative. It was subsequently found through knockdown studies that full length *Ars2* conferred arsenic sensitivity (3). However, it is still unknown whether ARS2's role in RNA metabolism relates to the cellular response to arsenic.

In this study, we demonstrated that alternative splicing of intron 5 generates cytoplasmic isoforms of ARS2 (herein named ARS2c), that lack 270 amino acids from the N-terminal of the protein. Despite sharing the helical core, C-terminal domain, RNA and CBC binding regions with nuclear ARS2 (ARS2n), ARS2c is functionally distinct from ARS2n. We show that ARS2c and not ARS2n is specifically induced during arsenic stress, and it is responsible for arsenic sensitivity. Knock down of ARS2c is sufficient to abrogate proteomic remodeling and cellular response to arsenic, in a process that it is independent of the CBC complex. ARS2c mediated arsenic sensitivity is, to our knowledge, the

first described function for mammalian ARS2 that is potentially independent from the CBC complex. Differential biological roles of ARS2 isoforms are also seen in the regulation of NMD, where our data shows that both ARS2n and ARS2c work in tandem to modulate the pathway. ARS2n primarily functions in the nucleus with the CBC to regulate mRNA splicing, degradation and export (1–7). 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 NCBP1, promotes the interaction between UPF1 and ERF1/NCBP1, favoring the formation of the SURF complex at the PTC, SMG7 recruitment and transcript degradation through NMD. Thus, the cytoplasmic switch of ARS2 isoforms tailors the protein to function in a nonredundant manner in both nuclear and cytoplasmic CBC-dependent processes.

MATERIALS AND METHODS

Cell culture, transfection, and generation of stable cell lines

C₂C₁₂ (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% CO₂ buffered incubators. 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 48 h prior to selection with hygromycin (InvivoGen). Stable clones were pooled and tested for expression with the addition of 1 µg/ml tetracycline (Sigma) for 24 h. Transfections were initially performed using Lipofectamine 3000 (Invitrogen), but it was switched to JetPrime reagent (VWR) due to toxicity. Small interfering RNAs (siRNAs) were purchased from Integrated DNA Technologies (IDT) and All-Stars 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 was used as control. All siRNA/shRNA sequences are reported in Supplementary Table S1.

Plasmids

Expression vectors for stable integration and tetracycline-inducible expression, were generated by subcloning synthesized ARS2n/ARS2c1/ARS2c2-3xFLAG (GenScript) into a modified pcDNA5/FRT/TO vector containing BirA, kindly provided by Nelson (23). BioID2-ARS2n/ARS2c1/ARS2c2-3xFLAG vectors were generated by synthesizing 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 (24). 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. λ N was synthesized and cloned into pcDNA3.1 (-) ARS2n (25) or pcDNA3.1 (-) alone generating ARS2n- λ N and λ N control respectively. λ N-RNPS1 plasmid was generated by synthesis of RNPS1 and cloning on λ N pcDNA3.1 (-) plasmid. pNMD+ and pNMD- reporter plasmids were kindly provided by K. Lukyanov (26). RNT1-GFP was a gift from H. Dietz (Addgene plasmid # 17708) (27). pAc5.1C-FLuc-Stop-5BoxB was a gift from Elisa Izaurralde (Addgene plasmid # 21301) (28). CBP20-3flag was a gift from Torben Heick Jensen and John LaCava (29).

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

To ensure data robustness and consistency, three individually treated wells from a six-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, three to six biological replicates were included, leading to 9–18 datapoints per condition that represent 9–18 individually treated wells. Total RNA was isolated using RNeasy plus kit (Qiagen). To ensure extensive elimination of genomic DNA, RNA samples were additionally treated with DNaseI (ThermoFisher). 1 μ g 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 Supplementary Table S1.

Confocal microscopy

HEK 293T cells were seeded on glass coverslips (neuVITRO) and transfected with BioID2-ARS2n/ARS2c1/ARS2c2-3xFLAG or control during 48 h in complete DMEM media. Cells were washed with PBS and fixed for 15 min with 4% paraformaldehyde in PBS at RT. Next, cells were permeabilized in 0.25% Triton X-100 for 15 min at RT and blocked with 1% BSA in PBS/Tween (0.01%), 1 h 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 h at RT, respectively. Hoechst 3334 (1:2000) was added at RT for 5 min to label nuclei. Images were acquired using a confocal microscope (Nikon) and processed in ImageJ.

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 h 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 h 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).

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), NCBP1 (1:1000, Cell Signalling), Phospho-(Ser/Thr) ATM/ATR Substrate (1:1000, Cell Signalling).

Protein fractionation

C₂C₁₂ cells were washed with buffer H (20 mM HEPES pH8, 2 mM MgCl₂, 0.1 mM EGTA, 1 mM EDTA) and lysed for 10 min at 4°C in buffer I (2× buffer H, 0.2% NP-20, protease inhibitor cocktail). Cell's nuclei were pelleted at 16 000 × g for 10 min 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 min on ice. Cell debris was pelleted at 16 000 × g for 20 min and supernatant was stored as nuclear fraction. Nuclear and cytoplasmic fractions were quantified by BCA Protein Assay kit (ThermoFisher). Equal amounts of total proteins (30 μ g) 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.

Affinity capture of biotinylated proteins

Purification of biotinylated proteins was performed as outlined by (30). Flp-In T-REx cells stably expressing tetracycline inducible ARS2n-BirA/control or BioID2-ARS2n/ARS2c1/ARS2c2/control, were treated for 24 h with 10–200 ng/ml of tetracycline. HEK 293T cells were transfected with BioID2-ARS2n/ARS2c1/ARS2c2/-3xFLAG or control and incubated for 24 h. All the conditions were incubated an additional 24 h in complete DMEM media supplemented with 50 μ M of biotin (Sigma). Five individually treated 10 cm² dishes, corresponding to around 50 × 10⁶ total cells were pooled together and considered one biological replicate. One biological replicate (5 × 10 cm² dishes) was used for tetracycline inducible BioID2-ARS2n/ARS2c1/ARS2c2/control and three biological replicates (15 × 10 cm² 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 500 μ l Lysis Buffer (50 mM Tris pH 7.4, 500 mM NaCl, 0.2% SDS, 1 mM 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 s with 2 min rest, using a sonic dismembrator (Fisher Scientific). Lysates were diluted in chilled 50 mM Tris pH 7.4 and subjected to a final round of sonication. Insoluble cellular debris was cleared by centrifugation at 16 000 \times g for 20 min at 4°C. Cleared extracts were incubated with 200 μ l 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, 500 mM NaCl, 1 mM EDTA, 50 mM HEPES pH7.5 and 1% Triton X-100), once in Wash Buffer 3 (10 mM Tris pH 7.4, 250 mM LiCl, 0.5% NP-40, 0.5% deoxycholic acid and 1 mM EDTA), and finally two times with Wash Buffer 4 (50 mM Tris pH 7.4 and 50 mM NaCl), pelleting beads using a magnetic rack. For western blotting, beads were resuspended in Laemmli sample buffer supplemented with 50 μ M biotin, incubated for 10 min, and boiled for 10 min before loading on an SDS-PAGE gel. Total protein was detected using Silver Stain kit (ThermoFisher) or Coomassie 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.

Protein identification by mass spectrometry

BioID samples were processed for mass spectrometry as outlined by (23). Briefly, streptavidin beads were washed with 50 mM NH_4HCO_3 , resuspended in 50 mM NH_4HCO_3 containing 5 mM dithiothreitol, and heated at 75°C for 10 min. Iodoacetamide was added to a final concentration of 10 mM to each sample followed by incubation in the dark at RT for 1 h. Afterwards, 1 mM CaCl_2 and 1 μ g 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 $\frac{1}{2}$ in water, lyophilized and store at -80°C until use.

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

Digested samples (6 μ l) 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 μ m I.D., 2.5 cm length, 5 μ m, 100 Å) and an in-house prepared reverse phase nano-analytical column Magic C-18AQ (75 μ m I.D., 15 cm length, 5 μ m, 100 Å,

Michrom BioResources Inc., Auburn, CA), at a flow rate of 300 nl/min. The chromatography system was coupled on-line with an Orbitrap Fusion Tribrid mass spectrometer (ThermoFisher Scientific, San Jose, CA) equipped with a Nanospray Flex NG source (ThermoFisher Scientific). Solvents were A: 2% acetonitrile, 0.1% formic acid; B: 90% acetonitrile, 0.1% formic acid. After a 348 bar (\sim 3 μ l) pre-column equilibration and 348 bar (\sim 3 μ l) nanocolumn equilibration, samples were separated by a 60-min 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 s, with the software determining 'Automatic' number of MS/MS acquisitions during this period (31).

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 two 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) (32). To identify significant interactors with the different ARS2 isoforms, we tested for differences in unique peptide counts across different baits using triplicate replicates of each bait condition, as well as additional negative controls from (33,34). 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 (35) within the Cytoscape framework (36). 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 (37). Saint analyses are reported in Supplementary Tables S2-S4. Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier (PXD026453).

Protein immunoprecipitation

GFP or Flag IP. Cells co-transfected with BioID2-ARS2n/ Δ NLS-ARS2n/ARS2c/control vectors and GFP-UPF1 or GFP control were lysed in RIPA lysis buffer for 30 min at 4°C. Cell debris was removed by centrifugation (20 min at 16 000 \times 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/ Δ NLS-ARS2n/ARS2c/control plasmids and lysed in RIPA lysis buffer for 30 min at 4°C. Cell debris was removed by centrifugation (20 min at 16 000 \times 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 1 μ g 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 min and interactors were detected with specific antibodies. For all the performed IPs, 500 μ g of total proteins were used.

