UNIVERSITY OF CALIFORNIA

SANTA CRUZ

GLOBAL ANALYSIS OF U2AF2-RNA INTERACTIONS

A thesis submitted in partial satisfaction of the

requirements for the degree of

MASTER OF ARTS

in

MOLECULAR, CELLULAR, DEVELOPMENTAL BIOLOGY

by

Garam Kim

June 2018

The Thesis of Garam Kim

is approved:

_____

Professor Jeremy Sanford

_____

Professor Grant Hartzog

_____

Professor Alan Zahler

_____

Tryus Miller

Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Global Analysis of U2AF2-RNA Interactions

by

Garam Kim

Alternative pre-mRNA splicing is a dynamic mechanism responsible for the diversification of the transcriptome in eukaryotes. This robust mechanism is regulated by the interactions between *trans* acting RNA binding proteins (RBPs) and numerous *cis*-regulatory elements which exist on the pre-mRNA sequence. The interplay between RBPs and cis elements are integral to the definition of the protein coding sequences (exons) and mediating the spliceosome assembly on the non-coding regions (introns). One of the earliest events in spliceosome assembly on a pre-mRNA transcript occurs at the 3' splice site (ss), where U2 small auxiliary factor 2 (U2AF2) recognizes the AG-dinucleotide, in part, defining the 3' splice site. hnRNP A1 has also been shown to play a role in the definition of the 3' ss through the direct interaction with U2AF2 and providing a proofreading mechanism through the binding of CG-dinucleotides displacing U2AF2 to the correct 3' ss. Our lab has previously shown that hnRNP A1 plays a direct role in exon definition through the displacement of U2AF2 to decoy sites further upstream of the 3' ss causing changes in splicing. We have also identified the decoy sites to which U2AF2 relocates to as *Alu* elements which are short interspaced nuclear elements (SINEs). *Alu* elements have long been characterized as disruptive agents in gene expression, however here we propose that *Alu* derived RNA elements may play a role as functional splicing regulators which might be responsible for the species-specific alternative splicing events we observe in the primate lineage.

This thesis is dedicated to my mother Miran Choi, my father Young Min Kim, my younger sister Jenna Kim, and my aunt Elli Kim. I thank you for your never-ending support and guidance throughout my life and academic career. To my friends for always being a source of entertainment and mitigating difficult situations which arose throughout my academic career.

Acknowledgment

First and foremost, I would like to thank my advisor Jeremy Sanford for providing me the opportunity to grow as a student, a scientist, and as a person. Jeremy has supplemented me with endless guidance since my freshman year, on the rugby pitch as well as my academic career. Thank you for recognizing my potential as a scientist, developing my scientific training, and expanding my thought process to great heights. I would like to thank Jolene Draper for her support as a lab mate and friend, she has taught me many technical skills and helped solve problems that would arise in my experiments. I would not have been well acclimated to the lab if it were not for her kindness and generosity. I would also like to thank Marija Dargyte for the hours of endless banter and dialogue which have made long hours in lab much more enjoyable. Thanks to Michael Blanco for being a great friend, housemate, lab mate, and teammate for the last 5 years. Lastly, I would like to thank Jon Howard for providing me with the basis for my thesis and consultation about my project.

Chapter 1

Introduction

1.1 Eukaryotic gene expression

Gene expression in eukaryotic organisms is a highly regulated process which requires a series of sequential steps to derive a functional protein from a genetic blueprint such as DNA. The flow in which genetic information is transformed into functional proteins starts with a process called transcription where an assembly of highly conserved proteins such as RNA Polymerase II form the initiation complex alongside general transcription factors at the promoter region of a gene.

Through unwinding of the helical structure of DNA, RNA polymerase II converts genomic DNA into an intermediate template, RNA (Yin & Steitz, 2009). This precursor form of RNA (pre-mRNA) contains both the protein coding regions as well as non-coding segments which consists of regulatory factor binding sites. Pre-mRNAs are further processed to excise all the non-coding segments and ligate the coding regions together to produce a mature messenger RNA (mRNAs).

The mature mRNA is then exported out from the nucleus into the cytoplasm where ribosomes bind to the start codon and starts the translation of the messenger RNA (Dever & Green, 2012). The nascent polypeptide is then further processed through folding and modifications to allow correct function.

1.2 Pre-mRNA Splicing

Pre-mRNA splicing is a pivotal step that links transcription and translation. Pre-mRNA splicing consists of the excision of stretches of noncoding sequences from the pre-mRNA transcripts by the spliceosome to form a mature mRNA. The spliceosome is a large ribonucleoprotein particle composed of small nuclear RNAs (snRNAs) and myriad proteins, which are divided into the two classes: the Sm-class RNAs and the Sm-like class RNAs. The Sm-class of spliceosomal snRNA is comprised of the U1, U2, U4, $U4_{atac}$, U5, U11, and U12 whereas the Sm-like class of snRNAs is composed of U6 and $U6_{atac}$. The Sm-class of spliceosomal snRNAs are transcribed in the nucleus and shuttled to the cytoplasm for maturation to become functional small nuclear ribonucleoproteins (snRNPs) (Matera et al, 2007).

The excision of the intronic sequences from a pre-mRNA begins with the assembly of the spliceosome. The GU-dinucleotide of the 5' splice site is recognized by serine-arginine rich (SR) proteins and base paring interactions of U1 snRNP (Eperon et al, 1993; Matera & Wang, 2014). Likewise, the AG-dinucleotide of the 3' splice site is recognized by SR proteins and the U2 snRNP auxiliary factor (U2AF). The recognition of the 3' splice site is mediated by the two subunits of U2AF: U2AF1 and U2AF2, which recognize the AG-dinucleotide and the polypyrimidine tract, respectively (Shao et al, 2014).

The interaction of U2AF and the 3' splice site mediates the binding of splicing factor 1 (SF1) to the branchpoint sequence along with base pairing interactions with U2 snRNP (Wu & Fu, 2015). These combinatorial effects lead to the aggregation of the U4/U6.U5 tri-snRNP to the pre-spliceosome converting it into a catalytically active spliceosome. The mature

spliceosome is then able to remove the intron lariats and ligate the ends of the exons together through a two-step transesterification reaction (Wahl et al, 2009).

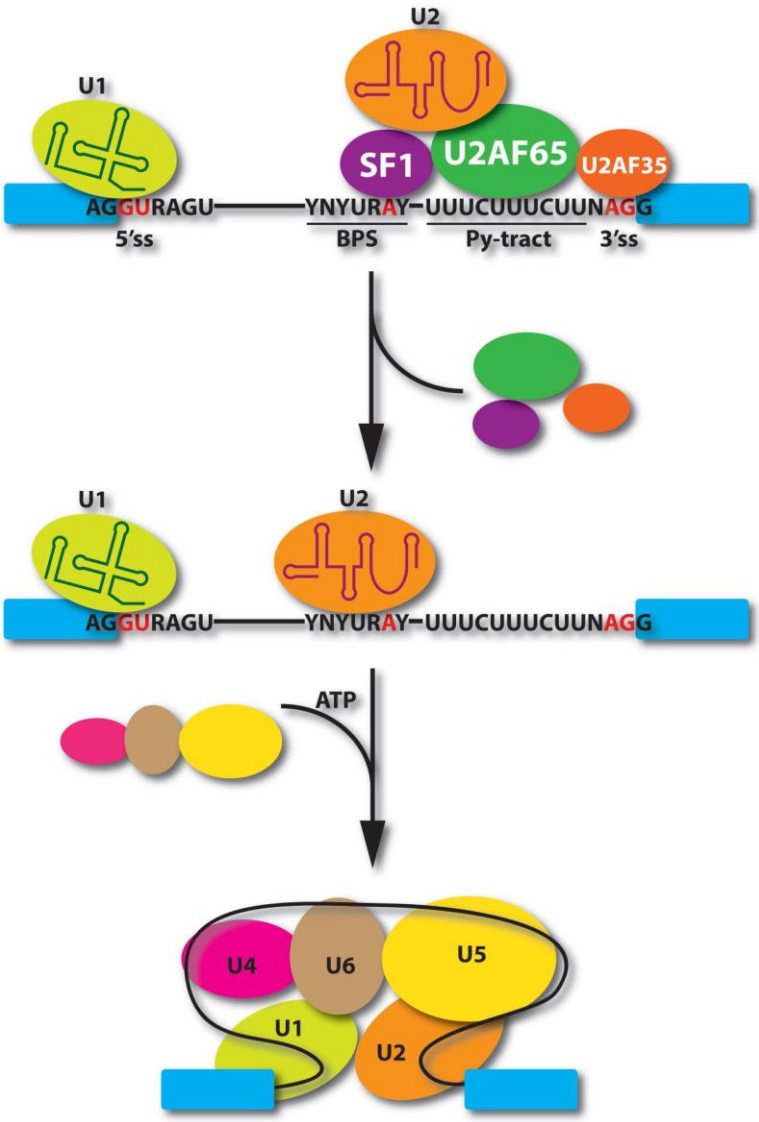

Figure 1.1: Early spliceosome assembly on a pre-mRNA transcript. The 5' splice site denoted by a red GU sequence is recognized by the U1 snRNP. The branch point sequence denoted b a red A is recognized by SF1, the polypyrimidine tract shown in black U's and C's

