Published online 28 November 2016 Nucleic Acids Research, 2017, Vol. 45, No. 6 **3473–3486**
doi: 10.1093/nar/gkw1140

An Ultraconserved Element (UCE) controls homeostatic splicing of *ARGLU1* mRNA

Stephan P. Pirnie, Ahmad Osman, Yinzhou Zhu and Gordon G. Carmichael*

Department of Genetics and Genome Sciences, UCONN Health Center, 400 Farmington Avenue, Farmington, CT 06030, USA

Received January 13, 2016; Revised October 25, 2016; Editorial Decision October 28, 2016; Accepted October 31, 2016

ABSTRACT

Arginine and Glutamate-Rich protein 1 (ARGLU1) is a protein whose function is poorly understood, but may act in both transcription and pre-mRNA splicing. We demonstrate that the *ARGLU1* gene expresses at least three distinct RNA splice isoforms – a fully spliced isoform coding for the protein, an isoform containing a retained intron that is detained in the nucleus, and an isoform containing an alternative exon that targets the transcript for nonsense mediated decay. Furthermore, *ARGLU1* contains a long, highly evolutionarily conserved sequence known as an Ultraconserved Element (UCE) that is within the retained intron and overlaps the alternative exon. Manipulation of the UCE, in a reporter minigene or via random mutations in the genomic context using CRISPR/Cas9, changed the splicing pattern. Further, overexpression of the ARGLU1 protein shifted the splicing of endogenous *ARGLU1* mRNA, resulting in an increase in the retained intron isoform and nonsense mediated decay susceptible isoform and a decrease in the fully spliced isoform. Taken together with data showing that functional protein knockout shifts splicing toward the fully spliced isoform, our data are consistent with a model in which unproductive splicing complexes assembled at the alternative exon lead to inefficient splicing and intron retention.

INTRODUCTION

Splicing of RNA transcripts is a complex process in which primary RNA transcripts are assembled into mature, processed mRNAs through interactions with the spliceosome as well as other proteins, notably SR and hnRNP proteins (1). Alternative splicing is the process by which multiple mRNA isoforms of a precursor transcript may be generated by choosing to use, or not use, different splice sites. These choices are driven by the interactions between *cis*-RNA sequences and *trans*-acting proteins that can promote or inhibit the usage of a particular splice site (1,2). The expression of *trans*-acting factors in a developmental and tissue specific manner results in regulated splicing that is often cell specific. Most notably, *trans*-acting proteins such as the NOVA (3–6), the RBFOX (7) and SR protein (8–11) families, and a number of hnRNP (12,13) proteins compete to bind nascent RNAs at specific motifs and drive regulation of alternative splicing in a tissue and developmentally specific manner. Alternative splicing is an important process that is seen in at least 95% of multi-exon genes in the human transcriptome (14). Furthermore, alternative splicing can serve multiple functions, including increasing proteome diversity, regulating mRNA localization and changing mRNA stability or turnover.

One specific type of alternative splicing (AS) that has been poorly understood is alternative intron retention. These events are characterized by the inclusion of one or more ‘introns’ in an otherwise mature (capped and polyadenylated) mRNA. Functionally, this can lead to a number of outcomes for these transcripts, including alternative protein coding potential, nuclear localization (15) or degradation through the nonsense mediated decay (NMD) pathway (16,17). Retained intron events are relatively rare in animals, but are a common type of AS in plants (18), unicellular eukaryotes (19,20) and viruses (21). A recent study has described a class of transcripts labeled detained introns (15). These are polyadenylated transcripts that contain one or more intronic sequences, are localized to the nucleus and are stable and not subject to NMD (15). Another recent study revealed that retained introns, like other alternative splicing events, occur in a cell- and tissue-dependent manner and allow for regulation of a large number of transcripts in mammals (22). Other recent studies have found that regulation of intron retention is important for the expression of genes involved in neuron (23) and granulocyte (24) differentiation and during terminal erythropoiesis (25).

Animal models have been helpful in determining the function of many of the *trans*-acting proteins that have been found to regulate alternative splicing. However, the specific events that are evaluated can only be applied to an understanding of human biology if there is event level conservation between humans and the model in question

*To whom correspondence should be addressed. Tel: +1 860 679 2259; Fax: +1 860 679 8345; Email: carmichael@uchc.edu

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

(26,27). It has been noted that alternative splicing events that are evolutionarily conserved are positively correlated with nucleotide level conservation surrounding these alternative splice sites (26). Indeed, a subset of DNA elements that are 100% conserved for over 200 bases between humans and rodents, termed Ultraconserved Elements (UCEs) (28), were found to be correlated with auto-regulatory splicing of SR and hnRNP splicing regulatory proteins, and these alternative splicing events are often associated with NMD (16,17), or with nuclear detention of intron containing isoforms (15,29).

This study focuses on the alternative splicing of arginine and glutamate-rich protein 1 (*ARGLU1*). *ARGLU1*, previously known as FLJ10154, is a gene with poorly defined cellular function(s). The most in depth study described ARGLU1's interaction with a protein called MED1, a component of the mediator complex (30). The mediator complex functions to bring the transcriptionally poised RNA pol II complex into contact with transcription factors, thus allowing transcription to proceed. In their 2011 study, Zhang *et al*. demonstrated that in the context of estrogen receptor (ER) dependent MCF7 cells, ARGLU1 potentiated increased expression of the ER responsive genes *MYC* and *pS2*, but not the housekeeping gene *GAPDH,* after treatment with estradiol (30). Furthermore, shRNA-mediated knockdown of ARGLU1 led to a decrease in the growth and adhesion-independent survival of MCF7 cells, with and without estrogen stimulation (30).

Interestingly, other reports indicate a possible role for ARGLU1 in a different cellular context, namely in RNA splicing. ARGLU1 protein has been found in multiple preparations of spliceosome complexes, including complex A (31), the active step 1 spliceosome (32) and purified complex B (33). Additionally, ARGLU1 has been found to interact with (34) and be phosphorylated by (35), multiple members of the SRPK family of protein kinases, which regulate a family of proteins important for splicing, known as SR proteins (36–38). Additionally, the *Drosophila* homolog of ARGLU1, CG31712, has been shown to associate with U1-70K (39), a key player in the assembly of the spliceosome that has previously been shown to interact with SR and SR-like proteins (40–42). Taken together, these interactions indicate that ARGLU1 interacts with components of the spliceosome, although there has been no direct evidence that ARGLU1 interacts directly with RNA (43).

The current study addresses a number of questions about the regulation of function of the ARGLU1 gene. First, we used site directed mutagenesis to systematically detail the function of an UCE within ARGLU1 that leads to autoregulatory alternative splicing of the *ARGLU1* transcript. By using a splicing reporter that is amenable to manipulation, as well as the CRISPR/Cas9 system (44) to introduce more random mutations in the endogenous gene locus, we show that many alterations within the UCE dramatically alter splicing of *ARGLU1*. Finally, we show using overexpression and functional ARGLU1 knockout cells that *ARGLU1* splicing is regulated in response to changes in ARGLU1 protein levels.

MATERIALS AND METHODS

Nuclear/Cytoplasmic fractionation

Nuclear/Cyotplasmic fractionation was carried out as in (45) with some modifications. 10 cm dishes of cells were washed twice with ice-cold PBS. These cells were then harvested by scraping in ice-cold PBS and pelleted by centrifugation at 1000 × *g* for 5 min at 4°C. Cells were then gently resuspended in 250 μL of lysis buffer A (10 mM Tris (pH 8.0), 140 mM NaCl, 1.5 mM $MgCl_2$, 0.5% Igepal, 2 mM vanadyl ribonucleoside complex (VRC; Invitrogen)) and incubated on ice for 5 min. During this incubation 1/10 of the total cell lysate was retained for Western blotting and 1/5 of the whole cell lysate was added to 1 ml of Trizol reagent. The rest of the lysate was centrifuged for 3 min at 4°C to pellet the nuclei. The supernatant was reserved as the cytoplasmic fraction, and was spun a second time to remove any remaining nuclei. One-fifth of the cytoplasmic fraction was then used for Western blot, while the remaining fraction was added to 1 ml of Trizol reagent. The remaining nuclear pellet was washed 2X with 300 μl lysis buffer A, then the nuclei were resuspended in 100 μl of lysis buffer A, 1/5 of the nuclear lysate was used for Western blot and the remaining nuclear fraction was added to 1 ml of Trizol reagent.

Preparation of RNA

RNA was prepared by lysing cells in an appropriate volume of Trizol (Life Technologies). Chloroform was added, mixed vigorously, and incubated on ice for 5 min. These samples were centrifuged at maximum speed for 15 min. The aqueous phase was then either precipitated with equal volume of 100% isopropanol and washed with 70% ethanol, or mixed with equal volume of 70% ethanol and processed with the Purelink RNA mini kit according to the manufacturer's instructions. RNAs were treated with Purelink On-column DNase I or Turbo DNA Free DNase I (Life Technologies) according. RNA concentration was assessed using absorbance at 280 nm on a nanodrop spectrophotometer.

RT-PCR and RT-qPCR

RT was performed on DNase I treated RNA using the Protoscript II cDNA synthesis kit (NEB) with random primers. Semi-quantitative RT-PCR was carried out using OneTaq 2X mastermix (NEB) according to the manufacturer's instructions. RT-PCR products were resolved in 2% agarose TAE gels stained with ethidium bromide. RT-qPCR was carried out using iTaq Universal SYBR Green Supermix on the CFX96 qPCR platform. Primers used in RT-PCR and RT-qPCR are listed in Supplementary Table S1.

Northern blotting

Ten micrograms of RNA was used for loading onto a 1% agarose gel in sodium phosphate buffer. RNAs were resuspended in an RNA gel loading buffer containing glyoxal, bromophenol blue and xylene cyanol. RNAs were denatured at 75°C for 3 min, then loaded into the gel. The RNAs were resolved at 90 volts for 1 h. RNA was then transferred to positively charged nylon membrane using capil-

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

lary transfer in 20X SSC for 4 h. After transfer, the membrane was crosslinked to the membrane using 1200 mJ UV 285 nm irradiation. The membrane was then blocked using Ambion Northern Max hybridization buffer. The membrane was then hybridized overnight with an RNA probe against a region common to both the sliced and intron-retained isoforms that was internally-labeled with biotin. The membrane was then washed first with low stringency wash buffer 1X, and high stringency buffer 2X. The probes were detected with alkaline phosphatase conjugated streptavidin and illuminated with CDP-star chemiluminescent reagent.