Quantification. Four types of quantifications were performed, as indicated in the figure legends. (i) 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. (ii) 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. (iii) 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). (iv) 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.

Cell survival assays: WST1 and crystal violet

C₂C₁₂ cells were transfected with *Ars2c/Ars2(all)/Ncpb1/Ars2c + Ncpb1* or control RNAi for 24 h. Media was changed to Arsenic (35 μ M/40 μ M) and cells were incubated for 24, 48, 72, 96 or 120 h after the arsenic treatment. Cells transfected with the described siRNA for 48 h but untreated with arsenic were defined as point zero. Arsenic media was changed every 3 days. For WST1 assays, WST1 solution (Sigma) was added to the supernatant of

the cells and incubated for 4 h at 37°C in 5% CO₂ 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 min at RT. After washing the plates with water, crystal violet solution was added, and plates were incubated on a rocker for 20 min at RT. Crystal violet solution was removed with several washes of water and plates were dried 24 h 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.

NMD tethering assay

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 λ N or λ N-ARS2n. Firefly and Renilla luminescence were analyzed 72 h after the first 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.

NMD reporter assay

HeLa and C₂C₁₂ cells were transfected with *ARS2si/ARS2csi/UPF1si* or control. Media was changed 16 h post-transfection and each condition was transfected again with either pNMD+ or pNMD- vectors. Flow cytometry and RT-qPCR analysis were performed 24 or 48 h after the second transfection. For flow cytometry 1.0 \times 10⁵ 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 (38). Flow cytometry data was deposited on the FlowRepository with the identifier FR-FCM-Z3V5.

Immunofluorescence

Cells were transfected with the indicated GFP-ARS2n/ARS2c/ Δ nls-ARS2n/control constructs. Twenty-four hours post-transfection, cells were fixed and imaged using a 20 \times 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 min to label nuclei. Images were processed in ImageJ. GFP intensity in nuclei and cytoplasm was quantified for fifty individual cells across 5 image panels.

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.

RESULTS

Cytoplasmic isoforms of *Ars2*

The *Ars2* (*Srrt*) locus produces twelve transcripts in mouse and human, some of which encode different proteins isoforms (NCBI Gene database). In mouse, the transcripts can be organized into three groups based on the potential major protein products. For simplicity, only representative transcripts from each group are shown in Figure 1A. The first group represents canonical ARS2, which initiates translation within exon 2, encodes similar isoforms of 864–875 amino acids, and has a molecular weight of ~130 kDa (Figure 1A). 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 (RNAP II) transcript biogenesis and turnover. The second group (2*) comprises one isoform, initiates translation within exon 4 and encodes a protein of 778 amino acids. The third group (3*) is comprised of seven transcripts and involves alternative splicing of intron 5. Group 3* transcripts retain a portion of intron 5, are predicted to begin translation within either intron 5 or exon 6 and encode putative isoforms of 710 or 648 amino acids (predicted molecular weight ~100 kDa, based on ARS2n size). Group 3* isoforms are especially intriguing since they are missing up to 227 amino acids from the functionally important N-terminus of ARS2n. This missing region contains a structural N-terminal helix-turn-helix, conserved tyrosine phosphorylation sites, and a nuclear localization signal. We designated the isoforms in group 3* as ARS2c (1,2), for their potential to encode cytoplasmic isoforms of unknown function. Intron 5 is highly conserved across vertebrates, and it is striking that 7 of the 12 *Ars2* isoforms retain portions of this intron (Figure 1A).

To investigate the expression of the group 3* transcripts, we first performed RT-PCR using combinations of primers flanking intron 5 (Figure 1B). Indeed, we detected expression of transcripts retaining intron 5 in C₂C₁₂ myoblasts. Controls lacking reverse transcriptase did not show any amplification products confirming that our results are not a consequence of genomic contamination (Figure 1C). Using a forward primer within intron 5 and reverse primer within the 3'UTR of *Ars2*, a full-length amplicon of ~2150 bp was amplified (Figure 1D). Sequencing showed that this transcript 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 until the 3'UTR. All amplicons detected in Figure 1C, and D were sequenced and were ≥99% identical to the sequences within the NCBI Gene Database (Supplementary sequence).

We next determined if the *Ars2c* isoforms are translated. A cartoon of the reported protein structure of ARS2n and ARS2c1/c2 isoforms is represented in Figure 1F (20). Using an antibody (XL14.1) raised against the C-terminus of ARS2 (676–871), which is shared between ARS2n and ARS2c isoforms, we detected approximately 130 and 100

kDa products in lysates from C₂C₁₂ cells. To confirm the identity of the products, we performed RNAi. Unique siRNA targeting regions in common to all transcripts were designed and their sequences are included in Supplementary Table S1. When combined, the siRNAs strongly knocked down the expression of both protein products (Figure 1E). These results were confirmed using the Human Atlas antibody HPA042858, previously validated for detecting ARS2 by western blot. HPA042858, like XL14.1, detected 2 protein products (130–100 kDa) that are knocked down when shRNA targeting all *Ars2* forms is used (Supplementary Figure S1A–C). These data suggest that both ARS2n and ARS2c isoforms are translated in C₂C₁₂ cells.

To determine the localization of ARS2c isoforms, we expressed the cDNAs fused in-frame with a 3xFlag epitope or a GFP tag (Figure 1G and Supplementary Figure S1D, respectively). ARS2n isoforms have been shown to be predominantly nuclear but capable of shuttling between the nucleus and the cytoplasm (3). As expected, ARS2n was predominantly nuclear in localization. In contrast, ARS2c1 and ARS2c2 localized predominantly to the cytoplasm (Figure 1G, Supplementary Figure S1D–F). Similarly, nuclear/cytoplasmic fractionation and immunoprecipitation from C₂C₁₂ cell lysates showed that endogenous ARS2n and ARS2c are preferentially enriched in the nuclear and cytoplasm fractions, respectively (Figure 1H, Supplementary Figure S1G–I).

The expression of cytoplasmic ARS2 isoforms is conserved in humans. In humans, an isoform analogous to ARS2c is encoded by XM-024446794.1 (SRRT-X8) (Supplementary Figure S2A). Using RT-PCR, Western Blotting, and shRNAs specific for all *ARS2* isoforms or intron 5 exclusively, we demonstrated that ARS2c is expressed in the human cell line HeLa (Supplementary Figure S2B–E). Additionally, the Human Protein Atlas shows that ARS2 localizes to both the nucleus and cytoplasm in several tissues, including muscle, implying expression of both ARS2n and ARS2c isoforms (Supplementary Figure S2F). While we cannot rule out some of the cytoplasmic immunofluorescence signal is due to shuttling of the ARS2n, taken together, our data strongly indicates the presence of cytoplasmic isoforms of ARS2.

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 (33,39). 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 (Figure 2A). Biotinylated proteomes were normalized against the ligase control and enriched proteins were detected by Western Blot or LC-MS/MS (Figure 2B). We confirmed that fusion of ARS2 isoforms to BioID2 ligase did not affect their localization or functionality: BioID2-ARS2n and BioID2-ARS2c1/c2 localized in the nucleus and cytoplasm respectively, and BioID2-ARS2n maintains its interaction with the CBC