is bound by U2AF2, and the 3' splice site denoted by a red AG dinucleotide is recognized by U2AF35. All these RBPs in conjunction with the U1 snRNP, recruits the remaining snRNPs to form a spliceosome on a pre-mRNA transcript leading to a two-step transesterification of the intron lariats. Adapted from Wu and Fu, 2015.

1.3 Alternative Splicing

Although many multicellular eukaryotes have an abundance of homologous genes, most have vastly differing transcriptomes due to a process called alternative splicing. Humans are estimated to have two-thirds of their genes contain multiple exons and it is believed that around 95% of multiexon genes in humans are alternatively spliced and gives rise to around 100,000 alternative splicing events (Pan et al, 2008; Johnson et al, 2003).

Alternative splicing arises from the spliceosomal machinery being influenced by the strength of the splice site. The more divergent the splice site is from the consensus sequence, the more variability in affinity for known splicing factors which aid in constitutive splicing (Kornblihtt et al, 2013). The conditions which affect the overall strength of a splice site are dependent on the cis-regulatory sequences that lie on the primary transcript along with the trans-acting factors which interact with it (Wang & Wang, 2014). The cis-regulatory elements involved in alternative splicing are comprised of exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs), and intronic splicing silencers (ISSs) (Lee & Rio, 2015).

Two well characterized groups of RNA binding proteins (RBPs) which interact with the cis-elements to enhance or silence splicing reactions, are the heterogeneous ribonucleoprotein particles (hnRNPs) and the serine-arginine (SR) proteins. Generally, the SR and the hnRNP proteins have opposite effects of enhancing or repressing traits dependent on where they are bound (Fredericks et al, 2015).

An example of exon enhancement/silencing can be observed in the regulation of Fas exon 6 (Conti et al, 2012). Two trans acting factors TIA1 and PTB regulate the splicing of this exon antagonistically. TIA1 stimulates an inclusion event by inducing a positive effect on the U1 snRNP binding at the 5' splice site, in turn causing a stabilizing effect for U2AF binding at the upstream 3' splice site. PTB has been shown to promote exon skipping by binding an exonic splicing silencer and inhibit the molecular mechanism necessary for defining the exon through the binding of U2AF (Izquierdo et al, 2005).

One of the most abundant nuclear proteins within the hnRNP family, hnRNP A1 has been shown to have proofreading capabilities at the 3' splice site, forming a ternary complex with the U2AF heterodimer, and ultimately leading the U2 snRNP to a correct 3' splice site (Jean-Philippe et al, 2013; Tavanez et al, 2012). Through iCLIP assays of U2AF2 in HEK293T cells, we have shown that the overexpression of hnRNP A1 promotes the eviction of U2AF2 from the canonical binding site to intronic decoy sites, including antisense *Alu* elements, leading to splicing repression (Howard et al, 2018). We hypothesize that primate specific elements, *Alu* elements, may drive evolution of splicing regulation by modifying the recognition of bona fide exons.

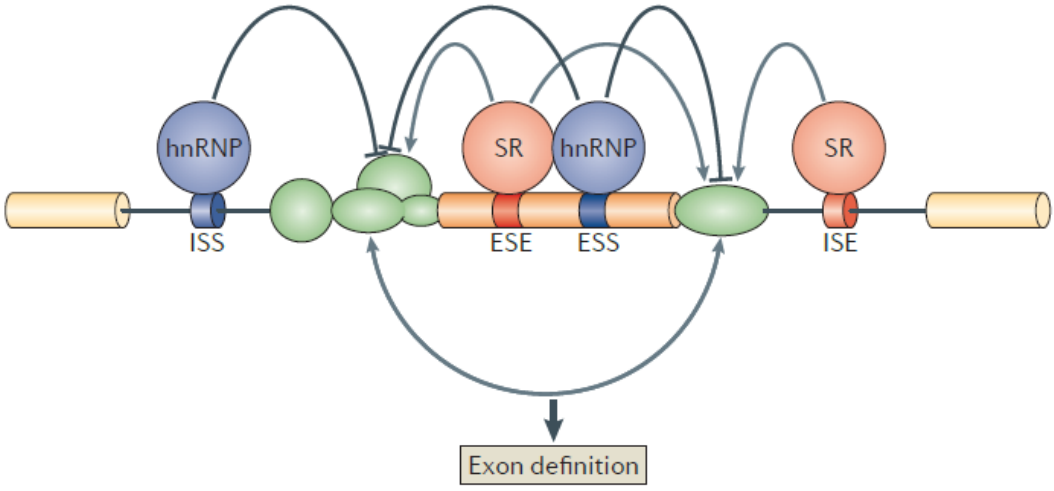

Figure 1.2: A diagram of a pre-mRNA with respective RBPs to its cis-regulatory sites. SR proteins bind to intronic and exonic enhancers to promote the inclusion of the alternative exon, leading to an inclusion isoform. While hnRNP proteins bind to intronic and exonic silencers to repress the inclusion of the alternative exon, leading to a skipped isoform. Adapted from Kornblihtt et al, 2013.

1.4 Role of *Alu* Elements in Primate Evolution

Alu elements are a family of short interspersed nuclear elements (SINEs) in the primate lineage, making up about 10% of the human genome. The expansion of *Alu* elements in the primate genomes are thought to have occurred early in primate evolution around 30 – 50 million years ago and characterized as an abundant source of intraspecies variation in primate genomes (Salem et al, 2003).

Genomic analysis performed on different primate lineages have identified each phylogeny as having differing amounts of *Alu* insertions in their genomes. *Alu* elements in the

orangutan genome are largely quiescent, in contrast, humans have been characterized to have one of the highest amounts of Alu insertions (Locke et al, 2011). *Alu* elements exist in the genome as a 300-base pair fragment of repetitive sequences commonly found in the 3' untranslated regions of genes and intergenic regions of the genome (Batzer & Deininger, 2002).

Through their presence in the cis-regulatory segments of genes, *Alu* elements have been classified to have a myriad of effects in gene expression, some of which have served as a driving force of evolution amongst primate lineages. The antisense orientation of *Alu* elements have been identified to contain potential cryptic 5' and 3' splice sites which could lead to exonization events by the spliceosome (Ast, 2004). Certain insertion events of *Alu* elements during meiosis are thought to cause mis-matches between sister and non-sister chromatids leading to deletion events called *Alu* recombination-mediated deletions (ARMD). The phenotypic differences between humans and chimpanzees are believed to have arisen during the divergence of the two lineages around 6 million years ago, in which the chimpanzee genome experienced ARMD events. This resulted in large deletion around 32 kilobases relative to human ARMD events (Lee et al, 2007).