Vectors

A fragment consisting of ARGLU1 exon 2, intron 2 and exon 3 was PCR amplified using primers noted in Supplementary Table S1 and TA cloned into the TOPO 2.1 vector. This fragment was subsequently subcloned into the pcDNA3.1 vector (Invitrogen) for use as the splicing reporter pcA23. Modifications to pcA23 were performed by site directed mutagenesis using the Q5 Site-Directed mutagenesis kit (New England Biolabs). See Supplementary Table S1 for the primers used for mutagenesis.

The ARGLU1 cDNA was cloned into pcDNA5/FRT/TO for use in 293 Flp-In T-Rex cells. Stable cell lines were created by cotransfection of this vector with the pOG44 vector, which expresses Flp recombinase. Stable integration of pcDNA5 confers hygromycin B resistance, while simultaneous removing the resistance cassette for Zeocin. Cotransfected cells were then selected and maintained with 15 μg/ml Blastcidin and 150 μg/ml hygromycin B.

Western blotting

Cells were lysed for 30 min in ice cold RIPA buffer supplemented with protease inhibitors and benzonase. Protein was quantified and cell lysates were mixed with 2X SDS loading buffer with 10 mM DTT. Thirty micrograms of protein was loaded into SDS-PAGE gels. Transfer to nitrocellulose membrane was carried out in Tris-glycine buffer with 10% methanol at 4°C at 240 mA for 2 h. Membranes were blocked with Licor blocking buffer for 1 h at room temperature. Blots were then incubated overnight at 4°C with primary antibody diluted (TBP 1:1000, GAPDH 1:2000, ARGLU1 (Thermo-Fisher item# PA5-26298) 1:1000, Alpha Tubulin (Cell Signaling Technology, clone DM1A, item # 3873) 1:1000) in Licor blocking buffer supplemented with 0.1% tween-20. Membrane was then washed with TBS-0.1% tween 3 times for 5 min. Appropriate florescent labeled secondary antibodies diluted in Licor blocking buffer with 0.1% tween and 0.01% SDS were incubated with the membrane for 45 min at room temperature with rocking. Membrane was then washed with TBS-0.1% tween 3 times for 5 min. Membranes were imaged on the Licor Odyssey system.

CRISPR/Cas9 mediated mutagenesis of endogenous ARGLU1 locus

sgRNAs were designed using the optimized CRISPR design tool at http://crispr.mit.edu/. Two different sgRNAs were chosen that targeted the UCE and off-target sites with low scores and low probability of targeting annotated genes (see Supplementary Table S1 and Supplementary Figure S7A). sgRNAs were cloned into the pX330-U6-Chimeric_BB-CBh-hSpCas9 vector (46) available from Addgene (Plasmid #42230). An sgRNA mediated cleavage reporter was created by cloning sgRNA target sequences between RFP and GFP ORFs, with GFP out of frame. The sgRNA expressing Cas9 plasmids were then co-transfected into 293 Flp-In T-Rex cells using lipofectamine 2000. One day later, cells were trypsinized and sorted by FACS into 96 well plates based on dual RFP and GFP fluorescence, allowing for clonal selection of cells.

Sanger sequencing

DNA or cDNA prepared by reverse transcription were subjected to PCR using primers noted in the figure legends. These PCR products were then purified, and the samples and appropriate primers were sent to Genewiz for sequencing. Chromatograms were visualized using FinchTV. For the deconvolution of cDNAs with inclusion of the alternate exon, the Mutation Surveyor (47) program was used, with the WT chromatogram as a reference.

RESULTS

ARGLU1 is alternatively spliced and partitions to the nucleus or cytoplasm depending on the presence of a retained intron

We initially examined ARGLU1 transcripts through RNA-Seq data from both ENCODE data sets, or from HeLa cells that had been fractionated in our laboratory into cytoplasmic and nuclear pools and then poly(A) selected. A polyadenylated *ARGLU1* transcript containing a single retained inton, but otherwise fully spliced, was found primarily in the nuclear fraction, with some cell-line specific variations in subcellular location. The fully spliced isoform was found in both the nucleus and cytoplasm, as seen by RNA-Seq in HeLa cells (Figure 1A). Similar results were seen upon examination of ENCODE RNA-seq data from a variety of cultured cell lines (for example, see Supplementary Figure S1), consistent with recently published findings (48). This was confirmed in HeLa, HEK293 Flp-In TRex and MCF7 cells by RT-PCR and Northern blot (Figure 2A and B, Supplementary Figure S1B). We have carried out the remainder of the experiments using 293 Flp-In TRex cells, as they allowed for inducible over-expression of ARGLU1 protein in later experiments.

ARGLU1 contains an UCE

In order to identify functional elements within this alternatively spliced intron we hypothesized that evolutionary conservation could serve as a useful marker. Based on this hypothesis, we queried databases of highly evolutionarily conserved sequences. Indeed, a region within the *ARGLU1* intron has been previously identified as an UCE by the UCNEbase, a database of Ultraconserved Noncoding Elements which identified UCEs as stretches of nucleotides 95% conserved for over 200 bases between human and

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

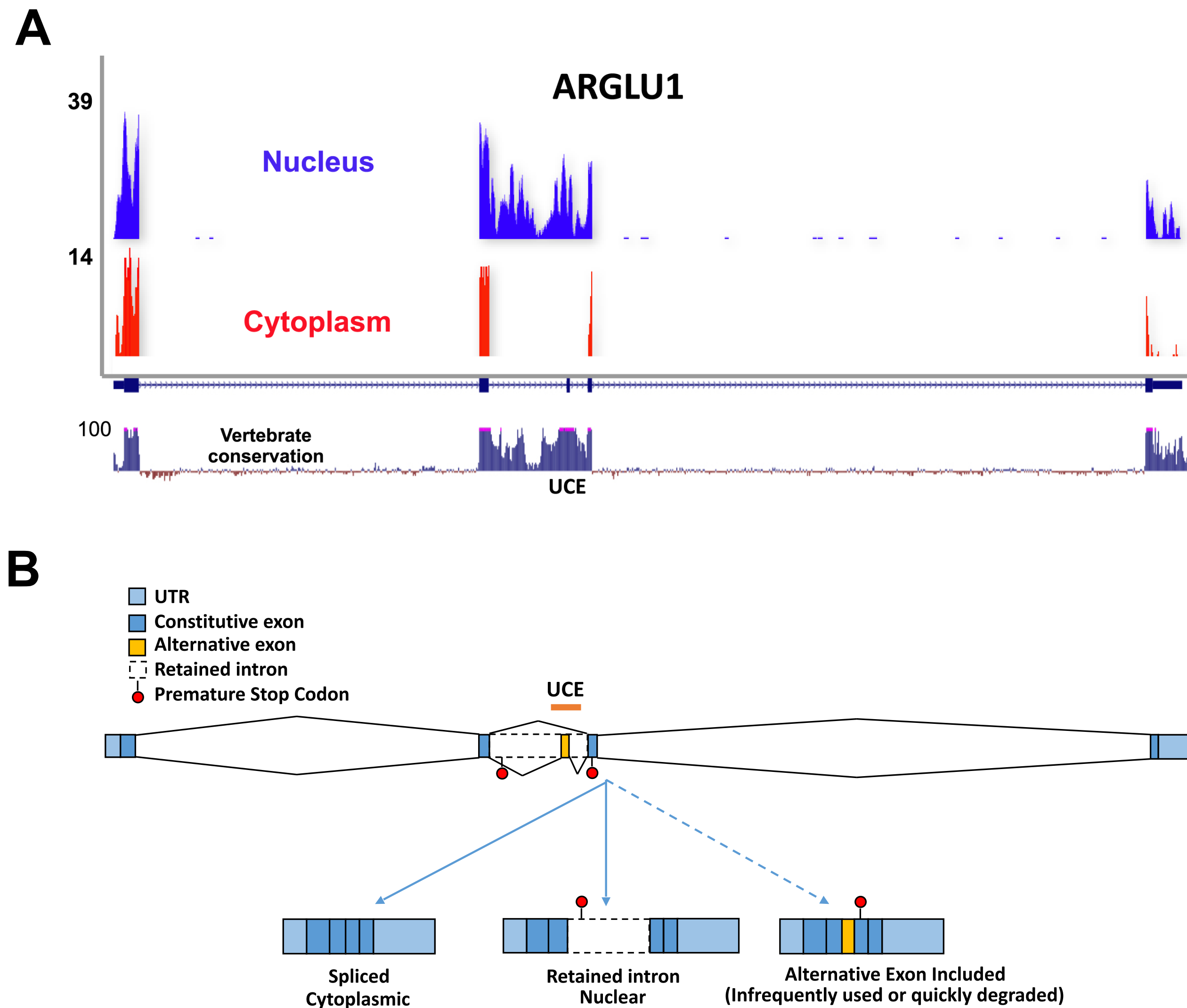

Figure 1. ARLGU1 is alternatively spliced and contains an Ultraconserved Element. (**A**) RNA-Seq tracks from HeLa nuclear and cytoplasmic fractions are shown. Nuclear reads are in blue and cytoplasmic reads are in red. Normalized expression levels (RPKM) are shown on the y-axis. Below the tracks vertebrate conservation is shown and the UCE is noted. (**B**) An overview of the three splice isoforms of arginine and glutamate-rich protein 1 (*ARGLU1*).

chicken (49). This represents a base substitution rate of approximately 1% per 100 million years, and sequences fulfilling these stringent requirements have been previously found to only exist in vertebrates (50). Using UCNEbase and the Evolutionary Conserved Browser (51), we were able to identify an ultraconserved region of 500 bases that fit these requirements. This region is 95% conserved at the nucleotide level between human and chicken for 500 bases, and 95% conserved with the frog *Xenopus tropicalis* for 265 bases (Figure 1A). Notably, using the more strict definition of a UCE set out by Bejarano *et al.* (28), which required 100% conservation for over 200, the *ARGLU1* UCE is 100% conserved between human and chicken for 228 bases.