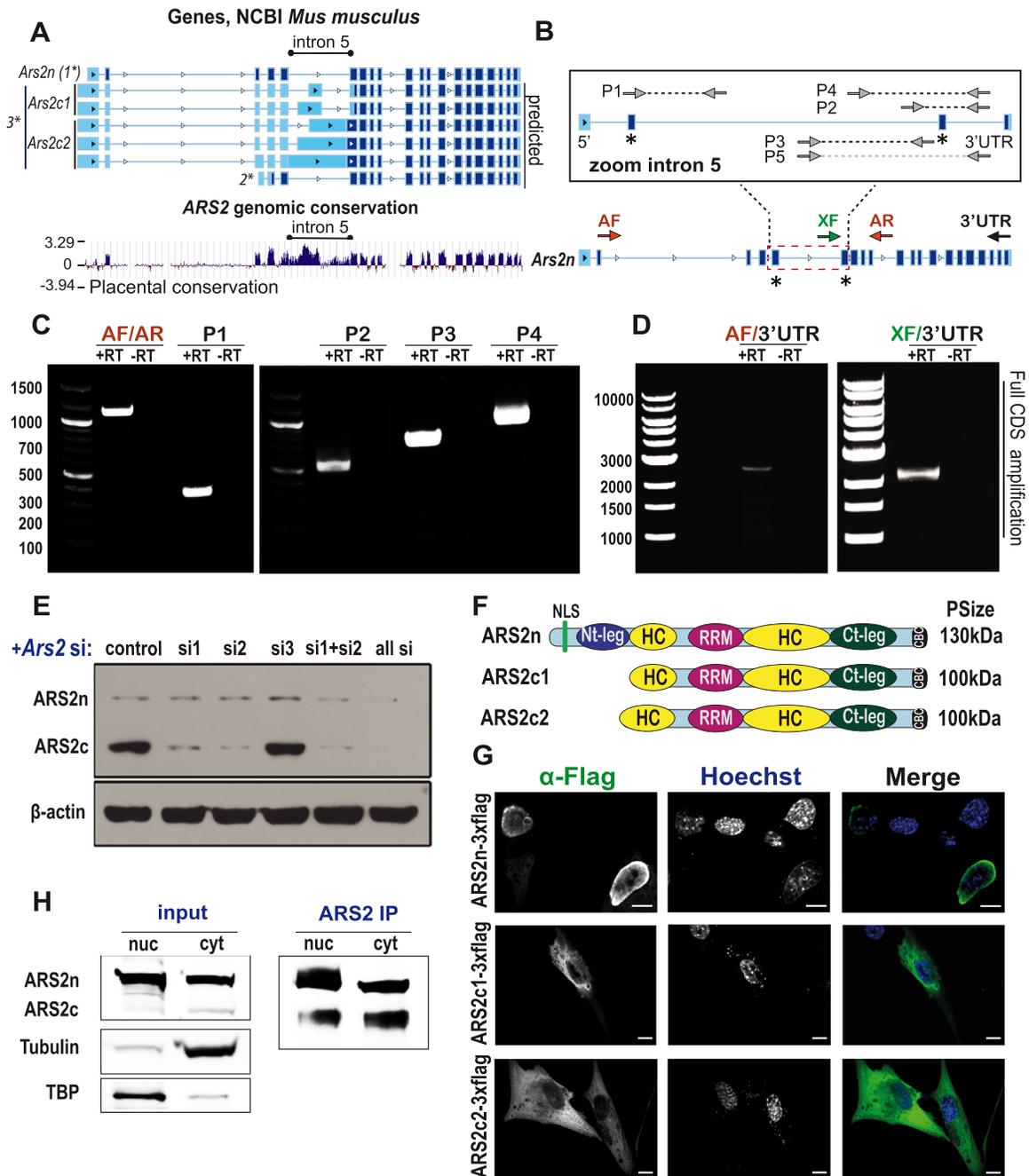


Figure 1. Cytoplasmic isoforms of *Ars2*. (A) Representative *Ars2* sequences of major isoforms in NCBI Gene database (*mus musculus*), exons are represented in dark blue boxes. Bottom: conservation analysis of *Ars2* between mouse and 60 vertebrates. The alignments were generated using Multiz and UCSC/Penn State Bioinformatics comparative genomics alignment pipeline. Evolutionary conservation was measured using PhyloP and PhastCons. (B) Diagram of primers used in (C) and (D) to amplify intron 5 (C) or the full CDS of the new isoforms (D), in C₂C₁₂ cells. Predicted amplicon sizes: AF/AR:1100 bp, P1: 332 bp, P2: 545 bp, P3: 750 bp, P4: 915 bp, AF/3'UTR: 2645 bp, XF/3'UTR: 2150 bp. (E) Western Blot of C₂C₁₂ whole cell lysates transfected with *Ars2* si or control. ARS2 was detected using an anti-ARS2 antibody and β-actin is shown as loading control. (F) Protein motif structure of ARS2n and ARS2c1/c2 isoforms, based on the reported structure of ARS2n (20), predicted protein sizes (PSize) are included. (G) HEK 293T cells were transfected with ARS2n/c1/c2-3xflag. ARS2 was detected with anti-flag antibodies (green) and nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μM. (H) Nuclear/cytoplasmic fractionation of C₂C₁₂ lysates. Endogenous ARS2 was immunoprecipitated and detected with anti-ARS2 antibodies. TBP and Tubulin detection are used as fractionation quality control.

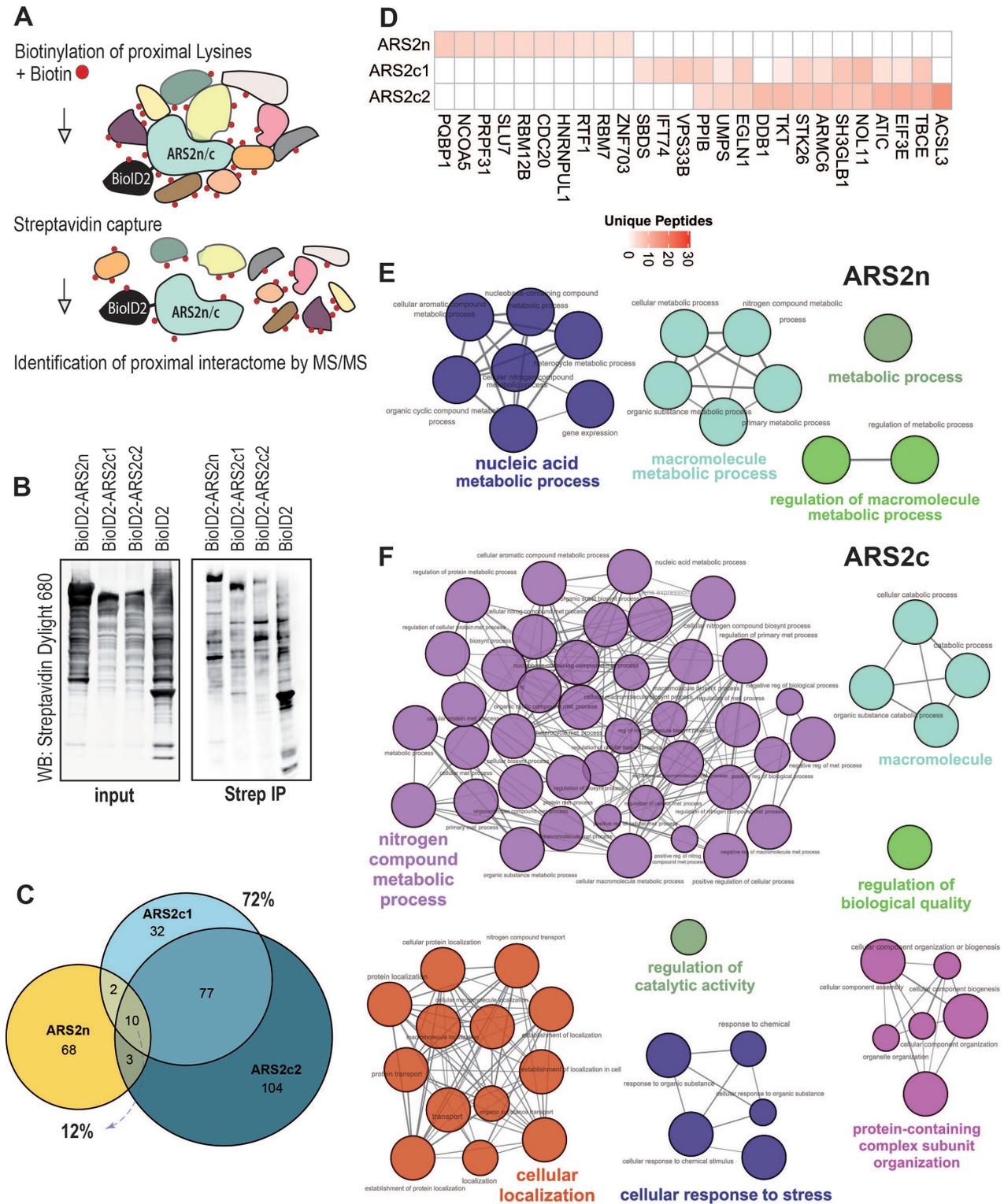


Figure 2. Comparison of ARS2n and ARS2c interactomes. (A) Schematic representation of BioID. Biotinylated proteins are immunoprecipitated by streptavidin capture and interactors are detected by Western Blot or LC/MS-MS. (B) HEK 293T cells were transfected with BioID2-ARS2n/c1/c2-3xflag or control, and treated with biotin for 24 h. 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. (C–F) Samples were treated as in (B), protein interactomes were detected by LC/MS-MS. For each condition, 15 biological replicates were pooled and ran in three independent experiments. ARS2 interactomes were normalized against BioID2 control and analyzed by SAINT. Only proteins with a SAINT score (SP > 0.7) were included in the subsequent analysis. (C) Venn diagram comparing ARS2 forms. (D) Top ten interactors for each isoform. (E, F) Gene ontology (GO) analysis of ARS2 isoform interactomes using ClueGO (35). 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.

complex (Supplementary Figure S3A, B). 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 (data not shown). We 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 we could not adequately control the levels of the biotin ligase control relative to ARS2c1,c2 expression. However, we 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 (20) we overexpressed ARS2 isoforms and control. An equal amount of protein for each condition, as quantified by silver stain, was used for LC-MS/MS.

The proximity interactome of ARS2n confirmed known interactors including NELF, NCBP3, PHAX, XRN2 and RBM7, previously detected by AP-LC-MS/MS (19,20) (Supplementary Figure S3C). As expected, given the difference in localization, ARS2n shared only 12% of its interactome with ARS2c isoforms. In contrast, 72% of the ARS2c1 interactome is shared by ARS2c2 (Figure 2C). Indeed ARS2c isoforms share most of their top 10 interactors (Figure 2D). 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 (Figure 2E). The ARS2c interactome, on the other hand, is enriched in proteins that participate in the biological processes of nitrogen compound metabolic processes, cellular localization, and cellular response to stress (Figure 2F). This last group is intriguing since the *Ars2* gene has been reported to confer sensitivity to the cellular stressor arsenic (3).

ARS2c and the cellular response to arsenic stress

Arsenic treatment suppresses global transcription and translation, and in our case caused a notable shift in the proteome of cells and a dramatic reduction of protein diversity. After arsenic treatment, there is an accumulation of proteins between 50 and 75 kDa (Figure 3A). 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: NCBP1 and UPF1, are downregulated (Figure 3B, C). The induction of ARS2c expression appears to be at the transcription or RNA stability level and was inhibited by siRNA against ARS2c (Figure 3D). Full sequences of RNAi and shRNAs used in the study, validation, and targeting diagrams are included in Supplementary Table S1, Supplementary Figures S4A–C and S2C. We tested other cellular stressors such as tunicamycin, serum starvation, and translation inhibition (Puromycin) (Supplementary Figure S5A–D). Interestingly, only puromycin treatment induced

ARS2c expression (Supplementary Figure S5C–E). Intriguingly, we also saw an increase in ARS2c expression when using early passage C2C12 cells and transfection with Lipofectamine (Figure 1E). 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.