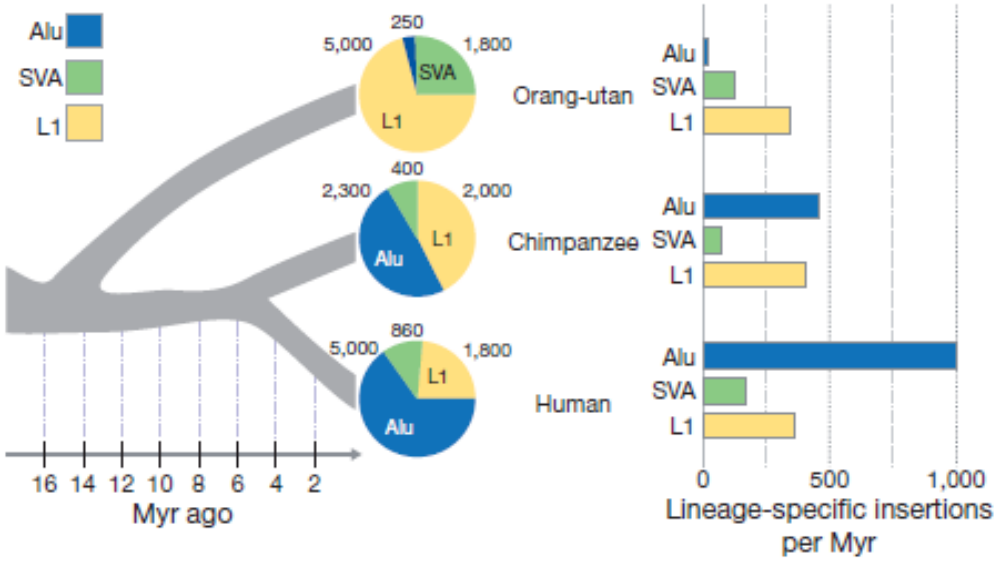

Figure 1.3: Quantification of retrotransposable elements within orangutan, chimpanzee, and human genomes. Shown in blue are the number of *Alu* insertion events which have occurred within the three primate genomes throughout their lineages. The most notable differences are the number of *Alu* insertion events between the orangutan genome and the human genome. Adapted from Locke et al, 2011.

1.5 Comparative Transcriptomics

Sixty six percent of *Alu* elements are found within the intronic sequences in the human genome. Given that the consensus sequence of antisense *Alu* elements contain seven potential 5' splice sites and twelve potential 3' splice sites, it is no surprise that 5% of alternatively spliced exons are believed to be from Alu derived sequences (Elbarbary et al, 2016). It was long believed that *Alu* containing exons that are present within the coding regions of mRNA transcripts can cause frameshift mutations or premature terminations, suggesting that *Alu* elements in internal sequences have a deleterious effect on mRNA transcripts (Sorek, 2002).

However, recent studies suggest that *Alu* elements may provide more regulatory functions in gene expression by generating novel enhancers, promoters, and polyadenylation signals to their residing genes (Häsler, & Strub, 2006).

Expression of *Alu* elements in our genes involves a tightly controlled mechanism in which two competitive molecules, hnRNP C and U2AF2 are involved in combative binding of the cryptic splice sites located within *Alu* elements. Through iCLIP experiments in HeLa cell lines, hnRNP C has been characterized to have high binding affinity for intronic sequences, including associations with poly-uridine sequences within alternative exons which it represses (König et al, 2010). The polypyrimidines tracts within introns have also been extensively characterized as a binding site for U2AF2, a functional protein involved in the recruitment of the U2 snRNP during spliceosome formation on a pre-mRNA transcript (Zamore & Green, 1989).

However, the majority of hnRNP C binding has been identified in "deep" intronic sequences void of exons and about 200 nucleotides away from any annotated exons, indicating repression of a cryptic splice site (*Alu* elements). Knockdown assays of hnRNP C have shown that without the competitive binding of hnRNP C, exonization of *Alu* elements is possible through an inclusion splicing event by U2AF2 (König et al, 2010).

Our lab has shown that the overexpression of hnRNP A1 in HEK293T cells shows relocation of U2AF2 from the orthodox polypyrimidine tract to an *Alu* element which lead to an inclusion event altering the canonical skipped splicing event. To further analyze this interaction between U2AF2 and *Alu* elements, we are using the three primate cell lines (ATRX stable line, EPI 8919-1A, and JOS 2A4). RNA-seq was performed on total RNA within these cell lines to identify cis-variation of *Alu* elements in the different primate genomes, followed

by iCLIP-seq on cell lysates to identify possible transcripts arising from an *Alu* element mediated splicing by U2AF2. This data will be used to generate splicing reporter constructs containing *Alu* elements to confirm their functionality during splicing. Using this approach, we hope to characterize functional *Alu* elements which may have been involved in the splicing events which lead to the cladogenesis of the primate lineages.

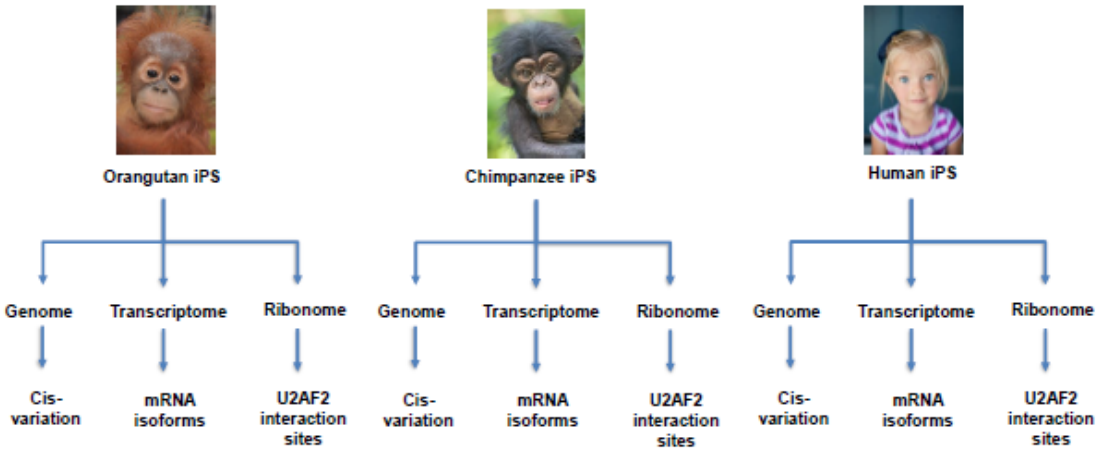

Figure 1.4: Our lab's strategy to determine the functional role of *Alu* elements in lineage specific alternative splicing events is to compare the genomes of orangutan, chimpanzee, and human iPS cells from 10X sequencing data, the transcriptome of each species through Fraq-seq analysis, and the effects of *Alu* elements in splicing events through iCLIP-seq of U2AF2 in the different primate cell lines.

Chapter 2

Materials and Methods

2.1 Cell Lines and Cell Culture

HEK293T cells were grown in Dulbecco Modified Eagle Medium (Sigma-Aldrich). For the validation of hnRNP A1 regulated splicing, transfections were performed using Lipofectamine 2000 (Invitrogen) following the manufacturer's protocol on 6-well plates (Greiner Bio-One) seeded with ~0.3 x 10^6 cells 18 hours prior to transfection. Cells were grown to ~60% to 70% confluence and transfected with 2.5 µg of T7 epitope tagged hnRNP A1 expression plasmid or a control plasmid. Approximately 24 hours after the transfection, the cells were harvested, and cell extracts were prepared.

For the β-globin splicing reporter assay, HEK293T cells were prepared as mentioned above and co-transfected with 4.5 µg of a designed beta-globin-based (HBB1) splicing reporter gene construct and 0.5 µg of T7 epitope tagged hnRNP A1 expression plasmid or a control plasmid using Lipofectamine 2000. Approximately 24 hours after co-transfection, the cells were harvested, and extracts were prepared.

For iCLIP of U2AF2 in primate cell lines, three cell lines were used: human iPSC, chimpanzee iPSC, and orangutan iPSC. ATRX stable line (human iPSC) were grown on 10 cm plates, fed with mTeSR media (Stemcell Technologies) daily, and passaged with EDTA. EPI 8919-1A (chimpanzee iPSC) were grown on 10 cm plates, fed every other day with Essential8

Flex media (Gibco), and passaged manually. JOS 2A4 (orangutan iPSC) were grown on 10 cm plates, fed every other day with Essential8 Flex media, and passaged with EDTA. All the iPSCs were trimmed manually as needed before undergoing crosslinking for iCLIP experiments.

