4-26-2017

An Ultraconserved Element Controls Homeostatic Splicing of ARGLU1 mRNA

Stephan P. Pirnie
University of Connecticut School of Medicine, spirnie@uchc.edu

An Ultraconserved Element Controls Homeostatic Splicing of *ARGLU1* mRNA

Stephan P Pirnie, PhD

University of Connecticut, 2017

Abstract

ARGLU1 is a well conserved protein whose function is not well understood. I will show that *ARGLU1* is an alternatively spliced gene, with at least three alternatively spliced isoforms. The main isoform codes for a nuclear protein that has been associated with the mediator transcriptional regulation complex as well as components of the spliceosome. One alternative isoform of *ARGLU1* retains a single unspliced intron, even though all other introns have been removed, and is localized exclusively in the nucleus. A second alternative isoform causes inclusion of a premature termination codon, and is quickly degraded by the nonsense mediated decay quality control pathway. Interestingly, the retained intron contains an ultraconserved element, which is more than 95% conserved between human and chicken for over 500 bases. I will show that this ultraconserved element plays a key role in the alternative splicing of *ARGLU1*. Furthermore, I will show that exogenous overexpression of ARGLU1 leads to dramatic changes in alternative splicing of its own endogenous mRNA, causing a decrease in the protein coding isoform, and an increase in the retained intron and nonsense mediated decay targeted isoforms. Additionally, overexpression of ARGLU1 causes changes in mRNA levels and alternative splicing in a number of genes. Taken together, these results indicate that ARGLU1 can regulate its own splicing to regulate cellular protein levels. Furthermore, these results suggest that ARGLU1 plays a role in cellular alternative splicing.

An Ultraconserved Element Controls Homeostatic Splicing of *ARGLU1* mRNA

Stephan P Pirnie

B.S., University of New Hampshire, 2007

PhD, University of Connecticut, 2017

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

Approval Page

Doctor of Philosophy Dissertation

An Ultraconserved Element Controls Homeostatic Splicing of *ARGLU1* mRNA

Presented by:

Stephan P Pirnie, B.S.

Major Advisor:

Gordon G Carmichael

Associate Advisor:

Stormy J Chamberlain

Associate Advisor:

Brenton R Graveley

Associate Advisor:

Arthur Günzl

University of Connecticut

2017

Dedication

I would like to dedicate this dissertation to my family, for always being there to support me through this long and difficult process. My sister, Claire Campbell, was my first and longest friend. As we've both grown up and gone through graduate school (surprisingly both at University of Connecticut) we have only grown closer. My parents, Peter Pirnie and Diane Arsenault, have been a constant supportive force throughout my life. They have encouraged me to pursue the things that I'm passionate about, be it science or theater or traveling. They have always fostered an environment of education at home, and encouraged me to attend the program that sent me down the path of molecular biology and genetics at St. Paul's School.

St. Paul's played an important role in meeting the person I'd like to thank most, my wife Dr. Loreen Pirnie. We met at St. Paul's in the summer course on Molecular Biology almost 15 years ago. At that time, I liked her already, but was too shy to let her know. Luckily, we both ended up at University of New Hampshire, and she was willing to go out with me. Since then, our paths have taken twists and turns, but they are always headed the same direction, even if they did take some detours at times. She has been so patient and so flexible throughout this long trek through graduate school. It has not always been easy, or fun, and I've not always been the most pleasant. Together we've made it through, because of her support. We'll look back on this as a difficult time, but I know that it has brought us closer together because of the miles that have been between us at times.

It has been a great 8 years, but I'm very excited to see where our paths lead us next. We have a new person on our journey together, tiny Reid. He's come along just as this part of this adventure is ending, but I know that we'll all be able to tell him good stories about our time at UConn.

Acknowledgements

I would like to thank all of the members of the Carmichael lab for their tireless support through this difficult process. Thanks to Joe Autoro for many stimulating and fun conversations, as well as for bring a sounding board for new idea. I would like to thank Seth Garren for welcoming me to the lab when I started, for providing wonderful scientific advice, and for always being up for an interesting conversation. I would like to thank Yinzhou Zhu for his help in the lab, as well as carrying on work on 2'-O RNA methylation that we began in the lab together, as well as his insightful thoughts. I would like to thank Ahmad Osman for his help with continuing the work on understanding the function of ARGLU1, as well as for his good humor and ready smile. I would like to thank Na Zhang for her great troubleshooting ideas, as well as her wonderful company in the lab. I would especially like to thank Kim Morris for her endless emotional and technical support. She has been such a pillar of support for generations of graduate students in Gordon's lab. She has been there for me through every failed experiment and not only helped me feel better, but provided suggestions to improve. She's also been a trusted friend to help steer a steady course through the personal difficulties that are inevitable in grad school. Finally, I would like to thank Gordon. He has been everything I could have hoped for in an advisor and mentor. From the first day, he has had trust in me to decide how to approach this project. He has given me space to learn and gain confidence in my ability to ask questions and carry out scientific research, but has always had his door open to have candid conversations and excellent technical advice.

Thank you also to my committee members, Brent Graveley, Stormy Chamberlain and Arthur Günzl for your invaluable insight, guidance and support. I am very appreciative to this wonderful team of mentors.

Thank you all so much.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 ARGLU1

This paper will focus on the regulation of a protein called Arginine and Glutamate rich protein 1, hereafter referred to as ARGLU1. The function of ARGLU1 has been poorly described, but a number of factors indicate that it is of biological importance. First, it has been shown that depletion of ARGLU1 protein leads to defects in cellular responsiveness to exogenous stimuli such as estrogen (Zhang et al., 2011). Second, ARGLU1 protein sequence is well conserved through evolutionary time, indicating a conserved function (Figure 1.1). In addition to this conservation, *ARGLU1* also has a region of extraordinarily high non-coding DNA sequence conservation, suggesting a consistent selective pressure not only at the level of amino acid sequence (Dimitrieva and Bucher, 2013). Such highly conserved elements have been shown to have a variety of important regulatory effects, including transcriptional enhancer activity and regulation of alternative splicing, which will be discussed below.

ARGLU1 is encoded by a gene located at human chromosome 13q33.3. Initial reports involving *ARGLU1*, which was previously annotated as FLJ10154, focused on rare deletions of the 13q33-34 region that overlap *ARGLU1* and are associated with microcephaly, developmental delay, and genitourinary malformations in males (Andresen et al., 2010; Ballarati et al., 2007; Brown et al., 1993, 1995; Quélin et al., 2009; Walczak-Sztulpa et al., 2008). While the adjacent gene *EFNB2* was posited to be the gene responsible for these malformations, *ARGLU1* is also deleted in all reports of genitourinary malformations associated with deletions from this region.

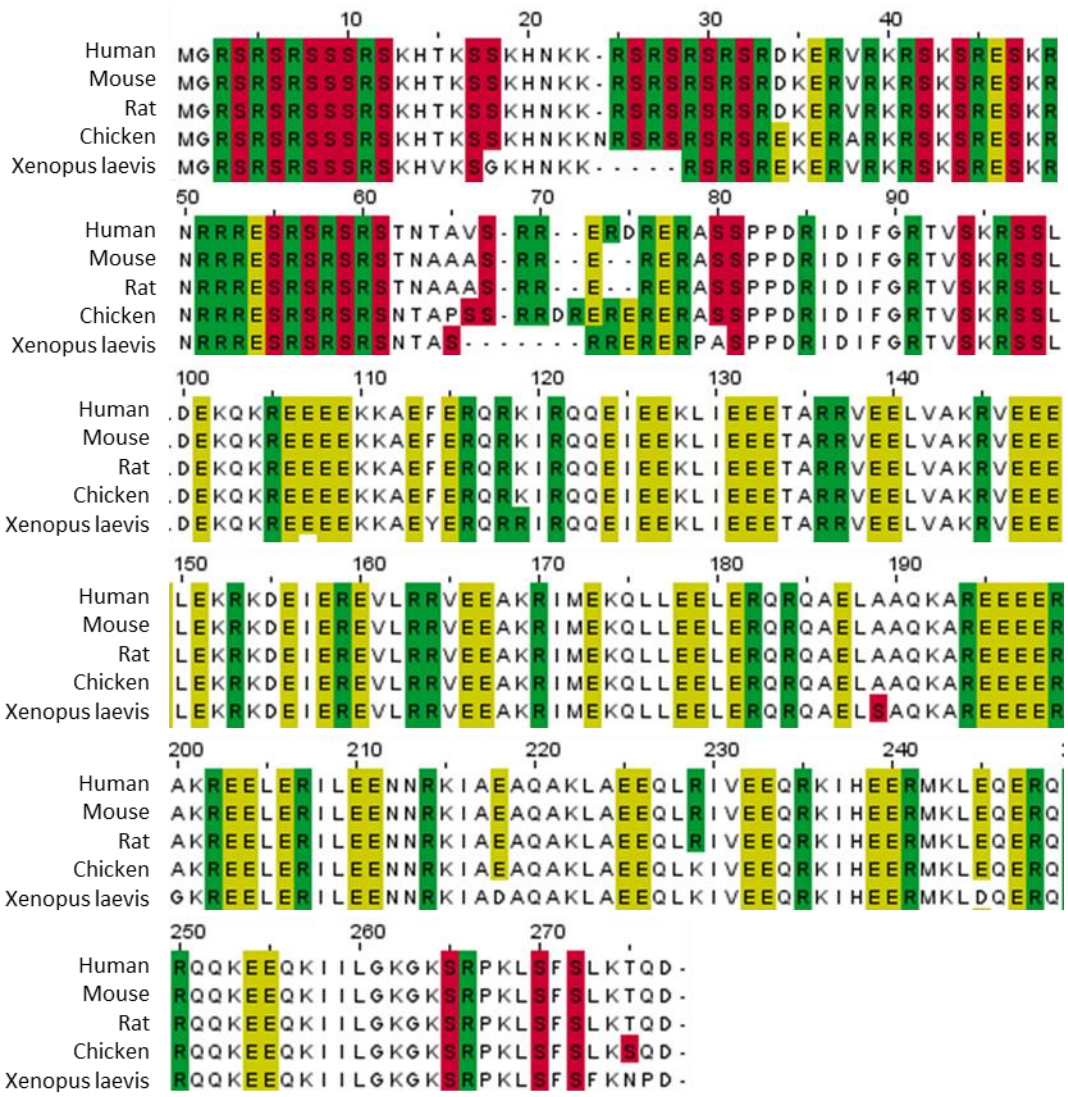

Figure 1.1 ARGLU1 Protein Conservation. Multiple-alignment view of ARGLU1 proteins from a variety of species, including the frog *Xenopus laevis.* Serine residues are highlighted in red, arginine residues in green, and glutamate residues in yellow. Note the enrichment of RS dipeptides in the N-terminal region, and the enrichment of glutamate in the c-terminal portion of the protein.

A 2009 study investigating glomerulus specific gene expression in the human kidney indicated that *ARGLU1* was expressed exclusively in the glomerulus, but not in the cortex or medulla, of the kidney (Cuellar et al., 2009). This study used an RT-PCR based assay to demonstrate this exclusive expression. Interestingly, this study used primers that amplify an alternatively spliced form of the *ARGLU1* transcript that retains an intron between exons 2-3 (Cuellar et al., 2009), which will be the focus of much of this dissertation. Therefore, this report would seem to indicate that the intron retained isoform is expressed in a very tissue specific manner. Other reports have also indicated that the retained intron isoform of *ARGLU1* is alternatively spliced in a tissue dependent manner (Braunschweig et al., 2014).

The most in depth investigation of the function of ARGLU1 protein was made by Zhang et al. in 2011. This study sought to find novel protein interactions with the Mediator complex protein MED1 (Zhang et al., 2011). This study identified ARGLU1 as a MED1 interacting protein through affinity purification of MED1 from MCF7 cells, an estrogen receptor positive breast cancer cell line, followed by mass spectrometry (Zhang et al., 2011). They further characterized the protein, finding that it was nuclearly localized, and directly interacts with the C-terminus of MED1 through its N-terminal arginine-rich domain (Zhang et al., 2011). Interestingly, a different report analyzing the subcellular localization of proteins throughout the cell cycle demonstrated that ARGLU1's nuclear localization was highly dependent on cell cycle stage, peaking in mid-S-phase, and with the lowest nuclear localization during the G1 phase (Sigal et al., 2006). As Zhang et al. were interested in new regulators of MED1's role in transcriptional licensing in response to estrogen, they showed that ARGLU1 was preferentially co-recruited with MED1 to the promoter of the estrogen responsive genes *MYC* and *pS2*, but not the control gene *GAPDH*, following estradiol (E2) treatment (Zhang et al., 2011). They further found that shRNA mediated knock

down of *ARGLU1* or *MED1* decreased expression of these E2 responsive genes, but not *GAPDH*. Finally, they showed that shRNA-mediated knockdown of ARGU1 mRNA and protein caused decreased basal and E2 responsive cellular proliferation and anchorage independent colony formation (Zhang et al., 2011). Taken together, this study indicates that ARGLU1 interacts specifically with the MED1 component of the mediator complex, and functions to regulate the expression of at least two important estrogen responsive proto-oncogenes.

1.2 Mediator Complex

Mediator is a large, multi-subunit complex that is involved in licensing of RNA polymerase II in transcription. Its co-activator function is mediating the interaction of enhancer-bound transcription factors and activators, which may be bound to distal enhancer regions, with the core RNA pol II pre-initiation complex located at the core gene promoter region (Poss et al., 2013). Models of the interaction between the mediator complex and RNA pol II indicate multiple points of interaction, but perhaps the most interesting is the interaction with the carboxy-terminal domain of RPB1, the largest subunit RNA pol II, which is composed of 52 heptad repeats that are extensively and dynamically modified during the transcription cycle (Bentley, 2014). This subunit of RNA pol II serves as a hub for transient interactions of many proteins, including mediator, splicing factors, histone modifiers. It serves as a way to functionally link these factors involved in transcription (Bentley, 2014). Indeed, the MED23 subunit of mediator was recently found to interact with a number of proteins involved in mRNA splicing and maturation (Huang et al., 2012). This study focused on the interaction between MED23 and hnRNPL, a gene involved in mRNA splicing and stability, and found that in the absence of MED23 splicing of a subset of hnRNPL-regulated genes was altered (Huang et al., 2012). This mediator-splicing factor interaction, possibly potentiated by the CTD of RNA pol II, may serve as a model the interplay between

transcriptional activators and splicing regulation. Given ARGLU1's interaction with mediator and findings, discussed later, that ARGLU1 is involved in alternative splicing, this type of interaction will be of future interest.

1.3 Ultraconserved Elements

Advances in sequencing technology over the last two decades have laid the foundations for a number of advances in our understanding of genome biology (Lander et al., 2001; Venter et al., 2001) including comparative genomic studies that have furthered our understanding of evolution at the nucleotide level (Waterston et al., 2002). One of the goals of comparative genomics is to use evolutionary selection as a marker of biological function (Alföldi and Lindblad-Toh, 2013). An important initial study comparing the human genome with that of mice and rats revealed hundreds of ultra-conserved elements (UCEs) that they defined as longer than 200 bases of 100% sequence identity (Bejerano et al., 2004). Subsequent analysis found that a set of similarly highly conserved non-coding elements functioned as enhancers associated with development (Woolfe et al., 2005), or clustered near developmentally regulated genes (Lindblad-Toh et al., 2005), indicating that such conservation can indicate regulatory sequences. Further studies have taken a slightly different approach, looking for ultra-conserved non-coding sequences over longer evolutionary time scales; for example sequences that are more than 95% identical and over 200 bases between human and chicken (Dimitrieva and Bucher, 2013). My work has used this standard of over 95% conserved nucleotide identity over more than 200 bases between human and chicken to define a UCE.

While one set of highly conserved sequence seems to function in the context of transcriptional regulatory regions, another subset of UCEs are postulated to have an RNA regulatory function. One method of analysis of conserved function is to use computational

predictions of RNA folding. Indeed, reports using computational folding predictions based on highly conserved non-CDS regions are able to predict transcribed non-coding RNAs (ncRNA) with fairly high confidence (Washietl et al., 2005). Another report by the same group was able to identify ncRNA structural conservation even without filtering based on sequence conservation, albeit with roughly twice the false positive rate for truly transcribed ncRNAs when compared with conservation based prediction methods (Washietl et al., 2007). Other studies have found that sequence conservation is not always a predictor for conserved ncRNA function, good examples being certain miRNAs, *Xist*, *Air*, and *H19*, which have known and conserved functions but poor sequence conservation (Bentwich et al., 2005; Pang et al., 2006). So while a high degree of conservation may predict ncRNA structure and function in some cases, it is not the only predictor.

Another RNA function for non-coding UCEs may be in RNA splicing, processing and localization. In this context, these highly conserved sequences would be predicted to serve as RNA *cis*-elements which would interact with *trans*-elements, namely proteins and protein complexes. A notable example of this is the *β-actin* mRNA, which is localized to the dendrites of hippocampal neurons (Tiruchinapalli et al., 2003) or other subcellular localizations by an interaction between a 3' UTR *cis*-element called zipcode and the RNA recognition motif (RRM) and hnRNP K homology (KH) domains of zipcode binding protein 1 (Farina, 2002; Ross et al., 1997). However, these types of *cis/trans* interactions are usually on the order of tens of bases. Indeed the zipcode sequence itself is 54 bases long (Kislauskis et al., 1994), whereas ultraconserved elements require at least 200 bases of conservation. One could thus hypothesize that such long *cis*-elements may be concatenations of binding sites for multiple *trans*-factors, or that they coordinate multiple types of regulation, such as splicing, sub-cellular localization, nuclear-cytoplasmic trafficking and translation.

Indeed, highly- and ultra-conserved sequences have been associated with conserved and tissue specific alternative splicing events (Baek and Green, 2005; Lareau et al., 2007; Ni et al., 2007; Sugnet et al., 2006). Of particular interest are the association between highly conserved sequences and alternative splicing of many SR and hnRNP proteins, which are themselves regulators of splicing (Lareau et al., 2007; Ni et al., 2007). It is now well established that these alternative splicing events are linked with a cellular quality control mechanism known as nonsense mediated decay (NMD) (Baek and Green, 2005; Lareau et al., 2007; Lewis et al., 2003; Ni et al., 2007); however another recent study has linked UCE-related intron retention alternative splicing to a different regulator mechanism that they labeled 'detained introns' (Boutz et al., 2015). While I want to highlight this association between highly conserved sequences and alternative splicing, I will further discuss the critical splicing-associated proteins regulated by these elements, their roles in splicing, as well as the mechanisms and outcomes of their alternative splicing in a later section.

1.4 RNA splicing—Coordination of spliceosome assembly by *cis*- and *trans*-acting factors

RNA splicing is a process by which exon sequences from a primary RNA transcript are brought together, and the intervening introns are removed. This process was first appreciated in adenovirus (Berget et al., 1977; Chow et al., 1977). Sharp and colleagues observed R-looping of DNA following hybridization of polyribosome-associated viral RNA to viral DNA treated with restriction enzymes, and proposed that this was the result of bringing together and joining specific sequences within large, nuclear viral transcripts (Berget et al., 1977). In a report published just a month later, Richard Roberts and colleagues reported similar results mapping R-loop interactions, and concluding that these sequences were not coded linearly by the DNA genome of the virus (Chow et al., 1977). The report by the Sharp group (Berget et al., 1977), and a separate report by

Daniel Klessig (Klessig, 1977), proposed a model that is in line with our current understanding of RNA splicing, whereby the 3'-end of one segment is brought together with the 5'-end of the next segment and the intervening sequences are looped out and removed. Shortly thereafter, a number of groups reported splicing of a wide variety of metazoan genes, including the rabbit (Jeffreys and Flavell, 1977) and mouse (Tilghman et al., 1978) β-globin genes, the chicken ovalbumin gene (Breathnach et al., 1977), and the immunoglobulin light chain (Tonegawa et al., 1978).

Development of a method to splice RNA molecules *in vitro* (Grabowski et al., 1984; Kole and Weissman, 1982; Krainer et al., 1984; Padgett et al., 1983; Ruskin et al., 1984) allowed for further dissection and identification of the component sequences of a splicing reaction, namely the 5'-splice site and 3'-splice site (Mount, 1982), and branch point (Konarska et al., 1985). The splicing reaction came to be understood to be two sequential transesterification reactions. In the first step, the 5'-splice site, with a consensus sequence of MAG|GURAGU where M is A/C and R is A/G, is brought into contact with the branch point, an AU upstream of the 3'-splice site, and the 5' G of the GU dinucleotide is joined by a 2'-5' phosphodiester linkage to adenosine of the branch point (Konarska et al., 1985), forming a structure known as a lariat (Padgett et al., 1984). In the second step of splicing, the 3'-splice site, which has a consensus YAG|G, where Y is a pyrimidine, is cleaved following the AG, and the 5'-end of the 3' exon is joined to the 3'-end of the of the upstream exon by a 3'-5' phosphodiester bond (Grabowski et al., 1984).

During this same time, it was recognized that snRNP complexes, composed of small nuclear RNAs (snRNAs), as well as protein components interact with specific sequences on the RNA (Chabot et al., 1985) and function to mediate the splicing reaction by dynamically bringing together the mRNA components of the splicing reaction. These snRNP complexes were first

identified using anti-nuclear antibodies produced by patients with Systemic Lupus Erythematosus (Lerner and Steitz, 1979).

This dynamic macromolecular complex, called the spliceosome, is characterized by the formation of a number of distinct complexes that coordinate the two transesterification reactions of RNA splicing. The first complex formed, complex E, begins with U1 snRNP binding through U1 mediated base pairing to the 5'-splice site. Additionally during complex E formation, a protein called splicing factor 1 (SF1) or branch point binding protein (BPB) is recruited to the branchpoint, and the subunits of the U2 auxiliary factor A2, U2AF65 and U2AF35 (Zamore et al., 1992), bind to the polypyrimidine tract and the characteristic 3'splice site AG, respectively (Wahl et al., 2009). Prior to and during this stage proteins, including the SR and hnRNP families, are recruited to the RNAs through *cis*-elements and can promote or repress binding of these complex E factors. These regulatory proteins will be discussed in greater detail in later sections. Complex A is then formed, with the displacement of SF1/BPB by the U2 snRNP at the branch point (Wahl et al., 2009). Next, complex B is formed by the recruitment of the U4/U5/U6 tri-snRNP, with U5 base-pairing to the 5'-splice site (Wahl et al., 2009). This leads to destabilization of the complex, and the U1 and U4 snRNPs exit the complex. At the same time, U6 base pairs with the 5'-splice site, displacing U5. With the hydrolysis of ATP, the catalytic spliceosome is generated and catalyzes the first transesterification reaction in which the 2' hydroxyl of the branch point adenosine attacks the 5' phosphate of the 5'-splice site guanidine, forming the lariat. The complex then rearranges again, with U5 binding to the 3'-splice site, bringing the upstream and downstream exons together. With U2/U5/U6 still bound, the free 3'-hydroxyl of the 5'-exon attacks the 5' phosphate bond of the 3'-ss, and the ligation reaction is carried out following ATP hydrolysis. (Will and Lührmann, 2011). This process is well reviewed elsewhere (Wahl et al., 2009; Will and Lührmann, 2011).

There is a large body of work that has sought to define the complement of proteins involved in the spliceosome. One productive system has been affinity purification of proteins and RNAs associated with different stages of spliceosome assembly followed by mass spectrometry. Interestingly, ARGLU1 protein has been found in multiple preparations of spliceosome complexes, including complex A (Behzadnia et al., 2007), the active step 1 spliceosome (Bessonov et al., 2008), and purified complex B (Deckert et al., 2006). Based on the finding that ARGLU1 was associated with the spliceosome, it was included in a yeast 2-hybrid screen for protein-protein interactions between spliceosome proteins. In that experiment, ARGLU1 was found to interact with itself, CHERP, PRMT5, SRPK2, and ZCCHC10 (Hegele et al., 2012). Another high throughput protein-protein interaction screen, this time in drosophila, found that the ARGLU1 homolog CG31712 was associated with spliceosome components, including the U1-70K protein and CG4119 (Guruharsha et al., 2011). A high throughput assay of protein kinase targets identified ARGLU1 as a phosphorylation target of both SRPK1 and SRPK2 protein kinases (Varjosalo et al., 2013). A list of proteins interacting with ARGLU1 is found in Table 1.1. These interactions are especially interesting in the context of our experimental findings, as the SRPK protein kinases are known to regulate SR proteins, an important class of splicing regulators that will be covered in a later section in more detail. Taken together, these interactions indicate that ARGLU1 interacts with components of the spliceosome. However, proteomic studies designed to determine the full array of RNA binding proteins have not provided evidence that ARGLU1 interacts directly with mRNA (Baltz et al., 2012; Castello et al., 2012).

Official Symbol Interactor A	Official Symbol Interactor B	Experimental System	Experimental System Type	Author	Pubmed ID
IKBKG	**ARGLU1**	Reconstituted Complex	physical	Fenner BJ (2010)	20098747
ELAVL1	**ARGLU1**	Affinity Capture-RNA	physical	Abdelmohsen K (2009)	19322201
SF3A2	**ARGLU1**	Affinity Capture-MS	physical	Behzadnia N (2007)	17332742
CDK2	**ARGLU1**	Affinity Capture-MS	physical	Neganova I (2011)	21319273
ARGLU1	**APP**	Reconstituted Complex	physical	Olah J (2011)	21832049
CHERP	**ARGLU1**	Two-hybrid	physical	Hegele A (2012)	22365833
ARGLU1	**CHERP**	Two-hybrid	physical	Hegele A (2012)	22365833
ARGLU1	**ARGLU1**	Two-hybrid	physical	Hegele A (2012)	22365833
ARGLU1	**PRMT5**	Two-hybrid	physical	Hegele A (2012)	22365833
ARGLU1	**ZCCHC10**	Two-hybrid	physical	Hegele A (2012)	22365833
ARGLU1	**SRPK2**	Two-hybrid	physical	Hegele A (2012)	22365833
PRMT5	**ARGLU1**	Two-hybrid	physical	Hegele A (2012)	22365833
ZCCHC10	**ARGLU1**	Two-hybrid	physical	Hegele A (2012)	22365833
SRPK2	**ARGLU1**	Two-hybrid	physical	Hegele A (2012)	22365833
SRPK2	**ARGLU1**	Biochemical Activity	physical	Varjosalo M (2013)	23602568
SRPK1	**ARGLU1**	Biochemical Activity	physical	Varjosalo M (2013)	23602568
CUL7	**ARGLU1**	Affinity Capture-MS	physical	Hanson D (2014)	24711643
EED	**ARGLU1**	Affinity Capture-MS	physical	Cao Q (2014)	24457600

Table 1.1 Known Protein-Protein interactions for ARGLU1

1.5 Intron and Exon Definition

There are two models for how initial assembly of the spliceosome components occurs: intron definition and exon definition. The intron definition model posits that components of the spliceosome assemble and interact across small introns (Talerico and Berget, 1994), where the 5' and 3' splice sites are close to each other in space. *In vitro* and transfection splicing assays have shown that as intron length rises above 250 nucleotides, splicing becomes quite inefficient (Fox-Walsh et al., 2005). This same study found that increased inclusion of human alternatively spliced exons was associated strongly with a shorter upstream intron (Fox-Walsh et al., 2005). Furthermore, this study found that inclusion of an RNA *cis*-element known as an exonic splice enhancer (ESE) could dramatically increase splicing efficiency of even longer introns (Fox-Walsh et al., 2005). While intron definition mediated splicing seems to be the exception rather than the rule in mammals, it is seen almost exclusively in plants, fungi and invertebrates (Talerico and Berget, 1994; Xiao et al., 2007). Consequently, vertebrates have developed a complex set of regulatory elements to enforce the correct usage of exons within the context of long introns that often contain pseudoexons with paired 3'- and 5'-splice sites (Xiao et al., 2007). These elements will be discussed further below.

The second model for assembling spliceosome components is called exon definition, and occurs by the assembly of snRNP complexes to either end of an exon (Robberson et al., 1990; Sterner et al., 1996). These assembled mRNA/snRNP complexes can then form associations across introns mediated by interactions between snRNPs and other splicing associated proteins. The exon definition model is supported by evidence that U1 snRNP binding to the 5'ss of an exon stimulated splicing of the upstream 5'ss to the 3'ss of that exon (Kuo et al., 1991). It was found that polypyrimidine tracts play an important role in potentiating the exon definition model by

recruiting important protein factors such as U2AF (Dominski and Kole, 1992; Talerico and Berget, 1994). Similar to the intron definition model, there appears to be a size limit to the exon definition model, with failure of exon definition by *in vitro* splicing inclusion of exons larger than approximately 300 bases (Robberson et al., 1990), while exons larger than approximately 500 bases are poorly included in transfection splicing studies, especially in the context of long introns (Sterner et al., 1996). Thus, there seem to be limits on the size of either exons or introns for their definition in splicing.