To evaluate the importance of ARS2c upregulation during arsenic treatment, we knocked down ARS2c and examined cellular survival in response to arsenic, by measuring metabolically active cells using the WST-1 assay. Depletion of cytoplasmic ARS2 (*Ars2c* si, blue) significantly increases the resistance of C₂C₁₂ myoblast cells to arsenic treatment (Figure 3E). Depletion of all ARS2 isoforms (*Ars2all* si, gray), also increases the resistance of C₂C₁₂ to arsenic but to a lesser extent than ARS2c knockdown alone. Since ARS2c is similarly knocked down by *Ars2* si (all) and *Ars2c* si (Figure 3D), 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. We confirmed the effect of ARS2c on arsenic sensitivity using a crystal violet assay (Figure 3F, Supplementary Figure S5F).

Surprisingly, the ARS2c mediated arsenic response seems to be independent of the CBC complex. Downregulation of NCBP1 does not increase cellular survival following arsenic treatment, and co-depletion of ARS2c and NCBP1 results in an arsenic resistant phenotype comparable to depletion of ARS2c alone (Figure 3G, Supplementary Figure S5H, I). Additionally, *Ars2c* knockdown had no effect on the survival following puromycin treatment (Supplementary Figure S5G), indicating that the effect of ARS2c on arsenic sensitivity is not due to the ability of both arsenic and puromycin to block translation.

The wholesale changes in the proteome, which occur in response to arsenic, suggest that this change is important for the cellular response to the chemical. We 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 (Figure 3H). Previous studies implicating the role of ARS2 in arsenic sensitivity examined the effects of *Ars2* knockdown on sensitivity (3). 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. The effects of ARS2c on arsenic sensitivity appear to be independent of the CBC complex and the ability of arsenic to inhibit translation. Instead, ARS2c is upregulated in response to arsenic and is necessary for the cellular response and sensitivity to arsenic stress.

The role of ARS2n and ARS2c in NMD

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 shows that ARS2n and ARS2c isoforms share interactions with components of the mRNA-RNA catabolic process:

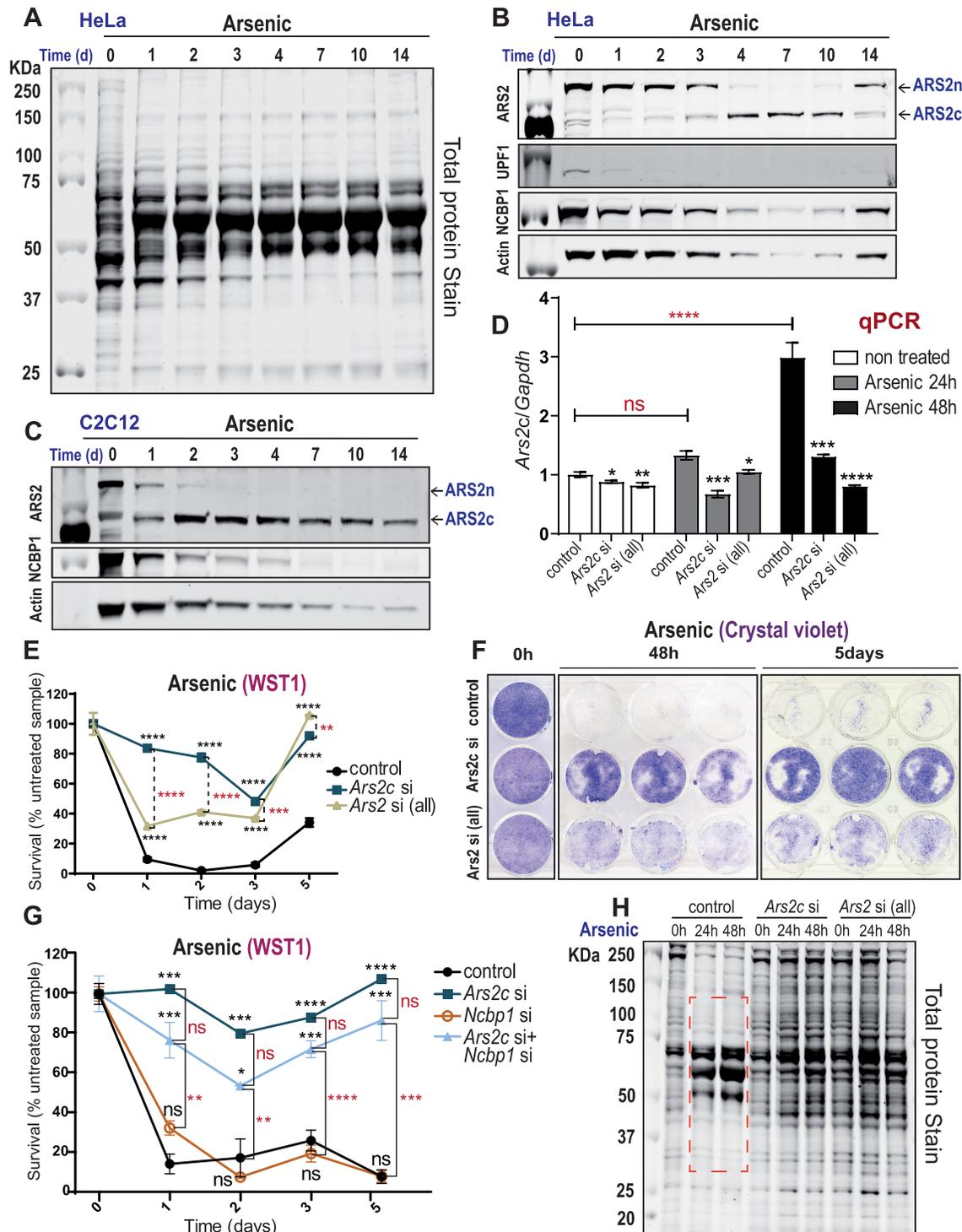


Figure 3. ARS2c is required for arsenic stress response. (A–C) HeLa and C₂C₁₂ cells were treated with 40μM arsenic trioxide for 14 days. Western blots of the cell lysates are shown with the indicated antibodies. Total protein was detected by Revert™ 700 Total Protein Stain. (D) C₂C₁₂ cells transfected with *Ars2c*/*Ars2*(all)/control RNAi were treated with arsenic trioxide for 48 h. *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). (E–G) C₂C₁₂ cells transfected with *Ars2c*/*Ars2*(all)/*Ncbp1*/*Ars2c* + *Ncbp1*/control RNAi were treated with arsenic trioxide (35μM/40μM) for 1–5 days. Survival (% treated/untreated sample) was measured by WST1 (E, G) or crystal violet (F). Data are represented as mean ± SEM (E, G) 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). (H) C₂C₁₂ cells were transfected with *Ars2c*/*Ars2*(all)/control RNAi and treated with arsenic trioxide for 48 h. Cell lysates were analyzed by western blot and total protein was detected as in (A). Each lane represents three pooled biologically independent samples.

NMD (Figure 4A, Supplementary Figure S6A, Supplementary Table S5). Consistent with previous reports, we detected proximity interactions of ARS2n with core EJC components MAGOH, MAGOHB and CASC3 (Figure 4B, left). In contrast, ARS2c isoforms are enriched for NMD components UPF1 and ERF1 (Figure 4B, right). We 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 (Figure 4C, D). From previous work, we know that ARS2n interacts with all four core components of the EJC: CASC3, MAGOH, RBM8A and EIF4A3, the EJC peripheral factors: UPF3B, RNPS1, PNN and ACINUS as well as UPF1 and the phosphatases PPP2R1A and PPP2R2A (Figure 4E) (5,19,20,40–43). Additionally, the CBC complex (NCBP1/2) has been shown to stabilize the interaction between UPF1-ERF1/ERF3 and the EJC, promoting NMD (9,18). However, to our knowledge, no role for ARS2 in NMD has been described and this is the first report of specific interactors for ARS2c isoforms with this pathway.

To evaluate the role of ARS2n in NMD, we initially performed a tethering assay in which the protein of interest is fused with a λ N tag and binds to a 5x-B-Box in the 3' UTR of a luciferase reporter (Figure 4F, top, Supplementary Figure S6B). The tethering assay mimics the deposit of an EJC downstream of a stop codon and measures the ability of a protein of interest to recruit components of the NMD pathway and trigger decay of the reporter mRNA. Tethering of ARS2n to the luciferase reporter resulted in reporter degradation (Supplementary Figure S6C). To test whether the observed degradation was dependent on NMD, we knocked down *UPF1*, a key component of the NMD pathway. As shown in Figure 4F, knockdown of *UPF1* rescued the expression of the reporter, indicating that the effect of ARS2n is partially dependent on UPF1. ARS2n has well described roles in targeting RNA to the nuclear exosome (44,45) and our results do not exclude ARS2n participation in this process. However, the partial dependency of the reporter degradation on NMD component UPF1, suggests that in addition to its role in exosomal degradation, ARS2n also participates in NMD. Although the tethering assay suggests a role for ARS2n on NMD, it does not indicate the function of ARS2n in the pathway. The assay forces the interaction of ARS2n with the 3'UTR of the reporter and simply tests the ability of the protein to recruit NMD components to the 3'UTR to trigger 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.