The UCE in ARGLU1 intron 2 contains a *cis*-regulatory element controlling intron retention

The observation of the UCE located within the retained intron led us to hypothesize that this element is regulating alternative splicing of this gene. A minor alternative splicing product of the ARGLU1 gene has been observed and annotated previously in the ENSEMBL database (52). This isoform includes an alternative exon that is within the retained intron, and is completely contained within the UCE (Figure 1B). The inclusion of this exon would be predicted to cause degradation by NMD, as it introduces a premature termination codon. Our initial RNA-seq and RT-PCR experiments showed little evidence of inclusion of this isoform (Figure 1). However, analysis of ENCODE Nuclear-Cytoplasmic RNA-Seq indicates that inclusion of this alternative exon is increased in certain cells (Supplementary Figure S1). Furthermore, treatment with the translation inhibitor emetine,

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

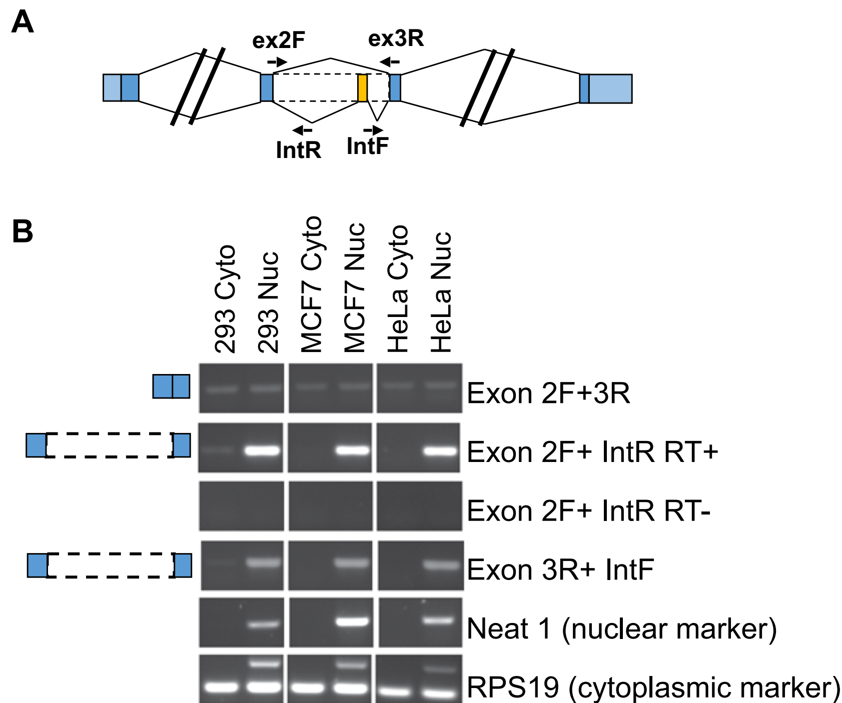

Figure 2. ARGLU1 isoforms are differentially localized between the nucleus and cytoplasm. (**A**) A gene schematic of ARGLU1 with the primers used for RT-PCR in 2B. (**B**) RT-PCR of nuclear/cytoplasmic fractionated cells. Primers for *ARGLU1* are indicated in 2B and noted to the right of the gels. Diagrams depicting the splice product are noted to the left of the gels. NEAT1 was used as a representative nuclear RNA, and RPS19 was used as a representative RNA with a spliced cytoplasmic isoform.

which is known to inhibit NMD (17), stabilizes transcripts containing this alternative exon (Supplementary Figure S2).

To test whether ARGLU1 alternative splicing is regulated by sequences in the UCE, including the alternative exon or flanking intron, deletion analysis was carried out in the context of a splicing reporter, A23 (Figure 3A). When transfected into HEK-293T cells, A23 RNA is spliced similarly to endogenous ARGLU1, with roughly equal amounts of spliced and retained intron RT-PCR product (Figure 3B). The largest deletion of 504 bp corresponds to sequence that is 95% conserved among vertebrates (Figure 1). Three smaller deletions of 168 bp (168-1, -2 and -3) are subdeletions of the same conserved region, while a control deletion of the same size was generated in an unconserved region of the intron. RNase protection assays revealed similar levels of intron retention for the wild type and control deletion reporter as well as the 168-1 deletion (Supplementary Figure S3), suggesting the smaller 168-1 portion of the region does not contribute to regulation. In contrast, detection of exon 2-exon 3 ligated RNA by RT-PCR increased upon deletion of the 504 bp UCE as well as 168-2 and -3, with a corresponding decrease in intron retention (Figure 3B), indicating that sequences within the UCE are indeed regulatory.

To further determine specific sequences regulating ARGLU1 alternative splicing, a series of 25 bp deletions were created spanning the 168 bp deletions 2 and 3 (Figure 3A and Supplementary Figure S4), which had previously been shown to contain regulatory sequences. To compare spliced and retained intron isoforms of our splicing reporter, we designed a set of primers for competitive RT-PCR, in which we used a forward and reverse primer complementary to transcribed plasmid specific sequences spanning the insert, combined with a reverse primer within the retained intron (Figure 3A), yielding 3 predicted amplicons: exon 2 spliced to exon 3, a product with the retained intron and an exon 2-alternate exon–exon 3 splice product. The 25 bp deletions that overlapped the alternative exon (25 bp deletions 6–8) were found to increase splicing of exons 2 and 3 (Figure 3B), and decrease the amount of retained intron transcript from the reporter. Unexpectedly, 25 bp deletion 13, beginning 111 bp downstream of the alternative exon 5′SS (Supplementary Figure S4), demonstrated decreased exon 2–3 splicing.

We noted that deletion 168-3, which was downstream of this annotated alternative exon, caused increased splicing with a shifted band, indicating the inclusion of the alternative exon (Figure 3B and Supplementary Figure S6). Upon deletion of other sequences downstream of the alternative exon (25 bp deletions 9–11), we observed the inclusion of the normally skipped exon, with aberrant splicing (deletion 9) or usage of different 5′-splice sites on the now included exon depending on the deletion (Figure 3B and Supplementary Figure S5). This indicated that there was some role for the sequences downstream of the normally skipped exon. The observation that different 5′-splice sites could be used following deletion of these sequences downstream of the alternative exon suggested that there might be alternative 5′-splice sites competing within this region. Analysis of the predicted downstream 5′-splice sites using different computational models (53) indicates that the annotated alternative exon 5′-splice site is actually predicted to be weaker

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

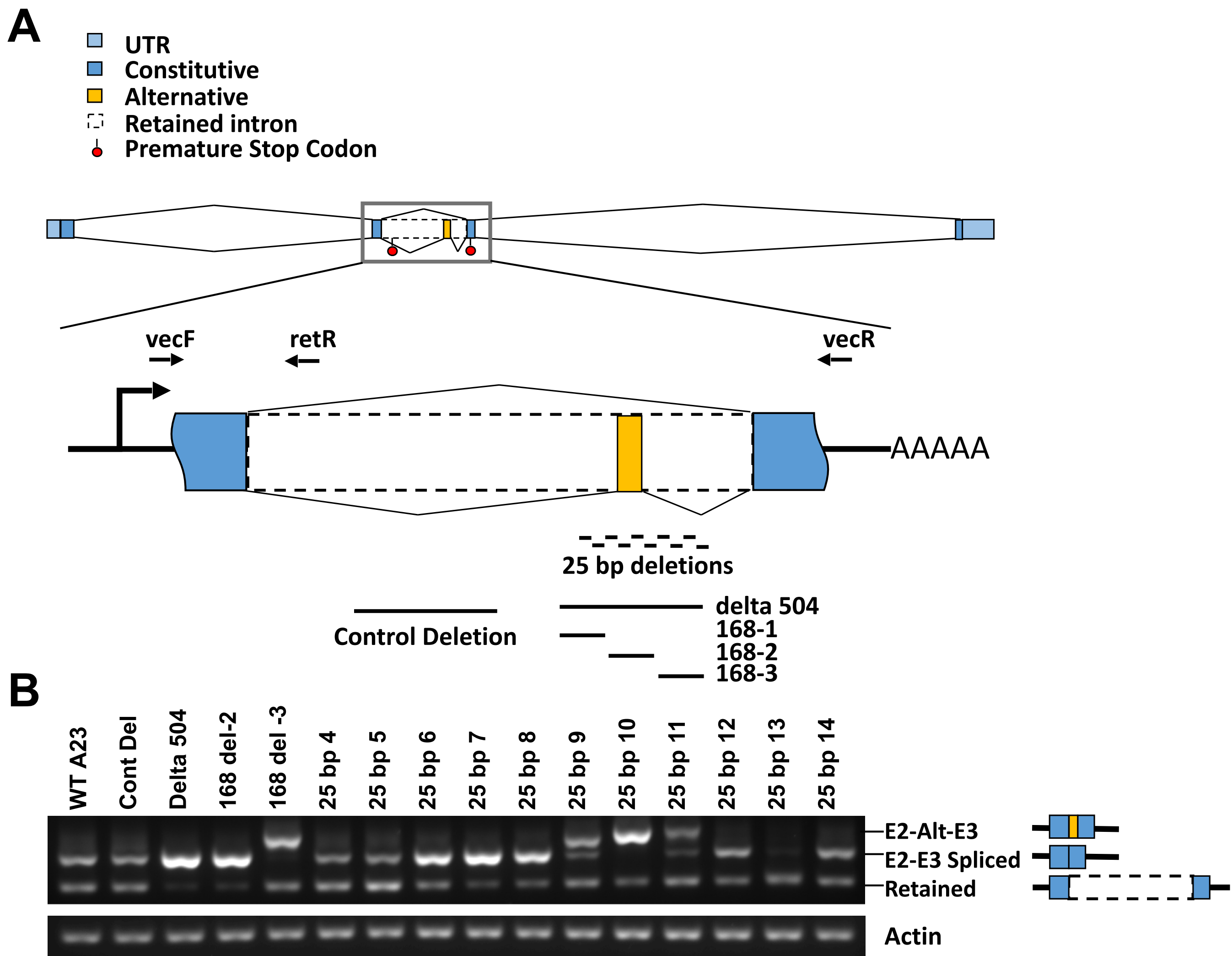

Figure 3. ARGLU1 splicing is regulated by sequences within the Ultraconserved Element (UCE). (**A**) Diagram of the splicing reporter plasmid A23, which is derived from the pcDNA3 vector with an insert of the ARGLU1 sequence from exon 2 through exon 3. Multiple deletions were made in this vector to test the contribution of that sequence to splicing regulation. Notably, delta 504 corresponds to the sequence 95% conserved with chicken. The control deletion is in a poorly conserved region of the intron. The 168-bp deletions 1–3 span the 504 bp deletions. The 25 bp deletions depicted here cover 168-bp deletions 2 and 3. (**B**) RT-PCR products using primers defined in Figure 2, panel A. Three primers, one forward and two reverse, were used simultaneously in a competitive PCR. The bottom band is the PCR product from a retained intron reporter RNA. The middle band results from splicing of the reporter transcript from exon 2–exon 3. The upper bands are products that include the alternative exon in the reporter transcript. These results are representative of three independent experiments. Results from Sanger sequencing of some of these products are found in Supplementary Figure S5.