2.2 Cell Extract Preparation

Cells were washed using Dulbecco's Phosphate Buffered Saline (Sigma Aldrich) before the addition of lysing agent. Whole cell extracts were prepared by using RSB-100 (10mM Tric-HCL Solution pH 7.4, 100 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP-40, 0.5% Triton X-100, and 1 tablet of cOmplete Mini, EDTA-free [Sigma-Aldrich]) to lyse the cells. An aliquot of the cell lysate suspended in RSB-100 was saved for western blotting, and the remaining lysate was used in RNA isolation.

2.3 Western Blot Analysis

An equal volume of 2X sample buffer (100mM Tris-HCl solution pH 6.8, 4% SDS, 0.2% Bromophenol blue, 20% glycerol) was added to the reserved aliquot of cell lysate. The western blot sample was sonicated using a Branson Sonifier 450 (VWR Scientific) at max output for 1 minute before incubation at 80º C for 5 minutes.

The proteins were separated by a SDS-PAGE stacking gel constituted of a 4% upper stacking gel and a 10% lower resolving gel before being transferred onto a nitrocellulose membrane (Bio-Rad Laboratories). The membrane was incubated in a 5% milk-TBST solution containing the primary hnRNP A1 antibody (Santa Cruz Biotechnology) overnight in 4º C.

After washing, the membrane was incubated with a 2.5% milk-TBST solution containing HRP conjugated secondary antibodies at room temperature for 1 hour. The bound antibodies were detected using SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Scientific).

2.4 RNA Isolation and RT-PCR Analysis

Total RNA was extracted from transfected cell lysate using Direct-zol RNA Miniprep Plus kits (Zymo Research) following the manufacturer's protocol. For the validation of hnRNP A1 regulated splicing and β-globin splicing reporter assay, 1 μg of total RNA was reverse transcribed using Multiscribe Reverse Transcriptase (Applied Biosciences) with random primers provided in the kit, following the manufacturer's protocols. The cDNA from these reactions were diluted to 1/200 dilution before being used as templates for RT-PCR.

For the validation of hnRNP A1 regulated splicing, RT-PCR was performed 28 primer sets designed for hnRNP A1 regulated splicing events depicted in RNA-seq data (Howard et al, 2018) under the following PCR program: denaturation at 95º C for 2 minutes, 25 to 28 cycles of 95º C for 30 seconds, 55º C to 65º C for 30 seconds, 68º C for 30 seconds, and a final extension at 68º C for 1 minute.

For the β-globin splicing reporter assays, RT-PCR was performed using globin assay forward (5' -GCAACCTCAAACAGACACCA- 3') and reverse (5' -AGCTTGTCACAGTGCAGCTC- 3') primers designed to the flanking β-globin exons within the plasmid following the PCR program mentioned above.

The RT-PCR products were run on a 1% agarose gel, stained with SYBR green for visualization, and quantified using a DNA 1000 kit (Agilent Technologies) following the

manufacturer's protocol. The statistical significance was analyzed using a ratio-paired t-test. The P values are represented in the figures in the form of asterisks (* $P < 0.5$) and the absence of an asterisks indicates that the change was not statistically significant.

2.5 Validation of hnRNP A1 Regulated Splicing

2.5.1 Primer design

Through the RNA-seq analysis of HEK293T cell overexpressing hnRNP A1, approximately 250 hnRNP A1 dependent skipped splicing events with detectable U2AF2 crosslinking were observed. Out of the 250 skipped splicing events, 41 events exhibited changes of U2AF2 binding (Howard et al, 2018). To validate the hnRNP A1 regulated splicing by U2AF2, primers were designed for 27 genes. The primers were designed to encompass the alternatively skipped exon and anneal to sequences in the flanking constitutively spliced exons and out of the 27 genes (Table 2.1). The transfection and RT-PCR analysis were performed as mentioned above.

Gene	Forward Primer (5' - 3')	Reverse Primer (5' - 3')
EDRF1	GGTGATATCCAACTAATGCTGGC	GCAAGATTTCATTAGCAGCCTC
PIEZO1	GCTGGTGCCGTTCCTGGTGG	CTCATAGCCGCCCAGCTTCAG
EHMT2	GCCATGCCACAAAGTCATTCC	CTGAGGTCACCTTTCCCAGTG
SLC25a16	GTTCTGTAAGTTGGCAACCTAG	CGATCCAATGGAGCAACTGTTG
SKA2	CCAGAAAGCTGAGTCTGATCTGG	GACTAAGTGTGTGTGTGCATGCG
IST1	GCAGAAGCTCCTCCTGGGGTA	GCTGGCACCAGCAGATGCAGT
TRNAU1AP	GATCCCAGCTGGCTACTGCTT	GGTTGTTTCCCGTAAGTGGCA
COG4	GGATTGCCCGCATTGTGGAGA	CAGGTTGTTCTGAACATGCCGG
AK2	GCGTGCGTGCGTTGACCTGG	GCCTGCCTCACAGTCCGAGG
FAM192A	GAGCGAGCGGCTCTGCG	GCGCCGTTCATCTAGTTCTGC
TRA2A	GGCCAGGAGATTTCTGGCGG	GGAAACCCTTGATGGACTACG
KIF23	GCAGCCAAACAGCTGGAGATG	GAACAGATTGTCCACCACCC
SREK1	CGCACCTCTTGCTTTTTCCTC	GGACGCTTTCCCAGGACGGTG
ZNF711	CGCAGAGTAGATTGTGATTGGC	GCATAATCATGGTATGGGCCATTC
ITSN1	GGATTCTGGTTCTTCAGAGAGTCC	GCAGCTCTCCTTCCCACCATC
SRSF11	GCCTTTGCCAGTCTCATCTCG	GGTAAGTGGGTTAGGAGTAGGC
TIAL1	CACTTCCATGTGTTTGTTGGG	GTCTTTAACTACCCGGGCATC
PRPF39	CACTTGATGGCTGCCAGGAAG	GTTTGTCTCAGGATCACCAGG
WNK1	GGAATACAGCAGACAGCCC	CAGAGGAAGCCAAAGTGGTC
SRSF6	GTGGAGTTCGAGGACTCCCGC	CTTGCCAACTGCACCGACTAG
HMGN1	CCTCCTGCAAAAGTGGAAGCG	CTTCGTTTCCCCGTTTTCCGC
SON	CCTGGGCTCAGCTGAACTCTA	CATCAAAACGGCTCCCATTCC
DCAF8	CTATGGCCCCAAGAGTGAGTT	GTGATGGCGTCTCTGTCTCAG
HNRNPL	GGGACTACACAAACCCCAATC	GCAGAAGACATTGAAGACTCGG
ACIN1	GCCAATACAATAGAGCCAGCC	GACAACGGAGATCTTCCTCTTG
NAP1L4	GAAGGTGACGAGGAGGGAGAA	CCACCAGCCGCCAAGCGGTCA

Table 2.1: Primers used for the validation of hnRNP A1 regulated genes

2.6 β – Globin Splicing Reporter Assay

2.6.1 Reporter construct design

To characterize the non-canonical binding patterns exhibited in the RNA-seq data, we constructed splicing reporters from 2 genes (*SRSF6* and *TRANU1AP*) which validated the hnRNP A1 regulated splicing by U2AF2.

The gene fragments used in the construction of the reporters were generated using primers which encompassed the alternatively skipped exon and the non-canonical binding sites identified previously (Howard et al, 2018). The amplified gene fragments were cloned into beta globin based (HHB1) splicing reporter. The co-transfection and RT-PCR analysis of the splicing reporters and the hnRNP A1 expression were performed as mentioned above.

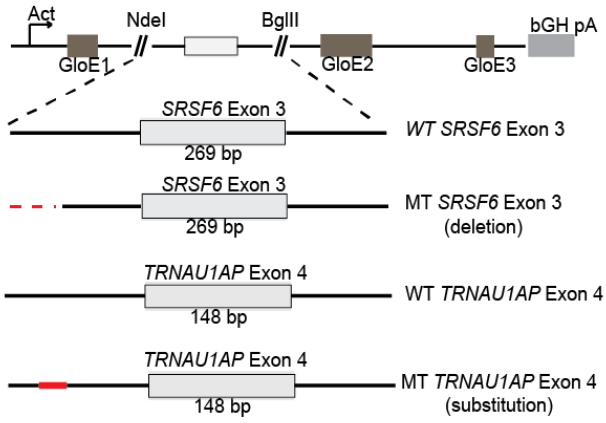

Figure 2.1: Splicing reporter gene construct design. The non-canonical binding site of U2AF2 for SRSF6 was deleted and substituted for TRANU1AP.