To gain a better understanding of the role of the nuclear and cytoplasmic ARS2 isoforms in NMD, we used a NMD dual fluorescence reporter (26). 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– vec-

tor is a negative control. In this plasmid, exon 2 is only 35 nucleotides long. After splicing, the distance between the termination codon of TagGFP2 and the EJC will be <50 nucleotides and therefore insensitive to NMD. To normalize for transfection efficiency, the vectors also contain an expression cassette for the far-red protein *Katushka* (TurboFP635) (Figure 5A, B, Supplementary Figure S7A–C). For each sample, TagGFP2 is normalized to *Katushka* and then expressed as the ratio of pNMD– ve/pNMD+ ve. The pNMD– ve/pNMD+ ve ratio represents NMD activity, where higher ratio corresponds to higher 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 system, we found that ARS2c and ARS2n have opposing impacts on NMD. Specific knockdown of *ARS2c*, promotes a decrease in the NMD activity, as measured by flow cytometry quantification of TagGFP2 protein, and RT-qPCR quantification of the TagGFP2 transcript at 48 h (Figure 5C, Supplementary Figure S7F). A similar effect was observed with downregulation of UPF1 (Figure 5E, Supplementary Figure S7H), suggesting that ARS2c isoforms are required for NMD. We confirmed the role of ARS2c in NMD is conserved in mouse cells by testing this reporter assay in C₂C₁₂ cells. A similar NMD activity decrease was seen in these cells (Supplementary Figure S7D). There are no sequences unique to *ARS2n* to specifically target this isoform by RNAi. Thus, to test the impact of ARS2n on NMD we used RNAi to exon 4, which disrupts expression of both *ARS2n* and *ARS2c* (Supplementary Figure S2C). 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 (Figure 5D, Supplementary Figure S7G). 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, we 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 the endogenous NMD targets (Figure 5F), an effect that was counteracted by downregulation of UPF1 (Supplementary Figure S7I). Conversely, both UPF1 and ARS2c depletion inhibited the activity of the NMD pathway and favored the accumulation of endogenous NMD targets (Figure 5H, Supplementary Figure S7J). 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 (Figure 5G). Consistent with an inhibitory role, downregulation of ARS2n increased NMD activity and promoted the degradation of endogenous NMD targets (Supplementary Figure S7K). These results were observed in both human HeLa and mouse C₂C₁₂ cells, confirming the effects of the isoforms on endogenous targets are conserved between the species.

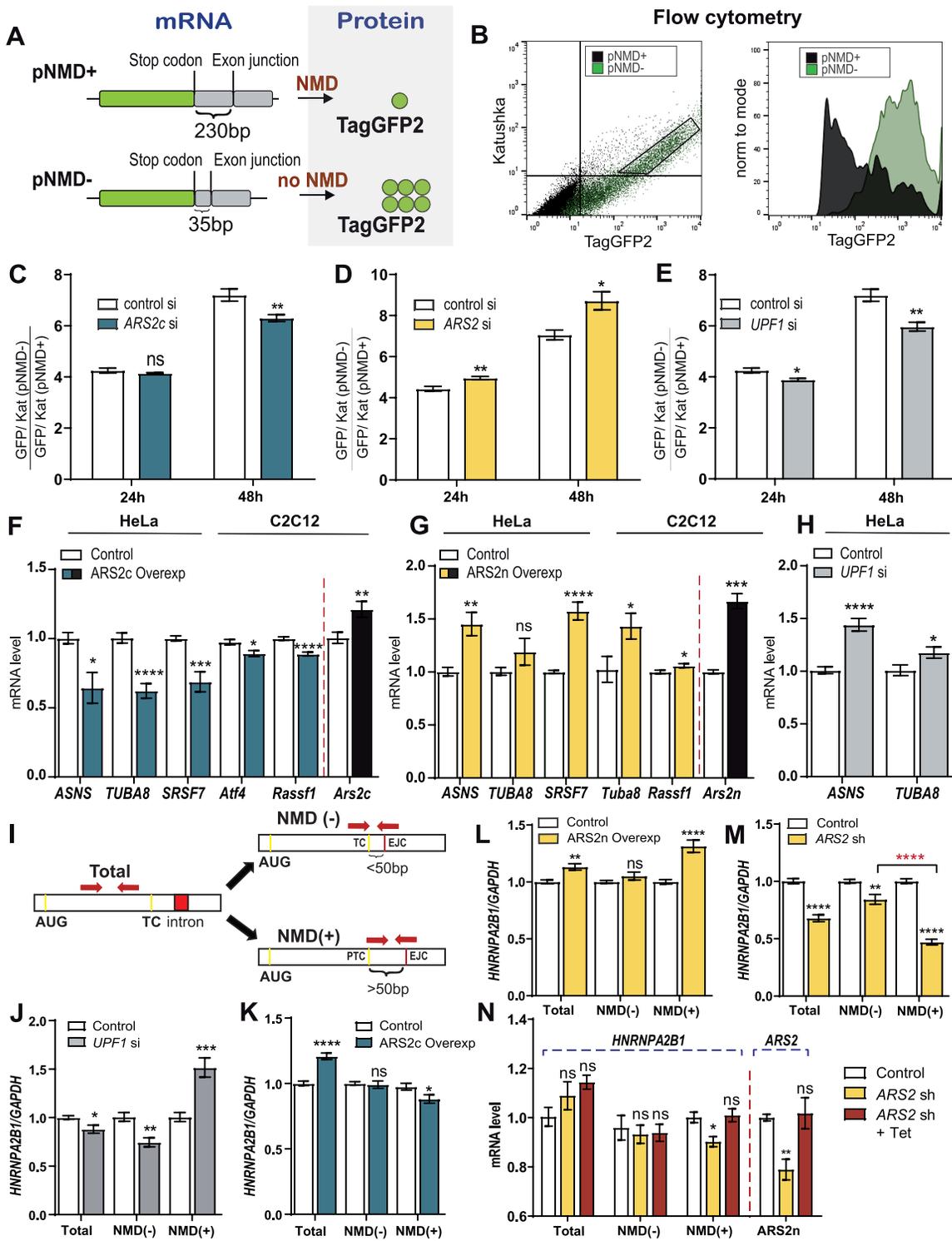


Figure 5. The role of ARS2n and ARS2c in NMD. (A) Schematic representation of pNMD+/pNMD- reporters. pNMD+ is degraded by NMD, while pNMD- is insensitive to NMD degradation. (B) Flow cytometry showing pNMD+ degradation and reporter functionality. (C-E) HeLa cells were transfected with pNMD+/pNMD- and ARS2c/ARS2/UPF1 or control RNAi. Expression levels were detected by Flow cytometry. TagGFP2 levels were normalized to Katushka, 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 (*P ≤ 0.05; **P ≤ 0.01). (F-M) RT-qPCR performed after transfection of HeLa or C2C12 cells with ARS2c-GFP/ARS2n-GFP/GFP or ARS2sh/UPF1si/negative controls for 48 h. Each gene was normalized against GAPDH. (I) Schematic representation of primers used (J-N). (N) RT-qPCR of Flp-In T-REx 293 cells inducibly expressing ARS2n were transfected with ARS2 sh or control for 48h and treated with 2.27 ng/ml of tetracycline for 36 h to induce ARS2n to endogenous levels. Data are represented as mean ± SEM for n = 3 to n = 6 biologically independent samples (see Materials and Methods section). Statistical analysis (F-M) two-tail unpaired t-test (*P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001, ****P ≤ 0.0001), (N) one way ANOVA, post-test Dunnett's (*P ≤ 0.05; **P ≤ 0.01).

We next demonstrated that the effect of ARS2 isoforms on the pathway was NMD dependent and not due to off-targeting. We used a set of primers to amplify alternatively spliced transcripts of the *HNRNPA2AB1* gene (46) in HeLa or in a Flp-In T-REx 293 inducible cell line, in which 3xFlag-ARS2n has been knocked into an *FRT* locus within the cell line. The endogenous *HNRNPA2AB1* gene generates two alternative spliced transcripts represented here as NMD⁻ and NMD⁺ (Figure 5I). Similar to the pNMD^{-/+} reporter, 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 (Figure 5J). Supporting its NMD promoting role, overexpression of ARS2c preferentially induces the degradation of the NMD⁺ transcript (Figure 5K). In contrast, overexpression of ARS2n promotes the accumulation of the NMD⁺ transcript, while downregulation of ARS2n preferentially induces NMD⁺ degradation (Figure 5L, M). To confirm that the effects of *ARS2n* knockdown on the pathway were specific to *ARS2n*, we repeated the experiment but induced expression of 3xFlag-ARS2n with the addition of tetracycline. Restoring *ARS2n* levels rescued the expression of the *HNRNPA2AB1* NMD⁺ transcript, showing that the inhibitory effect of *ARS2n* is specific. (Figure 5N). In summary, these results show ARS2n inhibits while ARS2c promotes the NMD pathway.

ARS2 isoforms 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. We narrowed the nuclear localization signal region from 105 amino acids (45) to 12, and generated an ARS2n mutant which lacks amino acids (73-LSPQKRMRRD-84). As shown in Figure 6A, 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 (Figure 6B).

To evaluate the mechanism of the isoforms and mutant Δ nls-ARS2n on NMD, we next looked at their ability to interact with endogenous 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 Δ NLS mutant (Figure 6C). To test whether the interaction with UPF1 takes place in the cytoplasm, we performed endogenous UPF1 pull-downs on cytoplasmic fractions of cell lysates. Both ARS2 isoforms and the mutant were detected in cytoplasmic UPF1 pull-downs (Supplementary Figure S8A). We adjusted protein concentrations (Supplementary Figure S8B, C) and compared ARS2 isoforms/mutant expression in the input and the supernatant (SN), obtained after UPF1 pull-down. Consistent with Figure 6C, 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 (Supplementary Figure S8D).

Furthermore, ARS2c interaction with UPF1 potentially affects the cytoplasmic binding of UPF1 to NCBP1. As shown in Figure 6D, expression of ARS2c, but not ARS2n or the mutant, promoted the cytoplasmic binding of endogenous UPF1 to NCBP1. Since NCBP1 plays a critical role in NMD induction, these results suggest that ARS2c affects NMD through NCBP1.