compared with other 5′-splice sites analyzed in the same region (Table 1). Based on the observation that deletion of sequences overlapping and downstream of the UCE alternate exon altered intron retention and alternate exon inclusion, we hypothesized that downstream 5′-splice sites mediate this effect.

To test whether the annotated and downstream putative 5′ splice sites might mediate the ARGLU1 intron retention, we mutated the these individually or in combination (Figure 4A and Supplementary Figure S4). Mutating the annotated 5′SS resulted in larger RT-PCR bands, suggesting the use of different 5′-splice sites downstream of the normal one but without affecting the ratio between spliced to intron retained isoforms of the reporter (Figure 4B). We next mutated all six possible 5′-splice sites downstream the annotated 5′SS (Figure 4A). Mutation of putative splice sites 5 and 6 decreased intron retention (Figure 4B). A similar effect was seen upon mutating the annotated 3′-splice site of the alternative exon. Mutation of putative splice sites 2–4 did not alter the ratio of spliced to retained isoform. Interestingly, however, these mutations resulted in the inclusion of alternative exons of a greater size (Figure 4B middle panel). Sequencing RT-PCR products from cells transfected with mutant plasmids revealed that mutation of SS5 and SS6 caused exclusive use of the annotated 5′ site (Supplementary Figure S6) while mutants SS3 and SS4 redirected alternative splicing to the SS5 site (Supplementary Figure S6).

To further analyze the role of the splice sites flanking the retained intron and skipped exon, we improved them to match consensus 5′ or 3′-splice sites (Figure 4C). Improving the 5′-splice site of exon 2 or the 3′-splice site of exon 3

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

Table 1. Analysis of 5′-splice site strength

	9-mer sequece	MAXENT:	MDD:	MM:	WMM:
>Ex2	gag|GTaacg	9.65	12.18	7.79	7.1
>Alt 5′	aca|GTaaat	1.91	2.98	3.33	3.74
>SS1	aat|GCtgat	-20.87	-14.68	-17.9	-9.94
>SS2	gga|GTatag	-7.36	1.08	0.86	-0.91
>SS3	tag|GTaaag	6.93	11.08	5.49	6.65
>SS4	tac|GTaggt	4.81	9.18	4.92	5.81
>SS5	tag|GTatac	4.69	8.18	3.68	3.41
>SS6	ttg|GTgaga	6.29	11.68	6.58	6.27

Calculations based on the MaxEntScan model (53).

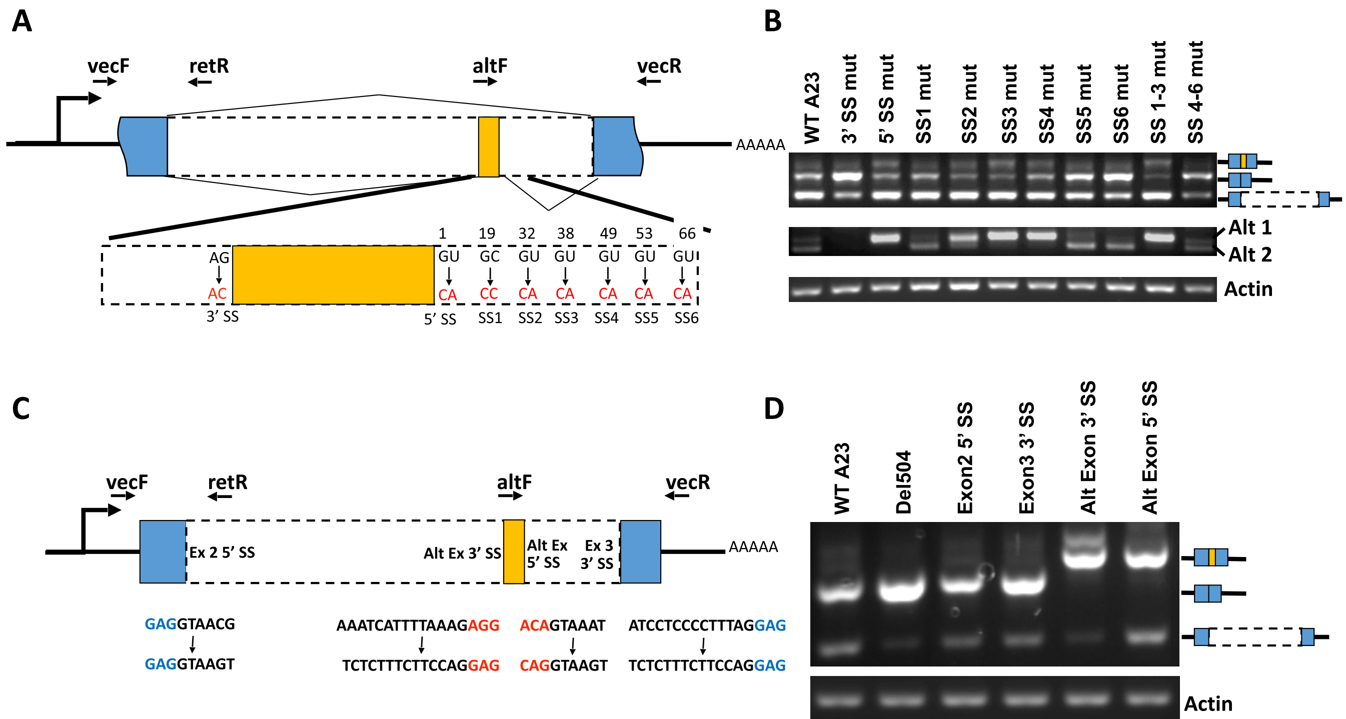

Figure 4. Manipulation of alternative exon splice sites or downstream sequences alters splicing of the ARGLU1 reporter. (**A**) Detailed view of the alternative exon and splice sites that were mutated. The alternative exon is in yellow. Sequences of annotated and potential splice sites that were mutated are shown, with the wild-type sequences in black and the mutated sites in red. The numbers represent the nucleotide distances from the annotated 5′SS of the alternative exon to the potential downstream sites. (**B**) As in Figure 2B, the top panel indicates the spliced and intron retained reporter transcript. The bottom panel indicates the amplicon from the alternative exon forward primer to the vector reverse primer. (**C**) Detail of the mutations that improved 5′ and 3′-splice sites to ideal consensus sequence. Below the cartoon of the reporter transcript are the altered sequences. The top sequence is the endogenous sequence, while the bottom sequence is the mutated consensus sequence. (**D**) RT-PCR of the A23 splicing reporter with site-directed improvement of the splice sites noted in panel D.

did not appreciably increase the ratio of spliced to retained product (Figure 4D). However, improving the 3′-splice site and polypyrimidine tract of the alternative exon dramatically increased its inclusion, as did improving its 5′SS (Figure 4D). By sequencing RT-PCR products from the mutated alternative 5′SS mutant, we confirmed usage of the improved site (Supplementary Figure S6).

Modification of the endogenous ARGLU1 UCE disrupts intron retention

We next sought to validate the ability of mutations and deletions within the UCE to cause changes in splicing in an endogenous context. To accomplish this we used the CRISPR/Cas9 system (44) to generate random mutations by non-homologous end joining. We were able to isolate a number of clones with heterozygous or compound heterozygous insertions and deletions within the UCE (Supplementary Figure S7A and S7B). RT-PCR (Figure 5A) confirmed that all clones selected demonstrated increased exon 2-exon 3 splicing and decreased levels of *ARGLU1* intron retention compared with wild-type cells (Figure 5B and Supplementary Figure S7B). Clone 4 showed the largest increase in the ratio of spliced/retained intron (Figure 5 and Supplementary Figure S7B), and contained a homozygous single nucleotide insertion within the alternate exon (Supplementary Figure S7B). Interestingly, clones 3 and 14, which contained either a homozygous single nucleotide deletion (clone 14) or a large heterozygous insertion (clone 3) within the alternate exon that shifted the reading frame