2.7 Individual-Nucleotide Resolution Crosslinking and Immunoprecipitation

2.7.1 Preparation of crosslinked cell extracts

Cells were washed with ice-cold PBS before crosslinking with 150 mJ/cm^2 at 254 nm in a UV Stratalinker 2400 (Stragene).

The crosslinked cells were suspended in CLIP lysis buffer (50 mM Tris-HCL solution pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% sodium dodecyl sulfate, 0.5% sodium deoxycholate) before sonication at max output for 1 minute with the Branson sonifier 450 (VWR Scientific). $CaCl_2$ was added to the lysates to 5 mM before the addition of DNase I (Promega) and incubation for 20 minutes at 37º C. After incubation, a 1/5 dilution was made from the lysates and nanodropped to determine the total RNA concentration.

A low concentration of MNase (Worthington Biochemicals) was added to the crosslinked lysates. The MNase treatment was optimized to a 1X reaction which consisted of 63 µg of total RNA and 1.5 units/µL of MNase which was further diluted to a 1/1000 dilution and added to the lysates. The lysates containing the appropriate amount of MNase were incubated at 37º C for 3 minutes precisely and EGTA was added immediately after incubation to 10 mM.

The MNase treated lysates were then spun in a microcentrifuge at 16,000 x g at 4º C for 15 minutes. Two small aliquots of the supernatant were saved for size matched input preparation and a pre-immunoprecipitation control for western blots. The remainder of the

supernatant was used in the immunoprecipitation for antibody selection of the protein-RNA complexes.

2.7.2 Bead preparation and immunoprecipitation

Protein A Dynabeads (Invitrogen) and 1 µg of U2AF2 antibody (Santa Cruz Biotechnology) were used per experiment. The beads were washed three times with a 0.1 M sodium phosphate solution pH 8.1. The washed beads were resuspended with the sodium phosphate solution and the U2AF2 antibody was added before incubation at 4º C for 1 hour while rotating.

The incubated beads were washed three times with the CLIP lysis buffer and the supernatant from the crosslinked lysates were added. The samples were incubated for 1 hour at 4º C while rotating.

A small aliquot of the supernatant was saved as the post-immunoprecipitation sample for western blotting from the immunoprecipitated crosslinked sample. The remainder of the supernatant was discarded, and the beads were washed twice with a high salt wash buffer (50 mM Tris-HCl solution pH 7.4, 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% sodium dodecyl sulfate, and 0.5% sodium deoxycholate). The second wash was rotated at 4º C for 1 minute. The beads were washed twice with PNK buffer (20 mM Tris-HCl solution pH 7.4, 10 mM $MgCl_2$, and 0.2% Tween-20) before on bead RNA 3' end adapter ligation.

2.7.3 Phenol-chloroform extraction of size matched input

The protein-protein complexes were isolated form the reserved aliquot of supernatant from the crosslinked lysate during the immunoprecipitation. Deionized water and an equal volume of Acid Phenol-Chloroform was added to the supernatant before incubation in a

thermomixer for 5 minutes at 37º C and 1100 rpm. The solution was spun in a microcentrifuge to separate the phases and the aqueous top layer was extracted and underwent ethanol precipitation (Green and Sambrook, 2016). 2X LDS sample buffer (Novex) was added to precipitated sample before incubation at 80º C for 5 minutes. The incubated sample was run on a SDS-PAGE gel alongside the crosslinked samples.

2.7.4 On bead 3' end adapter ligation

The immunoprecipitated RNA was treated with T4 PNK (Promega) in a 5X PNK pH 6.5 buffer (350 mM Tris-HCl solution pH 6.5, 50 mM $MgCl_2$, and 5 mM dithiothreitol) for 20 minutes at 37º C to remove the 3' end phosphate group. The incubated samples were washed with PNK buffer and washed with high salt wash buffer while rotating at 4º C. The samples were washed twice more with PNK buffer before the ligation of the adapter.

After the dephosphorylation of the 3' end of the RNA, 20 μM of NEXTflex 3' 4N adenylated adapter (BIOO Scientific) was ligated using T4 RNA ligase (Clonetech) in a reaction containing 4X ligation buffer (200 mM Tris-HCl solution pH 7.8, 40 mM $MgCl_2$, and 4 mM dithiothreitol) overnight at 16º C. The adapter ligated protein-RNA complexes were washed twice with a high salt wash buffer, rotating each wash for 5 minutes at 4º C. The samples were washed twice with PNK buffer before radiolabeling.

2.7.5 Radiolabeling 5' end of RNA

The adapter ligated U2AF2-RNA complexes were 5' end labeled using γ^{32}-ATP (PerkinElmer) with T4 PNK (Promega) for 5 minutes at 37º C in a thermomixer at 1100 rpm. The supernatant was discarded appropriately, and the beads were suspended in 1X LDS buffer (Novex) and incubated for 5 minutes at 70º C before being separated on the SDS-PAGE gel.

2.7.6 SDS-PAGE and nitrocellulose transfer

The samples were loaded on a 4-12% NuPAGE Bis-Tris gel (Novex) and run in 1X MOPS running buffer for 1 hour and 30 minutes at 120V. Pre-stained Protein Standards (Novex) were used to visualize the sizes of the RNA-protein complexes and the gel was run until the 110 kDa band was at the middle of the gel.

The U2AF2-RNA complexes were transferred on to a Protran Nitrocellulose Membrane (Amersham) using a GENIE wet transfer apparatus (Idea Scientific) at 24V for 30 minutes.

The membrane containing the transferred RNA-protein complexes were wrapped in saran wrap and exposed on to an autoradiography film (UltraCruz) for 5 min and 10 minutes at -80º C.

2.7.7 RNA isolation and reverse transcription

The exposed autoradiography film was used as a guide to extract the desired size range of U2AF2-RNA complexes (70 kDa to 110 kDa). The U2AF2-RNA complexes were Proteinase K (Ambion) treated for 20 minutes at 37º C to abolish the bound U2AF2 protein. The RNA was further isolated through phenol-chloroform extraction as mentioned above.

The isolated RNA was reverse transcribed using Superscript III (Invitrogen) with RT primers with NextFLEX compatibility (Table 2.2) following the thermal program: 25º C for 5 minutes, 42º C for 20 minutes, 50º C for 40 minutes, 80º C for 5 minutes, and a 4º C hold. The cDNA from the reverse transcription reaction was ethanol precipitated (Green and Sambrook, 2016).

Primer Name	Sequence (5' - 3')
RT# 1	/5Phos/NNAACCNNNGATCGTC
RT# 2	/5Phos/NNACAANNNGATCGTCG
RT# 3	/5Phos/NNATTGNNNGATCGTCG

Table 2.2: Reverse transcriptase primers used for each replicate of the iCLIP experiment.

2.7.8 cDNA size selection

TBE-urea sample buffer was added to the cDNA and run on a 6% TBE-urea gel at 180V until the dark blue dye front was 0.5 cm from the bottom of the glass plate. Generuler Ultra Low Range Ladder (Thermo Scientific) was used to approximate the sizes of desired cDNA ranges and were cut from the gel. The low size range of cDNA were cut between 60 base pairs to 75 base pairs and the high size range of cDNA were cut between 80 base pairs to 100 base pairs.

RNA elution buffer (20 mM Tris HCl solution pH 7.5, 0.25 M sodium acetate, 1 mM EDTA, 0.25% SDS) was added to the gel fragments and crushed using a 1 mL syringe plunger. The gel solution was incubated for 1 hour at 1100 rpm at 37º C, placed in dry ice for 2 minutes and incubated once more as previously mentioned. The incubated samples were placed into SpinX columns (Costar) and centrifuged for 1 minute at max speed. The cDNA was phenol-chloroform extracted as mentioned above.