Early studies indicated that there is also a lower size limit for exon definition of approximately 50 nt (Dominski and Kole, 1991), although inclusion of small exons could be modulated by altering the splice site strength or upstream polypyrimidine tract and branch point (Dominski and Kole, 1991, 1992). The use of high throughput RNA sequencing techniques has allowed for the identification of a large number of 3-27 nt small exons, which they have termed microexons, which are used in neurons and are misregulated in the autistic brain (Irimia et al., 2014). Interestingly, there have been a number of groups that have identified zero length exons, which function in the splicing of exceptionally long exons in drosophila (Burnette, 2005; Duff et al., 2015; Hatton et al., 1998) and humans (Duff et al., 2015; Sibley et al., 2015) through a process called recursive splicing where the removal of an intron regenerates a 5'ss that can be used for further splicing. These types of events have challenged the classical concept of an exon, as they are simply the juxtaposition of a 3' and 5'-splice site.

1.6 Alternative splicing

One of the earliest reported examples of RNA splicing was the human immunoglobulin light chain (Tonegawa et al., 1978). Soon after this discovery, the immunoglobulin gene played an important role in understanding how RNA splicing lead to the incredible diversity of molecules

derived from this locus (Alt et al., 1980; Early et al., 1980) through a process known as alternative splicing (AS), whereby a precursor RNA (pre-mRNA) transcript can be processed to include or exclude exons in the final mature mRNA. Alternative splicing plays a number of important roles in cells, including expanding the protein repertoire of cells (Nilsen and Graveley, 2010), allowing temporal and tissue specific control of protein expression (Wang et al., 2008), and response to cellular stimuli (Braunschweig et al., 2013). AS seems to be an important and widely used mechanism, as recent studies have concluded that approximately 95% of all genes in the human genome generate more than one spliced isoform (Pan et al., 2008; Wang et al., 2008). Technological advances in the past two decades such as microarray and more recently RNA-Seq technologies have allowed for mapping and quantitation of an ever increasing number of alternatively spliced RNAs across a host of species.

There are a number of common types of alternative splicing (Fig 1.2). Most common is the cassette exon, which is a simple inclusion/exclusion event. Other types include alternative 5'-splice site, alternative 3'-splice site, mutually exclusive exon usage, and intron retention. More complex sets of alternative splicing can be created by combining these types of alternative events. An example relevant to our study is the combination of an intron retention event and a cassette exon with alternative 5'-splice sites on the cassette exon. It is easy to see how this can rapidly increase the number of isoforms derived from one pre-mRNA.

Alternative splicing is regulated by the interactions between *cis*-elements within RNA molecules and *trans*-acting proteins that bind to these RNA sequences. We will discuss the details of these regulatory elements below.

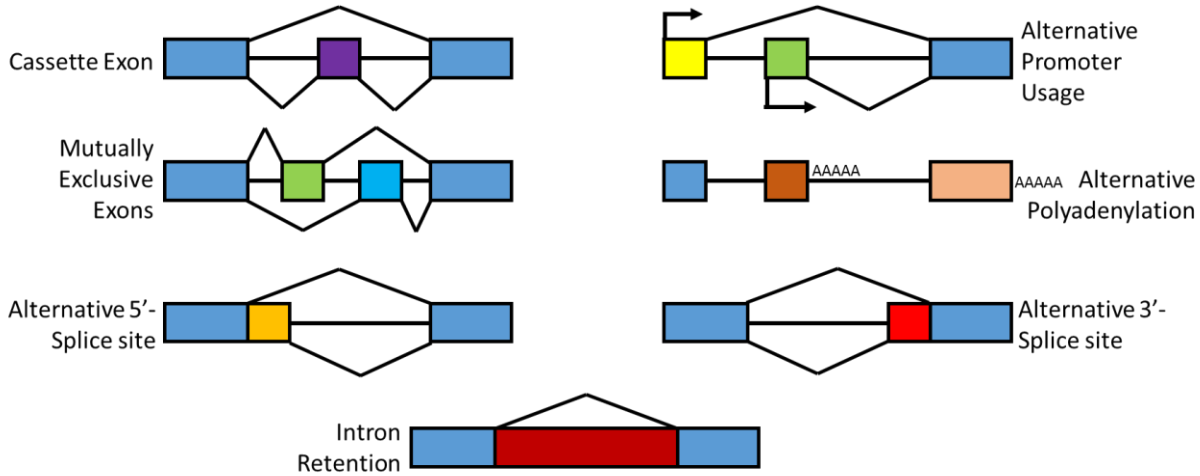

Figure 1.2 Types of alternative splicing. The various types of alternative splicing are depicted in cartoon form. Constitutive exons are colored blue, while alternative exon are colored to differentiate them.

1.7 *Cis*-elements regulating splicing

The exon definition process, discussed earlier, is modulated through the interaction of *cis*-acting RNA elements with *trans*-acting proteins, such as SR or hnRNP proteins (Caceres et al., 1994). The *cis*-elements that inform exon definition, generally known as Splicing Regulatory Elements (SREs), can be defined as exonic splice enhancers of suppressors (ESEs or ESSs) or intronic splice enhancers or silencers (ISEs or ISSs), based on their location and propensity to promote or suppress inclusion of the associated exon (Wang and Burge, 2008). The 5'- and 3'-consensus splice sites are only 9 bases and 4 bases, respectively, and thus found quite frequently in the genome. Furthermore, the introns separating these sequences are often hundreds or thousands of bases long, making these *cis*-elements critical for more specificity in the recruitment of the core spliceosomal components (Wang and Burge, 2008).

Many studies have defined *cis*-elements using a variety of techniques. Classic studies identified individual *cis*-elements through molecular biology techniques such as mutagenesis and deletion of specific sequences proximal to regulated exons within minigene reporters (Schaal and Maniatis, 1999; Sun and Chasin, 2000). Other SREs have been deduced from pathogenic SNPs linked to modification of splicing, notably in the CFTR gene (Pagani et al., 2003), and the SMN2 gene (Cartegni and Krainer, 2002). However, high throughput screens and computational approaches have led to more general identification of SREs. For example, cloning of random sequences into the middle exon of multi-exon reporters followed by selection for inclusion of that exon has been used to identify ESE sequences in both *in vitro* (Tian and Kole, 1995) and *in vivo* splicing assays (Coulter et al., 1997), both of enriched for purine-rich sequences, similar to analysis of individual ESEs that had been found to interact with SR proteins (Dirksen et al., 1994; Lavigueur et al., 1993; Lynch and Maniatis, 1995; Tanaka et al., 1994). Another useful high

throughput approach has been to split green fluorescent protein with an internal exon that disrupts its fluorescence. The inclusion or exclusion of this internal exon can be modulated by insertion of random sequences within either the exon or in adjacent introns. This system was first used to screen random decamers for their ability to cause skipping of the internal exon and thus green fluorescence, indicating their ability to function as ESSs (Wang et al., 2004). More recently, this same group has used a similar approach to screen both ISSs (Wang et al., 2013) and ISEs (Wang et al., 2012). The information gained from such high-throughput screens can then be applied to predict splicing regulation of other exons.

1.8 *Trans*-acting Splicing Regulators

These *cis*-elements act through interactions with *trans*-acting proteins that serve to promote or inhibit spliceosome subunit binding, based on the RNA element and the protein recruited (Wang and Burge, 2008). These *trans*-acting proteins mostly fall into two large classes of proteins, although there are other splicing regulatory proteins that do not easily fit into either class (Fu and Ares, 2014). Early studies showed that stereotypical proteins within these two families of proteins, SR proteins and hnRNP proteins, had generally antagonistic effects on the inclusion of alternatively spliced exons (Caceres et al., 1994; Mayeda and Krainer, 1992; Mayeda et al., 1994), although more recent studies have indicated a more nuanced effect of these proteins on exon inclusion/exclusion (Huelga et al., 2012; Pandit et al., 2013).

The first class of *trans*-acting regulators are the hnRNP proteins, which are generally held to have an inhibitory role in exon inclusion. The most well studied hnRNP proteins have been shown to have splicing inhibitory roles, and include hnRNPA/B family (Mayeda and Krainer, 1992; Mayeda et al., 1994; Yang et al., 1994) as well as polypyrimidine tract-binding protein (PTB) (Singh et al., 1995). The mechanism by which these proteins cause inhibition of exon

inclusion is still unclear, although recent studies have begun to explain their effects. Studies have proposed that hnRNP A/B initially bind to ISS or ESS sequences and subsequently oligomerize to cover nearby RNA sequences, thereby preventing binding of SR proteins that promote splicing (Okunola and Krainer, 2009; Zhu et al., 2001). A recent study sought to understand the genes regulated by a variety hnRNP proteins in human cells. They used both siRNA mediate knockdown to identify alternative splicing events regulated by these proteins, as well as a technique known as CLIP-seq to map the interactions between the hnRNP proteins and their target RNAs, which revealed (Huelga et al., 2012). This study produced a number of interesting results, including the fact that knockdown of hnRNP proteins caused the increased or decreased inclusion of a large number of cassette exons, indicating that hnRNP proteins are not exclusively inhibitory, but play a more complicated role in balancing the usage of certain exons. Furthermore, they presented evidence for hnRNPs cross-regulating other RNA binding proteins, including other hnRNPs and SR proteins (Huelga et al., 2012).

PTB is another RNA binding protein generally thought to be inhibitory to exon inclusion which has been shown to interact with intronic polypyrimidine tracts upstream of 3'-splice sites (Singh et al., 1995) and is proposed to block cross-intron interactions between the 5'- and 3'-splice sites after assembly of the A complex and exon definition (Sharma et al., 2005, 2008). Additionally, PTB has been shown to interact directly with U1 at the 5'ss and stabilizing the interaction (Sharma et al., 2011). This particular study may be of further interest, as it suggests a mechanism for inhibiting splicing of upstream and downstream exons over long distances, by locking them in a conformation that does not allow progression beyond the spliceosomal E complex, thus stalling the splicing reaction (Roca et al., 2013). Deletion of the mouse homolog of PTB, Ptbp1, was shown to regulate a number of neuron specific intron retention events, causing

the nuclear retention of these transcripts (Yap et al., 2012). Other studies have shown that binding of SR protein SRSF7 to intronic regions downstream of exons will inhibit their splicing by preventing progression beyond the spliceosomal E complex, while still allowing recruitment of U1 to the 5'-splice site (Erkelenz et al., 2013), perhaps suggesting a similar role to PTB. The role of hnRNPs seems to be complicated, and may range from negatively modulating the recruitment of spliceosome components through competition with SR proteins to inhibiting the rearrangement of these components into a committed complex.

The second large family of splicing regulatory proteins are known as SR proteins (for serine/arginine-rich proteins). SR proteins take their name as the result of a characteristic C-terminal RS domain, which is enriched in clustered Serine-Arginine dipeptides. Furthermore, these proteins have one or two N-terminal RNA-recognition motifs (RRMs), similar to many other RNA binding proteins including hnRNP proteins (Bandziulis et al., 1989; Zahler et al., 1992). These proteins are known to localize to subnuclear bodies known as splicing speckles, which are repositories for splicing factors (Spector et al., 1991). Furthermore, the function and localization of SR proteins is regulated by the phosphorylation state of their RS domains (Cáceres et al., 1997), which is dependent on multiple protein kinases, including SRPK1 (Gui et al., 1994), SRPK2 (Wang et al., 1998), and the CLK kinases (Colwill et al., 1996), as well as long non-coding RNAs such as MALAT1 (Tripathi et al., 2010).

The presence and function of the SR family of proteins was first recognized in the *Drosophila* development system, where a cascade protein factors regulate sex determination through auto-regulatory splicing as well as alternative splicing of the downstream transcripts (Amrein et al., 1988, 1994; Bell et al., 1988; Boggs et al., 1987; Goralski et al., 1989). Shortly thereafter, a number of human proteins were found that shared protein sequence and functional

similarities to the splicing regulators found in *Drosophila*, such as ASF/SF2 (Ge and Manley, 1990; Ge et al., 1991; Krainer et al., 1991; Mayeda and Krainer, 1992), now known as SRSF1, and SC-35 (Fu and Maniatis, 1992a), now known as SRSF2. It was recognized that these SR proteins interact with *cis*-elements, the SREs discussed earlier, to carry out their regulatory role (Dirksen et al., 1994; Sun et al., 1993). SR proteins have been shown to interact with each other, as well as other RS domain containing proteins such as the U1 snRNP (Wu and Maniatis, 1993). Furthermore, SR proteins seems to have overlapping ability to interact with other proteins to potentiate splicing, as they can functionally complement each other (Fu et al., 1992). Additionally, SR proteins can interact with U1 snRNP and U2AF proteins to recruit them to the 5' (Kohtz et al., 1994; Wu and Maniatis, 1993) and 3'-splice site (Fu and Maniatis, 1992b), respectively, and promote exon definition as discussed earlier. Interestingly, the RS domains of SR proteins can still function to potentiate splicing independent of the RRM domain, when tethered to pre-mRNAs (Graveley and Maniatis, 1998; Graveley et al., 1998). Thus, a general model of SR proteins has their RRMs mediating RNA binding specificity, while their RS domains allow for protein-protein interactions important for exon definition and spliceosome assembly.

Other proteins involved in regulation of splicing are expressed in tissue specific patterns, allowing for activation of alternative splicing programs important for cellular specialization. Most notably, *trans*-acting proteins such as the NOVA (Licatalosi et al., 2008; Ule et al., 2003; Yano et al., 2010; Zhang et al., 2010), and RBFOX (Li et al., 2014; Weyn-Vanhentenryck et al., 2014; Yeo et al., 2009) protein families, and tissue specific SR proteins such as nSR100 (Calarco et al., 2009; Irimia et al., 2014) compete to bind nascent RNAs at specific motifs and drive regulation of alternative splicing in a tissue and developmentally regulated manner.

Several recent studies have employed a variety of techniques to identify splicing events regulated by SR proteins. The use of high throughput techniques such as RNA splicing-specific microarrays and RNA-seq experiments have helped to understand the subset of splicing events regulated by specific SR proteins. Furthermore, techniques such as CLIP-seq have allowed for high throughput identification of RNAs bound by these proteins. For example, the cellular RNAs bound by SRSF1 were mapped using CLIP-Seq, which revealed that this protein interacts with predominantly mRNAs at the exon-intron border, but also binds a variety of other cellular RNAs, including lncRNAs and intronless genes (Sanford et al., 2009). Another example of CLIP-seq being used to understand the regulatory landscape of SR proteins is the mapping binding sites of SRSF3 and SRSF4 (Ankö et al., 2010; Änkö et al., 2012), which showed that these two SR proteins bind distinct subsets of RNA, and preferentially regulate different types of splicing events.

1.9 NMD and splicing

One role of alternative splicing seems to be the regulation of transcript abundance through cellular RNA decay pathways, notably the Nonsense-Mediated Decay (NMD) pathway. NMD is a cellular quality control pathway that degrades RNA transcripts that contain premature termination codons (PTCs), which if translated result in truncated proteins. The ability of PTCs to cause RNA stability was first recognized in yeast (Losson and Lacroute, 1979). Subsequently, it was recognized that patients with thalassemia resulting from frame shifting mutations leading to PTCs caused rapid turnover of RNA, but this turnover could be repressed with the translational inhibitor actinomycin D (Kinniburgh et al., 1982; Maquat et al., 1981). PTCs may arise from a number of sources, including DNA mutations, inaccurate transcription leading to frameshifts or single nucleotide nonsense mutations, or from alternative splicing that puts stop codons in frame in an inappropriate position.

One of the key factors in NMD is a multi-subunit complex known as the Exon Junction Complex (EJC), which, as its name implies, marks the junction between two exons resulting from splicing. The EJC is deposited upstream of splicing junctions in a sequence-nonspecific manner mainly through the interaction between the spliced RNAs and eIF4AIII (Shibuya et al., 2004), a DExH/D RNA helicase which is deposited during the second step of RNA splicing (Zhang and Krainer, 2007) and acts as a scaffold for the further association of other proteins such as Magoh and Y14. The EJC is important for a number of cellular processes, such as nuclear export and subcellular localization (Schell et al., 2002) as well as translation (Nott et al., 2003, 2004; Wiegand et al., 2003). Most importantly for this discussion, however, it acts as an important scaffold for the NMD machinery, as the critical NMD factors UPF3 is added during splicing, and is joined by UPF2 in the cytoplasm (Chamieh et al., 2008; Le Hir et al., 2001; Kervestin and Jacobson, 2012).

During the pioneer round of translation of a transcript with a PTC, the ribosome will cause a change in the protein composition of the mRNP, displacing the compliment of proteins that accompanied the mRNA out of the nucleus, including the EJC (Ishigaki et al., 2001; Lejeune et al., 2002). If the ribosome recognizes a stop codon during translation, it recruits release factors, such as eRF1 and eRF3, which are associated with the NMD-associated protein UPF1. If the stop codon is more than ~50 nt upstream of the EJC (*i.e.*, a PTC), a stable interaction will take place between UPF1 and UPF2, bridging UPF1 with the EJC. This in turn leads to phosphorylation of UPF1 and recruitment of either SMG6, an endonuclease that cleaves the NMD targeted RNA, or the SMG5-SMG7 complex that is associated with uncapping and deadenylation of RNAs. Either of these pathways leads to rapid degradation of the targeted RNA.

As mentioned previously, NMD is coupled to splicing in a number of ways. First, it is the act of splicing that deposits the EJC on mRNAs. However, another important link may be that

alternative splicing can lead to the regulated inclusion of PTCs that lead to preferential degradation of certain gene transcripts over others. This may represent a method of regulation that is dependent on alternative splicing rather than transcriptional regulation. Interestingly, this mechanism seems to be especially common in the regulation of splicing factors including SR proteins and hnRNP proteins (Lareau et al., 2007; Ni et al., 2007; Saltzman et al., 2008).

This link between NMD and alternative splicing was first appreciated in a genome wide manner when a comprehensive study of expressed sequence tags (ESTs) found that a number of alternatively spliced genes were found to have isoforms that would be predicted to be subject to NMD (Lewis et al., 2003). These authors termed this phenomenon regulated unproductive splicing and translations (RUST). They extended this line of inquiry further, and showed that any of these alternative splicing events were also conserved between human and mouse (Baek and Green, 2005), and that AS of these predicted NMD targets was linked with the presence of ultra- or highly-conserved elements at the sites of alternative splicing, especially in SR and hnRNP proteins (Lareau et al., 2007; Ni et al., 2007). Further detailed experimental evidence also showed that specific SR proteins, including SRSF1 (Sun et al., 2010) and SRSF2 (Sureau et al., 2001), produced NMD-sensitive splice isoforms that were modulated by overexpression of the protein coding transcripts. Taken together, these data indicate that there is a conserved regulatory system of splicing factors that takes advantage of NMD to post-transcriptionally modulate protein levels.

1.10 Retained and Detained introns

Intron retention is a specific type of alternative splicing that has been poorly studied until recently. It is characterized by the lack of removal of an intron in an otherwise processed transcript. Retained introns can shape the fate of RNAs in a number of ways, including inhibiting nuclear export (Nott et al., 2003), causing NMD through the inclusion of an in-frame PTC (Ge and

Porse, 2014), as well as altering coding potential (Bell et al., 2008, 2010; Buckley et al., 2011; Dirksen et al., 1995). The regulation of intron retention seems to be complex, with the interactions of a number of SREs, and thus necessarily their interacting *trans*-factors. However, there are a number of important regulatory elements. Perhaps most obviously, informatics studies have indicated that weak 5' and 3'-splice sites are correlated with intron retention (Sakabe and de Souza, 2007). It has also been recognized that multiple 5'ss-like sequences (*i.e.*, complementarity to U1) in close proximity can compete with each other (Pagani et al., 2002; Siebel et al., 1992). Furthermore, a number of ESS sequences identified in high throughput screens have been associated with increased intron retention (Wang et al., 2004). One of the most important *trans*-factors important for intron retention is PTB, which has been shown in a number of studies to be associated with intron retention, both by *in vitro* experiments (Marinescu et al., 2007; Sharma et al., 2005, 2008, 2011), and in the effect of animal knock out models (Yap et al., 2012).

This form of alternative splicing was thought to be less common than other types in mammals (Galante et al., 2004; Michael et al., 2005), and more common in single celled eukaryotes (Parenteau et al., 2011) and plants (Ner-Gaon et al., 2004), and is critical for many viruses (Li et al., 2006). However, a number of recent studies have revealed that this type of alternative splicing is more widespread than initially thought. First, intron retention has been seen in a number of cancers (Dvinge and Bradley, 2015; Simon et al., 2014; Sowalsky et al., 2015; Zhang et al., 2014). There have been a number of proposed mechanisms for this increased intron retention, including global increases in transcription levels leading to saturation of the splicing machinery (Sowalsky et al., 2015) or changes in histone methylation as the result of *SETD2* mutation (Simon et al., 2014). However, intron retention seems to play other, more regulated roles as well. For example, recent studies have shown that intron retention is regulated in a tissue

specific and developmentally regulated manner (Braunschweig et al., 2014). T cells demonstrate intron retention in a specific set of genes that is regulated by hnRNPLL (Cho et al., 2014). Additionally, granulocyte development causes a regulated intron retention in 86 genes linked to NMD degradation of those mRNAs (Wong et al., 2013).

As mentioned previously, transcripts with retained introns often have PTCs that are associated with regulation by NMD. While this seems to be an important mechanism to regulate protein levels through alternative splicing, a recent study has demonstrated that a large number of polyadenylated transcripts with retained introns localize to the nucleus, are not degraded by NMD, and can undergo induced splicing following cellular stress or treatment with an inhibitor of the Clk family of kinases (Boutz et al., 2015). The authors label RNAs with these characteristics "detained introns," and suggest that they may represent a mechanism for storing transcripts for rapid splicing and export. Importantly to the current study, ARGLU1 was identified as a "detained intron" transcript in both human and mouse cells.

1.11 Conclusion

In this introduction, I have tried to present context for the rest of the dissertation. To summarize, ARGLU1 is a nuclearly localized protein. ARGLU1 transcripts have been previously reported to be alternatively spliced in a number of ways, including the presence of a retained intron. ARGLU1 protein interacts with members of the mediator complex as well as the spliceosome. In this dissertation, I will investigate the *cis*-regulatory landscape of ARGLU1 alternative splicing. Furthermore, I will present evidence that ARGLU1 protein can function to homeostatically regulate its own level by feeding back and altering splicing of its own transcript.

Chapter 2: *ARGLU1* Alternative Splicing is Regulated by the Presence of an Ultraconserved Element

2.1 Introduction

In this chapter I will explore the regulation of alternative splicing of transcripts from the gene *ARGLU1*. *ARGLU1*, previously known as FLJ10154, is a gene with a poorly defined cellular function. The most in depth study describes ARGLU1's interaction with a protein called MED1, a component of the mediator complex (Zhang et al., 2011). 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 (Zhang et al., 2011). 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 (Zhang et al., 2011).

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 (Behzadnia et al., 2007), the active step 1 spliceosome (Bessonov et al., 2008), and purified complex B (Deckert et al., 2006). Additionally, ARGLU1 has been found to interact with (Hegele et al., 2012), and be phosphorylated by (Varjosalo et al., 2013), multiple members of the SRPK family of protein kinases, which regulate a family of proteins important for splicing, known as SR proteins (Aubol et al., 2013; Wang et al., 1998; Zhong et al., 2009). Additionally, the *Drosophila* homolog of ARGLU1, CG31712, has been shown to associate with U1-70K (Guruharsha et al., 2011), a key player in the assembly of

the spliceosome that has previously been shown to interact with SR and SR-like proteins (Blencowe et al., 2000; Fu and Maniatis, 1992b; Park et al., 2004).

Alternative splicing is a process by which a precursor mRNA (pre-mRNA) can be assembled with the inclusion or exclusion of certain exons, leading to multiple mature mRNAs. As discussed previously, regulation of alternative splicing is mediated by interactions between *cis*-elements within the RNA molecules and *trans*-acting protein complexes that recognize and bind to these *cis*-elements. The *cis*-elements that inform exon definition, generally known as Splicing Regulatory Elements (SREs), can be defined as exonic splice enhancers of suppressors (ESEs or ESSs) or intronic splice enhancers or silencers (ISEs or ISSs), based on their location and propensity to promote or suppress inclusion of the associated exon (Wang and Burge, 2008). There have been a number of techniques used to identify *cis*-elements regulating splicing. Studies originally identified these sequences through pathogenic mutations in diseases such as cystic fibrosis, in the *CFTR* gene (Pagani et al., 2003), and spinal muscular atrophy, in the *SMN2* gene (Cartegni and Krainer, 2002). Other individual *cis*-elements were discovered through molecular biology techniques such as mutagenesis and deletion of specific sequences proximal to regulated exons within minigene reporters (Schaal and Maniatis, 1999; Sun and Chasin, 2000).

The alternative splicing of a number of SR proteins has been correlated with the presence of ultra-conserved or highly-conserved elements within the primary transcripts of a number of their genes, including *SRSF1, SRSF2, SRSF4, SRSF4, SRSF5, SRSF6, SRSF7, SRSF9, SRSF10,* and *SRSF11* (Lareau et al., 2007; Ni et al., 2007) as well as the genes for the hnRNP proteins *HNRPDL, HNRPH1, HNRPK* and *HNRPM* (Ni et al., 2007). These studies correlated these highly conserved sequences with alternative splicing isoforms that were subject to NMD, as knockdown of the core NMD factor *UPF1* (Lareau et al., 2007; Ni et al., 2007) or treatment with the

translational inhibitor emetine (Ni et al., 2007) led to increased levels of inclusion of these exons containing the premature termination codons (PTCs) that trigger this RNA decay pathway. A 2010 study found that overexpression of SRSF1 regulated splicing of its own transcript to increase the usage the NMD linked splice isoform, which was within a UCE (Sun et al., 2010). SRSF2 (SC35) was also shown to regulate its own alternative splicing by upregulating inclusion of a cassette exon or splicing out of a retained intron, both in the 3' UTR, and thereby inducing NMD (Sureau et al., 2001), both of which are associated with highly conserved sequences. These studies concluded that the presence of exceptionally highly conserved sequences is associated with alternative splicing coupled to NMD allowing for autoregulation.

One characteristic of *ARGLU1* alternative splicing I will explore in this chapter is a retained intron, where the levels of one intron are much higher than all other introns in the gene. Additionally, I will show that the transcripts with this retained intron are localized to the nucleus. A recent study has defined a whole class of RNAs that have these same characteristics, retained introns and nuclear localization, which they have defined as 'detained introns' (Boutz et al., 2015). Notably, the majority of these detained intron transcripts were not regulated by NMD (Boutz et al., 2015). ARGLU1 was identified as a 'detained intron' containing transcript in both mouse and human (Boutz et al., 2015). This study also found that the levels of intron retention in a subset of these detained intron transcripts could be altered through certain treatments, such as DNA damaging agents or chemical inhibition of Clk kinases (Boutz et al., 2015), which phosphorylate SR proteins in the nucleus (Aubol et al., 2013; Colwill et al., 1996). Furthermore, it has been reported that the SR proteins *SRSF1* and *SRSF2* have a number of retained intron isoforms that localize to the nucleus and are thought to contribute to homeostatic control of protein abundance of that gene (Sun et al., 2010; Sureau et al., 2001).

The recent development of tools for targeted genome manipulation, including the Cas9/CRISPR system, has led to methods for manipulating genomic DNA sequences with ease. Cas9/CRISPR is a bacterial adaptive defense against foreign nucleic acids that functions through an RNA guided nuclease. This system has been co-opted for use in mammalian systems, and in the most basic iteration allows for simple targeting of double stranded DNA (dsDNA) breaks. The most common repair pathways for dsDNA breaks is non-homologous end joining, a low fidelity repair system that often leads to insertions or deletions of nucleotides at the site of the dsDNA break.

The hypothesis explored in this chapter is that the ultraconserved element within the retained intron of *ARGLU1* has a regulatory role in alternative splicing of this RNA. I will explore the role of this UCE in a number of different ways, including direct mutagenesis of a splicing minigene, Cas9/CRISPR mediated mutagenesis of the genomic loci, and an assessment of the role of splice site strength in the alternative splicing of this gene.

2.2 Results

2.2.1 *ARGLU1* is alternatively spliced and partitions to the nucleus or cytoplasm depending on alternative splicing

I initially identified *ARGLU1* as partitioning to the nucleus through RNA-Seq data from HeLa cells that had been fractionated into cytoplasmic and nuclear pools and poly(A) selected (LLC, LY, and GC, data not shown). A polyadenylated *ARGLU1* transcript containing a single retained intron, but otherwise fully spliced, was found exclusively in the nuclear fraction of cells I investigated (Figure 2.1). The fully spliced isoform (see Figure 2.2 for an overview of the splice isoforms) is seen in both the nucleus and cytoplasm by RNA-Seq (Figure 2.1), RT-PCR (Figure 2.3) as well as northern blot (Figure 2.4). As mentioned previously, a number of transcripts with these characteristics have been identified, and have been termed detained introns (Boutz et al., 2015).