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. As expected, ARS2c (like ARS2n) interacts with the CBC component NCBP1 (Figure 6E). To explore the mechanism of ARS2c promotion of NMD, we examined whether ARS2c interacts with other components of the NMD SURF complex. ARS2c interacts with SURF complex components ERF1 (previously detected in our BioID data) and SMG1. As was observed with the ARS2c and UPF1 interaction, ARS2c is more strongly associated with ERF1, NCBP1 and SMG1 than either ARS2n or the mutant (Figure 6F). This data suggests that ARS2c promotes NMD through NCBP1 and the SURF complex.

To test this idea, we first evaluated the effect of ARS2 isoforms on UPF1 phosphorylation. ARS2 isoforms differentially affected the phosphorylation of UPF1: overexpression of ARS2n increased UPF1 phosphorylation, while ARS2c and the Δ NLS mutant overexpression decreased UPF1 phosphorylation (Figure 7A, B, Supplementary Figure S9A–D). We 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 NCBP1, ERF1 and SMG7, effect that was not observed during overexpression of either ARS2n or the Δ NLS mutant (Figure 7C, D, Supplementary Figure S10A–C). Interestingly, despite its cytoplasmic localization the Δ NLS mutant behaved like ARS2n with respect to UPF1/ERF1/NCBP1/SMG1 binding and the inability to promote the association of UPF1 with NMD components (Figures 6C, D, F, 7C, D, Supplementary Figures S8A–E, S10A–C). These results suggest the Δ NLS 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.

DISCUSSION

In the current study, we have identified dedicated cytoplasmic isoforms of ARS2 in human and mouse cell lines that function in the regulation of NMD, and in the cellular response to arsenic induced stress. We propose a model in which ARS2n and ARS2c work in tandem to regulate NMD (Figure 7E). As previously described, ARS2n primarily functions in the nucleus in conjunction with the CBC to promote mRNA splicing, degradation and export (1–7). 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 that enhances SURF complex formation and transcript degradation. While we did not test the order of these interactions, our data shows that ARS2c interacts

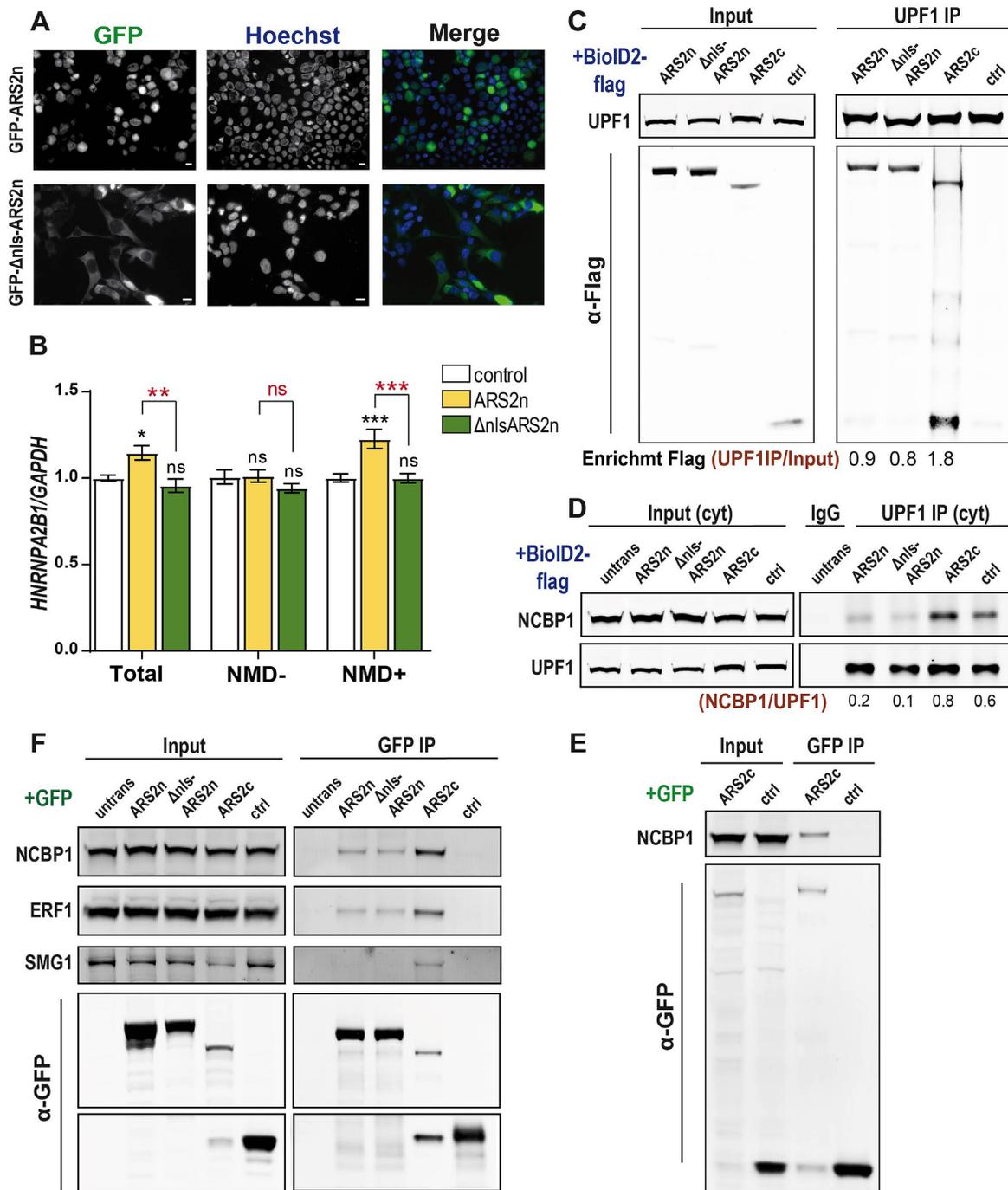


Figure 6. ARS2 isoforms are functionally distinct. (A) HEK 293T cells were transfected with GFP-ARS2n/ GFP-Δnls-ARS2n. GFP fusion proteins are shown in green and nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μM. (B) RT-qPCR performed after transfection of HeLa cells with GFP-ARS2n/ GFP-Δnls-ARS2n or GFP control for 48 h. *HNRNPA2B1* gene isoforms were detected using Figure S1 primers. 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 ($*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$). (C) HEK 293T cells were transfected with BioID2-ARS2n/Δnls-ARS2n/ARS2c/control-3xflag for 24 h. Endogenous UPF1 was pulled down with anti-UPF1 antibodies. 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. (D) HEK 293T cells were transfected with BioID2-ARS2n/Δnls-ARS2n/ARS2c/control-3xflag for 48 h. Cytoplasmic fractions were extracted and endogenous UPF1 pulled down with an anti-UPF1 antibody. Cytoplasmic fraction of untransfected cells, pulled down with IgG mouse control antibody were used as a control. NCBP1 and UPF1 were detected using anti-NCBP1 and anti-UPF1 antibodies respectively. Expression levels of the prey (NCBP1) in the IP were normalized to immunoprecipitated UPF1 and represented as NCBP1/UPF1 ratio. (E) HEK 293T cells were transfected with GFP-ARS2c/ GFP control. Samples were pulled down using anti-GFP beads and GFP-ARS2c/ GFP control and NCBP1 were detected with anti-GFP and anti-NCBP1 antibodies, respectively. (F) 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.