2.7.9 Ligation of primer to 5' end of the cDNA

The cDNA was circularized by CircLigase (Epicentre) through incubation at 60º C for 1 hour. 10 μM of BAMHI cut oligo (5' – CCAAGGGATCCGTTCAGAGTT – 3') was annealed and the cDNA was linearized by BAMHI through the following program: 95º C for 2 minutes, successive cycles of 20 seconds, starting from 95º C and decreasing the temperature by 1º C each cycle until 25º C, and a 25º C hold. The linearized cDNA was ethanol precipitated as mentioned above.

2.7.10 PCR and barcoding of primary amplified PCR products

A small amount of the eluted cDNA was used to optimize the PCR reaction to determine the best cycle number for producing a well-represented library. The PCR was performed using the NextFLEX Small RNA Kit v3 (BIOO Scientific) following the manufacturer's protocols. The optimizations were done using barcode primers not essential for sequencing experiments.

The libraries for sequencing were prepared using the barcode primers presented in Table 2.3 and quantified using DNA High Sensitivity Chips (Agilent Technologies) following the manufacturer's protocol.

Sample	Barcode Number	Index Sequence (5' - 3')
HCLIP Pool 1 Low	1	ATCACG
HCLIP Pool 1 High	2	CGATGT
HCLIP Pool 1 SMI Low	3	TTAGGC
HCLIP Pool 1 SMI High	4	TGACCA
CCLIP Pool 1 Low	5	ACAGTG
CCLIP Pool 1 High	6	GCCAAT
CCLIP Pool 1 SMI Low	8	ACTTGA
CCLIP Pool 1 SMI Low	7	CAGATC
OCLIP Pool 1 Low	21	GTTTCG
OCLIP Pool 1 High	1	ATCACG
OCLIP R2 SMI Low	31	CACGAT
OCLIP R2 SMI High	25	ACTGAT
CCLIP Pool 2 Low	33	CAGGCG
CCLIP Pool 2 High	9	CATCAG
CCLIP Pool 2 SMI Low	27	ATTCCT
CCLIP Pool 2 SMI High	28	CAAAAG
HCLIP R1 Low	12	CTTGTA
HCLIP R1 High	29	CAACTA
HCLIP R1 SMI Low	19	GTGAAA
HCLIP R1 SMI High	32	CACTCA
HCLIP R2 Low	34	CATGGC
HCLIP R2 High	35	CATTTT
HCLIP R2 SMI Low	46	TCCCGA
HCLIP R2 SMI High	47	TCGAAG

Table 2.3: Barcodes used for multiplexing the libraries generated from iCLIPs. Pool 1 refers to the first batch of iCLIPs performed with libraries from 3 biological replicates. Pool 2 refers to the second round of iCLIP performed with libraries from 2 biological replicates. R1 and R2 refers to samples prepared in the second round of iCLIP experiments with no pooling.

Chapter 3

Results

3.1 Identification of hnRNP A1 regulated exons

Through a series of iCLIP experiments performed on HEK293T cells for U2AF2, our lab has shown that U2AF2 is affected by hnRNP A1 during the process of 3' splice site selection. RNA-seq analysis of the U2AF2 iCLIP data has identified approximately 250 genes that display an alternative exon skipping event mediated by hnRNP A1. Further analysis of CLIPper data revealed, U2AF2 peaks shifting from its canonical binding site to a cryptic binding site further upstream within the intron under hnRNP A1 overexpression. This data suggested that the abundance of hnRNP A1 was responsible for promoting the redistribution of U2AF2 (Howard et al, 2018).

3.1.1 Validations of endogenous splicing event

To validate the hnRNP A1 dependent exon skipping events observed in the RNA-seq analysis, we chose 27 genes that exhibited the displacement of U2AF2 from its canonical binding site under hnRNP A1 overexpression. Out of the 27 genes, we were able to recapitulate the endogenous hnRNP A1-dependent alternative exon skipping event in 3 genes (Figure 3.1).

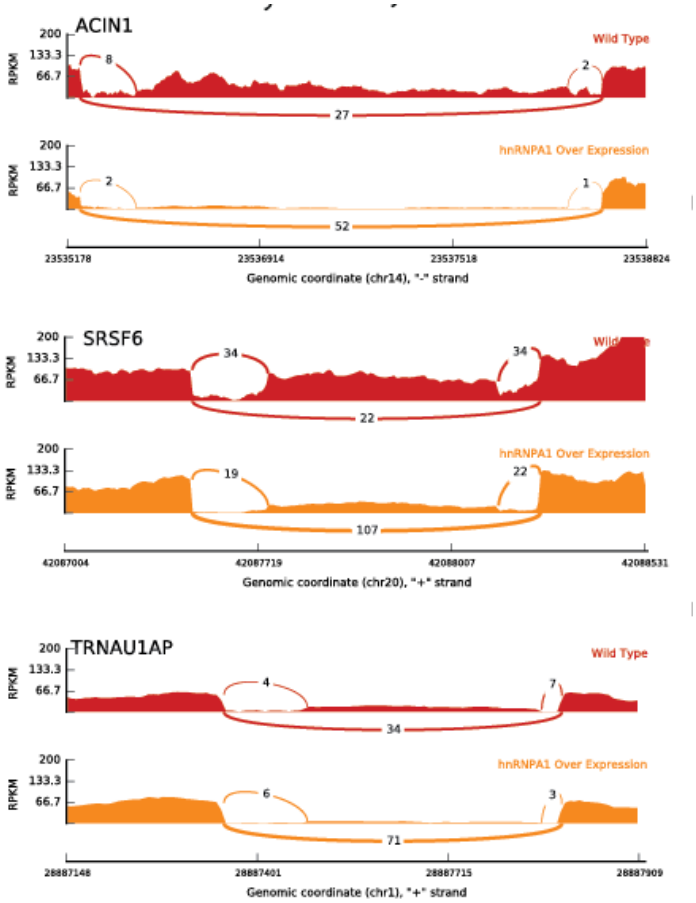

Figure 3.1: Sashimi plots showing read and junction coverage for three genes: ACIN1, SRSF6, and TRNAU1AP.

Transient transfection of T7-epitope tagged hnRNP A1 expression plasmid and RT-PCR was used to validate several splicing events which exhibited hnRNP A1 dependent reduction of the alternative exon. Quantification of the RT-PCR amplicons revealed approximately a twofold reduction of the alternatively spliced exon in ACIN1 and SRSF6 whereas TRNAU1AP showed a 1.3-fold reduction of the alternatively spliced exon (Figure 3.2.A and B).

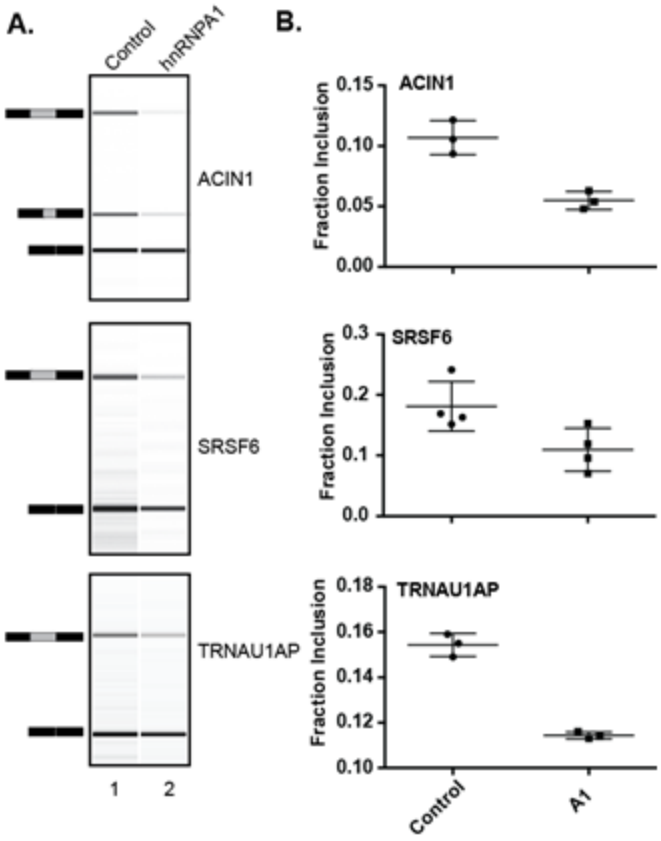

Figure 3.2: hnRNP A1 overexpression reduces use of alternative exon. (A) Virtual gel representation of Agilent Bioanalyzer 2100 DNA 1000 analysis of RT-PCR products of ACIN1, SRSF6, and TRNAU1AP. (B) Bar graph depicting mean exon inclusion for endogenous transcript expression of ACIN1, SRSF6, and TRNAU1AP under hnRNP A1 overexpression.