A minor alternative splicing product has also been observed and annotated previously in the Ensembl gene annotation database (Cunningham et al., 2014). This isoform includes an alternative exon that lies within the retained intron. The inclusion of this exon would be predicted to cause degradation by NMD, as it causes a premature termination codon (PTC). Initial RT-PCR experiments see little evidence of inclusion of this isoform (Figure 2.3), possibly the result of rapid degradation by NMD. I will address the degradation of transcripts including this exon in section 2.2.6. Furthermore, experiments below indicate that the presence of this alternative exon influences the level of intron retention. Furthermore, I see increased usage of this exon under conditions of ARGLU1 overexpression, which will be explored further later in this report.

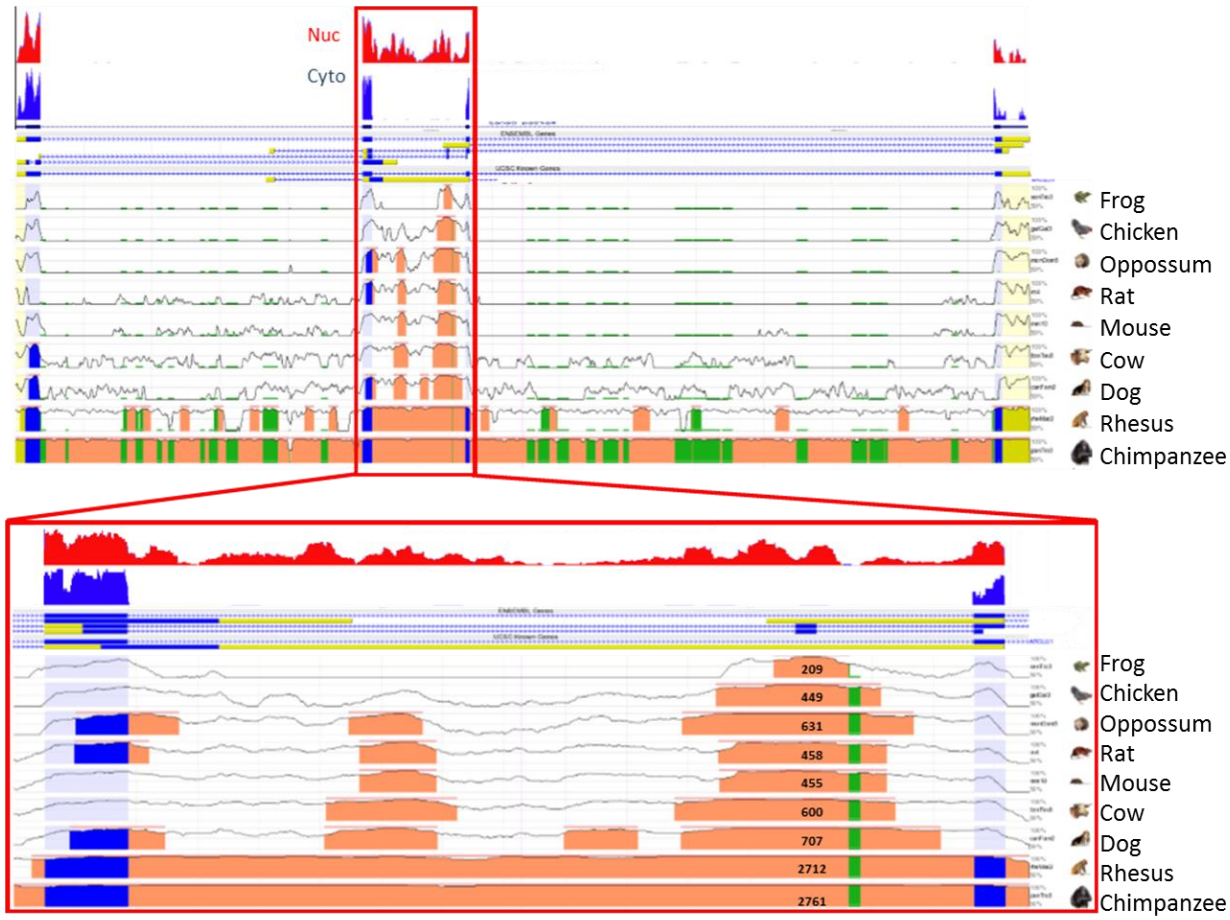

Figure 2.1 A transcript with a retained intron is localizes to the nucleus and is associated with an ultraconserved element. RNA-seq reads from nuclear and cytoplasmic fractions of HeLa cells are aligned to the RefSeq gene annotation for *ARGLU1*. Nuclear reads are presented in red, while cytoplasmic reads are in blue. Below the gene annotations is a view from the evolutionary conservation browser (Ovcharenko et al., 2004). In this view orange represents regions of nucleotides that are more than 97% conserved with human for over a 200 base window, dark blue indicates conserved coding regions, and green indicates simple repeats. I have noted the length of the regions that are 97% conserved between human and the species indicated.

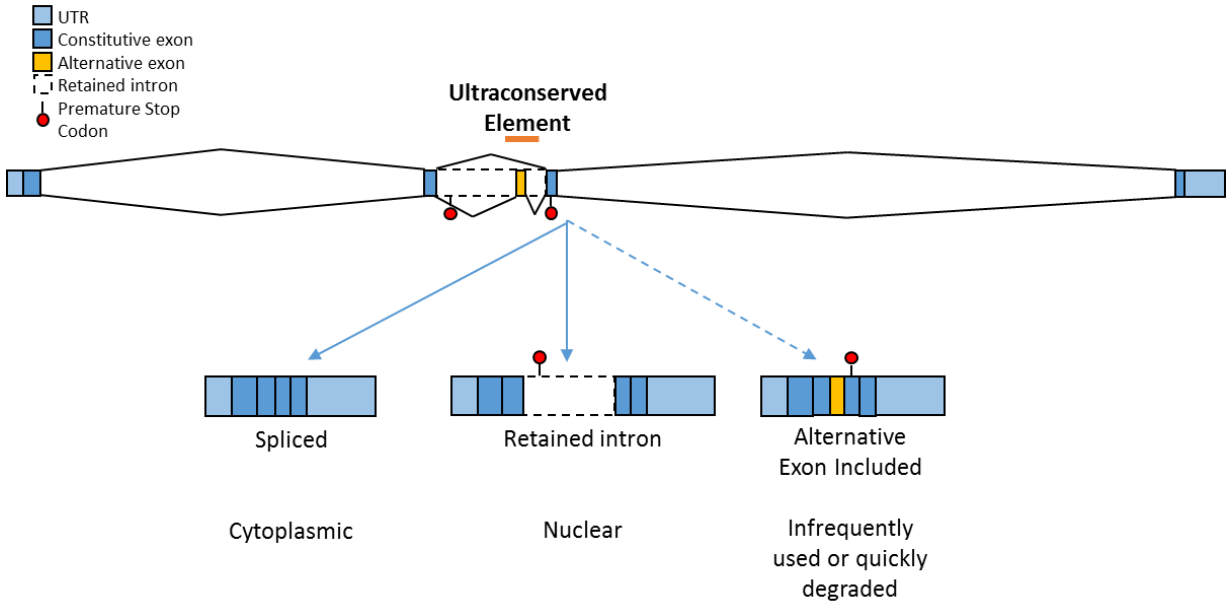

Figure 2.2 Overview of the *ARGLU1* isoforms studied in this dissertation. This cartoon view of the gene architecture of *ARGLU1* and the splice isoforms derived from the pre-mRNA. The fully spliced RNA has 4 exons, and is exported to the cytoplasm. The retained isoform contains 4 exons as well as the retained intron 2, and is located in the nucleus. The alternative exon included isoform is quickly degraded by NMD because it contains a premature termination codon, indicated by a red circle.

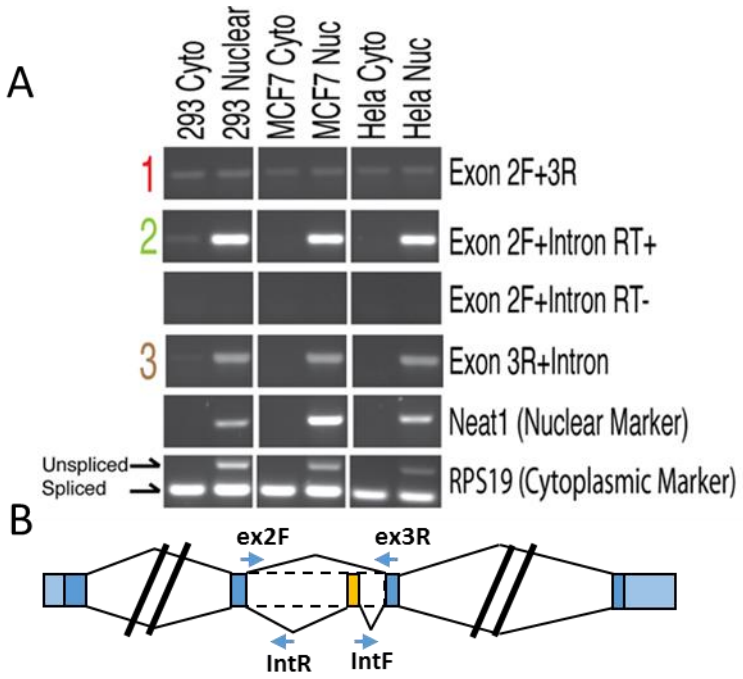

Figure 2.3 RT-PCR confirms that the *ARGLU1* intron retained isoform is present exclusively in the nucleus. A. RT-PCR was performed on nuclear and cytoplasmic fractions of the indicated cell types. NEAT1 is a lncRNA that is seen exclusively in the nucleus, and RPS19 primers cross a splice boundary and indicate the fully spliced isoform is localized exclusively in the cytoplasm. **B.** Primers used for RT-PCR of ARGLU1. Primer set 1 is Ex2+Ex3 primers. Primer set 2 is Ex2F+IntR. Primer set 3 is Ex3R+IntF.

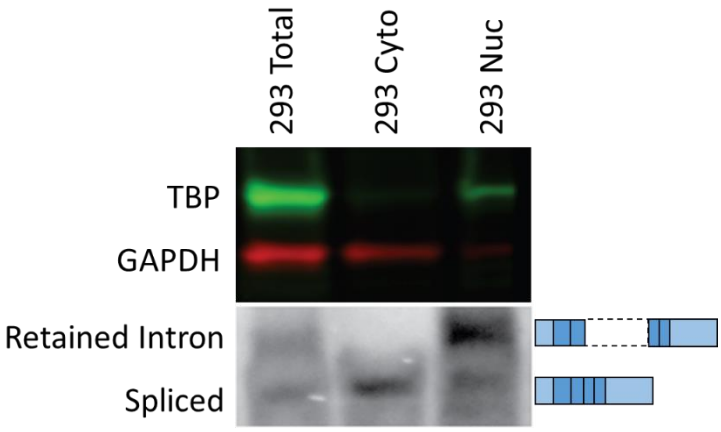

Figure 2.4 Western Blot and Northern Blot indicate that the *ARGLU1* retained intron isoform is localized exclusively in the nucleus. The top panel is a western blot against the nuclear marker TATA Binding Protein (TBP) and the cytoplasmic marker GAPDH in either total lysate or nuclear and cytoplasmic fractions of HEK-293 cells. Cells were fractionated as in Figure 2.2. The bottom panel is a northern blot using a probe that detects both the spliced and retained intron isoforms. The spliced and retained intron isoforms are indicated in cartoon form on the right.

2.2.2 ARGLU1 contains an Ultraconserved Element (UCE)

In order to identify sequences within this alternatively spliced intron I hypothesized that evolutionary conservation could serve as a marker for biological function. Based on this hypothesis, I queried databases of highly evolutionarily conserved sequences. Indeed, a region within the ARGLU1 intron has been previously identified as an ultraconserved element 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 chicken (Dimitrieva and Bucher, 2013). 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 (Retelska et al., 2007). Using UCNEbase and the Evolutionary Conserved Browser (Ovcharenko et al., 2004), I was able to identify an ultraconserved region of 500 bases that fit these requirements (Figure 2.1). 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 2.1). Notably, using the more strict definition of a UCE set out by Bejarano *et al* (Bejerano et al., 2004), which required 100% conservation for over 200, the ARGLU1 UCE is 100% conserved between human and chicken for 228 bases.

2.2.3 The UCE in ARGLU1 intron 2 is 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 this alternative splicing event. To test this in an experimentally tractable manner, I created a splicing reporter vector comprised of the retained intron as well as the flanking exons, which I called A23, for ARGLU1 exons2-3 (Figure 2.5). When transfected into HEK-293T cells, this vector is spliced similarly to the endogenous ARGLU1, with the retained intron mRNA representing approximately half of the mRNA derived from the reporter (Figure 2.6 and 2.7).

To test the role of the UCE in intron retention, I created a number of deletions in the splicing reporter vector. I created a large deletion of 504 bp, representing the sequence 95% conserved with chicken. I also made a series of three smaller deletions of 168 bp that were subdeletions of this large 504 bp deletion. A control deletion was also made of the same size in a non-conserved region of the intron. Using RNase protection assays, I observed that the wild type and control deletion reporter vector, as well as 168 bp deletion 1, showed mostly retained intron RNA when transfected (Figure 2.6). However, the 504 bp UCE deletion as well as 168 bp deletions 2 and 3 caused a dramatic increase in splicing of exons 2 and 3 (Figures 2.6 and 2.7).

To determine specific sequences regulating ARGLU1 alternative splicing, a series of 25 bp deletions were created spanning the 168 bp deletions 2 and 3 (Figure 2.5), which had previously been shown to contain regulatory sequences. To compare spliced and retained isoforms of the splicing reporter, I designed a set of primers for competitive RT-PCR, in which I used a forward and reverse primer annealed to transcribed plasmid specific sequences flanking the insert along with a primer within the retained intron (Figure 2.5). Deletions of sequences upstream and including the skipped alternative exon were found to increase splicing of exons 2 and 3 (Figure 2.6 and Figure 2.7). It should be noted that deletion 168-3, located downstream of the annotated alternative exon, caused increased splicing with a shifted band, indicating the inclusion of the alternative exon (Figure 2.7). Intriguingly, deletion of sequences downstream of the skipped exon, including 168-3 and 25 bp deletions 9-11, increased the inclusion of the normally skipped alternative exon, with usage of different 5'-splice sites on the now included exon depending on the deletion (Figure 2.7). The observation that different 5'-splice sites could be used following deletion of these sequences downstream of the alternative exon indicated that there might be alternative 5'-splice sites competing within this region as well as important regulatory sequences.

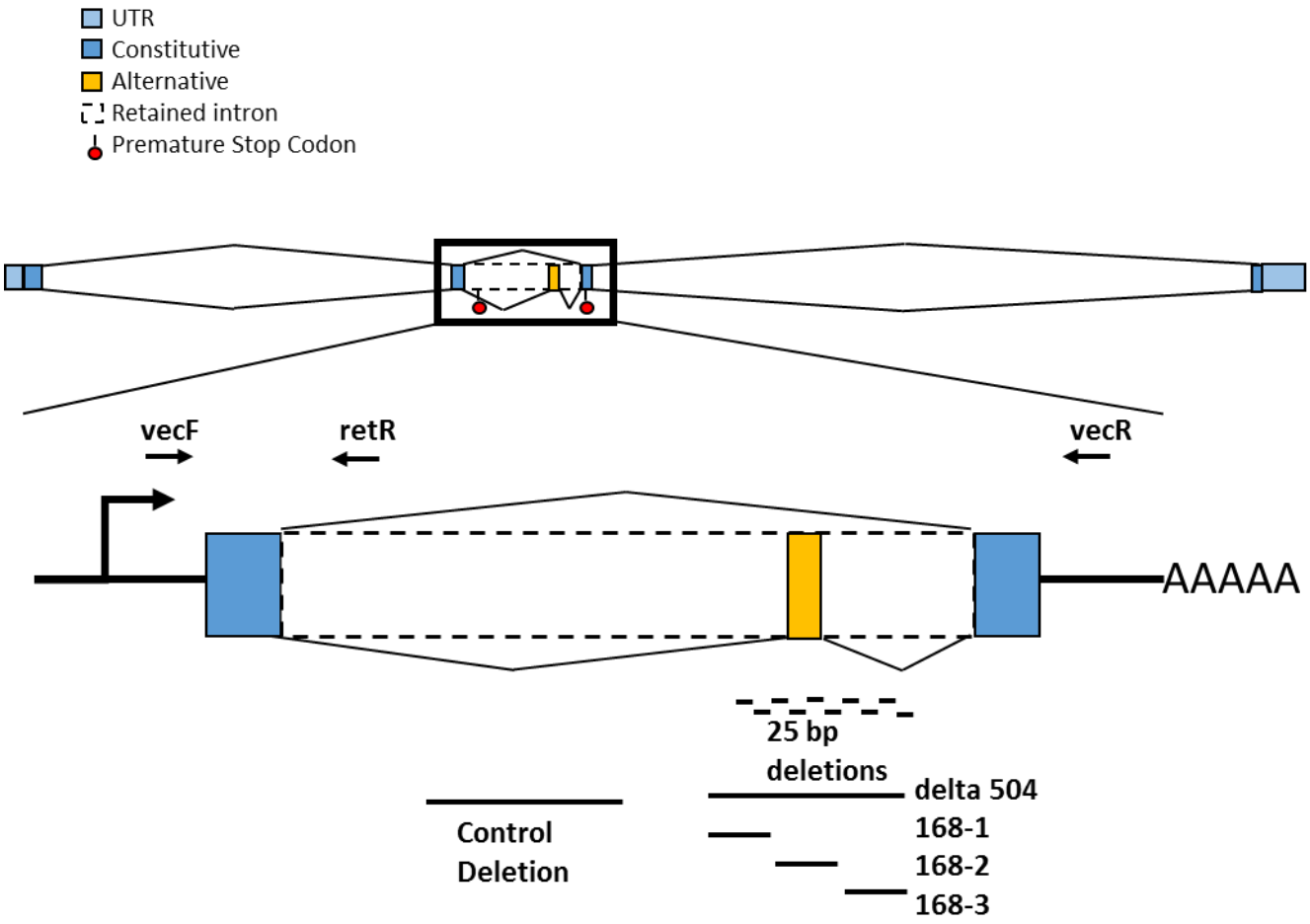

Figure 2.5 Diagram of the A23 splicing reporter vector. A PCR-derived fragment that includes Exon2-intron 2- Exon3 was cloned into the pcDNA3 plasmid, and it was named A23, after ARGLU1 exons2-3. I then confirmed that this splicing reporter recapitulated the splicing of the endogenous mRNA, generating both a spliced and retained intron mRNA. I then generated a number of deletion mutants of this plasmid, including a 504 bp deletion of the sequence 95% conserved between human and chicken, as well as a 500 bp control deletion in a non-conserved region of the intron. I then made three 168 bp deletions to subdivide the 504 bp deletion. These were tested by RPA (Figure 2.6), and it was determined that only deletions 168-2 and 168-3 altered splicing. Finally, I made a series of 25 bp deletions that spanned deletions 168-2 and 168-3. Primers used in subsequent RT-PCR assays are indicated as labeled arrows.

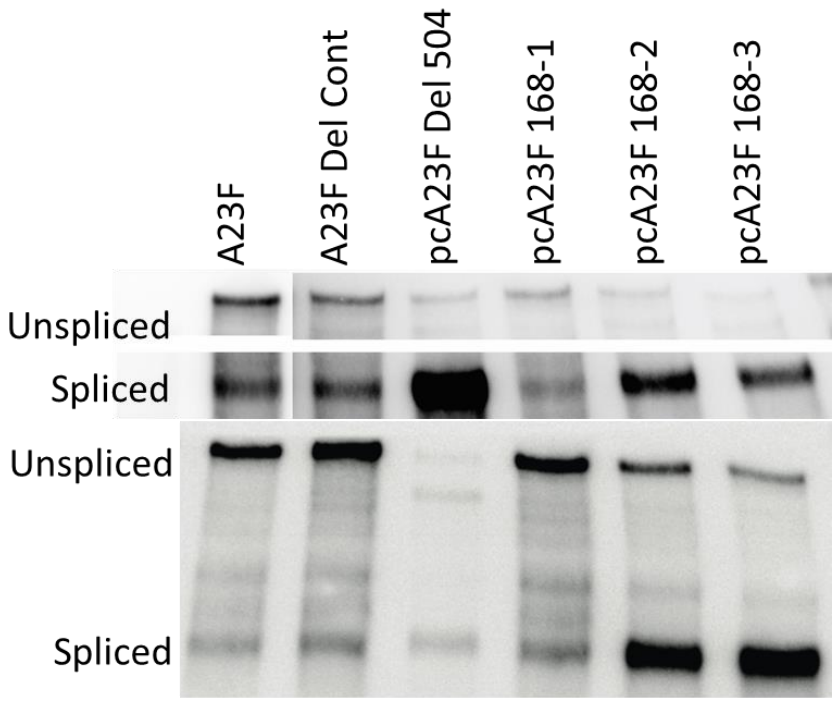

Figure 2.6 RNase Protection Assay indicates that elements within the UCE inhibit splicing of ARLGU1 exons 2 and 3. RNA from cells transfected with the A23 splice reporter vector or the derivatives noted was hybridized with a body labeled probe that annealed to exon 3 into the retained intron. Splicing of the reporter resulted in the lower band. This figure represents two replicates.

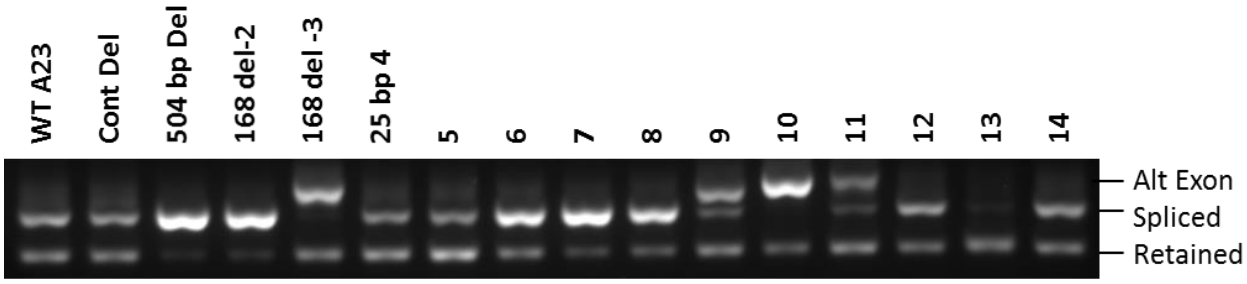

Figure 2.7 Competitive RT-PCR shows that specific elements within the ARGLU1 UCE regulate intron retention, Exon2-3 splicing, and alternative exon inclusion. The A23 splicing reporter plasmid as well as the deletion mutants were transfected into HEK-293 cells. RNA was harvested and used for competitive RT-PCR. Primers for RT-PCR are indicated in Figure 2.5. The bands produced from the primer sets are indicated on the right.

2.2.4 Analysis of putative splice sites bounding the alternative exon:

To test what, if any, role the splice sites bounding the UCE alternative exon and the possible 5'-splice sites downstream of this alternative exon play, I mutated these splice sites individually and in combination (Figure 2.8A). Mutation of the annotated 3' splice site of the alternative exon leads to decreased intron retention and increased splicing of Exon 2-Exon 3 (Figure 2.8B). Furthermore, plasmids with a mutation of this 3' splice site completely lacked splicing inclusion of the alternative exon (Figure 2.8B). Mutating the annotated 5'-splice site of the UCE exon did not drastically change the ratio of spliced to retained intron 2. Interestingly though, it did cause an increase in alternative exon inclusion; however the included exon was larger, indicating the usage of a different 5'-splice site downstream of the normal splice site (Figure 2.8B).

I also created mutations in six possible 5'-splice sites downstream (containing either a GT or GC dinucleotide), to determine which, if any, of these sequences might regulate inclusion of the UCE exon (Figure 2.8A). I found that mutation of putative splice sites 2-4 did not alter the ratio of intron retention to spliced product. These mutations did cause a change in the inclusion of the alternative exon, which was the larger of the two noted bands (Figure 2.8B). Interestingly, mutation of putative splice sites 5 and 6 did alter the ratio of retained and spliced reporter, causing an increase in splicing. Furthermore, these mutations caused a difference in splicing of the alternative exon as well, with exclusive inclusion of the smaller band (Figure 2.8B). By sequencing RT-PCR products from cells transfected with these plasmids, I determined that mutation of SS5 and SS6 caused exclusive use of the annotated 5'-splice site. Additionally, Sanger sequencing of the products from SS3 and SS4 indicated that the upper band corresponds to SS5 (Data not shown). Analysis for the predicted 5'-splice site using a variety of computational models

(Yeo and Burge, 2004) indicates that the annotated 5'-splice site is actually weaker compared to other 5'-splice sites analyzed, especially splice sites 5 and 6 (See Table 2.1).

To further analyze the role of the splice sites flanking the retained intron and skipped exon, I improved the splice sites to match strong consensus 5'- or 3'-splice sites (Figure 2.9A). I found that improvement of the 5'-splice site of exon 2 or the 3'-splice site of exon 3 did not appreciably increase the ratio of spliced to retained product (Figure 2.9B). However, mutation of the 3' splice site of the alternative exon dramatically increased the inclusion of the alternative exon, as indicated by a shift in the spliced band obtained by RT-PCR (Figure 2.9B). As seen in previous experiments, there seem to be two shifted bands (Figure 2.8B), indicating that two possible 5'-splice sites can be used upon inclusion of the alternative exon. Mutation of the annotated 5'-splice site of the alternative exon to a consensus 5'-splice site also leads to increased inclusion of this alternative exon by RT-PCR (Figure 2.9B). Interestingly, there is only one band in this situation, which indicates usage of only one 5'-splice site, which was confirmed by Sanger sequencing to be the improved alternative 5'-splice site.

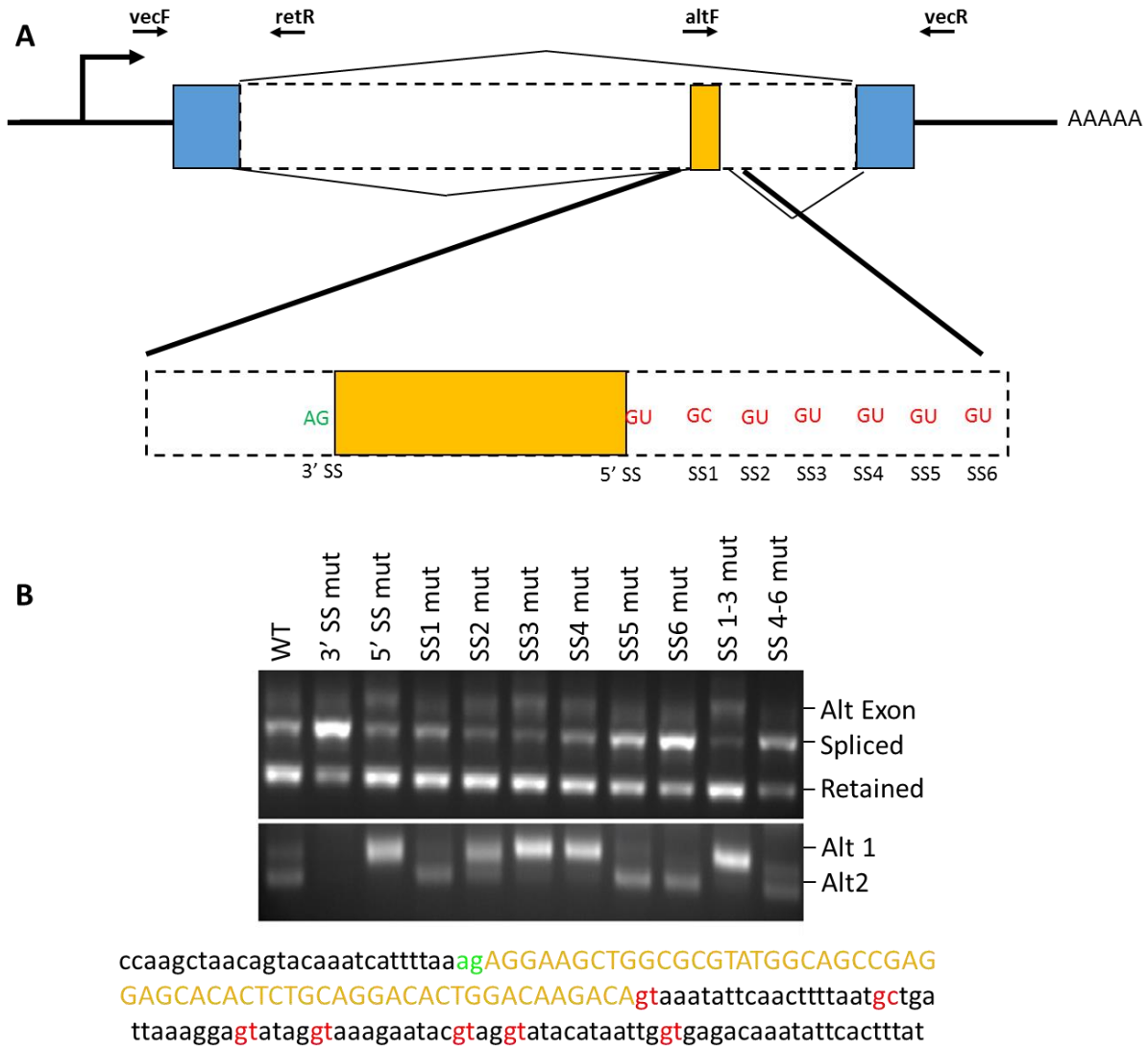

Figure 2.8 Mutagenesis of 3'- and putative 5'-splice sites downstream of the alternative exon modulates both intron retention and alternative exon inclusion. **A.** Splice site mutation derivatives of the A23 splicing reporter are detailed. The 3' splice site of the alternative exon is noted in green, and 5'-splice sites are noted in red. The 3' splice site was mutated. 5'-splice sites were mutated from GU to CA. **B**. RT-PCR was performed from HEK-293 cells transfected with the noted splicing reporter mutant. Two sets of PCR were performed, one with primers vecF+vecR+retR to assess splicing vs intron retention, and a second with altF+vecR to assess the 5'-splice site used upon alternative exon inclusion. At the bottom is the sequence of the region surrounding the alternative exon.