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

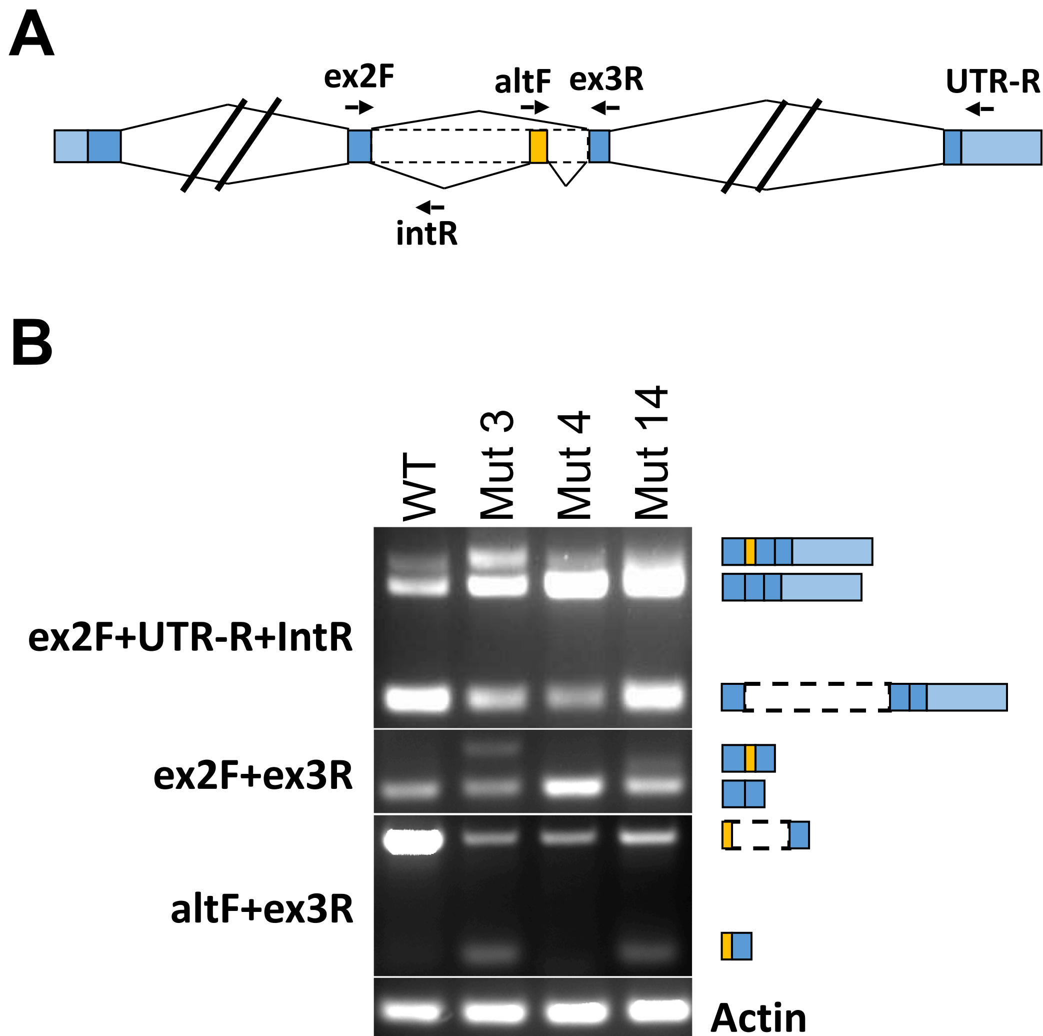

Figure 5. Mutations within the endogenous ARGLU1 UCE disrupt splicing. (**A**) Diagram of primers used to assess splicing of *ARGLU1* from cells with CRISPR/Cas9 induced mutations. (**B**) RT-PCR from cell lines with a variety of indels from sgRNA guided Cas9 endonuclease activity. RT-PCR was used by primer combinations noted to the left of the gel. A schematic representing the spliced products for each band is provided to the right of the gel.

to prevent NMD (Supplementary Figure S7C), showed stabilization of a band corresponding to inclusion of the alternative exon (Figure 5B). Sequencing of RT-PCR products from the WT and CRISPR mutants 3 and 14 further confirmed that insertions or deletions within the alternative exon that cause it to be in frame cause RNAs containing it to be stabilized (Supplemental Figure S8A–D). This also allowed for sequencing of the large heterozygous insertion in the alternative exon in clone 3 (Supplementary Figure S8D).

ARGLU1 regulates the splicing of its own mRNA

Since the ARGLU1 protein contains RS repeats and has been reported to be associated with spliceosomes, we hypothesized that one role of the UCE in ARGLU1 splicing, like the UCE elements found in a number of SR protein genes, could be to regulate protein expression levels through alternative splicing. To test this hypothesis, we examined the effect of a plasmid encoding an *ARGLU1* cDNA on the splicing of the A23 splicing reporter vector. Results (Figure 6A) showed that intron retention was responsive to ARGLU1 levels only when the UCE was present.

To test whether this finding also held true for the endogenous transcript, we created a stable cell line that overexpressed ARGLU1 in a tetracycline inducible manner. Using the HEK-293 Flp-In TRex cell line, we made cell lines expressing ARGLU1, ARGLU1 with a C-terminal HA tag, or the gene chloramphenicol acetyltransferase as a control (Figure 6B). Tetracycline induced overexpression of ARGLU1 for 24 h caused a dramatic overexpression of ARGLU1 protein (Figure 6B). Intriguingly, we found that overexpression of the HA-tagged version of ARGLU1, which migrates slightly slower by SDS-PAGE, led to a dramatic decrease in levels of the endogenous protein (red arrow, Figure 6B). This suggested to us that ARGLU1 protein leads to feedback that maintains ARGLU1 protein levels. Quantitative RT-PCR analysis of splicing of the endogenous transcript after 24 h of induction indicated that there was a 6-fold decrease in the level of spliced endogenous transcript, and 1.6-fold increase in the level of the retained intron transcript (Figure 6C and D). These changes were also

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

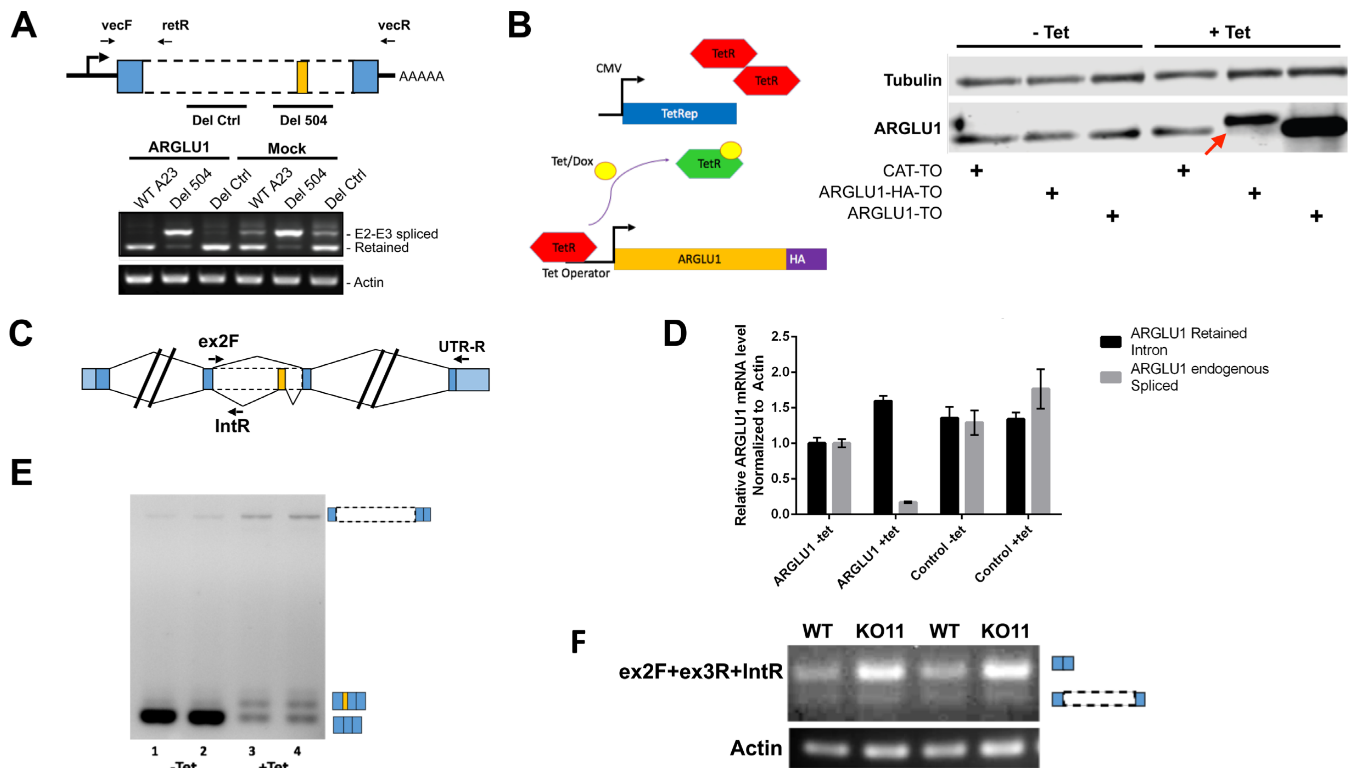

Figure 6. ARGLU1 protein regulates alternative splicing of its own mRNA. (**A**) A plasmid with ARGLU1 cDNA was cotransfected with the A23 splice reporter, the 504 bp deletion or the control deletion. RT-PCR results are shown using the indicated primers. (**B**) A T-Rex tetracycline inducible ARGLU1/ARGLU1-HA system was used to overexpress ARGLU1 or HA-tagged ARGLU1 in FlpIn Trex cells. On the left is a cartoon of the experimental system. On the right is a Western blot against alpha tubulin and ARGLU1 with the noted tetracycline inducible constructs. -TO stands for tet-operator. Tetracycline (1 μg/ml) was added to culture media for 24 h for tetracycline induction. Note that ARGLU1-HA migrates slightly slower than endogenous ARGLU1. The red arrow denotes endogenous ARGLU1, whose level is lowered upon overexpression of HA-tagged ARGLU1. (**C**) Primers used in Figures 6D, E and 7A and D. (**D**) qRT-PCR of endogenous spliced and retained intron ARGLU1 in inducible ARGLU1 or chloramphenicol acetyltransferase (CAT) cell lines. The samples are ARGLU1-TO without tetracycline, ARGLU1 with tetracycline, CAT-TO with and without tetracycline. PCR primers were designed from the endogenous UTR to either exon 2 or the intron, as seen in Figure 6C. Values (±s.d) are derived from the mean of three samples. (**E**) Semi-quantitative RT-PCR using the Ex2-F and UTR-R primers in 6C. The cartoons to the right indicate retained intron, alternative exon inclusion and full splicing, respectively. (**F**) Semi-quantitative RT-PCR from WT and ARGLU1 KO11 cells using primers for actin as well as the Exon 2F+Exon 3R+IntR noted in Figure 5A. Duplicate reactions are shown.

observed by semi-quantitative RT-PCR using primers that captured both the spliced and retained transcripts (Figure 6E). Intriguingly, overexpression of ARGLU1 caused not only an increase in intron retention and a decrease in the spliced isoform, but also appeared to increase the inclusion of the alternative exon (Figure 6E). To more specifically test the hypothesis that a lack of ARGLU1 would cause increased productive splicing of exons 2 and 3, we used the CRISPR/Cas9 system to create 293 Flp-In TRex cells in which frame shift mutations were induced, causing a lack of ARGLU1 protein while maintaining expression of *ARGLU1* mRNA (Supplementary Figure S9). We noted that ARGLU1 knockout resulted in an overall increase in ARGLU1 RNA levels, suggesting that this protein may influence its expression both transcriptionally and post-transcriptionally.