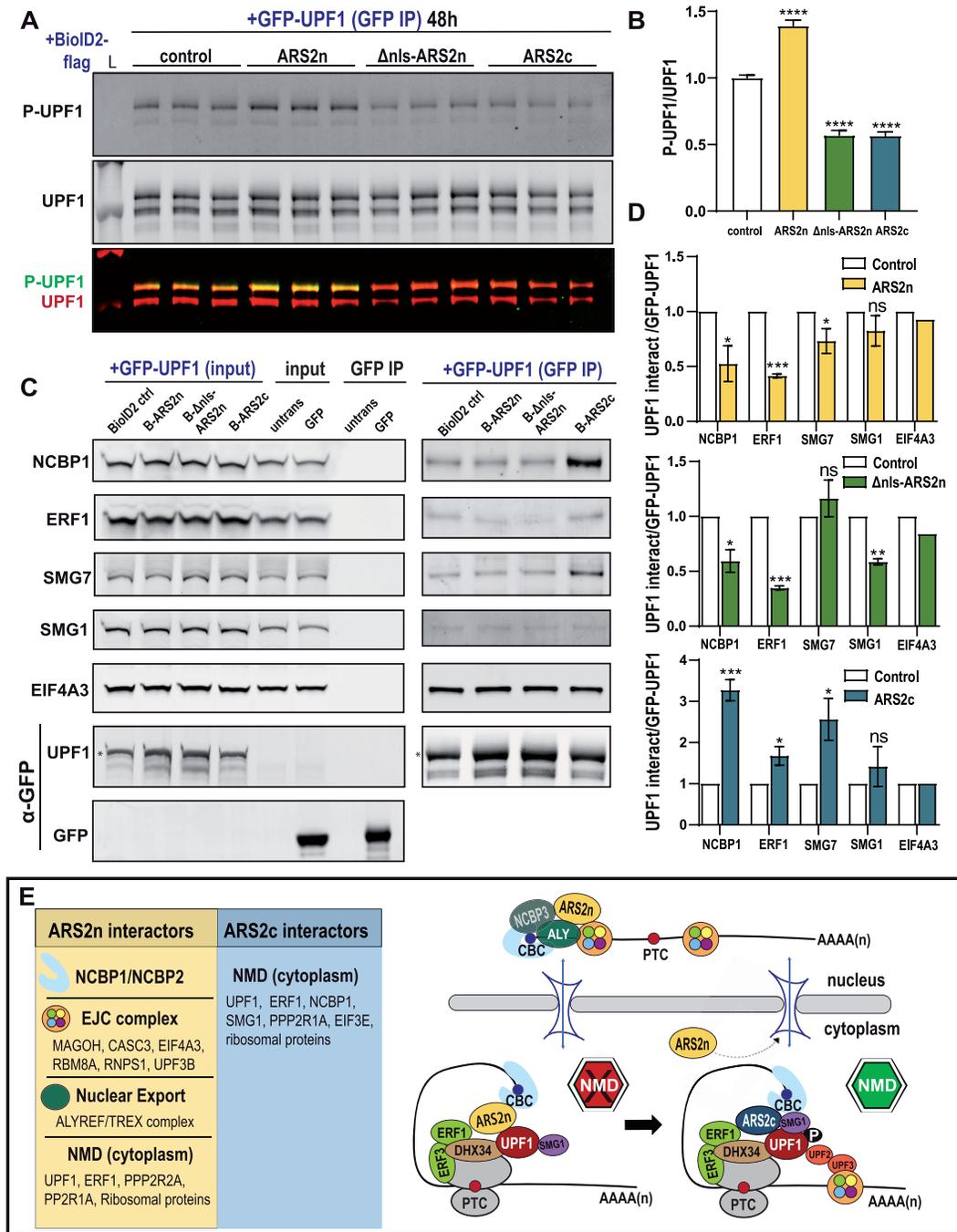


Figure 7. ARS2 isoforms work in tandem to regulate NMD. (A, C) HEK 293T cells were co-transfected with BioID2-ARS2n/Δnls-ARS2n/ARS2c/control-3xflag and GFP-UPF1. GFP-UPF1 was pulled down using anti-GFP beads and specified proteins were detected in GFP-UPF1 pull-downs. Full gels images of anti-GFP and anti-Flag detections are included in Supplementary Figure S10A. (A) 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 (B) for $n = 3$ biologically independent samples, repeated in three independent western blots. Statistical analysis: one-way ANOVA, post-test Dunnett's ($****P \leq 0.0001$). (C) 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 (D). 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 ($*P \leq 0.05$; $**P \leq 0.01$; $***P \leq 0.001$). (E) ARS2n and ARS2c role in NMD (model). ARS2n, bound to NCBP1/2/3, 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 NCBP1, UPF1, ERF1 and SMG1, enhances the binding of UPF1 to NCBP1/ERF1, promoting the formation of the SURF complex. The presence of an EJC more than 50 nt downstream of the termination codon favors the formation of the DECID complex and transcript degradation through NMD. Left panel: ARS2n and ARS2c interactors reported in our study and/or the literature.

with NCBP1/UPF1/ERF1/SMG1 and promotes the interaction between UPF1 and NCBP1/ERF1, favoring the formation of the SURF complex. Since ARS2c contains the CBC binding site within its C-terminus, these interactions and effect on the SURF complex is likely to be mediated through the CBC. Consistent with the ability of ARS2c to promote NMD, ARS2c expression affects UPF1 phosphorylation, and promotes SMG7 recruitment and degradation of PTC containing transcripts. Our model is consistent with previous literature showing that NCBP1 promotes both UPF1 interaction with ERF1 and ERF3 and the association of UPF1 with the EJC (9,18) 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. ARS2 isoform switching tailors the protein to function in a nonredundant manner in nuclear and cytoplasmic CBC-dependent processes.

BioID is a widely used method that allows the detection of transient, weak and temporally separated interactions (33,47). 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 significant enrichment of NCBP1 or NCBP2 in our data. We verified that NCBP2 was present in BioID2-ARS2n-3flag immunoprecipitations, confirming that the BioID2 tag did not affect ARS2n binding to NCBP1/2. Interestingly, we found a significant enrichment of NCBP3 in the ARS2n data. NCBP3 has been shown to interact with CBC-ARS2 and the EJC post-splicing, and functions to promote multi-exonic polyadenylated mRNA export (5,43). The presence of NCBP3, as well as PHAX, in this dataset suggests that the inability to detect NCBP1 and NCBP2 may be due to steric constraints from the placement of the biotin ligase at the N-terminus. Thus, consistent with other studies, our data illustrates that BioID is a complementary approach to AP-MS and does not replace it.

We identified the NMD pathway as one of the enriched biological processes that is shared between ARS2n and ARS2c interactomes. Using a tethering assay, NMD reporters, RT-qPCR of endogenous targets and both gain-of-function and loss-of-function experiments, we found that ARS2n inhibits while ARS2c promotes NMD. The role of ARS2n appears to be primarily in the nucleus through its interactions with the EJC and export machinery, whereas ARS2c interacts with SURF complex components, and promotes the binding of UPF1 to NCBP1/ERF1/SMG7 in the cytoplasm (Figure 6C–F, 7C, D, Supplementary Figures S8A–E, S10A–C). 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, we created a cytoplasmic version of ARS2n in which we deleted the NLS but left the N-terminal leg and remaining unstructured N-terminus intact. We show that deletion of the nuclear localization signal is sufficient to confer cytoplasmic localization to the full-

length isoform and abrogates the nuclear isoform's ability to inhibit NMD (Figure 6A, B). Despite the cytoplasmic localization, the Δ NLS mutant behaved like ARS2n with respect to UPF1/ERF1/NCBP1/SMG1 binding and the ability to promote the association of UPF1 with ERF1 and NCBP1 (Figure 6C, D, F, Figure 7C, D, Supplementary Figures S8A–E, S10A–C). Unlike ARS2n however, expression of Δ NLS mutant did not increase UPF1 phosphorylation (Figure 7A, B, Supplementary Figure S9A–D). Consistent with this finding, the Δ NLS mutant decreased the binding of UPF1 to SMG1, the kinase responsible for UPF1 phosphorylation (Figure 7D, Supplementary Figure S10C). 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 (20). 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 (24). 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 (48,49). 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.

ARS2 switching is important for the proteomic remodeling and cell sensitivity in response to arsenic treatment. Our findings show that ARS2c is upregulated in response to arsenic while ARS2n is downregulated. Both proteomic remodeling and arsenic sensitivity were lost upon ARS2c knockdown. 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. NCBP1 expression, like UPF1 or ARS2n, decreased following arsenic treatment, and depletion of NCBP1 had no effect on the survival of arsenic treated cells. Furthermore, co-depletion of ARS2c and NCBP1 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, we 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 (Sup-

plementary Figure S5G). It is also unlikely that the sensitivity to arsenic conferred by ARS2c is due to its role in NMD, since both UPF1 and NCBP1 are downregulated in response to arsenic treatment (Figure 3B, C). Mechanistically, arsenic generates genotoxic reactive oxygen species, ER stress, promotes the induction of the unfolded protein response (UPR) and triggers p53-regulated mitochondrial apoptosis (50). Interestingly, our BioID data showed interaction between ARS2c, and proteins involved in cell redox homeostasis, protein folding, and response to unfolded protein (Figure 4A, Supplementary Figure S6A). How such interactions regulate the cellular response to arsenic will require further study.

In summary, our study constitutes the first evaluation of multiple ARS2 isoforms in mammalian cells. By comparing differential and shared biological processes, our results show that ARS2 isoforms are non-redundant and work in tandem to regulate mRNA metabolism within the nucleus and cytoplasm. ARS2c promotes transcript degradation through NMD, and arsenic sensitivity, in CBC-dependent and CBC-independent manner, respectively. Furthermore, the involvement of ARS2c in arsenic sensitivity could have implications for cancer treatment. Arsenic trioxide is used in the treatment of 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 be useful in predicting patient response to arsenic treatment.

DATA AVAILABILITY

All codes used in this study are publicly available. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (51) partner repository with the dataset identifier PXD026453. The Flow cytometry data was deposited on the FlowRepository with the identifier FR-FCM-Z3V5.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Darryl Hardie and Jason Serpa and the staff of the UVIC Genome BC Proteomics Centre for the support in the LC/MS-MS data acquisition and analysis. We thank Ryan Erdman and Scott Scholz from BCMB department Biotech Support for equipment assistance. pNMD+ and pNMD- reporter plasmids were kindly provided by K. Lukyanov (26). RNT1-GFP was a gift from H. Dietz (Addgene plasmid # 17708) (27). pAc5.1C-FLuc-Stop-5BoxB was a gift from Elisa Izaurralde (Addgene plasmid # 21301) (28). CBP20-3flag was a gift from Torben Heick Jensen and John LaCava (29). Anti-ARS2 (XL14.1) was generously provided by the Ludwig Institute for Cancer Research.

FUNDING

NSERC Discovery Grant [NSERC DG-34656 to P.L.H.]. Funding for open access charge: NSERC [DG-34656].
Conflict of interest statement. None declared.