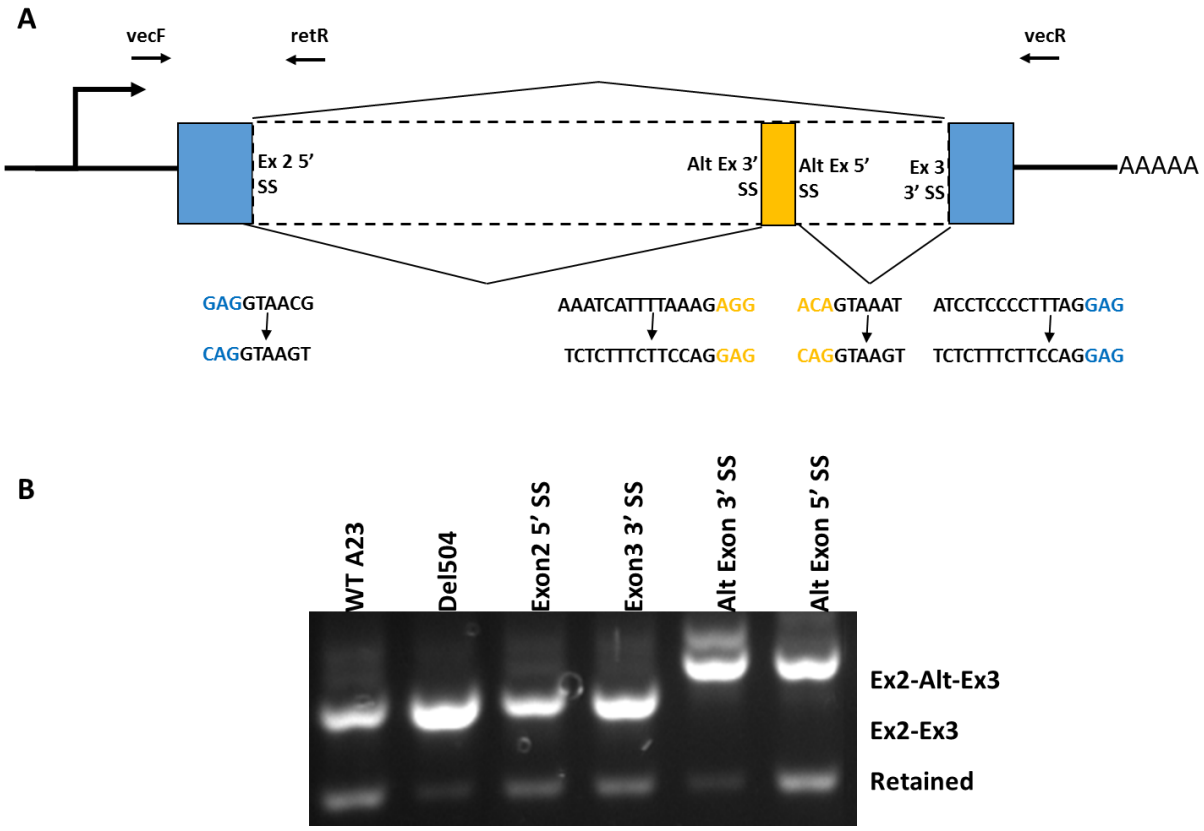

Figure 2.9 Improvement of the alternative exons, but not the constitutive exons, changes intron retention and alternative exon inclusion. **A.** Mutations were made in the A23 splicing reporter to improve the splice sites of the constitutive exons 2 and 3, or the splice sites flanking the alternative exon. Below the diagram is the specific sequence that was mutated. The top line is the endogenous sequence, and the bottom line is the mutant sequence. **B.** RT-PCR was performed from HEK-293 cells transfected with the noted splicing reporter plasmid. Primers are indicated in Figure 2.9A and the bands are indicated on the right.

Splice site	9-mer sequence	MAXENT:	MDD:	MM:	WMM:
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

Table 2.1 Analysis of 5'-splice site strength using various models. Predictions were made using the MAXENT tool (Yeo and Burge, 2004).

2.2.5 Modification of the endogenous ARGLU1 UCE disrupts intron retention:

I next sought to validate the ability of mutations and deletions within the UCE to cause changes in splicing. To accomplish this I used the Cas9/CRISPR system (Ran et al., 2013) to generate random mutations by non-homologous end joining. I designed two different guide RNAs targeting the UCE (appendix A). I inserted these guides into a plasmid co-expressing the sgRNA as well as Cas9 fused to GFP by a self-cleaving PP2A polypeptide (Ran et al., 2013). These plasmids were transfected into HEK-293 Flp-In T-Rex cells. I then sorted single cells expressing GFP into 96 well plates and allowed them to grow until the cells formed colonies. I was able to isolate a number of clones with heterozygous or compound heterozygous insertions and deletions within the UCE (data not shown), all of which demonstrated increased splicing and decreased levels of ARGLU1 intron retention compared with wild-type cells (Figure 2.10). Clone 4 showed the largest increase in the ratio of spliced/retained intron, and contained a homozygous single nucleotide insertion within the alternative skipped exon (Figure 2.10). However, other interesting clones were also obtained. For example, clone 14 contained a single nucleotide homozygous deletion within the alternative exon that caused inclusion of this alternative to be frame maintaining, thus stabilizing mRNAs that contained the alternative exon (Figure 2.10). Finally, clone 3 had a large heterozygous insertion within the alternative exon that caused increased inclusion of the alternative exon (Figure 2.10.

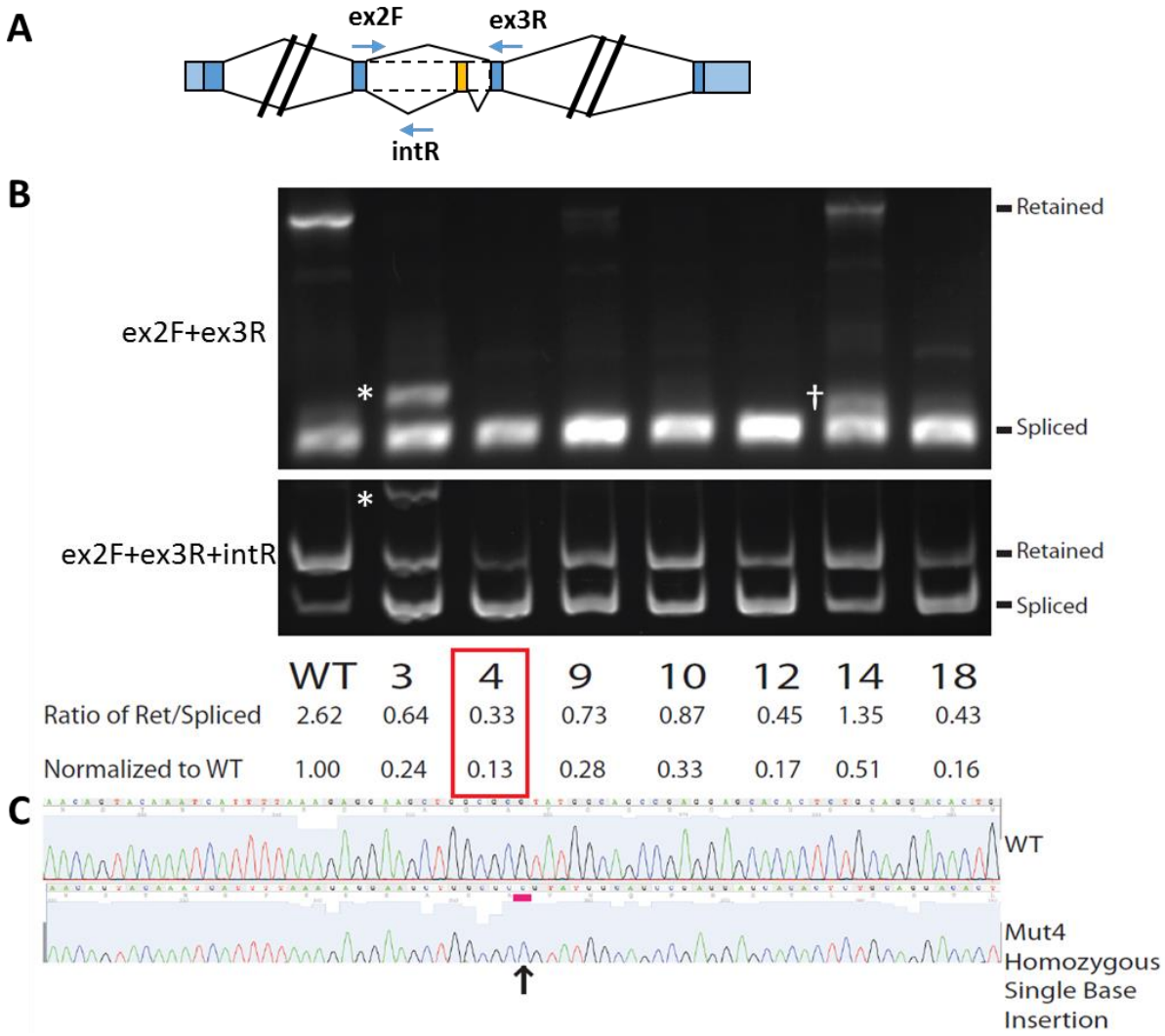

Figure 2.10 Disruption of the UCE at the endogenous chromosomal locus disrupts intron retention vs splicing. A. Primers used for RT-PCR. **B**. RT-PCR of RNA from clonal HEK-293 cell lines in which Cas9/CRISPR was used to target double stranded breaks to the alternative exon in the UCE. A number of clones were recovered that had a mixture of insertions and deletions. Clone 4 contains a homozygous single nucleotide insertion in the UCE that dramatically increased splicing and decreased intron retention. Clone 3 included a large insertion in the UCE that increases inclusion of the alternative exon including the insertion (noted by a *). Clone 14 contains a single nucleotide deletion that causes the alternative exon to be in frame, causing transcripts including the alternative exon to be stabilized. The included alternative exon is noted by a †). **C.** Chromatograph from Clone 4 indicating a single nucleotide insertion in the alternative exon in the UCE.

2.2.6 ARGLU1 intron 2 is sufficient to induce intron retention in an exogenous context

Previous experiments have shown that the UCE in the ARGLU1 intron 2 is necessary for intron retention or inhibition of splicing. However, we next wanted to ask whether the ARGLU1 intron could cause splicing inhibition in an exogenous context. To that end, I cloned the ARLGU1 intron 2 into the pzw1 vector in either the forward or reverse direction. The pzw1 vector consists of an EGFP cDNA with an intron containing a multiple cloning site in the middle of the GFP critical fluorescent site. The intron is inefficiently spliced, but the addition of sequences into the exon has been used to determine sequences that modify exon splicing (Wang et al., 2004). I found that the forward ARGLU1 insert, but not the reverse insert, was sufficient to generate a band consistent with the retained intron. However, the majority of the RT-PCR product indicated splicing, possibly because the 5'- and 3'-splice sites in the vector are quite strong.

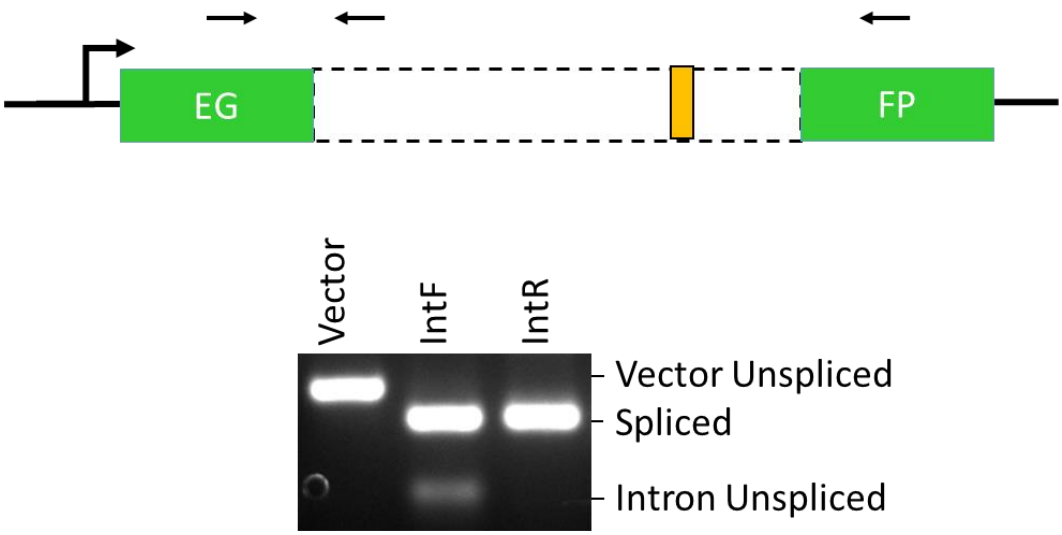

Figure 2.11 ARGLU1 intron 2 is sufficient to induce intron retention in an exogenous context. The pzw1 vector contains an intron with an MCS in the middle of EGFP. ARGLU1 intron 2 was inserted into this MCS in either a forward (intF) or reverse (intR) direction (i.e. the reverse compliment as a size-matched control). These plasmids were transfected into HEK-293 cells, and RNA was harvested 24 hours later. RT-PCR using competitive PCR with the primers indicated can detect a band produced by splicing as well as one consistent with intron retention.

2.3 Conclusion:

I have identified *ARGLU1* as a transcript that has a number of interesting characteristics. First, it has a retained intron, a form of alternative splicing relatively rare and poorly understood in mammals. Second, the *ARGLU1* isoform containing the retained intron is localized exclusively in the nucleus. These two characteristics have recently been used to define a class of mRNAs labeled 'detained introns' (Boutz et al., 2015). That study identified ARLGU1 as a detained intron in both human and mouse cells (Boutz et al., 2015). I have shown through multiple methods, including RNA-seq, RT-PCR and northern blot that the retained intron ARGLU1 transcript is abundant and exclusively nuclear. Additionally, it does not seem to be sensitive to NMD, as treatment with emetine did not increase the abundance of the retained intron isoform. Insensitivity to NMD was another characteristic used by Boutz et al. (Boutz et al., 2015) to functionally separate 'detained introns' from transcripts with retained introns that were exported from the nucleus and degraded by NMD. The third intriguing characteristic of ARGLU1 is that there is a UCE associated with the retained intron. As mentioned previously, the presence of UCEs has been linked to alternative splicing and homeostatic regulation of SR proteins (Lareau et al., 2007; Ni et al., 2007; Sun et al., 2010; Sureau et al., 2001).

In this chapter I have defined many of the *cis*-regulatory elements that control alternative splicing of *ARGLU1*. In line with previous reports that UCEs are associated with alternative splicing linked to NMD (Lareau et al., 2007; Ni et al., 2007), I have shown that there is an NMD sensitive alternative exon that lies completely within the *ARLGU1* UCE. However, unlike previous studies of these UCE-linked alternative splicing events, I have experimentally tested the role of sequences within the UCE to show this sequence is both necessary and sufficient to induce intron retention.

The presence of the alternative exon within the retained intron of *ARGLU1* seems to be central to the regulation of intron retention. Deletion of sequences that overlap the alternative exon cause a dramatic decrease in the level of intron retention relative to splicing of exon2-exon3. Mutation of the 3'-splice site of the alternative exon causes a dramatic increase in exon2-exon3 splicing and a decrease in intron retention. Furthermore, deletion or mutation of sequences downstream of the alternative exon also seem to regulate splicing of exon2-exon3. In particular, mutation of two 5'ss-like sequences downstream of the alternative exon (SS5/SS6) seem to dramatically increase splicing exon2-exon3. Conversely, mutation of two different 5'ss-like sequences in downstream of the alternative exon cause increased usage of the alternative exon, but changes the 5'ss of the alterative exon to SS5.

On the other hand, improving the annotated splice sites flanking the alternative exon causes a marked increase in the usage of the alternative exon. In particular, mutation of the 3'-splice site and addition of upstream pyrimidine caused a dramatic increase in the alternative exon usage, albeit with evidence of multiple 5'-splice sites, in line with previous results. Interestingly, improvement of the annotated 5'-splice site alone also increased inclusion of the alternative exon, but with usage of that splice site alone.

These data suggest a complicated *cis*-regulatory environment centered on the presence of an alternative exon. Based on the fact that improvement of either the 3' or 5'-splice site of the alternative exon shifts splicing almost exclusively to the inclusion of that exon, I propose that there is poor exon definition of the alternative exon. However, the mutagenesis of sequences downstream of the alternative exon indicate that there are other regulatory elements that seem to function as both intronic and exonic splice suppressors. The sequences at SS3 and SS4 seem to increase usage of the SS5 putative 5'ss at the expense of the annotated 5'-splice site, indicating

that these sequences may act as intronic splice silencers. Conversely, mutation of SS5 and SS6 sequences cause the sole use of the annotated 5'ss of the alternative exon, and also increase exon2-exon3 splicing. This indicates that the presence of at least one competing 5'ss downstream of the alternative exon is important for regulation of intron retention versus exon2-exon3 splicing. I therefore suggest that the combination of weak splice sites flanking the alternative exon as well as multiple putative 5'-splice sites downstream of this alternative exon causes poor exon definition, but is sufficient to act as a decoy for productive splicing of the upstream and downstream exons. This decoy exon may then form splicing complexes that are not able to be completed, and thus lead to intron retention.

This model seems to be in line with a previously proposed model of intron retention (Sakabe and de Souza, 2007). The improper exon definition model of intron retention is based on failure of complete exon definition. The exons flanking the retained intron are defined by interactions of *cis*- and *trans*-factors. However, after this exon definition step, these defined exons go on to associate and remove the intron, or fail to associate and thus retain the intron. This failure would be presumed to occur as a result of the spliceosome failing to transition from complex E to A, indicating that not all of the important exonic sites were recognized, or from a failure in transition from complex A to B, indicating that the 5'-and 3' splice sites were not brought together. A number of *trans*-factors have been shown to inhibit the assembly of the E complex, including hnRNPA1 (Yu et al., 2008). Interestingly, SRSF7, which normally stimulates exon inclusion, can play an inhibitory role if it binds intronic sequences proximal to exons, and was shown to inhibit progression past the spliceosomal E complex (Erkelenz et al., 2013). Other proteins have been shown to prevent the rearrangement of the spliceosome beyond the A complex, including PTB (Sharma et al., 2008, 2011).

2.4 Materials and Methods

Cell Culture

HEK-293 cells were maintained in DMEM supplemented with 10% fetal bovine serum and 100 U/mL Penicillin-Streptomycin. Cells were grown at 37°C, in an environment maintained at 5% CO_2.

Nuclear/Cyotplasmic Fractionation

Nuclear/Cyotplasmic fractionation was carried out as in (Chen et al., 2008) 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 x g for 5 minutes at 4°C. Cells were then gently resuspended in 250 uL 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 minutes. During this incubation $1/10^{th}$ of the total cell lysate was retained for western blotting and $1/5^{th}$ of the whole cell lysate was added to 1 mL of Trizol reagent. The rest of the lysate was centrifuged for 3 minutes 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. $1/5^{th}$ 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 uL lysis buffer A, then the nuclei were resuspended in 100 uL of lysis buffer A, $1/5^{th}$ of the nuclear lysate was used for western blot and the remaining nuclear fraction was added to 1 mL of Trizol reagent.

Northern Blotting

10 ug of RNA was used for loading onto a 1% agarose in sodium phosphate buffer. RNAs were resuspended in a RNA gel loading buffer containing glyoxal, bromophenol blue and xylene cyanol. RNAs were denatured at 75 C for 3 minutes, then loaded into the gel. The RNAs were resolved at 90 volts for 1 hour. RNA was then transferred to positively charged nylon membrane using capillary transfer in 20X SSC for 4 hours. 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 over night 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.

Plasmids

The A23 plasmid was created using a pcDNA3 vector backbone. A PCR fragment consisting of the genomic region spanning ARLGU1 exon 2 to exon 3, including the intron, was TA cloned into the TOPO 2.1 vector, thus the name A23. A clone was selected that did not show any mutations within the UCE or within 300 bases of either exon. The A23 fragment was excised from the TOPO 2.1 vector using EcoRI, which cleaves on both sides of the fragment, and cloned into the pcDNA3 expression vector also cut with EcoRI. This resulted in both forward and reverse inserts into the pcDNA3 vector. A number of clones were obtained for each direction, and clones were again screened with Sanger sequencing to ensure no mutations in the UCE or within 300 bases of either exon. The forward pcDNA A23 clone #3 was then used for future experiments. All other A23 mutants were derivatives of this initial plasmid.

Mutagenesis

Mutagenesis was carried out using the Q5 mutagenesis protocol. Briefly, PCR of circular plasmids was carried out with non-overlapping opposing forward and reverse primers that together contained the mutation or flanked the area to be deleted. The PCR products were then treated with a mixture of T4 PNK, DpnI nuclease, and T4 DNA ligase. This mixture was then transformed into DH5α *E. coli*. Mutations were confirmed by Sanger sequencing. Primers used for each mutagnesis are listed in appendix A.

RNase Protection Assay

RPAs were carried out using the Ambion RPA III kit according to the manufacturer's instructions. Labeled RNA probes were generated from an A23 plasmid with a reverse insert using the plasmid T7 promoter and internally labeled with Biotin-UTP. The RNA probes were then denatured and hybridized overnight to RNA from cells transfected with the A23 reporter plasmids. Non-hybridizes RNA was then digested with a mixture of RNase A and T1. The RNAs were then purified and resolved on a 5% TBE-urea polyacrylamide gel. The protected fragments were transferred to positively charged nylon membrane, and probed with an alkaline phosphatase conjugated streptavidin. The chemiluminescent reagent CDP-star was used to expose the bands. The membrane was then imaged on a Kodak Image Station machine.

Preparation of RNA

RNA was prepared by lysing cells in an appropriate volume of Trizol (Life Tech). Chloroform was added, mixed vigorously, and incubated on ice for 5 minutes. These samples were centrifuged at maximum speed for 15 minutes. The aqueous phase was then recovered and precipitated by the addition of equal volume 100% isopropanol with glycoblue carrier added. The RNA pellet was then washed 1X with 75% ethanol and resuspended in an appropriate volume of RNase free water.

RNA concentration was assessed using Absorbance at 280 on a nanodrop spectrophotometer. 5 ug of RNA was treated with Ambion Turbo-DNA free DNase I reagent, and collected into fresh tube following treatment with the Ambion DNase I inactivation reagent.

RT-PCR

Reverse transcription was carried out on DNase I treated RNA using the Protoscript II cDNA synthesis kit using random primers (New England Biolabs). PCR was then carried out using OneTaq quick load master mix (NEB) with the primers indicated for each experiment. RT- samples were also used as a control for all PCR sample to test for residual DNA. Primers can be found in Appendix A.

Cas9/CRISPR

I used the px458 plasmid (Ran et al., 2013) in order to generate site directed double stranded breaks in the UCE. This plasmid contains the Cas9 nuclease fused to GFP with a self-cleaving PP2A peptide, causing transfected cells to fluoresce when Cas9 is expressed. I cloned 2 separate guide RNAs into the expression plasmid, and transfected them into HEK-293 Flp-In TRex cells. After 24 hours, cells were FACS sorted and cells were plated into individual wells 96 well plates. After growing cells to obtain clonal colonies, cells were screened with PCR and sequencing of the genomic region surrounding the targeted sites. RT-PCR was carried out using primers in exon2F and exon3R, listed in appendix A.

Western blotting

Cells were lysed for 30 minutes in ice cold RIPA buffer supplemented with protease inhibitors and benzonase. Protein was quantified using the BioRad Cell lysates were then mixed with 2X Laemmli's SDS loading buffer with 10 mM DTT. 30 μg of protein was loaded into SDS-PAGE

gels and resolved at 100V. Transfer to nitrocellulose membrane was carried out in Tris-glycine buffer with 10% methanol at 4°C at 240 mA for 2 hour. Membranes were blocked with Licor blocking buffer for 1 hr at room temperature. Blots were then incubated overnight at 4°C with primary antibody diluted (TBP 1:1000, GAPDH 1:2000) in Licor blocking buffer supplemented with 0.1% tween-20. Membrane was then washed with TBS-0.1% tween 3 times for five minutes. 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 minutes at room temperature with rocking. Membrane was then washed with TBS-0.1% tween 3 times for five minutes. Membranes were imaged on the Licor Odyssey system.

Chapter 3

ARGLU1 Protein Regulates the Splicing of its own mRNA

3.1 Introduction

This chapter will focus on the homeostatic regulation of ARGLU1 mRNA splicing. As described in the previous chapter, an ultraconserved element (UCE) is associated with the complex alternative splicing of *ARGLU1* through the regulation of an alternatively spliced in exon as well as a retained intron.

Previous studies have indicated that UCEs and other highly conserved elements, are often associated with alternative splicing events (Baek and Green, 2005; Lareau et al., 2007; Ni et al., 2007). Interestingly, SR proteins, a class of proteins known to promote RNA splicing and implicated in alternative splicing, are enriched for highly conserved sequences around exons that are alternatively spliced. A number of studies have indicated that SR proteins can cause feedback loops that act to regulate splicing of their own mRNAs, and consequently protein levels, including SRSF1 (Sun et al., 2010) and SRSF2 (Sureau et al., 2001).

These highly-conserved element-associated alternative splicing events are often linked with a process known as nonsense-mediated decay (NMD). NMD is a cellular quality control mechanism that functions to degrade mRNA transcripts containing aberrant premature termination codons (PTCs). These PTCs are recognized by the presence of a complex known as the Exon Junction Complex (EJC), which is a group of proteins deposited 24 bases upstream an exon-exon junction (Le Hir et al., 2001; Shibuya et al., 2004). The EJC is important for nuclear export of spliced RNAs (Le Hir et al., 2001; Schell et al., 2002) and translation (Nott et al., 2004). However it also serves as a platform for the assembly of the NMD machinery. UPF1 is one of the major components of the NMD pathway and is recruited to stop codongs along with another NMD factor,

SMG1, and the release factors eRF1 and eRF3 (collectively known as the SURF complex) (Kashima et al., 2006). If an exon junction complex is present more than approximately 50 bases downstream of this stop codon, an interaction between the SURF complex and UPF2 in the EJC will lead to phosphorylation of UPF1 (Chamieh et al., 2008), leading to recruitment of the endonuclease SMG6 and consequently RNA degradation (Kervestin and Jacobson, 2012).

In this chapter, I will test the hypothesis that ARGLU1 protein has a regulatory role on the splicing of its own mRNA. Based on previous findings that the highly conserved sequences around the alternative exon are important for regulation of ARGLU1 spicing, I propose that the ARGLU1 protein itself may function to regulate this splicing event. To test this hypothesis, I have designed a cell line that allows expression of ARGLU1 protein in a tetracycline-inducible manner. Using this system, I have assessed the impact of ARGLU1 overexpression on the splicing of the endogenous *ARLGU1* mRNA transcripts, causing a decrease in exon2-exon3 splicing and an increase in inclusion of the alternative exon that is subject to NMD. Conversely, I will show that inhibition of protein synthesis is associated with splicing regulation opposite of that seen with ARGLU1 overexpression.

3.2 Results

3.2.1 ARGLU1 overexpression alters splicing of the A23 splicing reporter

I hypothesized that one role of the UCE in ARGLU1 splicing could be to regulate protein expression levels through alternative splicing. This would suggest that ARGLU1 protein might feedback to alter splicing of its own RNA. To test this hypothesis, I tested the effect of a plasmid encoding an *ARGLU1* cDNA on the splicing of the A23 splicing reporter vector (Figure 3.1A). I mock-transfected, or co-transfected the ARGLU1 cDNA expressing plasmid with the WT A23 reporter, the control deletion A23 reporter, or the 504 bp deletion reporter (Figure 3.1B). Similar to previous experiments, in cells not transfected with the ARGLU1 cDNA plasmid I saw that the 504 bp UCE deletion caused increased splicing of the reporter transcript compared with the WT reporter, while the control deletion did not (Figure 3.1B). However, in cells transfected with the ARGLU1 cDNA, the WT and control deletion reporter transcripts had a decreased ratio of spliced/retained compared with the cells not transfected with the ARGLU1 cDNA plasmid (Figure 3.1B). The 504 bp UCE deletion reporter had an increased ratio of spliced/retained transcript, and was not affected by the transfection of ARGLU1 cDNA (Figure 3.1B).