3.1.2 Identification of non-canonical binding site of U2AF2

To determine whether the upstream cryptic binding site was responsible for the observed changes in endogenous transcripts, we created hemoglobin subunit beta (HBB) splicing reporter gene constructs containing either the wild-type or mutated intronic sequences upstream of the hnRNP A1-sensitive exons found in the pre-mRNA (Figure 3.4). The SRSF6 and TRNAU1AP splicing reporters were transiently co-transfected into HEK293T cells with T7-epitope tagged hnRNP A1 expression plasmid or a control plasmid.

As shown in Figure 3.3 C and F, RT-PCR analysis of wild-type splicing reporters of SRSF6 and TNRAU1AP show the expected reduction of exon inclusion in response to hnRNP A1 overexpression compared to control conditions. Deletion/Substitution of the upstream non-canonical U2AF2 binding site led to a depletion of exon inclusion and a loss of hnRNP A1 dependent skipping.

Global analysis of the data from U2AF2 iCLIP under hnRNP A1 overexpression identified antisense *Alu* derived RNA as the non-canonical decoy sites that U2AF2 gets displaced to (Figure 3.5).

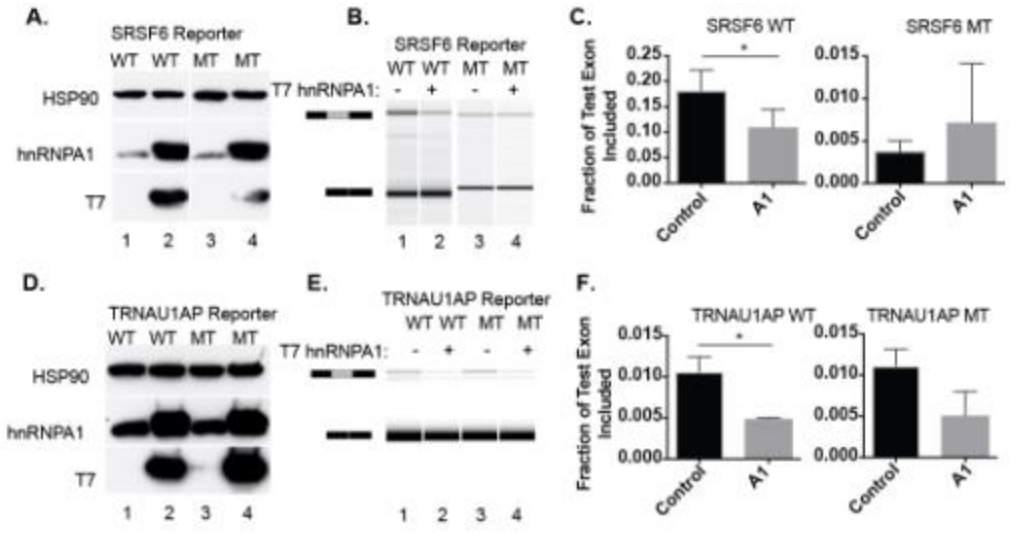

Figure 3.3: Non-canonical binding sites of U2AF2 influence hnRNP A1 dependent skipping. (A & D) Representative western blot probed against HSP90, hnRNP A1, and T7 epitope. Lysates were prepared from HEK293T cells transiently co-transfected with either wild-type (WT) or mutant (MT) splicing reporter gene constructs and T7-epitope tagged hnRNP A1 expression plasmid or control plasmid. (B & E) Virtual gel representation of an Agilent Bioanalyzer 2100 DNA 1000 assay of RT-PCR products from SRSF6 and TRNAU1AP splicing construct transfections. (C & F) Bar graph depicting mean exon inclusion for SRSF6 and TRNAU1AP splicing reporters quantified using an Agilent Bioanalyzer 2100 with standard deviation bars.

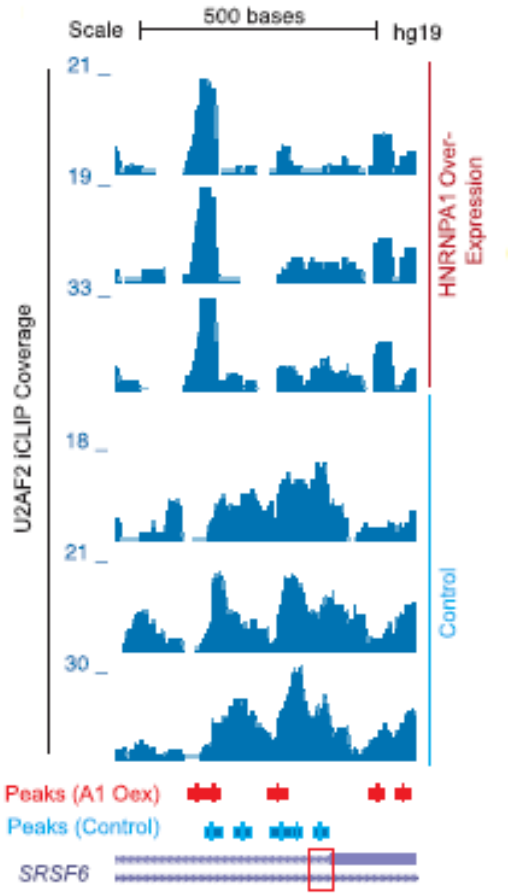

Figure 3.4: UCSC genome browser screenshot of U2AF2 iCLIP coverage at the SRSF6 exon 3 locus from control HEK293T cells (bottom 3 tracks) and hnRNP A1 overexpressing cells (top 3 tracks). The red box denotes the 3' splice site of SRSF6 exon 3 and the peaks called by CLIPper are shown below the coverage tracks.

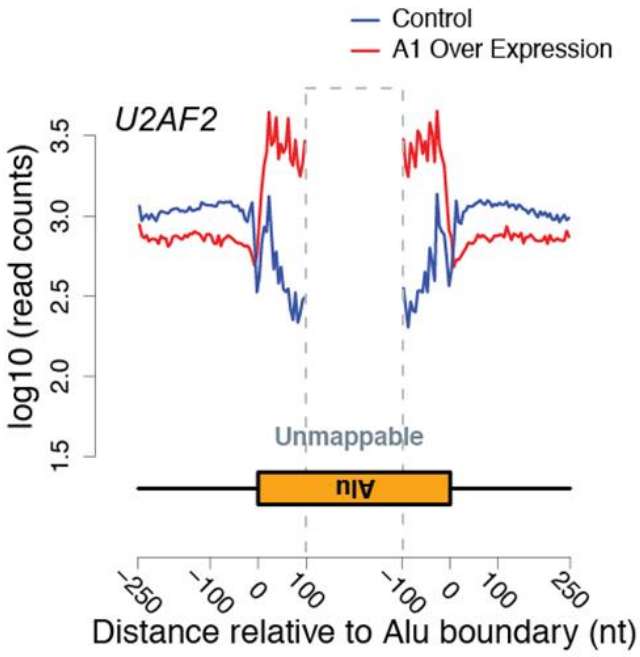

Figure 3.5: Aggregated crosslinking sites of antisense *Alu* elements as well as nearby regions of U2AF2. The blue line represents the wild-type binding of U2AF2 and the red line represents the hnRNP A1 overexpression of the $\log_{10}$ number of iCLIP read counts across all antisense *Alu* elements.