REFERENCES

- Müller-Mcnicoll, M. and Neugebauer, K.M. (2014) Good cap /bad cap: how the cap-binding complex determines RNA fate. *Nat. Publ. Gr.*, **21**, 9–12.
- Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Ratsch, G. and Weigel, D. (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 8795–8800.
- Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., Zong, W.X., Zhang, Z., Lau, C.K., Rawlings, J. *et al.* (2009) Ars2 links the nuclear cap-binding complex to rna interference and cell proliferation. *Cell*, **138**, 328–339.
- Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, N.E.H., Gostan, T., Vandermoere, F., Robert, M.C., Cusack, S. *et al.* (2013) CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.*, **20**, 1358–1366.
- Dou, Y., Barbosa, I., Jiang, H., Iasillo, C., Molloy, K.R., Schulze, W.B., Cusack, S., Schmid, M., Le Hir, H., LaCava, J. *et al.* (2020) NCBP3 positively impacts mRNA biogenesis. *Nucleic Acids Res.*, **48**, 10413–10427.
- Giacometti, S., Benbahouche, N.E.H., Domanski, M., Robert, M.C., Meola, N., Lubas, M., Bukenborg, J., Andersen, J.S., Schulze, W.M., Verheggen, C. *et al.* (2017) Mutually exclusive cbc-containing complexes contribute to RNA fate. *Cell Reports*, **18**, 2635–2650.
- Gromadzka, A.M., Steckelberg, A., Singh, K.K., Hofmann, K. and Gehring, N.H. (2016) A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res.*, **44**, 2348–2361.
- Choe, J., Oh, N., Park, S., Lee, Y.K., Song, O., Locker, N., Chi, S.G. and Kim, Y.K. (2012) 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.
- Hosoda, N., Kim, Y.K., Lejeune, F. and Maquat, L.E. (2005) CBP80 promotes interaction of upf1 with upf2 during nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.*, **12**, 893–901.
- Popp, M.W.L. and Maquat, L.E. (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.*, **47**, 139–165.
- Nickless, A., Bailis, J.M. and You, Z. (2017) Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell. Biosci.*, **7**, 26.
- Le Hir, H., Izaurralde, E., Maquat, L.E. and Moore, M.J. (2000) The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.*, **19**, 6860–6869.
- Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M.J. (2001) 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.
- Nott, A., Le Hir, H. and Moore, M.J. (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev.*, **18**, 210–222.
- Le Hir, H., Saulière, J. and Wang, Z. (2016) The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.*, **17**, 41–54.
- Isken, O. and Maquat, L.E. (2008) The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.*, **9**, 699–712.
- Durand, S., Franks, T.M. and Lykke-Andersen, J. (2016) Hyperphosphorylation amplifies UPF1 activity to resolve stalls in nonsense-mediated mRNA decay. *Nat. Commun.*, **7**, 12434.
- Hwang, J., Sato, H., Tang, Y., Matsuda, D. and Maquat, L.E. (2010) UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. *Mol. Cell*, **39**, 396–409.
- Thillainadesan, G., Xiao, H., Holla, S., Dhakshnamoorthy, J., Jenkins, L.M.M., Wheeler, D. and Grewal, S.I.S. (2020) Conserved protein pir2ars2 mediates gene repression through cryptic introns in lncRNAs. *Nat. Commun.*, **11**, 2412.
- Schulze, W.M., Stein, F., Rettel, M., Nanao, M. and Cusack, S. (2018) Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting. *Nat. Commun.*, **9**, 2183.

21. Iland, H.J. and Seymour, J.F. (2013) Role of arsenic trioxide in acute promyelocytic leukemia. *Curr. Treat. Options Oncol.*, **14**, 170–184.
22. Rossman, T.G. and Wang, Z. (1999) Expression cloning for arsenite-resistance resulted in isolation of tumor-suppressor *faucDNA*: possible involvement of the ubiquitin system in arsenic carcinogenesis. *Carcinogenesis*, **20**, 311–316.
23. Dilworth, D., Upadhyay, S.K., Bonnafous, P., Edo, A.B., Bourbigot, S., Pesek-Jardim, F., Gudavicius, G., Serpa, J.J., Petrochenko, E.V., Borchers, C.H. *et al.* (2017) The basic tilted helical bundle domain of the prolyl isomerase FKBP25 is a novel double-stranded RNA binding module. *Nucleic Acids Res.*, **45**, 11989–12004.
24. O'Sullivan, C., Christie, J., Pienaar, M., Gambling, J., Nickerson, P.E.B., Alford, S.C., Chow, R.L. and Howard, P.L. (2015) Mutagenesis of ARS2 domains to assess possible roles in cell cycle progression, MicroRNA and replication-dependent histone mRNA biogenesis. *Mol. Cell. Biol.*, **35**, 3753–3767.
25. Wilson, M.D., Wang, D., Wagner, R., Breysens, H., Gertsenstein, M., Lobe, C., Lu, X., Nagy, A., Burke, R.D., Koop, B.F. *et al.* (2008) ARS2 is a conserved eukaryotic gene essential for early mammalian development. *Mol. Cell. Biol.*, **28**, 1503–1514.
26. Pereverzev, A.P., Gurskaya, N.G., Ermakova, G.V., Kudryavtseva, E.I., Markina, N.M., Kotlobay, A.A., Lukyanov, S.A., Zaraisky, A.G. and Lukyanov, K.A. (2015) Method for quantitative analysis of nonsense-mediated mRNA decay at the single cell level. *Sci. Rep.*, **5**, 7729.
27. Mendell, J.T., Rhys, C.M.J. and Dietz, H.C. (2002) Separable roles for *rent1/hUpf1* in altered splicing and decay of nonsense transcripts. *Science*, **298**, 419–422.
28. Rehwinkel, J., Behm-ansmant, I., Gatfield, D. and Izaurralde, E. (2005) A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA*, **11**, 1640–1647.
29. Domanski, M., Molloy, K., Jiang, H., Chait, B.T., Rout, M.P., Jensen, T.H. and LaCava, J. (2012) Improved methodology for the affinity isolation of human protein complexes expressed at near endogenous levels. *BioTechniques*, **0**, 1–6.
30. Roux, K.J., Kim, D.I. and Burke, B. (2013) BioID: a screen for protein-protein interactions. *Curr. Protoc. Protein Sci.*, **91**, 19.23.1–19.23.15.
31. Senko, M.W., Remes, P.M., Canterbury, J.D., Mathur, R., Song, Q., Eliuk, S.M., Mullen, C., Earley, L., Hardman, M., Blethrow, J.D. *et al.* (2013) Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal. Chem.*, **85**, 11710–11714.
32. Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-kirby, A. and Lajoie, G. (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, **17**, 2337–2342.
33. Lambert, J.P., Tucholska, M., Go, C., Knight, J.D.R. and Gingras, A.C. (2015) Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J. Proteomics*, **118**, 81–94.
34. Youn, J.Y., Dunham, W.H., Hong, S.J., Knight, J.D.R., Bashkurov, M., Chen, G.I., Bagci, H., Rathod, B., MacLeod, G., Eng, S.W.M. *et al.* (2018) High-density proximity mapping reveals the subcellular organization of mRNA-Associated granules and bodies. *Mol. Cell*, **69**, 517–532.
35. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages, F., Trajanoski, Z. and Galon, J. (2009) ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, **25**, 1091–1093.
36. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **11**, 2498–2504.
37. Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M. *et al.* (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, 412–416.
38. Gurskaya, N.G., Pereverzev, A.P., Staroverov, D.B., Markina, N.M. and Lukyanov, K.A. (2016) Analysis of nonsense-mediated mRNA decay at the single-cell level using two fluorescent proteins. In: Filonov, G.S. and Jaffrey, S.R. (eds). *Methods in Enzymology*. Academic Press, Vol. **572**, pp. 291–314.
39. Gingras, A., Abe, K.T. and Raught, B. (2019) Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Curr. Opin. Chem. Biol.*, **48**, 44–54.
40. Tange, T., Shibuya, T., Jurica, M.S. and Moore, M.J. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA*, **11**, 1869–1883.
41. Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O' Connor, L., Li, M. *et al.* (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.*, **3**, 89.
42. Mabin, J.W., Woodward, L.A., Patton, R.D., Yi, Z., Jia, M., Wysocki, V.H., Bundschuh, R. and Singh, G. (2018) The exon junction complex undergoes a compositional switch that alters mRNP structure and nonsense-mediated mRNA decay activity. *Cell Rep.*, **25**, 2431–2446.
43. Dou, Y., Kalmykova, S., Pashkova, M., Oghbaie, M., Jiang, H., Molloy, K.R., Chait, B.T., Rout, M.P., Fenyó, D., Jensen, T.H. *et al.* (2020) Affinity proteomic dissection of the human nuclear cap-binding complex interactome. *Nucleic Acids Res.*, **48**, 10456–10469.
44. Andersen, P.R., Domanski, M., Kristiansen, M.S., Storvall, H., Ntini, E., Verheggen, C., Schein, A., Bunkenborg, J., Poser, I., Hallais, M. *et al.* (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Publ. Gr.*, **20**, 1367–1376.
45. Melko, M., Winczura, K., Rouviere, J.O., Oborski, M., Andersen, P.K. and Jensen, T.H. (2019) Mapping domains of ARS2 critical for its RNA decay. *Nucleic Acids Res.*, **48**, 6943–6953.
46. McGlinchy, N.J., Tan, L.Y., Paul, N., Zavolan, M., Lilley, K.S. and Smith, C.W.J. (2010) 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**, 565.
47. Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.P., St-Denis, N.A., Li, T., Miteva, Y.N., Hauri, S., Sardi, M.E., Low, T.Y. *et al.* (2013) The CRAPome: a contaminant repository for affinity purification – mass spectrometry data. *Nat. Methods*, **10**, 730–736.
48. Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V. and Sullivan, M. (2012) 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.
49. Amanchy, R., Zhong, J., Molina, H., Chaerkady, R., Iwahori, A., Kalume, D.E., Gronborg, M., Joore, J., Cope, L. and Pandey, A. (2008) Identification of c-Src tyrosine kinase substrates using mass spectrometry and peptide microarrays. *J. Proteome Res.*, **7**, 3900–3910.
50. Srivastava, R.K., Li, C., Ahmad, A., Abrams, O., Gorbatyuk, M.S., Harrod, K.S., Wek, R.C., Afaq, F. and Athar, M. (2016) ATF4 regulates arsenic trioxide-mediated NADPH oxidase, ER-mitochondrial crosstalk and apoptosis. *Arch. Biochem. Biophys.*, **609**, 39–50.
51. Perez-riverol, Y., Csordas, A., Bai, J., Bernal-linares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M. *et al.* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.*, **47**, 442–450.
52. Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P. *et al.* (2019) STRING v11: protein – protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, **47**, 607–613.