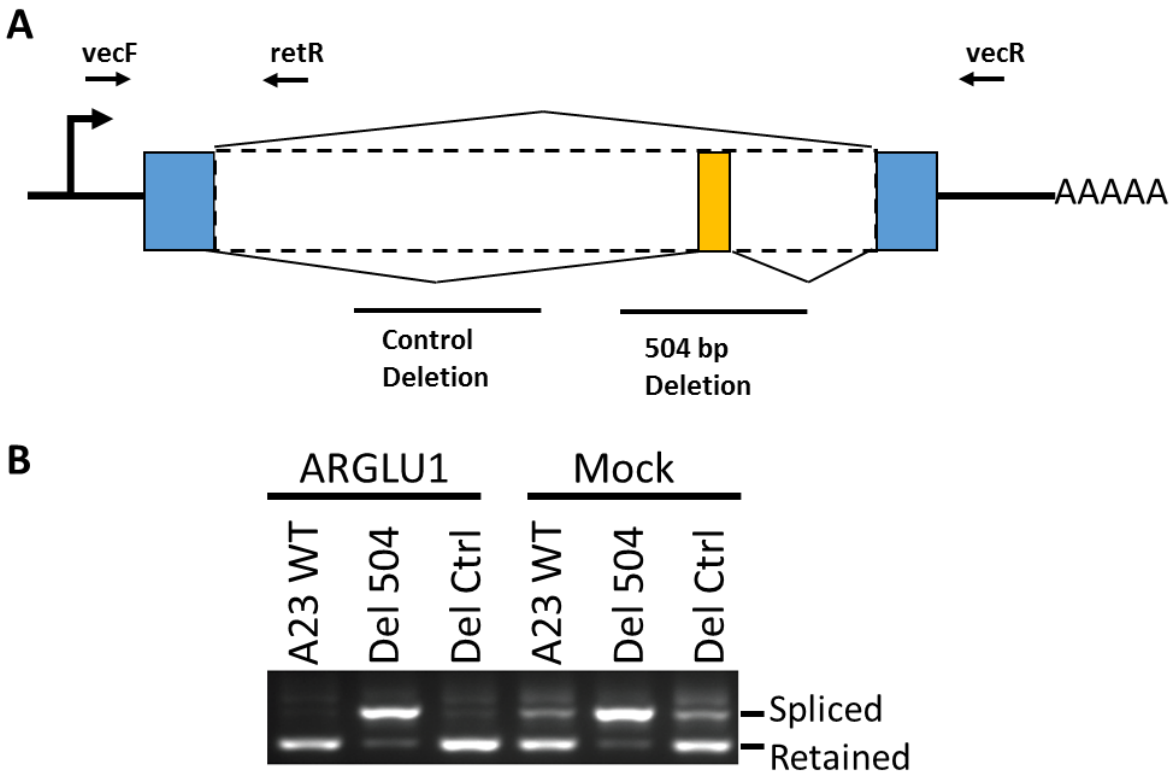

Figure 3. 1 Transient transfection of ARGLU1 protein alters splicing of the A23 splicing reporter. A. Diagram of the A23 splicing reporter vector with the 504 bp and control deletion noted. Primers for 3.1B are noted **B.** The A23 WT, A23 504 bp deletion, or Control Deletion splicing reporter plasmids were co-transfected with either a plasmid overexpressing ARGLU1 cDNA or empty vector. Splicing was assessed with the use of 3 primers, vecF, vecR, and RetR, seen in Figure 3.1A.

3.2.2 Inducible overexpression of ARGLU1 decreases endogenous ARGLU1 protein and alters splicing of the endogenous *ARGLU1* mRNA

To test whether this finding also held true for the endogenous transcript, I created a stable cell line that overexpressed ARGLU1 in a tetracycline-inducible manner. Using the HEK-293 Flp-In TRex cell line, I made cell lines expressing ARGLU1, ARGLU1 with a C-terminal HA tag, or the gene chloramphenicol acetyltransferase (CAT) as a control (Figure 3.2). Induction of ARGLU1 overexpression for 24 hours with 1 μg/mL of tetracycline caused a dramatic overexpression of ARGLU1 protein (Figure 3.3). Intriguingly, I found that overexpression of the HA-tagged version of ARGLU1, which migrates slightly slower by SDS-PAGE, led to a dramatic decrease in the level of the endogenous protein (Figure 3.3). This suggested to us that ARGLU1 protein leads to feedback that maintains ARGLU1 protein levels. Based on previous findings, I hypothesized that this could occur through regulation of alternative splicing. Indeed, quantitative RT-PCR analysis of splicing of the endogenous transcript after 24 hours of induction indicated that there was a 6-fold decrease in the fully spliced endogenous transcript, and 1.6-fold increase in the retained intron transcript (Figure 3.4). These changes were also observed by semi-quantitative RT-PCR using primers that captured both the spliced and retained transcripts (Figure 3.5). 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 3.4).

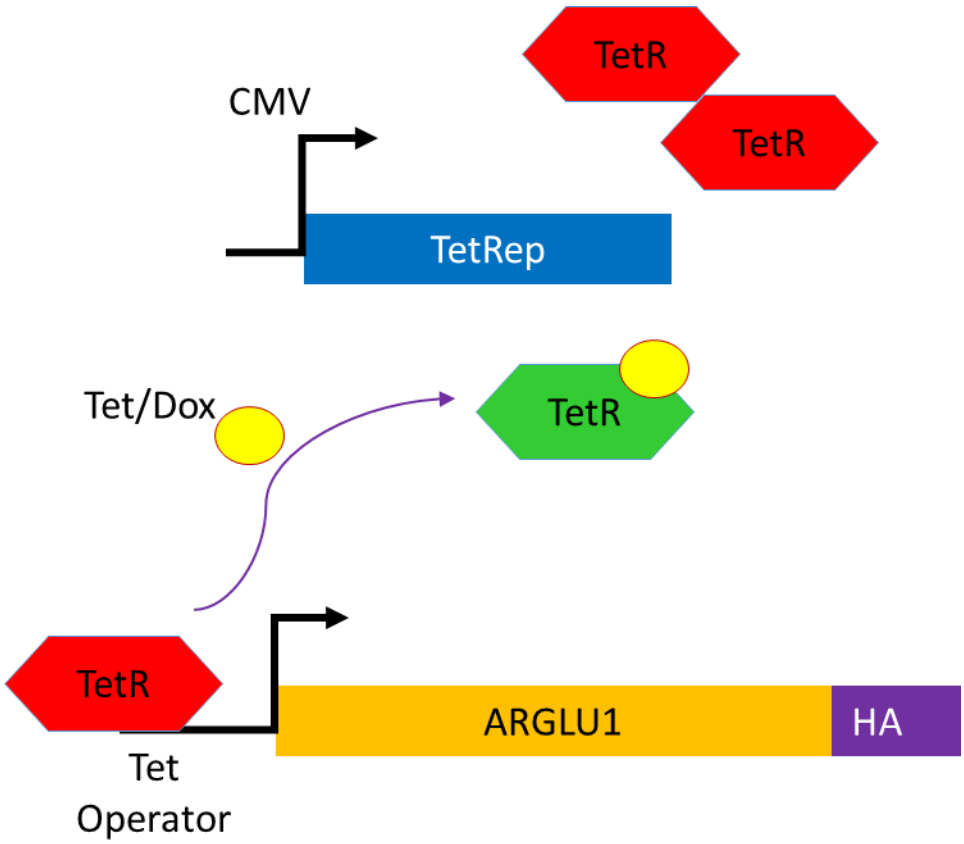

Figure 3.2 293 Flp-In TRex inducible expression of ARGLU1. This is a representation of the Flp-In T-Rex system (Life Technologies). The Flp-In component is created using the pcDNA5/FRT/TO plasmid containing the gene of interest fused to a CMV promoter with two tetracycline operator (TetO) sequences. The vector also has a FRT site adjacent to a hygromycin resistance gene. This plasmid can be recombined at a single stably integrated FRT site, which then drives the hygromycin resistance gene. There is a separate stably integrated component that constitutively expresses the tetracycline repressor (TetR) and is under blastcidin selection. The TetR gene will bind to TetO in the absence of tetracycline/doxacycline, repressing expression of the gene of interest. Addition of tetracycline will cause a conformational change in the TetR, releasing it from the TetO and allowing transcription of the gene of interest. I have made 3 stable cell lines with this system, expressing ARGLU1 cDNA (ARGLU1-TO), ARGLU1-HA cDNA (ARGLU1-HA-TO), and CAT cDNA (CAT-TO) as a control.

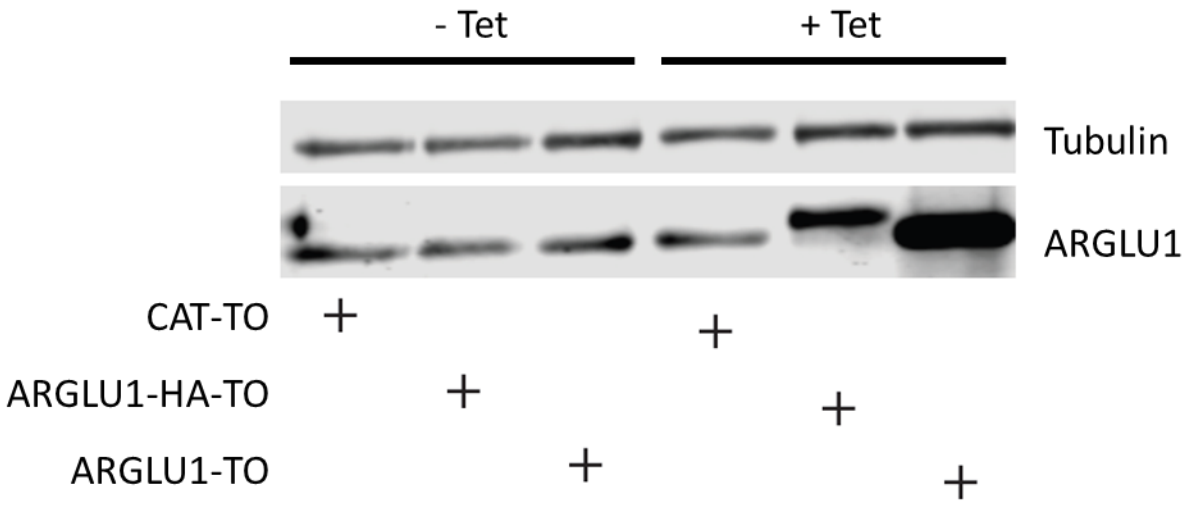

Figure 3.3 ARGLU1 overexpression from the 293 Flp-In T-Rex derived cell lines. Cell lines with inducible induction of either CAT (CAT-TO), HA tagged ARGLU1 cDNA (ARGLU1-HA-TO), or ARGLU1 cDNA (ARGLU1-TO) were either induced with 1 ug/ml tetracycline (+tet) or mock treated (-tet) for 24 hours. Western blot was performed with antibodies against alpha-tubulin as a loading control and ARGLU1. Note that the endogenous ARGLU1 band is not present following ARGLU1-HA overexpression.

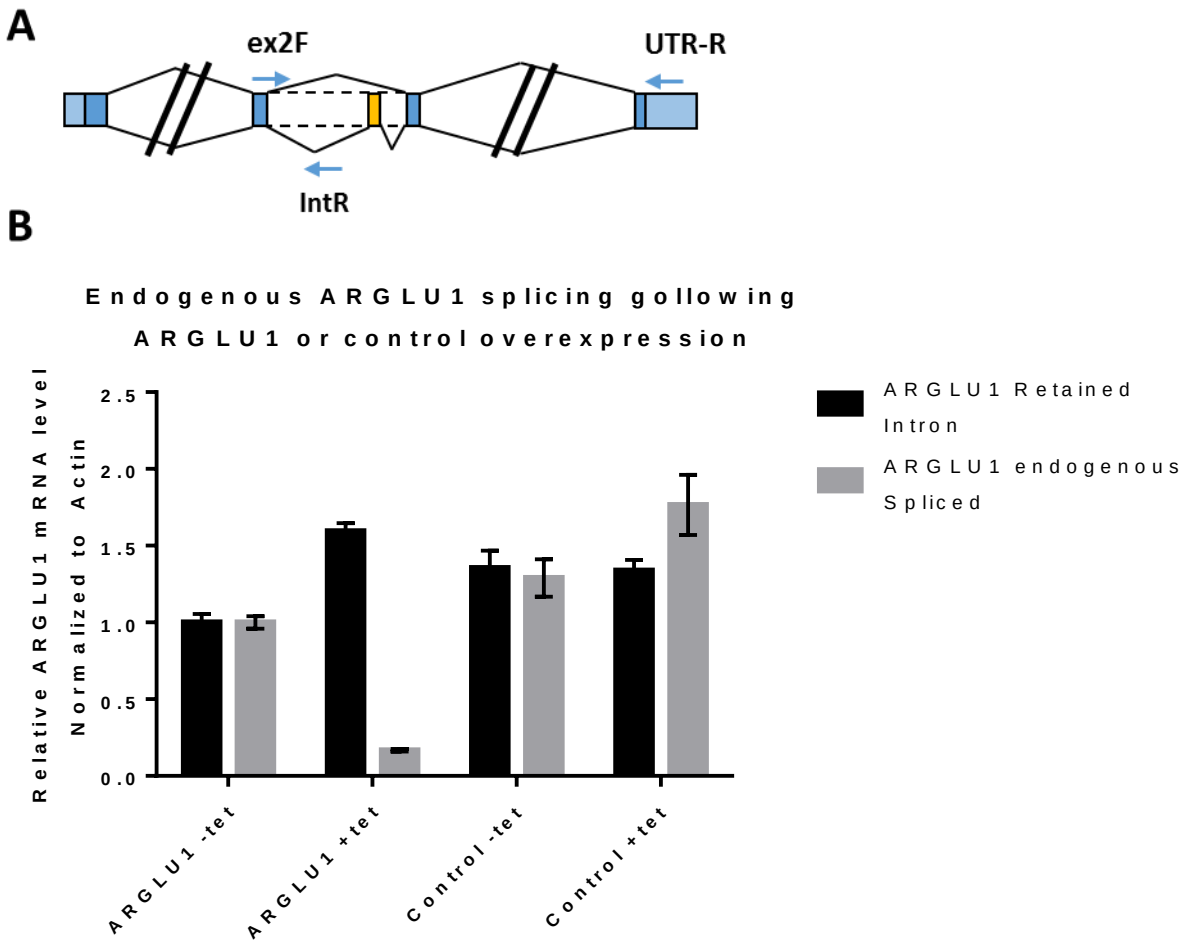

Figure 3.4 ARGLU1 overexpression increases levels of retained intron mRNA and decreases levels of spliced mRNA **A.** Primers used for RT-qPCR in B, as well as elsewhere. Ex2F+UTR-R are used to measure ARLGU1 endogenous spliced mRNA, and ex2F+IntR are used to measure ARGLU1 retained intron mRNA. **B**. RT-qPCR of the indicated HEK-293 Flp-In derived cell lines (Control is CAT-TO) either induced 1 µg/ml tetracycline (+tet) or mock treated (-tet) for 24 hours. mRNA levels are normalized to actin, and set relative to ARGLU1-TO –tet. Sample n=3, and error bars represent SEM.

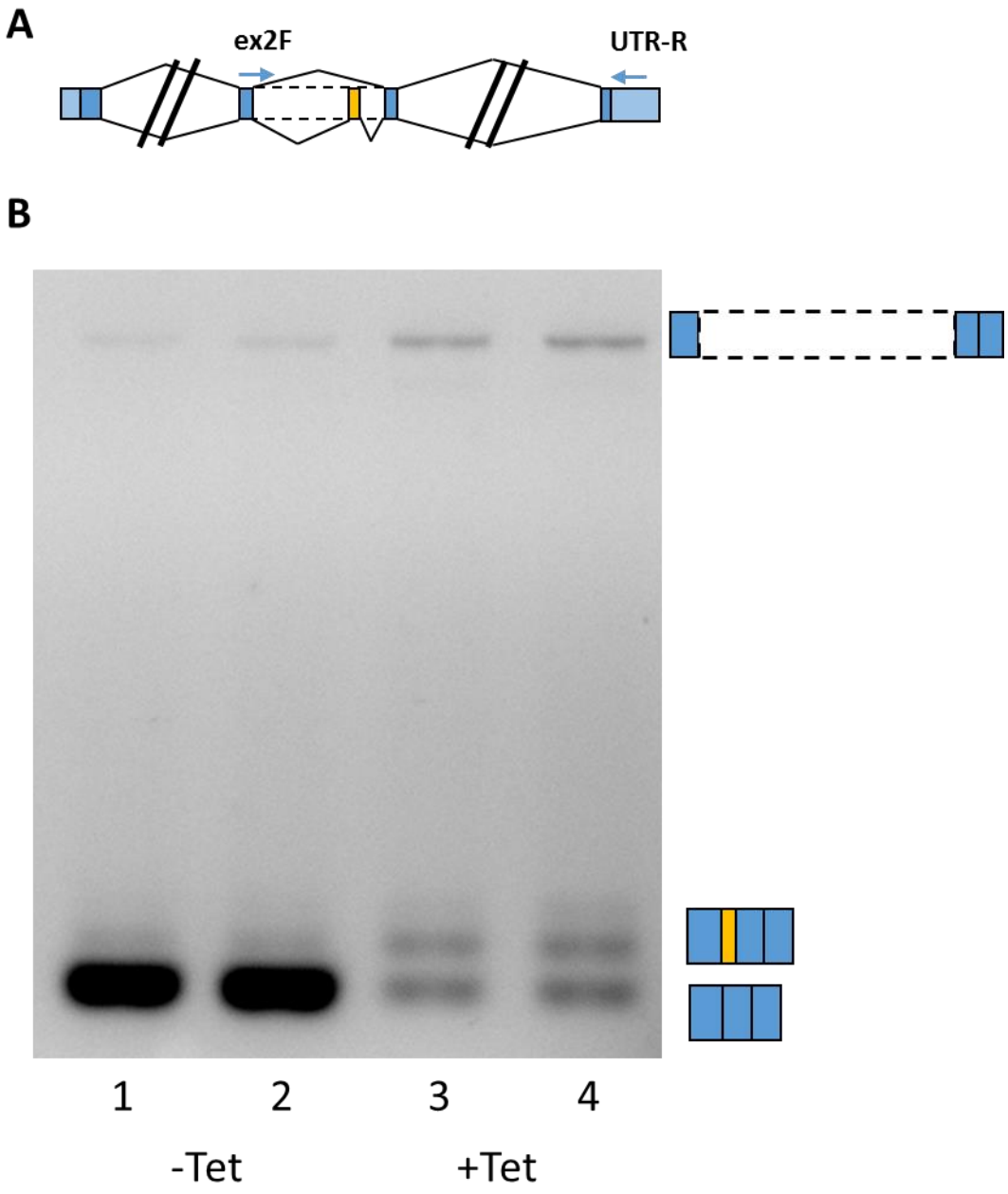

Figure 3.5 Induced overexpression of ARLGU1 causes a dramatic shift in endogenous ARGLU1 splicing A. Primers used for semi-quantitative RT-PCR. **B.** Semi-quantitative PCR was performed on RNA from ARGLU1-TO cells either induced with 1 µg/ml tetracycline (+tet) or mock treated (-tet) for 24 hours. RNA was then used for RT-PCR with a long extension time to capture all potential spliced isoforms, including fully spliced, retained intron, and alternative exon inclusion. Lanes 1 and 2 are replicates of mock treated cells. Lanes 3 and 4 are replicates of 1 ug/ml tetracycline treated cells.

3.2.3 Changes in *ARGLU1* splicing correspond closely with ARGLU1 protein levels

I next tested how quickly these splicing changes occurred in ARLGU1 following induced overexpression of the ARGLU1-HA protein. Using an HA-tagged ARGLU1 construct allowed me to assess and correlate ARGLU1 protein level in the cell with the splicing status of the endogenous ARGLU1 mRNA. I found that following induction of ARLGU1-HA the endogenous mRNA showed a significant decrease in the spliced isoform to approximately 35% of baseline levels, with no change in the retained intron isoform (Figure 3.6A). This was seen as early as 2 hours after induction, when ARGLU1-HA levels were only 75% of endogenous levels, and total ARGLU1 protein levels were approximately 1.8 times baseline (Figure 3.6B). With continued expression of ARGLU1-HA, levels of the endogenous spliced *ARGLU1* mRNA continued to decrease, with a concomitant decrease in the endogenous protein (Figure 3.6A, B). This indicated an inverse correlation between total ARGLU1 protein levels and the level of spliced *ARGLU1* mRNA.

To test the hypothesis that ARGLU1 protein levels correlate with splicing of ARGLU1 mRNA, I used emetine to inhibit translation of new proteins. This led to rapid turnover of ARGLU1, with a protein half-life of approximately 4 hours (Figure 3.7B). Using qRT-PCR to determine the levels of the spliced and retained intron isoforms of ARGLU1, I saw a dramatic 2.7-fold increase in the spliced product with a concurrent decrease in the retained intron isoform after 4 hours of treatment with emetine (Figure 3.7A). Taken together with the splicing changes seen following overexpression of ARGLU1 protein, this suggests that ARGLU1 maintains homeostatic control of its own protein level by regulating splicing of its *ARGLU1* mRNA.

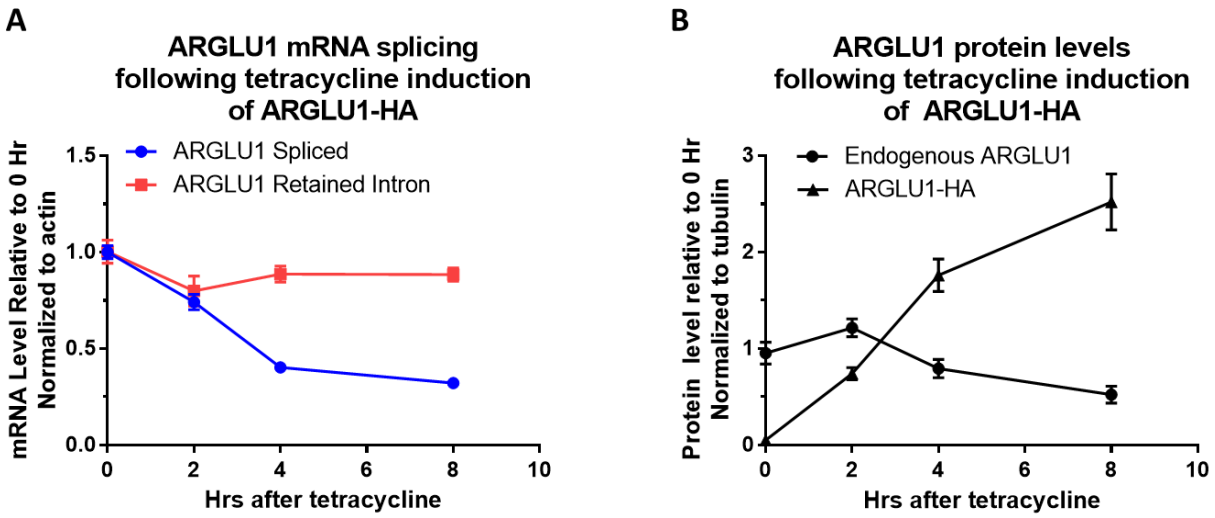

Figure 3.6 ARGLU1-HA overexpression rapidly alters endogenous ARGLU1 splicing **A.** RT-qPCR was performed using the primer sets detailed in figure 3.4A for the spliced and retained intron isoforms. RNA was harvested from ARGLU1-HA-TO cells at 2,4, and 8 hours following 1 µg/mL tetracycline induction, as well as from mock treated cells. For each time point n=3 and error bars represent the SEM. **B.** Protein quantification of ARGLU1 and ARGLU1-HA based on western blotting with an ARGLU1-antibody. Protein levels were normalized to tubulin and set relative to the endogenous protein level at time 0. For each time point n=3 and error bars represent the SEM.

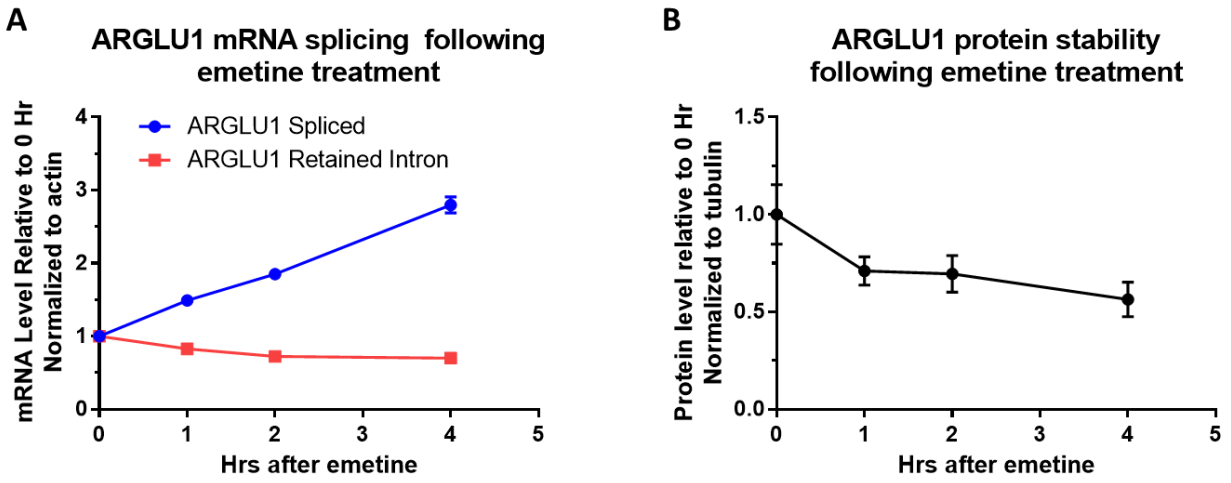

Figure 3.7 Inhibition of protein translation causes dramatic changes in ARLGU1 splicing **A.** RT-qPCR was performed using the primer sets detailed in figure 3.4A for the spliced and retained intron isoforms. RNA was harvested from HEK-293 cells at 1, 2, and 4 hours following 1 µg/mL tetracycline induction, as well as from mock treated cells. For each time point n=3 and error bars represent the SEM. **B.** Protein quantification of ARGLU1 based on western blotting with an ARGLU1-antibody. Protein levels were normalized to tubulin and set relative to the endogenous protein level at time. For each time point n=3 and error bars represent the SEM.

3.2.4 Inhibition of translation increases splicing of the protein coding isoform and stabilizes an isoform containing the alternative exon:

Emetine has been used to study transcripts subject to degradation through NMD (Ni et al., 2007; Noensie and Dietz, 2001), which is dependent on the pioneer round of translation. Given that emetine leads to decreased levels of NMD, it would be expected that mRNA levels of NMD targets should increase following treatment with emetine. I therefore performed semi-quantitative PCR on cells treated with 50 µg/mL emetine or mock treated for 4 hours to assess the stability of the spliced isoform, retained intron isoform, and alternative exon included isoform. Semi-quantitative RT-PCR showed that the spliced isoform is increased (Figure 3.8B), in agreement previous experiments shown in figure 3.7A. However, I found that the *ARGLU1* retained intron RNA levels decreased by 45% compared to untreated cells after 4 hours of treatment (Figure 3.7), in contrast to the expected result if this transcript was subject to NMD. Furthermore, 4 hour treatment with emetine caused stabilization of transcripts containing the alternative exon, indicating that it is subject to NMD as predicted (Figure 3.8B).