We next followed the dynamics of splicing change in ARGLU1 as a function of ARGLU1 protein expression. For that purpose, we used our tet-on inducible ARGLU1 cell line. Two hours following induction of ARLGU1-HA, the endogenous mRNA showed a significant decrease in the spliced isoform, with no change in the retained intron isoform (Figure 7A). At the same time point ARGLU1-HA levels were only 75% of endogenous levels, and total ARGLU1 protein levels were approximately 1.8 times baseline (Figure 7B). With continued expression of ARGLU1-HA, levels of the endogenous spliced ARGLU1 mRNA continued to decrease, with a concomitant decrease in the endogenous protein (Figures 7A and B). This indicated an inverse correlation between total ARGLU1 protein levels and the level of spliced ARGLU1 mRNA.

To further confirm the potential regulation of ARGLU1 mRNA splicing by its own protein, we used emetine to inhibit translation of new proteins. ARGLU1 protein showed a half-life of 4 h (Figure 7D). Interestingly, during that time frame, the ratio of ARGLU1 spliced/retained isoforms significantly increased (Figure 7C). Taken together with the splicing changes seen following overexpression and knockout of ARGLU1 protein, these results suggest that ARGLU1 maintains homeostatic control of its own protein level by regulating the splicing of its own mRNA.

Emetine has also been used to study transcripts subject to degradation through NMD since it inhibits this process (17,54) (55). We found that the ARGLU1 retained intron

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

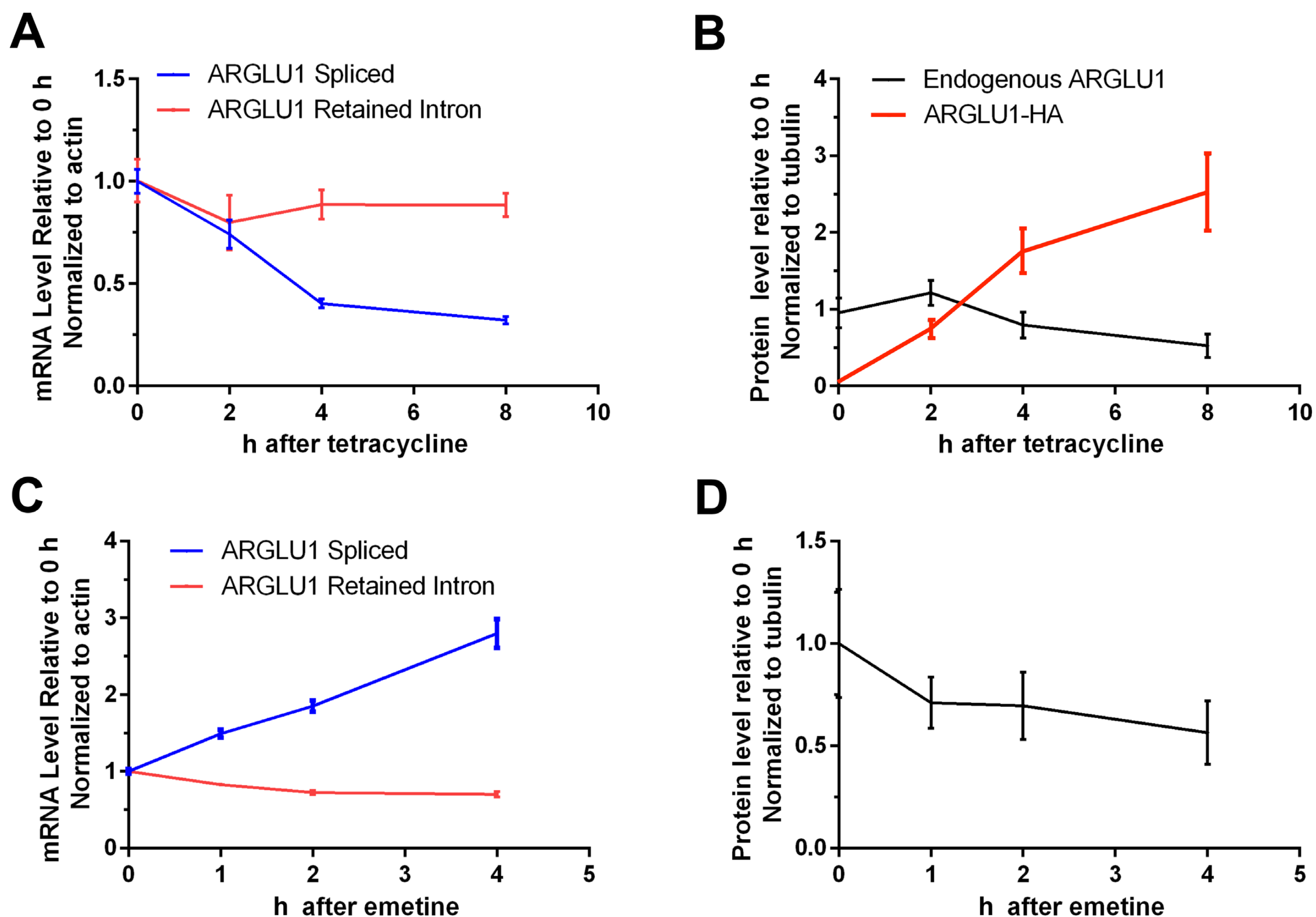

Figure 7. ARGLU1 splicing inversely correlates with ARLGU1 protein levels. (**A**) qRT-PCR of ARGLU1 spliced and intron retained isoforms following tetracycline induced expression of ARGLU1-HA. RNA levels were normalized to actin and set relative to time 0. Values (±s.d) are derived from the mean of three samples. (**B**) Quantitation of ARGLU1 protein levels following induction of ARGLU1-HA. Protein levels were normalized to tubulin, and set relative to time 0. Note that the ARGLU1-HA level was also set relative to endogenous ARGLU1, as they were distinguishable bands detected by the same antibody. Values (±s.d) are derived from the mean of three samples. (**C**) qRT-PCR of ARGLU1 spliced and retained intron isoforms following 50 μg/ml emetine treatment. RNA levels were normalized to actin and set relative to time 0. Values (±s.d) are derived from the mean of three samples. (**D**) ARGLU1 protein levels following treatment with 50 μg/ml emetine. ARGLU1 levels were normalized to tubulin and relative to time 0. Values (±s.d) are derived from the mean of three samples.

RNA levels decreased by 45% compared to untreated cells after 4 h of treatment (Figure 7C), in contrast to the expected result if this transcript had been subject to NMD. Furthermore, 4-h treatment with emetine did cause stabilization of transcripts containing the alternative exon (Supplementary Figure S2), indicating that it is subject to NMD.

Finally, since ARGLU1 regulates the splicing of its own mRNA, the likelihood exists that it also affects the splicing of a number of other cellular genes. This is indeed the case. While beyond the scope of the current story, we have investigated this possibility and will report comprehensive results elsewhere. As an example of the genomewide influence of ARGLU1 on RNA processing, however, Supplementary Figure S10 shows by RT-PCR analysis two cases of alternative splicing change induced by ARGLU1 overexpression.

DISCUSSION

The ARGLU1 protein has not been well characterized to date, and several functions have been suggested. A previous report indicated that ARGLU1 interacts with MED1, a component of the mediator complex (30). However, other reports have indicated that ARGLU1 interacts with components of the spliceosome (33,34,43,56) as well as SR protein kinases (34,35). While not ruling out a role for ARGLU1 interactions with the Mediator complex, much of the data we present here support the latter idea that ARGLU1 may be associated with the spliceosome and alternative splicing. Indeed, its interaction with the Mediator complex would bring it in contact with a transcriptionally active RNA pol II, possibly facilitating its role in splicing. Such a role has been previously proposed by the interaction of Med23 with hnRNP L to regulate alternative splicing of a subset of hnRNP L targets (57).

This study has focused on two important aspects regarding the regulation of ARGLU1: what alternative splicing regulates expression of this protein, and what role ARGLU1 plays in this regulation. First, in line with previous transcriptome-wide reports on retained introns (15,22) and nuclear/cytoplasmic RNA-Seq experiments (58), there is a

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

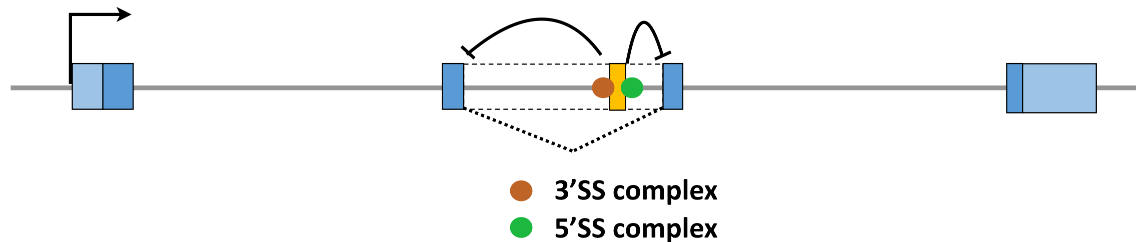

Figure 8. Model for ARGLU1 homeostatic regulation. Sequences within the UCE direct the assembly of unproductive splicing complexes which result in intron inclusion and nuclear retention. This regulation involves an alternative exon with a weak upstream 3′ splice site and multiple competing downstream 5′ splice sites. In order to ensure tight regulation of protein expression, NMD leads to mRNA degradation in instances where the alternative exon is productively spliced.

pool of ARGLU1 that is alternatively spliced and retains intron 2 and is stably localized to the nucleus. This same finding of nuclear detention of the retained intron isoform of *ARGLU1* was also recently demonstrated by RNA-seq and Northern blot (59). Our RNA-seq, RT-PCR and Northern blot results have also shown that this retained intron transcript is abundant and detained in the nucleus. However, the existence of such an isoform raises a number of questions, including the biological function for this alternative splicing event, and the *cis*- and *trans*-elements that control intron retention. Furthermore, what is the function of ARGLU1 protein that necessitates a tightly controlled alternative splicing event that is so highly evolutionarily conserved?