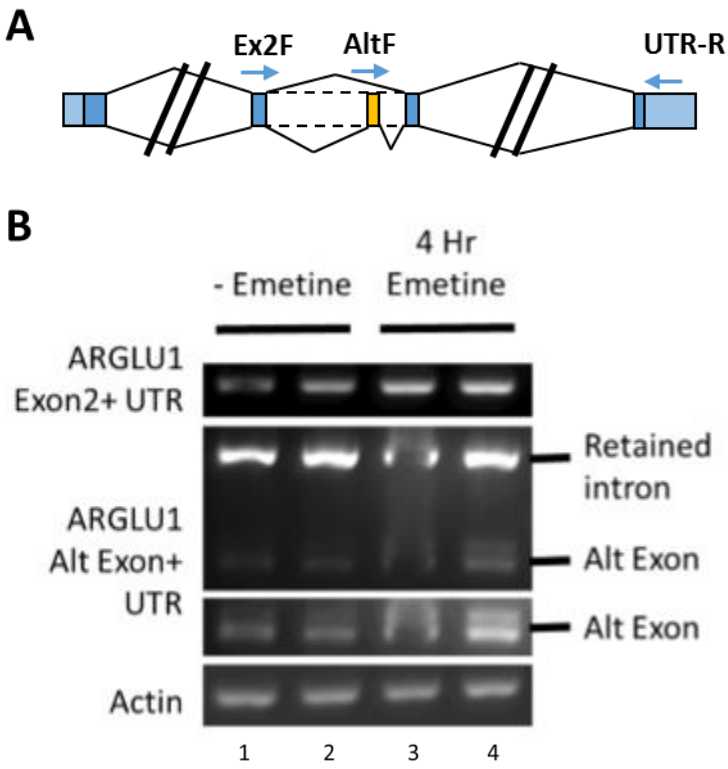

Figure 3.8 RT-PCR from cells treated with emetine shows stabilization of the alternative exon containing RNA. A. Primers used for semi-quantitative RT-PCR. Primers ex2F and UTR-R were used to look at the spliced mRNA. Primers UTR-R and altF were used to assess alternative exon inclusion, and also captured the retained intron isoform. **B.** RT-PCR was performed from cells treated with 50 µg/mL emetine (4 Hr emetine) or mock treated (-emetine) for 4 hours. ARGLU1 spliced isoform is seen in the top panel. The retained intron and alternative exon inclusion isoforms are seen in the middle panels, and actin is seen in the bottom panel. Lanes 1 and 2 are replicates of mock treatment and lanes 3 and 4 are replicates of emetine treatment.

3.3 Conclusion

In this chapter I have addressed the role of ARGLU1 protein in the regulation of its own splicing. As mentioned previously, splicing regulation takes place through *cis/trans* interactions. In the previous chapter I have described the role of a *cis* element that regulates splicing of *ARGLU1*. However, the *trans*-regulatory factors involved in this alternative splicing were not clear. I developed a hypothesis based on different lines of evidence. First, ARGLU1 contains an N-terminal domain that contains a number of SR dipeptides, an essential motif in SR and SR-like proteins. Second, ARGLU1 physically interacts with (Hegele et al., 2012), and is phosphorylated by (Varjosalo et al., 2013), SRPK1 and SRPK2, kinases that are known to regulate the activity of the SR splicing factor proteins. Third, as demonstrated in the previous chapter, *ARGLU1* has a UCE that is associated with the regulated inclusion of an alternative exon predicted to lead to NMD, a characteristic shared by many splicing factors (Lareau et al., 2007; Ni et al., 2007). Finally, a number of SR proteins have been shown to feedback and regulate the splicing of their own mRNAs to increase UCE associated exons containing PTCs that lead to degradation through NMD (Sun et al., 2010; Sureau et al., 2001). Based on these data, I formed the hypothesis that ARGLU1 may function as a regulator of splicing, and feedback to alter the splicing of its own mRNA.

To test this hypothesis, I used an inducible system to overexpress ARGLU1 cDNA to determine the impact of the ARGLU1 protein on splicing of the endogenous mRNA. Using this system I have shown that, in agreement with my hypothesis, ARGLU1 overexpression dramatically altered the splicing of the endogenous by decreasing the spliced mRNA expected to code for the protein, increasing the retained intron isoform, as well as increasing inclusion of the alternative exon that leads to degradation through NMD.

3.4 Materials and Methods

Cell Culture

HEK 293-Flp-In TRex cells were maintained in DMEM with 10% FBS and 100 U/mL Penicillin-Streptomycin. Cells in which the pcDNA5/FRT vector had been integrated were additionally maintained in media supplemented with 15 ug/mL Blastcidin and 150 ug/mL Hygromycin B. Induction of protein expression was performed by supplementing DMEM/Blast/Hygromycin media with 1 ug/mL tetracycline for the times indicated in the text.

Plasmids

A cDNA for ARGLU1 was cloned into the pcDNA3 expression vector using primers detailed in appendix A. This plasmid was used for the initial experiments for co-transfection with the A23 reporter plasmids, detailed in chapter 2. A C-terminal HA tag was added to the pcDNA3-ARGLU1 clone using Q5 site directed mutagenesis using primers found in appendix A. The ARGLU1 and ARGLU1-HA cDNA fragments were then cloned into the MCS of the pcDNA5/FRT/TO vector to generate pcDNA5/FRT/ARGLU1-TO and pcDNA5/FRT/ARGLU1-HA-TO plasmids. These plasmids, as well as the pcDNA5/FRT/CAT-TO plasmids, were individually transfected into HEK-293 Flp-In TRex cells along with the poG44 plasmid, which expressed FLP recombinase, to mediate stable integration into the genomic FRT site. These cells were then selected as noted above.

Emetine treatment

Cells were switched to DMEM 10% FBS supplemented with 50 ug/mL emetine for the time indicated in the text. Mock treated cells received fresh media with no emetine.

Western blotting

Cells were lysed for 30 minutes in ice cold RIPA buffer supplemented with protease inhibitors and benzonase. Protein was quantified using the BioRad protein quantification kit.

Cell lysates were then mixed with 2X Laemmli's SDS loading buffer with 10 mM DTT. 30 μg of protein was loaded into SDS-PAGE gels and resolved at 100V. Transfer to nitrocellulose membrane was carried out in Tris-glycine buffer with 10% methanol at 4°C at 240 mA for 2 hour. Membranes were blocked with Licor blocking buffer for 1 hr at room temperature. Blots were then incubated overnight at 4°C with primary antibody diluted (ARGLU1 1:1000, Tubulin 1:1000) in Licor blocking buffer supplemented with 0.1% tween-20. Membrane was then washed with TBS-0.1% tween 3 times for five minutes. 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 minutes at room temperature with rocking. Membrane was then washed with TBS-0.1% tween 3 times for five minutes. Membranes were imaged on the Licor Odyssey system.

Preparation of RNA

RNA was prepared by lysing cells in an appropriate volume of Trizol (Life Tech). Chloroform was added, mixed vigorously, and incubated on ice for 5 minutes. These samples were centrifuged at maximum speed for 15 minutes. The aqueous phase was then mixed with 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. RNA concentration was assessed using Absorbance at 280 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 appendix A.

Chapter 4

ARGLU1 Overexpression Alters the Expression Level and Alternative Splicing of a Subset of Genes

4.1 Introduction

In the previous chapter I have shown evidence that ARGLU1 protein levels are controlled through a negative feedback loop, where high levels of ARGLU1 protein induce complex alternative splicing of the *ARGLU1* transcript that ultimately results in lower levels of ARGLU1 protein. Overexpression of ARGLU1 increases the inclusion of an alternative exon that appears to lead to NMD degradation through the inclusion of a premature termination codon, increases intron retention of a transcript that is detained in the nucleus, while at the same time decreasing levels of the mature transcript that codes for the ARGLU1 protein. Additionally, ARGLU1 has been found to physically associate with a number of proteins involved in splicing, including in affinity purification of the spliceosome (Behzadnia et al., 2007; Bessonov et al., 2008; Hegele et al., 2012; Varjosalo et al., 2013). Taken together, these findings raise the question of whether ARGLU1 can regulate the splicing of other RNAs when overexpressed.

ARGLU1 protein has been shown to interact with MED1, a component of the mediator complex to potentiate a transcriptional and cellular response to estrogen in this context (Zhang et al., 2011). Mediator is a large, multi-subunit complex that brings together transcription factors and the core RNA pol II machinery, allowing for licensing of transcription. MED1 in particular is known to be recruited to enhancer sequences bound by nuclear receptors and potentiate transcriptional activation of regulated genes. This suggests an alternative role for ARGLU1 in transcriptional regulation.

In this chapter I will address the role of ARGLU1 protein in regulating transcription levels and alternative splicing in a transcriptome-wide manner. Based on the two probable functions for ARGLU1 suggested by previous studies, transcriptional regulation and the splicing machinery, it seems likely that ARGLU1 will play a role in RNA transcription or processing.

4.2 Results

4.2.1 RNA-Seq

To assess the role of ARGLU1 protein overexpression on the transcriptome of HEK 293 cells, I employed a previously described system to induce the overexpression of ARGLU1 in response to tetracycline. This system uses a tetracycline repressor to inhibit transcription of an ARGLU1 cDNA, which is released with the addition of tetracycline. Therefore, this system could be used to compare the mRNA compliment from cells overexpressing ARGLU1 against isogenic cells that had not been induced. Furthermore, to control for the effect of tetracycline, I created cell lines to express chloramphenicol acetyltransferase (CAT) following addition of tetracycline.

There are multiple control samples in this experimental setup. The main comparisons will be made against the ARGLU1 +Tet group, which is the ARGLU1 inducible cell line treated with tetracycline, which serves as the main control. The other control groups at the CAT +Tet group and the CAT –Tet, which are the CAT inducible cell line treated with or without tetracycline, respectively. In order to determine differences attributable to overexpression of ARGLU1, rather than tetracycline or CAT expression, I have required that differences recovered from informatics analysis are significant in either all three control samples compared with ARGLU1 +Tet, or at least significantly different in both the CAT +Tet and ARGLU1 –Tet groups.

I used two different conditions treatment conditions for both the ARGLU1 and CAT inducible cell lines. In duplicate for each condition, cells were treated with 1 µg/mL tetracycline, or mock treated, for 24 hours. During cell harvesting, half the cells were used to create a whole cell protein lysate, and the other half were treated with Trizol and processed using the Trulink RNA mini kit to extract RNA for RNA-seq and other experiments. The protein extracts were used

to assess the expression level of ARGLU1 in all samples (Figure 4.1). RNA from each sample was analyzed by the Tapestation instrument using the RNA Screentape (Figure 4.2A). After confirming that the RNA Integrity Number equivalent (RINe) (Schroeder et al., 2006) indicated the RNA was of high quality, it was used to generate a stranded cDNA library appropriate for sequencing on an Illumina platform using the Illumina Stranded mRNA TruSeq protocol. The libraries were run on a high sensitivity DNA screentape (Figure 4.2B) to determine the average size distribution of the libraries for molarity calculations, and as a check for aberrant products following PCR amplification. These libraries were then sequenced with 2x76 bp reads on the Illumina NextSeq platform.

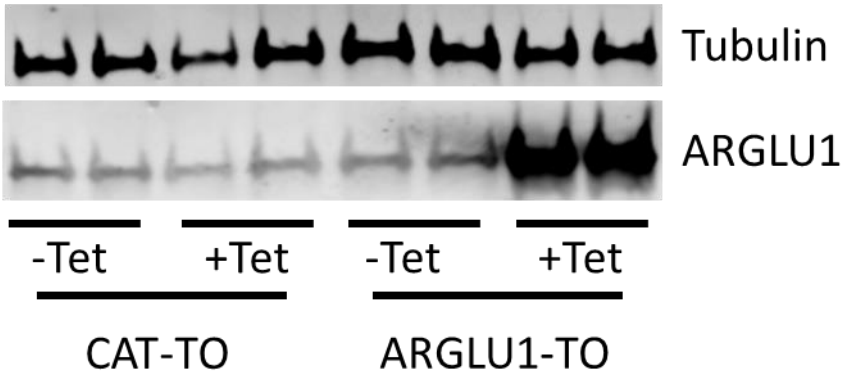

Figure 4.1 Western blot of cell extracts from the cells used for RNA-seq experiments. Western blotting was carried out on whole cell extracts from the samples used for RNA-Seq. Tubulin was used as a loading control. The cell lines are indicated at the bottom, and treatment with 1 ug/mL tetracycline or mock treatment is indicated below the individual samples.

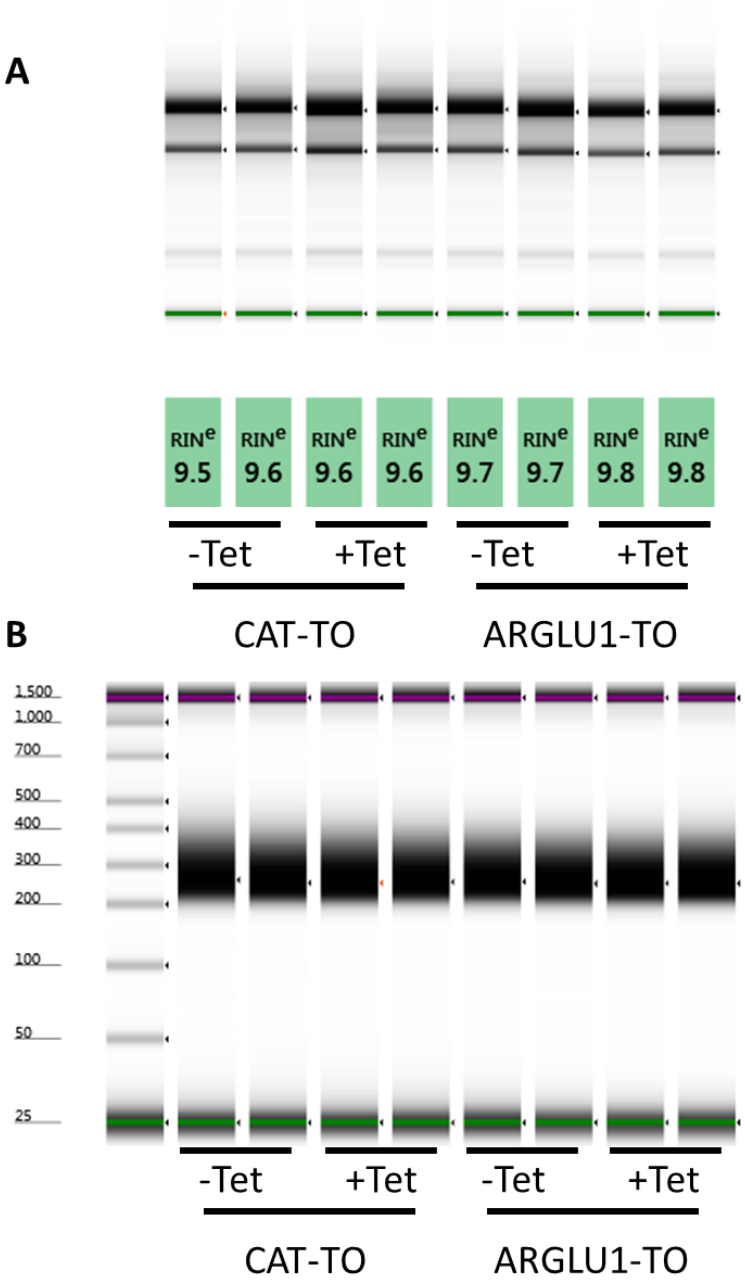

Figure 4.2 Quality control steps for RNA integrity and library size for RNA-seq libraries. **A.** RNA Screentape gel image showing the ribosomal RNA bands from the RNA used for library preparation. The RINe number, and indicator of RNA integrity, is shown at the bottom of the gel image. **B.** High Sensitivity DNA Screentape image of the RNA-seq libraries. The cell lines are indicated at the bottom, and treatment with 1 µg/mL tetracycline or mock treatment is indicated below the individual samples.

4.2.2 Alignment of RNA-Seq Data

TopHat2 (Kim et al., 2013) was used to align FASTQ files derived from RNA-Seq raw data to the hg19 human reference genome. Alignment statistics for the 8 individual samples are presented in Table 4.1. All samples showed a high rate of alignment to the human reference genome (See table 4.1). The output of Tophat2 is a .BAM file, which can be displayed graphically as alignments on the human genome. Furthermore, there .BAM files can be used to make a coverage track, which sums the number of reads aligning to a given base in the reference genome. These genome coverage tracks can be accessed through the UCSC genome browser.

Sample	ARGLU1 – Tet Rep 1	ARGLU1 - Tet Rep 2	ARGLU1 +Tet Rep 1	ARGLU1 +Tet Rep 2	CAT - Tet Rep 1	CAT - Tet Rep 2	CAT +Tet Rep 1	CAT +Tet Rep 2
Left Reads total	4.41E+07	7.07E+07	4.36E+07	4.37E+07	3.51E+07	5.16E+07	4.36E+07	4.07E+07
Left reads mapped	3.84E+07	6.12E+07	3.72E+07	3.80E+07	3.04E+07	4.52E+07	3.78E+07	3.56E+07
% of input	87.00%	86.58%	85.24%	86.91%	86.83%	87.53%	86.73%	87.50%
Left multiple alignment	4.30E+06	7.35E+06	4.69E+06	4.59E+06	3.34E+06	5.50E+06	4.46E+06	4.13E+06
% multiple alignment	11.21%	12.02%	12.61%	12.09%	10.96%	12.17%	11.80%	11.60%
Right reads total	4.41E+07	7.07E+07	4.36E+07	4.37E+07	3.51E+07	5.16E+07	4.36E+07	4.07E+07
Right Reads Mapped	3.81E+07	6.07E+07	3.69E+07	3.77E+07	3.02E+07	4.49E+07	3.76E+07	3.53E+07
% of input	86.36%	85.95%	84.65%	86.26%	86.10%	86.96%	86.17%	86.73%
Right multiple alignments	4.38E+06	7.45E+06	4.75E+06	4.65E+06	3.39E+06	5.60E+06	4.53E+06	4.20E+06
% multiple alignment	11.49%	14.27%	12.87%	12.34%	11.23%	12.47%	12.05%	11.89%
Aligned Pairs	3.74E+07	5.97E+07	3.63E+07	3.71E+07	2.97E+07	4.41E+07	3.69E+07	3.47E+07
Paired Multiple Alignments	3.17E+06	5.44E+06	3.51E+06	3.43E+06	2.45E+06	4.08E+06	3.33E+06	3.05E+06
Discordant Aligned Pairs	3.52E+05	5.96E+05	3.94E+05	3.80E+05	2.79E+05	4.45E+05	3.69E+05	3.30E+05
% Concordant Pair Alignment	84.09%	83.61%	84.24%	83.96%	83.85%	84.52%	83.77%	84.45%

Table 4.1 Alignment statistics for the 8 samples sequenced.

4.2.3 Overexpression of ARGLU1 alters the abundance of a number of genes

I utilized the RNA-seq statistical package Cuffdiff (Trapnell et al., 2013), which is incorporated in the Cufflinks suite of RNA-seq analysis tools (Trapnell et al., 2010), to quantitatively compare the alignments derived from TopHat on a gene and isoform level. Cuffdiff generates uses a measure known as Fragments Per Kilobase of transcript per Million reads mapped (FPKM), which allows for comparisons of transcript levels between samples. Based on the Cuffdiff output, 16 genes were significantly upregulated in the ARGLU1 +Tet group compared with all three control groups. A further 8 genes were upregulated in the ARGLU1 +Tet group compared with the CAT +Tet and ARGLU1 –Tet groups. There were substantially more genes downregulated following ARGLU1 overexpression, with 107 genes downregulated following in the ARGLU1 +Tet group compared with all three control groups, and 54 genes downregulated compared with the CAT +Tet and ARGLU1 –Tet groups. These genes are listed in Table 4.2.

To determine if there was any enrichment of genes with similar molecular function, I used the list of up- and downregulated genes to query the DAVID functional annotation tool (Table 4.3). I found that the most highly enriched molecular function was catalytic activity, indicating that ARGLU1 seems to be regulating the expression levels of enzymes. Two other enriched molecular functions were the categories lyase and hydrolase.

To validate the Cuffdiff findings of genes that were upregulated and downregulated, I performed semi-quantitative RT-PCR for selected. I found that when RT-PCR was performed on duplicate samples from ARGLU1 –Tet and ARGLU1 +Tet cells, these RT-PCR reactions followed the trends observed in by Cuffdiff analysis (Figure 4.3A and B). Interestingly, two genes involved in paraspeckles, PSPC1 and the lncRNA NEAT1, were both downregulated upon ARGLU1 overexpression (Figure 4.3B)

Upregulated Significantly in ARGLU1 +Tet compared to 3 groups				Upregulated in ARGLU1 +Tet compared to both CAT +Tet and ARGLU1 -Tet	
ABHD16A	BTN3A3	DDX39A	RIMKLB	CMTR2	TCP10L
ARGLU1	C21orf140	FBXL5	SNRPA1	COQ2	ZNF117
ATF7IP2	CCNH	N6AMT1	SRSF2	MOSPD2	ZNF223
BTN3A2	CHPF2	OVGP1	TNFRSF10D	NAPB	ZSCAN12P1
Downregulated Significantly in ARGLU1 +Tet compared to 3 groups				**Downregulated in ARGLU1 +Tet compared to both CAT +Tet and ARGLU1 -Tet**	
ABHD14B	FAM86B1	MIR4435-1HG	SAT1	ABCA2	MXRA8
ACYP1	FAM86FP	MPC1	SERPINB1	ABHD16B	NAGLU
AIG1	FAM86HP	MRPL1	SFXN2	ATHL1	NAT14
ANKRD29	FBXL4	NEAT1	SHISA4	C17orf70	NDUFV3
AP1S2	GALT	NIPSNAP3A	SIGMAR1	C1QTNF6	NPTXR
APTX	GATS	NPL	SLC35G1	CBS	NUDT18
ASB16-AS1	GDAP1	NRBP2	SLC5A3	CCNL2	NUDT8
ASIC1	GNE	PCYOX1L	SMARCD3	CENPM	OCEL1
ATP5G1	H6PD	PEX7	STYXL1	CYB5D2	PAPLN
ATP7B	HAGHL	PIGV	TCTA	D2HGDH	PCBP4
ATPAF2	HDHD2	PIGX	TCTN1	DNPH1	PCK2
C22orf29	HGSNAT	PM20D2	TDRP	DPYSL4	PHF7
C22orf39	HMGCL	PPCDC	THBS3	FAM132B	PIGQ
CBR4	HOXC-AS1	PPCS	THSD4	GALNT16	PPM1M
CCDC53	HSD17B8	PPFIA4	TMCC1-AS1	HINT2	PYCRL
CDHR1	IFT88	PPM1N	TMEM187	IDH2	RPUSD3
CLN5	IGDCC4	PPP1R21	TUSC3	IL17RC	SEMA6C
COL4A4	LINC00152	PRAF2	TXNRD3	KIAA0141	SH3GLB2
COL4A5	LINC00173	PRELID2	TYSND1	LOC100129534	SPATA20
DDIT4	LINC01004	PSPC1	UCP2	LOC101927667	SPC24
DDT	LINC01184	RBM3	UNC119B	LTBP4	ST3GAL3
DENND5B	LINC01355	RBPMS2	UNK	MAPK11	TMEM219
DHRS11	LYRM1	RHBDD2	UQCC2	MAPK12	TMEM42
DOK3	LYRM9	RNF216P1	UROS	MC1R	TPCN1
EMC3-AS1	LZTS3	RRAGB	WDR31	MFSD3	TRAPPC6A
EPHX1	METTL21B	RRM2	WEE1	MMP24-AS1	TTC38
EPHX2	MGC72080	RWDD2B		MORN1	ZNF586

Table 4.2 Genes significantly upregulated and downregulated in ARGLU1 +Tet group compared to controls.

Category	Term	Count	P Value	Fold Enrichment	Benjamini	FDR	Genes
Molecular Function	catalytic activity	67	1.83E-08	1.77E+00	5.90E-06	2.48E-05	HGSNAT, SAT1, SPATA20, D2HGDH, TUSC3, LTBP4, HINT2, PPCS, CHPF2, N6AMT1, ST3GAL3, MC1R, ACYP1, HMGCL, NUDT18, PCYOX1L, CCNH, RPUSD3, PIGV, APTX, CBR4, SIGMAR1, PIGQ, WEE1, COQ2, H6PD, RRM2, TYSND1, DDT, FBXL5, TXNRD3, PPM1N, HAGHL, PPM1M, NRBP2, PYCRL, STYXL1, NAGLU, GNE, UROS, ATHL1, ABCA2, ATP5G1, PPCDC, NUDT8, IDH2, NAT14, HSD17B8, OVGP1, PM20D2, DHRS11, NPL, GALT, EPHX2, EPHX1, DPYSL4, ABHD14B, PAPLN, MAPK11, PCK2, RIMKLB, NDUFV3, MAPK12, THSD4, HDHD2, ATP7B, CBS
Molecular Function	carbon-carbon lyase activity	5	2.30E-04	1.64E+01	3.64E-02	3.11E-01	DDT, NPL, PPCDC, PCK2, HMGCL
Molecular Function	lyase activity	7	7.75E-04	6.38E+00	7.99E-02	1.04E+00	DDT, NPL, UROS, PPCDC, PCK2, HMGCL, CBS
Molecular Function	manganese ion binding	7	8.59E-04	6.26E+00	6.68E-02	1.15E+00	NUDT18, NUDT8, PPM1N, IDH2, PPM1M, PCK2, RIMKLB
Molecular Function	cofactor binding	7	0.00922	3.87E+00	4.49E-01	1.18E+01	D2HGDH, TXNRD3, UROS, IDH2, CBR4, HMGCL, CBS
Molecular Function	hydrolase activity	27	0.0102	1.63E+00	4.26E-01	1.30E+01	STYXL1, NAGLU, HINT2, ATHL1, ABCA2, ATP5G1, MC1R, NUDT8, ACYP1, NUDT18, OVGP1, CCNH, PM20D2, EPHX2, DPYSL4, APTX, EPHX1, PAPLN, ABHD14B, H6PD, TYSND1, THSD4, PPM1N, HDHD2, HAGHL, PPM1M, ATP7B
Molecular Function	magnesium ion binding	9	0.0163	2.74E+00	5.31E-01	2.00E+01	NUDT18, MAPK12, NUDT8, EPHX2, PPM1N, IDH2, PPM1M, WEE1, ATP7B
Molecular Function	oxidoreductase activity	11	0.0266	4.20E+00	6.63E-01	3.06E+01	NDUFV3, PYCRL, D2HGDH, PCYOX1L, H6PD, RRM2, DHRS11, TXNRD3, IDH2, CBR4, HSD17B8
Molecular Function	coenzyme binding	5	0.0417	3.80E+00	7.83E-01	4.38E+01	D2HGDH, TXNRD3, IDH2, CBR4, HMGCL
Molecular Function	transferase activity	19	0.0754	1.49E+00	9.20E-01	6.53E+01	SAT1, HGSNAT, TUSC3, GNE, CCNH, LTBP4, HINT2, PIGV, GALT, MAPK11, CHPF2, PIGQ, WEE1, N6AMT1, ST3GAL3, COQ2, MAPK12, NAT14, NRBP2

Table 4.3 Molecular Function GO analysis of genes significantly upregulated and downregulated following ARGLU1 overexpression.

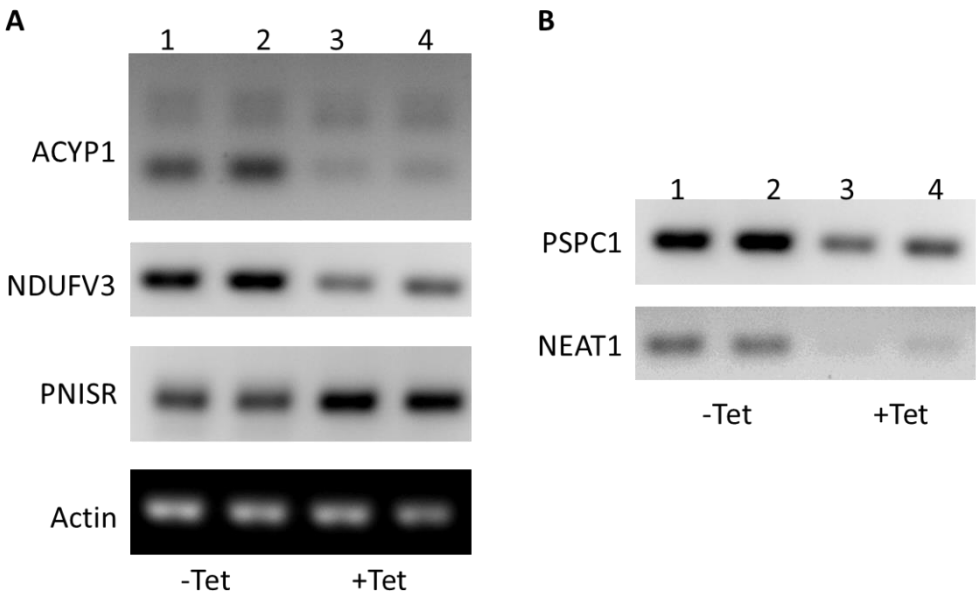

Figure 4.3 Semi-Quantitative RT-PCR confirms upregulated and downregulated genes following ARGLU1 overexpression. A. RT-PCR was performed for the indicated genes. **B.** RT-PCR was performed for the paraspeckle components PSPC1 and NEAT1. For all samples, lanes 1 and 2 are duplicate samples from mock treated ARGLU1-TO cells, while lanes 3 and 4 are duplicate samples from ARGLU1-TO cells treated with 1 ug/mL tetracycline.

4.2.4 Overexpression of ARGLU1 causes alternative splicing of a number of genes

Besides assessing gene level differences in expression, Cuffdiff is also able to create expression profiles for individual isoforms of expressed genes, thus allowing for an analysis of alternative splicing. To this end, I queried the differential splicing Cuffdiff output for genes that were significantly alternatively spliced. I have summarized these findings in Table 4.4. Notably this is a much smaller list. This may be explained by the fact that each group only contained two replicates. Cuffdiff generally requires three replicates for each group to make alternative splicing predictions. This program may not be able to generate robust alternative splicing results with only two replicates. Furthermore, there were many more alternative splicing events that were annotated between ARGLU1 +Tet and the CAT +/- Tet samples that may not have risen to the level of significance in the ARLGU1 –Tet sample because only two samples were used, or because there was some basal ARGLU1 expression in the absence of tetracycline that confounded these results.

I used the alternatively spliced Cuffdiff output to query the DAVID functional annotation tool to determine enrichment in molecular function (Table 4.5). Notable molecular functions included oxoreductase activity, cation binding, and cytoskeletal components.

In order to validate the alternative splicing predicted, I performed RT-PCR using primers that would include the alternative splicing event by crossing multiple exons, including alternative exons. RT-PCR indicated that alternative splicing found in the selected genes could be replicated using an alternative measure (Figure 4.4).

Significantly Alternatively Spliced in ARGLU1 +Tet compared to 3 control groups				Significantly Alternatively Spliced in ARGLU1 +Tet compared to both CAT +Tet and ARGLU1 -Tet		
AK2	FGFR2	NFATC1	SERPINB1	ADD3	ITPR1	SYTL5
BAG6	GPR98	NIN	SLC25A1	AGER	LAMTOR3	UBE2Q2
BUB3	HIBCH	NUDT7	STAG2	ALOX5	PCDH7	ZNF430
C8orf59	IL17RC	P4HA2	TCF19	AOC3	PMS1	ZNF438
CDC42EP3	KDM6A	PAX6	TUSC3	BTN3A3	RASSF7	FAM63A
CEP131	KIAA0319	PPARGC1B	UNK	CA12	RPH3AL	NUDT7
DEPDC1	MIS12	PRMT2	URI1	INCENP	SLC26A6	PLCB3
DPH3	MORF4L2	RALGPS2	YIPF1			
DTNA	MRPS6,SLC5A3	RFC5	ZNF207			
EPB41L4B	MYO1B	RFX3	ZXDC			
ETV1	NDUFV3	SAMD4A				

Table 4.4 Genes found to be significantly alternatively spliced in the ARGLU1 +Tet group.

Category	Term	Count	P Value	Fold Enrichment	Benjamini	FDR	Genes
Molecular Function	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	3	0.0254	1.19E+01	9.08E-01	27.27	KDM6A, P4HA2, ALOX5
Molecular Function	metal ion binding	23	0.0298	1.47E+00	8.45E-01	31.22	SLC5A3, KDM6A, ZNF430, CA12, ZXDC, RPH3AL, DPH3, PCDH7, ITPR1, GPR98, ZNF207, SLC26A6, PLCB3, P4HA2, NUDT7, SYTL5, TCF19, UNK, ALOX5, ZNF438, ADD3, DTNA, AOC3
Molecular Function	cation binding	23	0.0331	1.46E+00	7.89E-01	34.03	SLC5A3, KDM6A, ZNF430, CA12, ZXDC, RPH3AL, DPH3, PCDH7, ITPR1, GPR98, ZNF207, SLC26A6, PLCB3, P4HA2, NUDT7, SYTL5, TCF19, UNK, ALOX5, ZNF438, ADD3, DTNA, AOC3
Molecular Function	cytoskeletal protein binding	6	0.0376	3.15E+00	7.58E-01	37.75	MYO1B, RPH3AL, EPB41L4B, ADD3, CDC42EP3, GPR98
Molecular Function	ion binding	23	0.0388	1.44E+00	7.05E-01	38.73	SLC5A3, KDM6A, ZNF430, CA12, ZXDC, RPH3AL, DPH3, PCDH7, ITPR1, GPR98, ZNF207, SLC26A6, PLCB3, P4HA2, NUDT7, SYTL5, TCF19, UNK, ALOX5, ZNF438, ADD3, DTNA, AOC3
Molecular Function	calcium ion binding	8	0.0506	2.31E+00	7.46E-01	47.39	SLC26A6, PLCB3, PCDH7, ALOX5, ITPR1, GPR98, DTNA, AOC3
Molecular Function	transition metal ion binding	16	0.0736	1.52E+00	8.29E-01	61.15	KDM6A, ZNF430, CA12, ZXDC, RPH3AL, DPH3, ZNF207, P4HA2, NUDT7, SYTL5, TCF19, ALOX5, UNK, ZNF438, DTNA, AOC3

Table 4.5 Molecular function Gene Ontology of genes with significant alternative splicing following ARGLU1 overexpression.

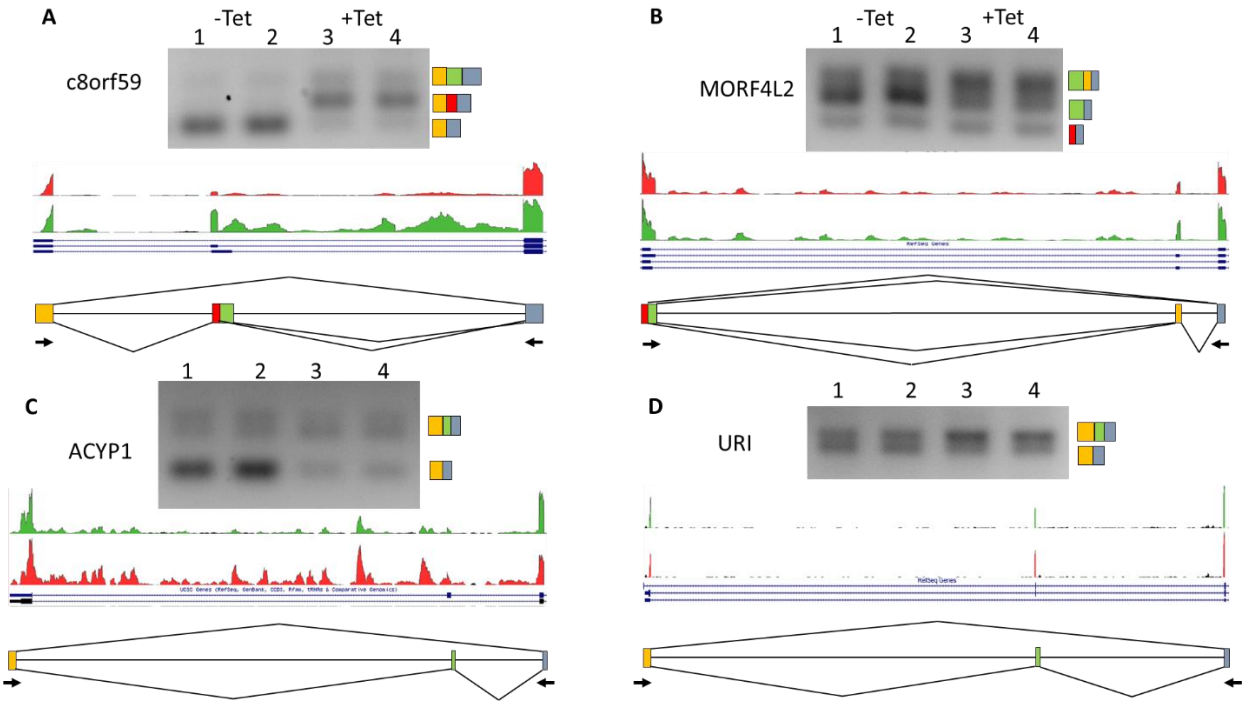

Figure 4.4 Validation of alternative splicing following ARGLU1 overexpression. **A.** Splicing sensitive RT-PCR for the c8orf59 transcript. **B**. Splicing sensitive RT-PCR for MORF4L2. **C.** Splicing sensitive RT-PCR for ACYP1. **D.** Splicing sensitive RT-PCR for URI. For all experiments, lanes 1 and 2 are duplicate samples from mock treated ARGLU1-TO cells, while lanes 3 and 4 are duplicate samples from ARGLU1-TO cells treated with 1 ug/mL tetracycline. For all tested events, the UCSC browser track for this alternative splicing event as well as a cartoon depiction of the alternative splicing event is noted. Primers are indicated by black arrows beneath the cartoon.

4.3 Conclusion

In this chapter I have assessed the transcriptome-wide effect of ARGLU1 overexpression. I found that a number of genes are upregulated and downregulated following inducible overexpression of ARGLU1. These genes are enriched for Gene Ontology categories including catalytic activity and ion binding. Furthermore, a number of genes were found to be alternatively spliced by Cuffdiff analysis, despite technical shortcomings resulting from the use of two RNA-seq replicates per group rather than three. I was able to validate the findings from RNA-Seq for a number of genes using RT-PCR to assess both expression level as well as alternative splicing.

Overexpression may not represent the ideal condition for assessing transcriptional and splicing events regulated by ARGLU1. A better system may be to look at cells in which ARGLU1 has been knocked out. I will address methods to generate these results in the next chapter.

4.4 Materials and Methods

Cell Culture

HEK 293-Flp-In TRex cells were maintained in DMEM with 10% FBS and 100 U/mL Penicillin-Streptomycin. Cells in which the pcDNA5/FRT vector had been integrated were additionally maintained in media supplemented with Blastcidin and Hygromycin B. Induction of protein expression was performed by supplementing DMEM/Blast/Hygromycin media with 1 ug/mL tetracycline.

Western blotting

Cells were lysed for 30 minutes in ice cold RIPA buffer supplemented with protease inhibitors and benzonase. Protein was quantified using the BioRad Cell lysates were then mixed with 2X Laemmli's SDS loading buffer with 10 mM DTT. 30 µg of protein was loaded into SDS-PAGE gels and resolved at 100V. Transfer to nitrocellulose membrane was carried out in Tris-glycine buffer with 10% methanol at 4°C at 240 mA for 2 hour. Membranes were blocked with Licor blocking buffer for 1 hr at room temperature. Blots were then incubated overnight at 4°C with primary antibody diluted in Licor blocking buffer supplemented with 0.1% tween-20. Membrane was then washed with TBS-0.1% tween 3 times for five minutes. 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 minutes at room temperature with rocking. Membrane was then washed with TBS-0.1% tween 3 times for five minutes. Membranes were imaged on the Licor Odyssey system.

Preparation of RNA

RNA was prepared by lysing cells in an appropriate volume of Trizol (Life Tech). Chloroform was added, mixed vigorously, and incubated on ice for 5 minutes. These samples were centrifuged at maximum speed for 15 minutes. The aqueous phase was then mixed with 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. RNA concentration was assessed using Absorbance at 280 on a nanodrop spectrophotometer. 1 uL of RNA for library preparation was run on the RNA Screentape on the tapestation machine.

RNA-Seq library generation

RNA was prepared for sequencing using the TruSeq Stranded mRNA Library Prep Kit from Illumina according to the manufacturer's direction. Briefly, poly(A) mRNA was selected with two rounds of oligo-d(T) selection. RNA was then fragmented, and reverse transcription was carried out using random primers. Second strand synthesis was then carried out, with dUTP in place of dTTP. The ds-cDNA was then end repaired and A-tailed. Adaptors were then ligated. Each sample was tagged with a unique barcode. Finally, libraries were amplified with 15 cycles of PCR and purified away from the primers using Axyprep PCR cleanup kit (Axygen). The final libraries were then quantified using Qubit 2.0 high sensitivity (Life Technologies). 1 uL of each library was also run on the Tapestation instrument using a high sensitivity DNA Screentape (Agilent). These results were used to calculate the molarity of the libraries. Sequencing libraries were diluted to 2 nM and pooled. PhiX sequencing library was added at 5% of the pooled libraries. Finally, the pooled library was prepared for sequencing on the NextSeq 500 using a 150 cycle high output flow cell, with 2x76 read length.

Alignment

After sequencing was complete, FASTQ files were generated and concatenated for each sample. FASTQs were then aligned with the hg19 Human Genome Reference from UCSC using Tophat 2 to output genomic alignments in the form of .bam files. Accepted_hits.bam file for each sample were then sorted using the samtools sort command, and indexed using the samtools index command. These .bam files were then turned into genome coverage tracks using the bedtools command genomeCoverageBed -bg -split –ibam. This genome coverage track was then used to make a BigWig track with the command bedGraphToBigWig. These files were then used for display on the UCSC genome browser.

Cuffdiff analysis

I ran Cuffdiff with the following command options:

cuffdiff -min-reps-for-js-test 1-L C_min,C_plus,A_min,A_plus -p 24 -library-type fr-firststrand

using the UCSC_hg19 genes.gtf file as a reference, which was obtained from the Tophat2 website at https://ccb.jhu.edu/software/tophat/igenomes.shtml, accessed on 30 November, 2014.

Chapter 5

Discussion and Future Directions

5.1 Discussion

In this dissertation, I have addressed a number of questions regarding the regulation and regulatory function of the protein ARGLU1. The first question I addressed is the role of an ultraconserved element (UCE) in the complex alternative splicing of *ARGLU1* mRNA. These experiments demonstrated that the *ARGLU1* UCE is necessary and sufficient to induce intron retention and inhibition of splicing. I next addressed the question of whether ARGLU1 protein can regulate alternative splicing of its own mRNA. Indeed, overexpression of ARGLU1 causes a rapid and dramatic shift in the splicing of endogenous ARGLU1 mRNA. Finally, using RNA-Seq, I asked whether ARGLU1 protein can cause changes in splicing of other genes. I will now discuss each of these findings to put them in the context of previous reports.

5.1.1 *Cis*-Element control of alternative splicing

Since alternative splicing was first described in the early 1980s (Alt et al., 1980; Early et al., 1980), there have been efforts to understand the regulatory elements present within RNA transcripts themselves that direct alternative splicing , and were known to interact with *trans*-factors. As described earlier, these splicing regulatory elements (SREs) can be found in both introns and exons, and can both enhance and suppress splicing (Braunschweig et al., 2013; Fu and Ares, 2014; Wang and Burge, 2008).

In *ARGLU1* alternative splicing, the *cis*-elements contained in the UCE seem to have a dual role. One function is to suppress inclusion of the alternative exon. These elements are located downstream of the alternative exon. The evidence for this is found in experiments in which three

25 bp deletion were made downstream of the alternative exon that increased the inclusion of the alternative exon. Interestingly, these deletions also caused usage of 5'-splice sites other than the annotated site for the alternative exon. Furthermore, mutation of GU dinucleotides (specifically SS5 and SS6) downstream of the alternative exon caused a dramatic increase in splicing of exons 2 and 3. Therefore, there are a number of sequences in the downstream region that are acting as intronic splicing silencers.

One model that I proposed for the inhibition of splicing between *ARGLU1* exon 2-3 is the presence of multiple competing 5'-splice sites following the alternative exon. It has been shown that multiple 5'-splice sites, each engaged with U1 can cause inefficient recognition of the true 5'-splice site and prevent the splicing reaction from proceeding past the E complex (Roca et al., 2013; Sharma et al., 2005). Therefore, it is possible that the alternative exon is able to undergo exon definition, but the presence of multiple 5'-splice sites interacting with the 3'-splice site of exon 3 causes stalling of the splicing reaction, which, if not resolved results in intron retention.

Another model, which is not mutually exclusive with the first model, is that the alternative exon has very poor splice sites, which can engage with the splice sites of the upstream and downstream exons, but result in an abortive splicing reaction, possibly because of the competition for 5'-splice sites proposed in model 1. There is certainly evidence that the alternative exon splice sites are weak, as mutation of these splice sites individually caused exclusive splicing to the alternative exon. Weak splice sites have been observed to be a common phenomenon in alternatively spliced exons, as it allows for regulation through the interaction between *cis*-regulatory elements and *trans*-regulatory factors (Sakabe and de Souza, 2007; Zhan, 2013). The possible regulatory role of *trans*-acting factors will be discussed in the next section.

5.1.2 ARGLU1 regulates its own alternative splicing—A new splicing *trans*-regulatory element?

In Chapter 3, I presented evidence that overexpression of an *ARGLU1* cDNA caused a dramatic and rapid shift in the splicing of *ARGLU1* mRNA away from exon 2-3 splicing toward increased intron retention as well as increased inclusion of the alternative exon. Furthermore, treatment with the translation inhibitor emetine caused an increase in the level of *ARGLU1* spliced isoform that correlated with a decrease in ARGLU1 protein levels. An additional insight into the emetine treatment experiments is that under translational inhibition, levels of mRNA including the alternative exon were increased, indicating that this transcript is most likely degraded by the nonsense mediated decay pathway (NMD), as predicted.

Interestingly, very similar alternative splicing regulation has been seen in a number of splicing factors, and are often associated with ultraconserved elements (Baek and Green, 2005; Lareau et al., 2007; Ni et al., 2007; Sun et al., 2010; Sureau et al., 2001). Indeed, experiments very similar to those I performed, in which SRSF1 (SF2/ASF) (Sun et al., 2010) or SRSF2 (SC35) (Sureau et al., 2001) were overexpressed under exogenous control, found that these splicing factors caused increased usage of alternative exons that induced NMD. Furthermore, SRSF1 has a number of retained intron isoforms that are localized in the nucleus (Boutz et al., 2015; Sun et al., 2010).

The exact mechanism of ARGLU1 autoregulation is not clear from our experiments. However, I will suggest a number of different models. First, ARGLU1 may bind directly to sequences within the UCE or the flanking exons and cause a decrease in exon 2-3 spicing. This could be caused by increased inhibition of this splicing event, or it could be the result of increased exon 2-alt exon-exon 3 splicing. In chapter 3 I presented evidence that overexpression of ARGLU1 protein causes increased splicing inclusion of the alternative exon which targeted it for

NMD, however it is unclear exactly how much this splicing contributes to regulation of exon 2-exon 3 splicing.

A second possible mechanism is that ARGLU1 is interacting with another splicing factor to mediate alternative splicing. Multiple efforts to define the full complement of mRNA proteins have not identified ARGLU1 as an RNA binding protein (Baltz et al., 2012; Castello et al., 2012). It is known that the RS domains of SR proteins interact with other non-SR proteins, such as U1-70K, through their own RS domains (Wu and Maniatis, 1993). As I have shown in chapter 1, ARGLU1 also contains an N-terminal domain multiple groups of SR dipeptides. Thus it is not unreasonable to assume that ARGLU1 can interact with SR proteins as well as other proteins with RS domains. In the context of ARGLU1 overexpression, this could have two effects. First, it could allow for increased ARGLU1 recruitment to the RNA, possibly mediating further interactions. Alternatively, it could interact with SR proteins or some other splicing regulator, causing them to be removed from the RNA, thus squelching their function and altering splicing.

Either of these possible mechanisms can be extended to the transcriptome wide splicing effects we observed in chapter 4. The small group of genes that were seen to be alternatively spliced in chapter 4 may suggest that ARGLU1's function is not on splicing. However, the analysis of alternative splicing by Cuffdiff may have been somewhat limited by only having two sequencing replicates while the Cuffdiff literature recommends three. Below, I will suggest a number of experiments that should help clarify the mechanism of *ARGLU1* alternative splicing regulation and its larger cellular role.

5.2 Future Directions

Based on the findings presented in this dissertation and the reports of others, I suggest that ARGLU1 is a new SR-like protein that plays a role in alternative splicing. To further understand the role that ARGLU1 plays in this process, I have described a number of lines of inquiry that should be fruitful.

5.2.1 Knock out the UCE in an animal model

In this study I have presented evidence that the UCE located in intron 2 is important for regulation of complex alternative splicing of *ARGLU1*, including the intron retention phenotype as well as the inclusion of the alternative exon. The high level of sequence conservation, and the demonstrated function of this sequence suggests that this regulation is quite important, and has been under selective pressure through evolutionary time. Strikingly, these two alternative splicing events seem to be conserved in other species, and most likely play a similar role. To address the function of this UCE at an organismal level, a conditional or global knock out of the UCE sequence in an animal model such as the mouse or even *Xenopus* would be of great interest to provide insight into the evolutionary selective pressure that has preserved this sequence. Alternately, this could be accomplished by the knock-in of a construct missing the UCE. Recent advances in genome editing techniques, such as Cas9/CRISPR, should make generation of this mouse line fast and efficient.

5.2.2 Knock out ARGLU1 in an animal model or cell line

As a complimentary experiment it will be of interest to examine the knock out mouse of ARGLU1 protein. The International Mouse Phenotyping Consortium has generated a number of constructs that would be of great use for understanding the expression of ARGLU1 as well as its

cellular and development role. These knockouts exist as ES cells that can be used different ways. The knock-out scheme used provides the ability to use LacZ as a gene trap to assess the distribution of expression as well as the conditional knock-out potential. This would provide a useful tool for understanding the developmental and tissue specific role that ARGLU1 plays, as well as answering the question of whether ARGLU1 is important for cellular viability or animal development. Furthermore, the presence of a floxed allele in an ES cell line would allow for the introduction of knock-in alleles, including the knock-in of a construct missing the UCE, as discussed previously. Finally, this could allow for the production of ARGLU1 knock-out cell lines, which could be useful for performing experiments such as RNA-seq by directly comparing identical cell types.

Instead of using a mouse system to knock out ARGLU1, another tractable method would be to use the Cas9/CRISPR technology discussed previously to induce targeted double-stranded breaks, which will then be repaired by non-homologous end joining to cause insertions and deletions, causing a knock out of the gene. These cell lines could then be used to assess the functional role of ARLGU1. One key experiment that should be performed with ARGLU1 cell knockouts is RNA-seq. Often, decreased expression of proteins is much more informative than overexpression, especially if the protein is already fairly highly expressed. Additionally, RNA-Seq from a knock-out cell line would allow for a comparison to RNA-Seq results from the ARGLU1 over-expressing cells. I am currently screening cells transduced with lentivirus expressing cas9 that can be selected with puromycin along with a number of sgRNAs targeting the first and second exon of ARGLU1. I am hoping to find cells in which the ARGLU1 protein is not expressed. These cells can then be used for further experiments to test the cellular response to the lack of ARGLU1, such as analysis of *ARGLU1* mRNA splicing or global changes in mRNA expression and splicing.

5.2.3 Experiments using the inducible ARGLU1 system

I have generated a number of resources that will be useful for further experiments to understand the role of ARGLU1. First, I have generated cell lines that allow for inducible expression of ARGLU1, as well as an HA tagged versions of the protein. This tag allows for the specific immunoprecipitation of this protein. I would suggest that this will be important for a number of experiments.

Deletion of protein domains or mutation of individual amino acids in the inducible ARGLU1

First, I would suggest a closer interrogation of the ARGLU1 protein domains necessary for regulation of its own splicing. I have already cloned an HA-tagged ARGLU1 cDNA into the pcDNA5/FRT/TO plasmid, thus making deletions in the protein and overexpressing them in cells should be feasible. Furthermore, a number of serines in the RS domain of ARGLU1 has been shown to be phosphorylated (Varjosalo et al., 2013). Phosphorylation of the RS domains in SR proteins are known to be important for their function and subcellular localization (Aubol et al., 2013; Cáceres et al., 1997). Thus mutation of these serines may give some insight into their function in ARGLU1.

RNA-Immunoprecipitation based experiments

Second, if ARGLU1 is indeed important in splicing, a key experiment will be to determine what RNAs ARGLU1 is targeted to. To accomplish this, it would be useful to perform RIP-seq (Townley-Tilson et al., 2006), in which immunoprecipitation of ARGLU1 would be followed by preparation of RNA-seq libraries to determine the population of RNAs that are associated with ARGLU1. This experiment should be greatly aided by having an HA tagged protein, as there are commercially available anti-HA antibodies that work well for immunoprecipitation. Possible

pitfalls of this approach are that the interaction between ARGLU1 and its targets may be transitory, as well as the possibility of non-physiological interactions taking place after cell lysis, leading to increased background (Mili and Steitz, 2004). Another approach to identifying ARGLU1-interacting RNAs is to use CLIP-seq, in which proteins and RNAs are cross-linked with either UV (Licatalosi et al., 2008; Ule et al., 2003), or formaldehyde (Silverman et al., 2014). Notably, UV crosslinking will only create protein-RNA and RNA-RNA crosslinks, and will cause protein-protein interactions. If ARGLU1 is not an RNA binding protein but instead functions through protein-protein interactions, UV crosslinking will not be a viable method of determining the population of bound RNAs.

Purification of ARGLU1 containing complexes

A third important use for a system to induce expression of tagged ARGLU1 is to affinity purify the proteins that ARGLU1 binds to. I have carried out initial experiments using a polyclonal antibody against ARGLU1, indicating that there is a limited number of proteins bound to ARGLU1. However, a high affinity tag or two-step purification process may represent a better system to purify ARGLU1-containing protein complexes. So far other studies have only tangentially tied ARGLU1 to either the mediator complex (Zhang et al., 2011) or the spliceosome (Behzadnia et al., 2007; Deckert et al., 2006; Hegele et al., 2012). Proteomic analysis of ARGLU1 protein-protein interactions should help resolve the uncertainty about ARGLU1's molecular function. Other biochemical studies of ARGLU1 will also be helpful in purifying the protein and its associated complexes. For example, use of a protein size exclusion columns or sucrose sedimentation, followed by immunoblotting, may further facilitate characterization and purification of ARGLU1-containing complexes.

I have identified ARGLU1 as regulating the splicing of a number of transcripts, including the homeostatic splicing of its own transcript. However, as alternative splicing is a highly complex process, it is highly unlikely that ARGLU1 alone is causing this alternative splicing. Finding other proteins that regulate the splicing may be accomplished by a number of methods. First, siRNA or shRNA mediated knock down of a large number of RNA binding proteins has been a useful tool for identifying splicing events regulated by those proteins (Pandit et al., 2013; Park et al., 2004; Saltzman et al., 2011). Indeed, Saltzman et al. knocked down SRSF1 and saw decreased levels of alternative exon and increased levels of exon2-exon3 splicing in ARGLU1 mRNA (Saltzman et al., 2011 supplmental table 1). Thus, further high throughput studies of RNA binding proteins may provide useful information on proteins regulating alternative splicing of ARGLU1. A second method may be to use RNA from the UCE to identify proteins bound to this region. I have previously carried out preliminary studies using an aptamer tagged version of portions of the ARGLU1 UCE to affinity purify proteins bound to this sequence, and identify them using mass spectrometry. However, these results have not yet been functionally validated.

Fluorescent tagging to identify subcellular localization

Yet another use of the inducible expression system will be to investigate the subcellular localization of ARGLU1. Previous studies have indicated that ARGLU1 can shuttle between the nucleus and cytoplasm in a cell cycle dependent manner (Sigal et al., 2006). I would propose an experiment in which an inducible ARGLU1 is tagged with a fluorescent protein. A pulse chase experiment could then be performed to track ARGLU1's cellular localization over time. Furthermore, if combined with mutation of the serines in the RS domain, this may be a useful system to identify residues important for this nuclear/cytoplasmic shuttling.

In vitro splicing of ARGLU1

It may be of interest to make a smaller splicing reporter that can be used for *in vitro* splicing. This will be useful for a number of reasons. It would allow for an analysis of the spliceosomal complex at which ARGLU1 splicing is stalled. However, this type of splicing reporter may be hard to construct, as I have shown that a large region of the UCE (approximately 150 nucleotides) is necessary for intron retention, and in vitro splicing is inefficient across introns larger than 250 bases.

Analysis of the other highly conserved region of the A23 intron

Besides the UCE I have investigated in this report, ARGLU1 intron 2 also contains another long, highly conserved region (although not as highly conserved as the UCE). It is possible that this sequence also plays a role in *ARGLU1* splicing.

Non-splicing related functions for ARGLU1 *UCE*

I have shown that there are sequences within the ARGLU1 UCE that do not appear to function in regulation of *ARGLU1* alternative splicing. It is not immediately clear what the function of these sequences is or why they are so highly conserved. One possible role may be as a nuclear rentention signal for the unspliced *ARGLU1* RNA. However, further experiments will be needed to understand fully the function of the *ARGLU1* UCE.

Bibliography

Alföldi, J., and Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. Genome Res. *23*, 1063–1068.

Alt, F.W., Bothwell, A.L., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3' ends. Cell *20*, 293–301.

Amrein, H., Gorman, M., and Nöthiger, R. (1988). The sex-determining gene tra-2 of Drosophila encodes a putative RNA binding protein. Cell *55*, 1025–1035.

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.

Andresen, J., Aftimos, S., Doherty, E., Love, D., and Battin, M. (2010). 13q33.2 deletion: a rare cause of ambiguous genitalia in a male newborn with growth restriction. Acta Paediatr. *99*, 784–786.

Ankö, 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.

Änkö, M.-L., Müller-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.

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.

Baek, D., and Green, P. (2005). Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. Proc. Natl. Acad. Sci. *102*, 12813–12818.