To answer these questions, we first used evolutionary conservation as a marker to identify functionally important sequences. A number of studies have shown that UCE and highly conserved nucleotide sequences are highly correlated with alternative splicing, and are often associated with splicing events that lead to transcript regulation through NMD (16,17). Intriguingly, these highly conserved elements are often found within SR proteins and other proteins involved in regulation of alternative splicing (16,17). We used databases and other evolutionary conservation resources to identify a 500 nucleotide Ultraconserved Element within the retained intron of *ARGLU1* (49,51).

We have tried to address the role that NMD may play in regulation of *ARGLU1* intron retention and alternative splicing through treatment with emetine to block translation. We found that the level of retained intron transcript was decreased by treatment with emetine, whereas it would be expected if this transcript were degraded by NMD the levels of this transcript would be expected to increase. Transcripts that include the alternative exon do seem to be susceptible to NMD, however, as treatment with emetine stabilized transcripts with this alternative exon (Supplementary Figure S2). Furthermore, mutations within the alternative exon that cause the alternate exon to be in frame also cause transcripts with the alternate exon to be stabilized.

To test the role of the UCE in regulating *ARGLU1* intron 2 retention versus splicing, we constructed a splicing reporter vector that consisted of the genomic region including the retained intron and flanking exons. By systematically deleting portions of the UCE within the splicing reporter plasmid, we were able to identify specific sequences important for intron retention. Intriguingly, deletion of sequences that overlapped an infrequently used alternative cassette exon caused a dramatic shift in splicing of the flanking exons. Furthermore, deletion of sequences downstream of this alternative exon led to increased inclusion of the exon, albeit with a variety of 5′-splice sites being used. This raised the possibility that there could be competition between putative 5′-splice sites downstream of the alternative exon with the annotated 5′-splice site of this exon. Indeed, mutagenesis of these splice sites support this hypothesis. Mutagenesis of the 3′-splice site led to dramatically increased splicing, decreased intron retention and no inclusion of the alternative exon. Mutagenesis of the 5′-splice site of the alternative exon did not lead to decreased inclusion of that exon, but instead lead to usage of an alternative downstream 5′-splice site. The annotated 5′-splice site for this alternative exon is not optimal. With a sequence of ACA|GTAAATAT, it deviates substantively from the ideal splice site CAG|GTAAGTAT at the highly conserved −1 G, as well as the well conserved +5 G (see Table 1). This seems to be in line with previous reports that have indicated that suboptimal splice sites are well conserved at alternatively spliced exons (26,27), and allow for dynamic regulation by adjacent intronic and exonic splicing regulatory elements.

Conversely, experiments in which we strengthened the splice sites of the alternative exon or the flanking exons indicated that the endogenous splice sites for the alternative exon are weak in comparison to the flanking exons. Indeed, strengthening of the 3′-splice site of the alternative exon caused the exclusive usage of this exon, decreased the level of intron retention and also led to the usage of multiple 5′-splice sites, as noted earlier. Similarly, strengthening the annotated 5′-splice site of this alternative exon led to increased inclusion of this alternative exon, however, with exclusive usage of the strengthened 5′-splice site. This suggests a model in which the alternative exon has weak splice sites in comparison to the flanking exon, but can be included if there is an increase in the strength of either splice site. Inclusion of splice sites seems to be balanced with intron retention, as increased alternate exon inclusion as a result

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

strengthening of the alternative exon 3′-splice site led to a decrease in intron retention.

In addition to determining the *cis*-elements responsible for *ARGLU1* alternative splicing regulation, we have investigated whether ARGLU1 protein plays a *trans*-regulatory role in this alternative splicing. Overexpression of ARGLU1 using an inducible system causes dramatic and rapid changes in the splicing of the endogenous transcript such that productive splicing is decreased and the intron retention and NMD sensitive isforms are increased. Conversely, a CRISPR induced nonsense frameshift mutation in the endogenous gene loci, causing functional knockout of ARGLU1 leads to increases productive splicing of exon 2 with exon 3, which would lead to increased ARGLU1 protein levels if not for the nonsense mutation. Future studies of ARLGU1 regulation should focus on what other *trans*-acting proteins regulate this alternative splicing. High throughput studies using siRNA based systems coupled with high-throughput sequencing (59) have allowed for rapid identification of *trans*-elements that regulate either specific alternative splicing or transcriptome wide changes. Using a similar system, it should be possible to define the complement of proteins regulating ARGLU1 alternative splicing.

Based on our data from mutagenesis of *cis*-elements important for alternative splicing of *ARGLU1* and the ability of either over- or under-expression of ARGLU1 to affect splicing, we have developed a model of the regulation of *ARGLU1* alternative splicing, which in turn functions to regulate ARGLU1 protein levels (Figure 8). Central to this model is the presence of the alternative exon, which is completely within the UCE. We propose that there are two levels of regulation on the level of *ARGLU1* mRNA that codes for the functional protein: intron retention coupled with nuclear detention, and NMD degradation of the transcript that includes the alternative exon. Furthermore, the fact that modulation of ARGLU1 protein levels feeds back to regulate splicing of its own transcript indicates that ARGLU1 functionally interacts with the splicing machinery. Interestingly, SR proteins such as SFRS3 (9) have been shown to interact with *ARGLU1* mRNA through CLIP-Seq experiments. ARGLU1 was recently shown to be alternatively spliced in chronic lymphocytic leukemia, which may indicate that this alternative splicing event has a role to play in some forms of malignancy (60). It is probable that ARGLU1 does not interact directly with RNA, as it does not have a recognizable RNA recognition motif. The most likely point of interaction is through the SR-like domain at the N-terminus of ARGLU1. RS motifs have been shown to mediate protein–protein interactions important for regulation of alternative splicing (61,62), and have been found in non-SR proteins, such as U1-70K (63,64). Interestingly, the *Drosophila* homolog of ARGLU1, CG31712, was seen to interact with U1-70K, as well as a number of other splicing related proteins (39). Additionally, ARGLU1 interacts with, and is phosphorylated by, the SR protein kinases SRPK1 and SRPK2 (34,35). Further study will be needed to define ARGLU1 protein–protein interactions and to understand what role it plays in splicing regulation.

Autoregulatory feedback loops that function through alternative splicing have been well described for a number of years and are found in many splicing regulatory proteins (29,65–67). Indeed, SRSF1 is regulated by almost identical alternative splicing and feedback loops, including alternative splicing and intron retention associated with nuclear localization (29). However, the completion of high quality genome assemblies has allowed for the use of comparative genomics leading to the finding that ultraconserved or highly conserved elements are often closely linked to these autoregulatory events (16,17). ARGLU1 seems to fit a feedback model whereby increased proteins levels increase inclusion of NMD sensitive exons or detained introns to maintain a protein steady state. This study has specifically defined the *cis*-element sequences that regulate this feedback loop within *ARGLU1* mRNA and shown that ARGLU1 plays a role in this regulation. Yet to be determined, however, are the other *trans*-elements that play a role in this feedback loop.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank all members of the lab for advice and Kim Morris for technical assistance and reading the manuscript. The authors also thank Li Yang and Lingling Chen for the sequencing results shown in Figure 1A. This work was supported by grant R01 HD072418 from the NIGMS.

FUNDING

National Institute of General Medical Sciences (NIGMS) [R01 HD072418]. Funding for open access charge: NIH [HD072418].
Conflict of interest statement. None declared.

REFERENCES

1. Fu,X.D. and Ares,M. Jr (2014) Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.*, **15**, 689–701.
2. Wang,Z. and Burge,C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
3. Licatalosi,D.D., Mele,A., Fak,J.J., Ule,J., Kayikci,M., Chi,S.W., Clark,T.A., Schweitzer,A.C., Blume,J.E., Wang,X. *et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464–469.
4. Ule,J., Jensen,K.B., Ruggiu,M., Mele,A., Ule,A. and Darnell,R.B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. *Science*, **302**, 1212–1215.
5. Yano,M., Hayakawa-Yano,Y., Mele,A. and Darnell,R.B. (2010) Nova2 regulates neuronal migration through an RNA switch in disabled-1 signaling. *Neuron*, **66**, 848–858.
6. Zhang,C., Frias,M.A., Mele,A., Ruggiu,M., Eom,T., Marney,C.B., Wang,H., Licatalosi,D.D., Fak,J.J. and Darnell,R.B. (2010) Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science*, **329**, 439–443.
7. Yeo,G.W., Coufal,N.G., Liang,T.Y., Peng,G.E., Fu,X.D. and Gage,F.H. (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.*, **16**, 130–137.