Ballarati, L., Rossi, E., Bonati, M.T., Gimelli, S., Maraschio, P., Finelli, P., Giglio, S., Lapi, E., Bedeschi, M.F., Guerneri, S., et al. (2007). 13q Deletion and central nervous system anomalies: further insights from karyotype-phenotype analyses of 14 patients. J. Med. Genet. *44*, e60.

Baltz, A.G., Munschauer, M., Schwanhäusser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., et al. (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol. Cell *46*, 674–690.

Bandziulis, R.J., Swanson, M.S., and Dreyfuss, G. (1989). RNA-binding proteins as developmental regulators. Genes Dev. *3*, 431–437.

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.

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.

Bell, L.R., Maine, E.M., Schedl, P., and Cline, T.W. (1988). Sex-lethal, a Drosophila sex

determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell *55*, 1037–1046.

Bell, T.J., Miyashiro, K.Y., Sul, J.-Y., McCullough, R., Buckley, P.T., Jochems, J., Meaney, D.F., Haydon, P., Cantor, C., Parsons, T.D., et al. (2008). Cytoplasmic BKCa channel intron-containing mRNAs contribute to the intrinsic excitability of hippocampal neurons. Proc. Natl. Acad. Sci. *105*, 1901–1906.

Bell, T.J., Miyashiro, K.Y., Sul, J.-Y., Buckley, P.T., Lee, M.T., McCullough, R., Jochems, J., Kim, J., Cantor, C.R., Parsons, T.D., et al. (2010). Intron retention facilitates splice variant diversity in calcium-activated big potassium channel populations. Proc. Natl. Acad. Sci. *107*, 21152–21157.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. *37*, 766–770.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. *74*, 3171–3175.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846–850.

Blencowe, B.J., Baurén, 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.

Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M., and McKeown, M. (1987). Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. Cell *50*, 739–747.

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.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R.R., and Blencowe, B.J.J. (2013). Dynamic Integration of Splicing within Gene Regulatory Pathways. Cell *152*, 1252–1269.

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. gr.177790.114 – .

Breathnach, R., Mandel, J.L., and Chambon, P. (1977). Ovalbumin gene is split in chicken DNA. Nature *270*, 314–319.

Brown, S., Gersen, S., Anyane-Yeboa, K., and Warburton, D. (1993). Preliminary definition of a "critical region" of chromosome 13 in q32: report of 14 cases with 13q deletions and review of the literature. Am. J. Med. Genet. *45*, 52–59.

Brown, S., Russo, J., Chitayat, D., and Warburton, D. (1995). The 13q- syndrome: the molecular definition of a critical deletion region in band 13q32. Am. J. Hum. Genet. *57*, 859–866.

Buckley, P.T., Lee, M.T., Sul, J.-Y., Miyashiro, K.Y., Bell, T.J., Fisher, S.A., Kim, J., and Eberwine, J. (2011). Cytoplasmic Intron Sequence-Retaining Transcripts Can Be Dendritically Targeted via ID Element Retrotransposons. Neuron *69*, 877–884.

Burnette, J.M. (2005). Subdivision of Large Introns in Drosophila by Recursive Splicing at Nonexonic Elements. Genetics *170*, 661–674.

Caceres, J., Stamm, S., Helfman, D., and Krainer, A. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science (80-.). *265*, 1706–1709.

Cáceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L., and Krainer, A.R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. J. Cell Biol. *138*, 225–238.

Calarco, J.A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinas, D., van der Kooy, D., et al. (2009). Regulation of Vertebrate Nervous System Alternative Splicing and Development by an SR-Related Protein. Cell *138*, 898–910.

Cartegni, L., and Krainer, A.R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat. Genet. *30*, 377–384.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Chabot, B., Black, D.L., LeMaster, D.M., and Steitz, J.A. (1985). The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science *230*, 1344–1349.

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

Chen, L.-L., DeCerbo, J.N., and Carmichael, G.G. (2008). Alu element-mediated gene silencing. EMBO J. *27*, 1694–1705.

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.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell *12*, 1–8.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., and Duncan, P.I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J. *15*, 265–275.

Coulter, L.R., Landree, M.A., and Cooper, T.A. (1997). Identification of a new class of exonic splicing enhancers by in vivo selection. Mol. Cell. Biol. *17*, 2143–2150.

Cuellar, L.M., Fujinaka, H., Yamamoto, K., Miyamoto, M., Tasaki, M., Zhao, L., Tamer, I., Yaoita, E., Yoshida, Y., and Yamamoto, T. (2009). Identification and localization of novel genes preferentially expressed in human kidney glomerulus. Nephrology (Carlton). *14*, 94–104.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2015. Nucleic Acids Res. *43*, D662–D669.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H.,

Urlaub, H., and Lührmann, 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.

Dimitrieva, S., and Bucher, P. (2013). UCNEbase--a database of ultraconserved non-coding elements and genomic regulatory blocks. Nucleic Acids Res. *41*, D101–D109.

Dirksen, W.P., Hampson, R.K., Sun, Q., and Rottman, F.M. (1994). A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. J. Biol. Chem. *269*, 6431–6436.

Dirksen, W.P., Sun, Q., and Rottman, F.M. (1995). Multiple Splicing Signals Control Alternative Intron Retention of Bovine Growth Hormone Pre-mRNA. J. Biol. Chem. *270*, 5346–5352.

Dominski, Z., and Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. Mol. Cell. Biol. *11*, 6075–6083.

Dominski, Z., and Kole, R. (1992). Cooperation of pre-mRNA sequence elements in splice site selection. Mol. Cell. Biol. *12*, 2108–2114.

Duff, M.O., Olson, S., Wei, X., Garrett, S.C., Osman, A., Bolisetty, M., Plocik, A., Celniker, S.E., and Graveley, B.R. (2015). Genome-wide identification of zero nucleotide recursive splicing in Drosophila. Nature *521*, 376–379.

Dvinge, H., and Bradley, R.K. (2015). Widespread intron retention diversifies most cancer transcriptomes. Genome Med. *7*, 45.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980). Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. Cell *20*, 313–319.

Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schöneweis, K., Hertel, K.J., and Schaal, H. (2013). Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA *19*, 96–102.

Farina, K.L. (2002). Two ZBP1 KH domains facilitate beta-actin mRNA localization, granule formation, and cytoskeletal attachment. J. Cell Biol. *160*, 77–87.

Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.-P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc. Natl. Acad. Sci. U. S. A. *102*, 16176–16181.

Fu, X.-D., and Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat. Rev. Genet. *15*, 689–701.

Fu, X.D., and Maniatis, T. (1992a). Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. Science *256*, 535–538.

Fu, X.D., and Maniatis, T. (1992b). 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.

Fu, X.D., Mayeda, A., Maniatis, T., and Krainer, A.R. (1992). General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5’ and 3' splice site selection. Proc. Natl. Acad. Sci. *89*, 11224–11228.

Galante, P.A.F., Sakabe, N.J., Kirschbaum-Slager, N., and Souza, S.J. De (2004). Detection and

evaluation of intron retention events in the human transcriptome. RNA *10*, 757–765.

Ge, H., and Manley, J.L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell *62*, 25–34.

Ge, Y., and Porse, B.T. (2014). The functional consequences of intron retention: Alternative splicing coupled to NMD as a regulator of gene expression. BioEssays *36*, 236–243.

Ge, H., Zuo, P., and Manley, J.L. (1991). Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. Cell *66*, 373–382.

Goralski, T.J., Edström, J.E., and Baker, B.S. (1989). The sex determination locus transformer-2 of Drosophila encodes a polypeptide with similarity to RNA binding proteins. Cell *56*, 1011–1018.

Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1984). Messenger RNA splicing in vitro: An excised intervening sequence and a potential intermediate. Cell *37*, 415–427.

Graveley, B.R., and Maniatis, T. (1998). Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. Mol. Cell *1*, 765–771.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO J. *17*, 6747–6756.

Gui, J.F., Lane, W.S., and Fu, X.D. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. Nature *369*, 678–682.

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.

Hatton, A.R., Subramaniam, V., and Lopez, A.J. (1998). Generation of Alternative Ultrabithorax Isoforms and Stepwise Removal of a Large Intron by Resplicing at Exon–Exon Junctions. Mol. Cell *2*, 787–796.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell *45*, 567–580.

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.

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.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Liu, P.P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnRNP proteins. Cell Rep. *1*, 167–178.

Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlon, D., et al. (2014). A Highly Conserved Program of Neuronal Microexons Is Misregulated in Autistic Brains. Cell *159*, 1511–1523.

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.

Jeffreys, A.J., and Flavell, R.A. (1977). The rabbit β-globin gene contains a large insert in the coding sequence. Cell *12*, 1097–1108.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev. *20*, 355–367.

Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. Nat. Rev. Mol. Cell Biol. *13*, 700–712.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Kinniburgh, A.J., Maquat, L.E., Schedl, T., Rachmilewitz, E., and Ross, J. (1982). mRNA-deficint β ° -thaladssemia results from a single nucleotide deletion. Nucleic Acids Res. *10*, 5421–5427.

Kislauskis, E.H., Zhu, X., and Singer, R.H. (1994). Sequences responsible for intracellular localization of beta-actin messenger RNA also affect cell phenotype. J. Cell Biol. *127*, 441–451.

Klessig, D.F. (1977). Two adenovirus mRNAs have a common 5’ terminal leader sequence encoded at least 10 kb upstream from their main coding regions. Cell *12*, 9–21.

Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Lührmann, R., Garcia-Blanco, M.A., and Manley, J.L. (1994). Protein-protein interactions and 5’-splice-site recognition in mammalian mRNA precursors. Nature *368*, 119–124.

Kole, R., and Weissman, S.M. (1982). Accurate in vitro splicing of human beta-globin RNA. Nucleic Acids Res. *10*, 5429–5445.

Konarska, M.M., Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature *313*, 552–557.

Krainer, A.R., Maniatis, T., Ruskin, B., and Green, M.R. (1984). Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell *36*, 993–1005.

Krainer, A.R., Mayeda, A., Kozak, D., and Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. Cell *66*, 383–394.

Kuo, H., Nasim, F., and Grabowski, P. (1991). Control of Alternative Splicing by the Differential Binding of U1 Small Nuclear Ribonucleoprotein Particle on JSTOR. Science (80-.). *251*, 1045–1050.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

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.

Lavigueur, A., La Branche, H., Kornblihtt, A.R., and Chabot, B. (1993). A splicing enhancer in

the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev. *7*, 2405–2417.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. EMBO J. *21*, 3536–3545.

Lerner, M.R., and Steitz, J.A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. Proc. Natl. Acad. Sci. U. S. A. *76*, 5495–5499.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc. Natl. Acad. Sci. U. S. A. *100*, 189–192.

Li, Y., Bor, Y.-C., Misawa, Y., Xue, Y., Rekosh, D., and Hammarskjöld, M.-L. (2006). An intron with a constitutive transport element is retained in a Tap messenger {RNA}. Nature *443*, 234–237.

Li, Y.I., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2014). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. Genome Res. *25*, 1–13.

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.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803–819.

Losson, R., and Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc. Natl. Acad. Sci. *76*, 5134–5137.

Lynch, K.W., and Maniatis, T. (1995). Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev. *9*, 284–293.

Maquat, L.E., Kinniburgh, A.J., Rachmilewitz, E.A., and Ross, J. (1981). Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. Cell *27*, 543–553.

Marinescu, V., Loomis, P.A., Ehmann, S., Beales, M., and Potashkin, J.A. (2007). Regulation of Retention of FosB Intron 4 by PTB. PLoS One *2*.

Mayeda, A., and Krainer, A.R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell *68*, 365–375.

Mayeda, A., Munroe, S.H., Cáceres, J.F., and Krainer, A.R. (1994). Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. EMBO J. *13*, 5483–5495.

Michael, I.P., Kurlender, L., Memari, N., Yousef, G.M., Du, D., Grass, L., Stephan, C., Jung, K., and Diamandis, E.P. (2005). Intron retention: A common splicing event within the human kallikrein gene family. Clin. Chem. *51*, 506–515.

Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA *10*, 1692–1694.

Mount, S.M. (1982). A catalogue of splice junction sequences. Nucleic Acids Res. *10*, 459–472.

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.

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.J. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. *21*, 708–718.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463.

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.

Nott, A., Meislin, S.H., and Moore, M.J. (2003). A quantitative analysis of intron effects on mammalian gene expression. RNA *9*, 607–617.

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.

Okunola, H.L., and Krainer, A.R. (2009). Cooperative-binding and splicing-repressive properties of hnRNP A1. Mol. Cell. Biol. *29*, 5620–5631.

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.

Padgett, R.A., Hardy, S.F., and Sharp, P.A. (1983). Splicing of adenovirus RNA in a cell-free transcription system. Proc. Natl. Acad. Sci. U. S. A. *80*, 5230–5234.

Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F., and Sharp, P.A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science *225*, 898–903.

Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dörk, T., and Baralle, F.E. (2002). A new type of mutation causes a splicing defect in ATM. Nat. Genet. *30*, 426–429.

Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., and Baralle, F.E. (2003). New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum. Mol. Genet. *12*, 1111–1120.

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.

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M., and Fu, X.-D. (2013). Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. Mol. Cell *50*, 223–235.

Pang, K.C., Frith, M.C., and Mattick, J.S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet. *22*, 1–5.

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.

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.

Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription regulation. Crit. Rev. Biochem. Mol. Biol. *48*, 575–608.

Quélin, C., Bendavid, C., Dubourg, C., de la Rochebrochard, C., Lucas, J., Henry, C., Jaillard, S., Loget, P., Loeuillet, L., Lacombe, D., et al. (2009). Twelve new patients with 13q deletion syndrome: genotype-phenotype analyses in progress. Eur. J. Med. Genet. *52*, 41–46.

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.

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.

Robberson, B.L., Cote, G.J., and Berget, S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. *10*, 84–94.

Roca, X., Krainer, A.R., and Eperon, I.C. (2013). Pick one, but be quick: 5' splice sites and the problems of too many choices. Genes Dev. *27*, 129–144.

Ross, A.F., Oleynikov, Y., Kislauskis, E.H., Taneja, K.L., and Singer, R.H. (1997). Characterization of a beta-actin mRNA zipcode-binding protein. Mol. Cell. Biol. *17*, 2158–2165.

Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317–331.

Sakabe, N.J., and de Souza, S.J. (2007). Sequence features responsible for intron retention in human. BMC Genomics *8*, 59.

Saltzman, A.L., Kim, Y.K., Pan, Q., Fagnani, M.M., Maquat, L.E., and Blencowe, B.J. (2008). Regulation of Multiple Core Spliceosomal Proteins by Alternative Splicing-Coupled Nonsense-Mediated mRNA Decay. Mol. Cell. Biol. *28*, 4320–4330.

Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. Genes Dev. *25*, 373–384.

Sanford, J.R., Wang, X., Mort, M., Vanduyn, N., Cooper, D.N., Mooney, S.D., Edenberg, H.J., and Liu, Y. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. Genome Res. *19*, 381–394.

Schaal, T.D., and Maniatis, T. (1999). Multiple Distinct Splicing Enhancers in the Protein-Coding Sequences of a Constitutively Spliced Pre-mRNA. Mol. Cell. Biol. *19*, 261–273.

Schell, T., Kulozik, A.E., and Hentze, M.W. (2002). Integration of splicing, transport and translation to achieve mRNA quality control by the nonsense-mediated decay pathway. Genome Biol. *3*, REVIEWS1006.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. *7*, 3.

Sharma, S., Falick, A.M., and Black, D.L. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol. Cell

19, 485–496.

Sharma, S., Kohlstaedt, L.A., Damianov, A., Rio, D.C., and Black, D.L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat. Struct. Mol. Biol. *15*, 183–191.

Sharma, S., Maris, C., Allain, F.H.-T., and Black, D.L. (2011). U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. Mol. Cell *41*, 579–588.

Shibuya, T., Tange, T.Ø., Sonenberg, N., and Moore, M.J. (2004). eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. Nat. Struct. Mol. Biol. *11*, 346–351.

Sibley, C.R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M.E., Hardy, J., et al. (2015). Recursive splicing in long vertebrate genes. Nature *521*, 371–375.

Siebel, C.W., Fresco, L.D., and Rio, D.C. (1992). The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. Genes Dev. *6*, 1386–1401.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Alaluf, I., Swerdlin, N., Perzov, N., Danon, T., Liron, Y., et al. (2006). Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. Nat. Methods *3*, 525–531.

Silverman, I.M., Li, F., Alexander, A., Goff, L., Trapnell, C., Rinn, J.L., and Gregory, B.D. (2014). RNase-mediated protein footprint sequencing reveals protein-binding sites throughout the human transcriptome. Genome Biol. *15*, R3.

Simon, J.M., Hacker, K.E., Singh, D., Brannon, A.R., Parker, J.S., Weiser, M., Ho, T.H., Kuan, P.-F., Jonasch, E., Furey, T.S., et al. (2014). Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res. *24*, 241–250.

Singh, R., Valcárcel, J., and Green, M.R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. Science *268*, 1173–1176.

Sowalsky, A.G., Xia, Z., Wang, L., Zhao, H., Chen, S., Bubley, G.J., Balk, S.P., and Li, W. (2015). Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer. Mol. Cancer Res. *13*, 98–106.

Spector, D.L., Fu, X.D., and Maniatis, T. (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. EMBO J. *10*, 3467–3481.

Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. Proc. Natl. Acad. Sci. *93*, 15081–15085.

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.

Sun, H., and Chasin, L.A. (2000). Multiple Splicing Defects in an Intronic False Exon. Mol. Cell. Biol. *20*, 6414–6425.

Sun, Q., Mayeda, A., Hampson, R.K., Krainer, A.R., and Rottman, F.M. (1993). General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes

Dev. *7*, 2598–2608.

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.

Sureau, A., Gattoni, R., Dooghe, Y., Stévenin, J., and Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. EMBO J. *20*, 1785–1796.

Talerico, M., and Berget, S.M. (1994). Intron definition in splicing of small Drosophila introns. Mol. Cell. Biol. *14*, 3434–3445.

Tanaka, K., Watakabe, A., and Shimura, Y. (1994). Polypurine sequences within a downstream exon function as a splicing enhancer. Mol. Cell. Biol. *14*, 1347–1354.

Tian, H., and Kole, R. (1995). Selection of novel exon recognition elements from a pool of random sequences. Mol. Cell. Biol. *15*, 6291–6298.

Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., Sullivan, M., Maizel, J. V., and Leder, P. (1978). Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. Proc. Natl. Acad. Sci. *75*, 725–729.

Tiruchinapalli, D.M., Oleynikov, Y., Kelic, S., Shenoy, S.M., Hartley, A., Stanton, P.K., Singer, R.H., and Bassell, G.J. (2003). Activity-dependent trafficking and dynamic localization of zipcode binding protein 1 and beta-actin mRNA in dendrites and spines of hippocampal neurons. J. Neurosci. *23*, 3251–3261.

Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O., and Gilbert, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc. Natl. Acad. Sci. *75*, 1485–1489.

Townley-Tilson, W.H.D., Pendergrass, S.A., Marzluff, W.F., and Whitfield, M.L. (2006). Genome-wide analysis of mRNAs bound to the histone stem-loop binding protein. RNA *12*, 1853–1867.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol. Cell *39*, 925–938.

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.

Varjosalo, M., Keskitalo, S., VanDrogen, 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.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell *136*, 701–718.

Walczak-Sztulpa, J., Wisniewska, M., Latos-Bielenska, A., Linné, M., Kelbova, C., Belitz, B., Pfeiffer, L., Kalscheuer, V., Erdogan, F., Kuss, A.W., et al. (2008). Chromosome deletions in 13q33-34: report of four patients and review of the literature. Am. J. Med. Genet. A *146A*, 337–342.

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.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, H.Y., Lin, W., Dyck, J.A., Yeakley, J.M., Songyang, Z., Cantley, L.C., and Fu, X.D. (1998). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. J. Cell Biol. *140*, 737–750.

Wang, Y., Ma, M., Xiao, X., and Wang, Z. (2012). Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. Nat. Struct. Mol. Biol. *19*, 1044–1052.

Wang, Y., Xiao, X., Zhang, J., Choudhury, R., Robertson, A., Li, K., Ma, M., Burge, C.B., and Wang, Z. (2013). A complex network of factors with overlapping affinities represses splicing through intronic elements. Nat. Struct. Mol. Biol. *20*, 36–45.

Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831–845.

Washietl, S., Hofacker, I.L., Lukasser, M., Hüttenhofer, A., and Stadler, P.F. (2005). Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nat. Biotechnol. *23*, 1383–1390.

Washietl, S., Pedersen, J.S., Korbel, J.O., Stocsits, C., Gruber, A.R., Hackermüller, J., Hertel, J., Lindemeyer, M., Reiche, K., Tanzer, A., et al. (2007). Structured RNAs in the ENCODE selected regions of the human genome. Genome Res. *17*, 852–864.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep. *6*, 1139–1152.

Wiegand, H.L., Lu, S., and Cullen, B.R. (2003). Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. Proc. Natl. Acad. Sci. U. S. A. *100*, 11327–11332.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb.

Perspect. Biol. *3*, a003707.

Wong, J.W.H.J.J.-L.L., Ritchie, W., Ebner, O.A., Selbach, M., Wong, J.W.H.J.J.-L.L., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., et al. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. Cell *154*, 583–595.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. *3*, e7.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061–1070.

Xiao, X., Wang, Z., Jang, M., and Burge, C.B. (2007). Coevolutionary networks of splicing cis-regulatory elements. Proc. Natl. Acad. Sci. *104*, 18583–18588.

Yang, X., Bani, M.R., Lu, S.J., Rowan, S., Ben-David, Y., and Chabot, B. (1994). The A1 and A1B proteins of heterogeneous nuclear ribonucleoparticles modulate 5' splice site selection in vivo. Proc. Natl. Acad. Sci. *91*, 6924–6928.

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.

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.

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.

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.

Yu, Y., Maroney, P.A., Denker, J.A., Zhang, X.H.-F., Dybkov, O., Lührmann, R., Jankowsky, E., Chasin, L.A., and Nilsen, T.W. (2008). Dynamic regulation of alternative splicing by silencers that modulate 5' splice site competition. Cell *135*, 1224–1236.

Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev. *6*, 837–847.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609–614.

Zhan, L.-L. (2013). Recent advances of studies on alternative intron retention. Trends Evol. Biol. *5*, e1.

Zhang, Z., and Krainer, A.R. (2007). Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. Proc. Natl. Acad. Sci. U. S. A. *104*, 11574–11579.

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.

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.

Zhang, Q., Li, H., Jin, H., Tan, H., Zhang, J., and Sheng, S. (2014). The global landscape of intron retentions in lung adenocarcinoma. BMC Med. Genomics *7*, 15.

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.

Zhu, J., Mayeda, A., and Krainer, A.R. (2001). Exon Identity Established through Differential Antagonism between Exonic Splicing Silencer-Bound hnRNP A1 and Enhancer-Bound SR Proteins. Mol. Cell *8*, 1351–1361.

Appendix A

Name	Sequence	
ARGLU1 Ex 2.2	GGAAGAAACAGCACGAAGAGT	
ARGLU1 3 UTR R2	GCAGAGCATAGCCCCTATTAGA	
ARGLU1 Exon 2 F	TGGTAGCAAAAAGGGTGGAG	
ARGLU1 Exon 3 R	CAGTTTGGCTTGTGCTTCTG	
ARGLU1 Exon 4 R	ACAGTTTTGGCCTGGACTTC	
ARGLU1 UCE 504 F	AATTGTCAACATCTGAATGTTAAGTCC	
ARGLU1 UCE 504 R	CTTTAATATCACACAAATCAACAAGG	
ARGLU1 Alt F	AAGCTGGCGCGTATGGCAG	
ARGLU1 ret F	TTTGGACTGGGAGGGATGA	
ARGLU1 ret R	AGGCAGAACCCAACACTGAA	
pcDNA3F	TCACTATAGGGAGACCCAAGC	
pcDNA3R	CTGATCAGCGAGCTCTAGCA	
Actin F	GGACTTCGAGCAAGAGATGG	
Actin R	AGCACTGTGTTGGCGTACAG	
CRISPR sgRNA sequences		
CRISPR sgRNA 1	AAAGAGGAAGCTGGCGCGTA	
CRISPR sgRNA 2	GCTGGCGCGTATGGCAGCCG	
Mutagenesis primers		
	F	**R**
A23mutAlt3’	CTTCCTGGCGCGTATGGCAGCC	GAGATTAAAATGATTTGTACTGTTAGCTTGGCAATACC
A23mutAlt5’	GGACAAGACACAAAATATTCAACTTTTAATGCTGATTAAAG	AGTGTCCTGCAGAGTGTG

SS1 GC-CC	AACTTTTAATCCTGATTAAAGGAGTATAG	GAATATTTACTGTCTTGTCCAG
SS2 GT-CA	GATTAAAGGACAATAGGTAAAGAATACGTAGGTATAC	AGCATTAAAAGTTGAATATTTACTG
SS3 GT-CA	AGGAGTATAGCAAAAGAATACGTAGGTATAC	TTAATCAGCATTAAAAGTTGAATATTTAC
SS4 GT-CA	TAAAGAATACCAAGGTATACATAATTGGTGAG	CCTATACTCCTTTAATCAGC
SS5 GT-CA	GAATACGTAGCAATACATAATTGGTGAAG	TTTACCTATACTCCTTTAATCAG
SS6 GT-CA	TACATAATTGCAGAGACAAATATTCACTTTATTTATATTTTATATATTATTTTTTTAATTTG	TACCTACGTATTCTTTACCTATAC
SS1-3 mut	GACAATAGCAAAAGAATACGTAGGTATACATAATTG	CTTTAATCAGGATTAAAAGTTGAATATTTACTGTCTTG
SS4-6 mut	ATAATTGCAGAGACAAATATTCACTTTATTTATATTTTATATATTATTTTTTTAATTTG	GTATTGCTTGGTATTCTTTACCTATACTCCTTTAATC
Ex2SS cons	aagtCTCGGTCGTTTGGAAAGTAG	acctgTCTAGCTTTTTGTGCGGC
Ex3SS cons	cttccagGAGGAAGAACGTGCAAAAC	aaagagaATGTTAAGATATTAGAAAAACAAAATG
crp3' cons	tccaggagAAGCTGGCGCGTATGGCA	agaaagagaGTACTGTTAGCTTGGCAATACCTG
crp5' cons	aagtATTCAACTTTTAATGCTGATTAAAGG	acctgCTTGTCCAGTGTCCTGCAG
	F	**R**
25 del 1 primers	AACATCTGCATTCCTCAG	AATTCATCATTCCTAAGGCAAAC
25 del 2 primers	TTCCTTGTATGTTGTTTCTTTATAAATG	TATGGTAAGAAGGAACAAAAAATG
25 del 3 primers	ATGGTTGAGCTGCTGATG	GGCTGAGCTGAGGAATGC
25 del 4 primers	TGCCAAGCTAACAGTACAAATC	TTATAAAGAAACAACATACAAGGAAG
25 del 5 primers	TTAAAGAGGAAGCTGGCG	ATACCTGCATCAGCAGCTC
25 del 6 primers	CAGCCGAGGAGCACACTC	AATGATTTGTACTGTTAGCTTGGC
25 del 7 primers	CACTGGACAAGACAGTAAATATTCAACTTTTAATG	CCATACGCGCCAGCTTCC
25 del 8 primers	ACTTTTAATGCTGATTAAAGGAGTATAG	TCCTGCAGAGTGTGCTCC
25 del 9 primers	TAGGTAAAGAATACGTAGGTATAC	TGAATATTTACTGTCTTGTCC
25 del 10 primers	TAATTGGTGAGACAAATATTCAC	TACTCCTTTAATCAGCATTAAAAG
25 del 11 primers	TATTTATATTTTATATATTATTTTTTTAATTTGGTAAATAC	TGTATACCTACGTATTCTTTAC
25 del 12 primers	TTAATTTGGTAAATACTATCCAGTTTTG	AAGTGAATATTTGTCTCACC

25 del 13 primers	TTGTAGTTGTCCTTGTTGATTTG	AAAAATAATATATAAAATATAAATAAAGTGAATATTTG
25 del 14 primers	TGATATTAAAGTATTAGTAATAATTGCCAG	AACTGGATAGTATTTACCAAATTAAAAAAATAATATATAAAATATAAATAAAG
25 del 15 primers	GCCAGGAAACTATCATTAG	CACAAATCAACAAGGACAAC
25 del 16 primers	TTTAGTTGGTTGCTGTTTG	AATTATTACTAATACTTTAATATCACAC
UCE del 504	TATTAGTAATAATTGCCAGGAAAC	GGAAGGCAGCTTATATTG
A23ContDel	CAAGCCATCAAGCAGTCTTC	ACACTCAGGGCTGTAAAG
A23 168 del 1	AACATCTGCATTCCTCAG	GGAAGGCAGCTTATATTG
A23 168 del 2	ATATTCAACTTTTAATGCTGATTAAAG	TATGGTAAGAAGGAACAAAAAATG
A23 168 del 3	TATTAGTAATAATTGCCAGGAAAC	TTACTGTCTTGTCCAGTG