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

8. Anko,M.L., Morales,L., Henry,I., Beyer,A. and Neugebauer,K.M. (2010) Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells. *Nat. Struct. Mol. Biol.*, **17**, 962–970.
9. Anko,M.L., Muller-McNicoll,M., Brandl,H., Curk,T., Gorup,C., Henry,I., Ule,J. and Neugebauer,K.M. (2012) The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.*, **13**, R17.
10. Han,J., Ding,J.H., Byeon,C.W., Kim,J.H., Hertel,K.J., Jeong,S. and Fu,X.D. (2011) SR proteins induce alternative exon skipping through their activities on the flanking constitutive exons. *Mol. Cell. Biol.*, **31**, 793–802.
11. Pandit,S., Zhou,Y., Shiue,L., Coutinho-Mansfield,G., Li,H., Qiu,J., Huang,J., Yeo,G.W., Ares,M. Jr and Fu,X.D. (2013) Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol. Cell*, **50**, 223–235.
12. Blanchette,M. and Chabot,B. (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.*, **18**, 1939–1952.
13. Cho,V., Mei,Y., Sanny,A., Chan,S., Enders,A., Bertram,E.M., Tan,A., Goodnow,C.C. and Andrews,T.D. (2014) The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA. *Genome Biol.*, **15**, R26.
14. Pan,Q., Shai,O., Lee,L.J., Frey,B.J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
15. Boutz,P.L., Bhutkar,A. and Sharp,P.A. (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.*, **29**, 63–80.
16. Lareau,L.F., Inada,M., Green,R.E., Wengrod,J.C. and Brenner,S.E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, **446**, 926–929.
17. Ni,J.Z., Grate,L., Donohue,J.P., Preston,C., Nobida,N., O'Brien,G., Shiue,L., Clark,T.A., Blume,J.E. and Ares,M. Jr (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.*, **21**, 708–718.
18. Ner-Gaon,H., Halachmi,R., Savaldi-Goldstein,S., Rubin,E., Ophir,R. and Fluhr,R. (2004) Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *Plant J.*, **39**, 877–885.
19. Munding,E.M., Igel,A.H., Shiue,L., Dorighi,K.M., Trevino,L.R. and Ares,M. Jr (2010) Integration of a splicing regulatory network within the meiotic gene expression program of Saccharomyces cerevisiae. *Genes Dev.*, **24**, 2693–2704.
20. Parenteau,J., Durand,M., Morin,G., Gagnon,J., Lucier,J.F., Wellinger,R.J., Chabot,B. and Elela,S.A. (2011) Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell*, **147**, 320–331.
21. Li,Y., Bor,Y.C., Misawa,Y., Xue,Y., Rekosh,D. and Hammarskjold,M.L. (2006) An intron with a constitutive transport element is retained in a Tap messenger RNA. *Nature*, **443**, 234–237.
22. Braunschweig,U., Barbosa-Morais,N.L., Pan,Q., Nachman,E.N., Alipanahi,B., Gonatopoulos-Pournatzis,T., Frey,B., Irimia,M. and Blencowe,B.J. (2014) Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.*, **24**, 1774–1786.
23. Yap,K., Lim,Z.Q., Khandelia,P., Friedman,B. and Makeyev,E.V. (2012) Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.*, **26**, 1209–1223.
24. Wong,J.J., Ritchie,W., Ebner,O.A., Selbach,M., Wong,J.W., Huang,Y., Gao,D., Pinello,N., Gonzalez,M., Baidya,K. *et al.* (2013) Orchestrated intron retention regulates normal granulocyte differentiation. *Cell*, **154**, 583–595.
25. Pimentel,H., Parra,M., Gee,S.L., Mohandas,N., Pachter,L. and Conboy,J.G. (2016) A dynamic intron retention program enriched in RNA processing genes regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res.* , **44**, 838–851.
26. Baek,D. and Green,P. (2005) Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12813–12818.
27. Sugnet,C.W., Srinivasan,K., Clark,T.A., O'Brien,G., Cline,M.S., Wang,H., Williams,A., Kulp,D., Blume,J.E., Haussler,D. *et al.* (2006) Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.*, **2**, e4.
28. Bejerano,G., Pheasant,M., Makunin,I., Stephen,S., Kent,W.J., Mattick,J.S. and Haussler,D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
29. Sun,S., Zhang,Z., Sinha,R., Karni,R. and Krainer,A.R. (2010) SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat. Struct. Mol. Biol.*, **17**, 306–312.
30. Zhang,D., Jiang,P., Xu,Q. and Zhang,X. (2011) Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. *J. Biol. Chem.*, **286**, 17746–17754.
31. Behzadnia,N., Golas,M.M., Hartmuth,K., Sander,B., Kastner,B., Deckert,J., Dube,P., Will,C.L., Urlaub,H., Stark,H. *et al.* (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.*, **26**, 1737–1748.
32. Bessonov,S., Anokhina,M., Will,C.L., Urlaub,H. and Luhrmann,R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature*, **452**, 846–850.
33. Deckert,J., Hartmuth,K., Boehringer,D., Behzadnia,N., Will,C.L., Kastner,B., Stark,H., Urlaub,H. and Luhrmann,R. (2006) Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.*, **26**, 5528–5543.
34. Hegele,A., Kamburov,A., Grossmann,A., Sourlis,C., Wowro,S., Weimann,M., Will,C.L., Pena,V., Luhrmann,R. and Stelzl,U. (2012) Dynamic protein-protein interaction wiring of the human spliceosome. *Mol. Cell*, **45**, 567–580.
35. Varjosalo,M., Keskitalo,S., Van Drogen,A., Nurkkala,H., Vichalkovski,A., Aebersold,R. and Gstaiger,M. (2013) The protein interaction landscape of the human CMGC kinase group. *Cell Rep.*, **3**, 1306–1320.
36. Aubol,B.E., Plocinik,R.M., Hagopian,J.C., Ma,C.T., McGlone,M.L., Bandyopadhyay,R., Fu,X.D. and Adams,J.A. (2013) Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. *J. Mol. Biol.*, **425**, 2894–2909.
37. Wang,J., Xiao,S.H. and Manley,J.L. (1998) Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. *Genes Dev.*, **12**, 2222–2233.
38. Zhong,X.Y., Ding,J.H., Adams,J.A., Ghosh,G. and Fu,X.D. (2009) Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. *Genes Dev.*, **23**, 482–495.
39. Guruharsha,K.G., Rual,J.F., Zhai,B., Mintseris,J., Vaidya,P., Vaidya,N., Beekman,C., Wong,C., Rhee,D.Y., Cenaj,O. *et al.* (2011) A protein complex network of Drosophila melanogaster. *Cell*, **147**, 690–703.
40. Blencowe,B.J., Bauren,G., Eldridge,A.G., Issner,R., Nickerson,J.A., Rosonina,E. and Sharp,P.A. (2000) The SRm160/300 splicing coactivator subunits. *RNA*, **6**, 111–120.
41. Fu,X.D. and Maniatis,T. (1992) The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3′ splice site. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 1725–1729.
42. Park,J.W., Parisky,K., Celotto,A.M., Reenan,R.A. and Graveley,B.R. (2004) Identification of alternative splicing regulators by RNA interference in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 15974–15979.
43. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
44. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
45. Chen,L.L., DeCerbo,J.N. and Carmichael,G.G. (2008) Alu element-mediated gene silencing. *EMBO J.*, **27**, 1694–1705.

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018

46. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
47. Minton,J.A., Flanagan,S.E. and Ellard,S. (2011) Mutation surveyor: software for DNA sequence analysis. *Methods Mol. Biol.*, **688**, 143–153.
48. Bresson,S.M., Hunter,O.V., Hunter,A.C. and Conrad,N.K. (2015) Canonical Poly(A) Polymerase Activity Promotes the Decay of a Wide Variety of Mammalian Nuclear RNAs. *PLoS Genet.*, **11**, e1005610.
49. Dimitrieva,S. and Bucher,P. (2013) UCNEbase–a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res.*, **41**, D101–D109.
50. Retelska,D., Beaudoing,E., Notredame,C., Jongeneel,C.V. and Bucher,P. (2007) Vertebrate conserved non coding DNA regions have a high persistence length and a short persistence time. *BMC Genomics*, **8**, 398.
51. Ovcharenko,I., Nobrega,M.A., Loots,G.G. and Stubbs,L. (2004) ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Res.*, **32**, W280–W286.
52. Cunningham,F., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
53. Yeo,G. and Burge,C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.*, **11**, 377–394.
54. Noensie,E.N. and Dietz,H.C. (2001) A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nat. Biotechnol.*, **19**, 434–439.
55. Ishigaki,Y., Li,X., Serin,G. and Maquat,L.E. (2001) 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.
56. Kuhn,A.N., van Santen,M.A., Schwienhorst,A., Urlaub,H. and Luhrmann,R. (2009) Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA*, **15**, 153–175.
57. Huang,Y., Li,W., Yao,X., Lin,Q.J., Yin,J.W., Liang,Y., Heiner,M., Tian,B., Hui,J. and Wang,G. (2012) Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Mol. Cell*, **45**, 459–469.
58. Tilgner,H., Knowles,D.G., Johnson,R., Davis,C.A., Chakrabortty,S., Djebali,S., Curado,J., Snyder,M., Gingeras,T.R. and Guigo,R. (2012) Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.*, **22**, 1616–1625.
59. Tejedor,J.R., Papasaikas,P. and Valcarcel,J. (2015) Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol. Cell*, **57**, 23–38.
60. Liao,W., Jordaan,G., Nham,P., Phan,R.T., Pelegrini,M. and Sharma,S. (2015) Gene expression and splicing alterations analyzed by high throughput RNA sequencing of chronic lymphocytic leukemia specimens. *BMC Cancer*, **15**, 714.
61. Amrein,H., Hedley,M.L. and Maniatis,T. (1994) The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by transformer 2. *Cell*, **76**, 735–746.
62. Kohtz,J.D., Jamison,S.F., Will,C.L., Zuo,P., Luhrmann,R., Garciablanco,M.A. and Manley,J.L. (1994) Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119–124.
63. Spritz,R.A., Strunk,K., Surowy,C.S., Hoch,S.O., Barton,D.E. and Francke,U. (1987) The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding. *Nucleic Acids Res.*, **15**, 10373–10391.
64. Theissen,H., Etzerodt,M., Reuter,R., Schneider,C., Lottspeich,F., Argos,P., Luhrmann,R. and Philipson,L. (1986) Cloning of the human cDNA for the U1 RNA-associated 70K protein. *EMBO J.*, **5**, 3209–3217.
65. Jumaa,H. and Nielsen,P.J. (1997) The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.*, **16**, 5077–5085.
66. Lejeune,F., Cavaloc,Y. and Stevenin,J. (2001) Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. Identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J. Biol. Chem.*, **276**, 7850–7858.
67. Sureau,A., Gattoni,R., Dooghe,Y., Stevenin,J. and Soret,J. (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.*, **20**, 1785–1796.

Downloaded from https://academic.oup.com/nar/article-abstract/45/6/3473/2605754
by guest
on 30 July 2018