3.2 Purification of U2AF2-RNA complexes from primate IPS cell lines

To try and substantiate our hypothesis of finding functional *Alu* elements that may have played a role in the evolution of the primate lineage, we performed individual nucleotide resolution crosslinking immunoprecipitation and high throughput sequencing (iCLIP-seq) on orangutan, chimpanzee, and human induced pluripotent stem cells (iPSCs) (Figure 3.6). All the

immunoprecipitated material were UV-dependent, antibody-dependent, and nuclease sensitive. The sequencing data from the iCLIP experiments revealed that most of the libraries we prepared, mostly mapped uniquely, however suffered from high duplication rates (Table 3.1).

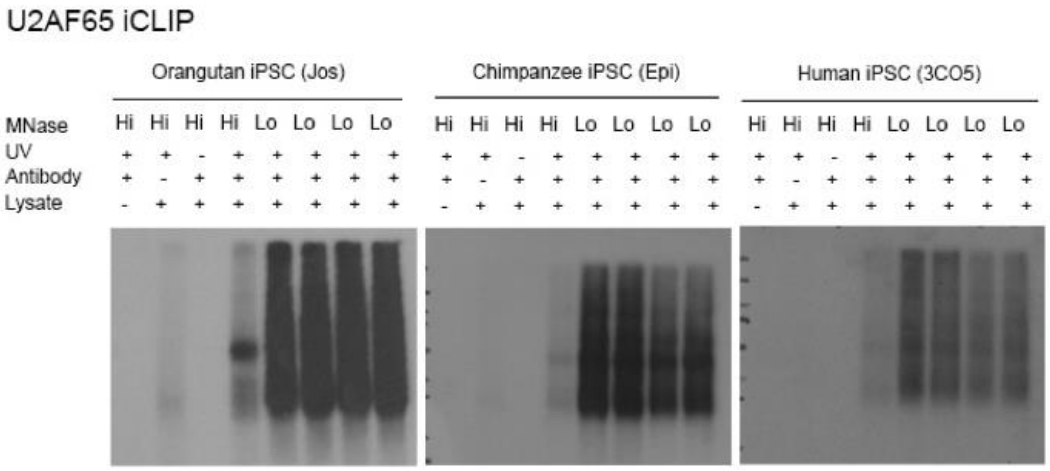

Figure 3.6: Examples of autoradiographs for U2AF2 iCLIP of orangutan, chimpanzee, and human induced pluripotent stem cells (iPSCs). Protein-RNA complex shifts are -lysate, -UV, -antibody, and micrococcal-nuclease sensitive. Protein-RNA complexes ranging from 70 kDA to 110 kDA were excised for RNA isolation and iCLIP library preparation. Micrococcal nuclease treatment at 15 U (high) and 0.0075 (low).

Sample	Total Read Count	Uniquely Mapped (%)	Duplication Rates
HCLIP Pool 1 Low	9 M	72.9	18.2
HCLIP Pool 1 High	15.16 M	88.8	31.47
HCLIP Pool 1 SMI Low	3.07 M	79.5	41.19
HCLIP Pool 1 SMI High	41.64 M	13.3	59.72
CCLIP Pool 1 Low	16.95 M	69.3	52.24
CCLIP Pool 1 High	81.87 M	83.1	19.86
CCLIP Pool 1 SMI Low	15.78 M	69.3	57.97
CCLIP Pool 1 SMI High	2.58 M	76.3	266.48
OCLIP Pool 1 Low	32.99 M	63.6	99.7
OCLIP Pool 1 High	37.41 M	83.6	93.8
OCLIP R2 SMI Low	0.1 M	56.4	1.88
OCLIP R2 SMI High	5.66 M	48	22.6
CCLIP Pool 2 Low	8.58 M	89.9	53.3
CCLIP Pool 2 High	12.87 M	85.1	44.3
CCLIP Pool 2 SMI Low	47.29 K	70.4	1.2
CCLIP Pool 2 SMI High	0.55 M	29.4	10.23
HCLIP R1 Low	4.27 M	91.7	17.04
HCLIP R1 High	74.62 M	90.1	62.25
HCLIP R1 SMI Low	0.26 M	65.1	3.9
HCLIP R1 SMI High	0.8 M	43	10.95
HCLIP R2 Low	11.39 M	91.1	24.14
HCLIP R2 High	110.25 M	86.4	99.17
HCLIP R2 SMI Low	0.61 M	49.1	16.17
HCLIP R2 SMI High	0.65 M	41.3	11.92

Table 3.1: Statistics of the sequencing results of the libraries prepared from the iCLIP experiments. M denotes millions of counts whereas K denotes thousands of counts.

Chapter 4

Discussion

The aim of this project began with a goal of validating splicing events due to changes in U2AF2 binding observed in RNA-seq data of HEK293T cells overexpressing hnRNP A1, which we approached using splicing reporters (Howard et al, 2018). However, upon analysis of our data, we were able to identify the new displacement sites of U2AF2 as repetitive elements called *Alu* elements. This discovery led to our most recent aim of trying to identify functional *Alu* elements involved in splicing that could have had an evolutionary impact in the primate lineage. To address our hypothesis, we have performed iCLIP analysis of U2AF2 on three primate iPSCs to generate libraries that can be used to compare the binding of U2AF2 to their transcriptomes.

Previous work in our lab has demonstrated that the overexpression of hnRNP A1 results in an altered binding profile for U2AF2 (Figure 3.4). Additionally, we have revealed that the change in U2AF2 binding under hnRNP A1 overexpression occurs globally throughout the transcriptome (Figure 3.5). Interestingly, approximately 250 genes showed a hnRNP A1 dependent change in splicing, 3 of which we were able to validate endogenously as well as through splicing reporter assays. The RNA-seq analysis of these three genes (SRSF6, ACIN1, and TRNAU1AP), showed similar exon skipping events under hnRNP A1 overexpression (Figure 3.1). Primers were designed to assay for the transcript levels of each of these genes under A1 overexpression. The endogenous splicing assay recapitulated the exon skipping events seen in the RNA-seq data which suggested that the effects of A1 overexpression are indeed palpable (Figure 3.2). The splicing reporter assays confirmed the functionality of

polypyrimidine tracts upstream of the canonical U2AF2 binding site identified by previous iCLIP experiments and suggest that they may play a role in hnRNP A1 dependent splicing silencing (Figure 3.3).

hnRNP A1 is involved in a myriad of splicing events regulating gene expression. Mechanisms such as proofreading of the 3' ss through the formation of a ternary complex with U2AF2 to aid the splicing machinery to discriminate cryptic and functional 3' splice sites have been identified to be functional roles of hnRNP A1 (Jean-Philippe et al, 2013; Tavanez et al, 2012). However, we believe that our data suggests that hnRNP A1 may play an additional role at the 3' splice site to influence the binding of U2AF2 from canonical 3' splice site to decoy 3' splice sites.

A1 has been shown to be required for the recognition of the correct AG dinucleotide over the CG dinucleotide at the 3' splice site through the displacement of U2AF2 from the CG dinucleotides. Our previous iCLIP data has revealed a global dissociation of U2AF2 at the 3' splice sites of alternatively spliced exons under hnRNP A1 overexpression while the 3' splice sites of constitutively splice exons showed little or no difference (Howard et al, 2018). Additionally, our data shows the displacement of U2AF2 from the 3' splice sites of alternatively spliced exons to polypyrimidine tracts further upstream in the intronic regions.

Differential binding locations within intronic regions by U2AF2 affects the splicing patterns of the target gene. The binding of U2AF2 to intronic sequences upstream of the alternative exon has been shown to lead to exon skipping, whereas the binding of U2AF2 downstream of the alternative exon leads to an increase in exon inclusion (Wu and Fu, 2015). Our results are well aligned with this observation, where we can detect an increase in hnRNP

A1 dependent displacement of U2AF2 to an upstream decoy splice site identified as an anti-sense *Alu* derived RNA element, ultimately leading to an increase in alternative exon skipping.

Alu elements have been identified to affect gene expression through various mechanisms (Ast, 2004; Lee et al, 2007; Locke et al, 2011). Much of the effects *Alus* have on gene expression have been characterized to have deleterious effects (AMRD) or induce exonization of intronic elements to change protein shape/function (Lee et al, 2007; Häsler, & Strub, 2006). hnRNP C has been shown to combat the disruptive effects *Alu* elements have on gene expression through direct competitive binding by U2AF2 which represses the inclusion of *Alu* elements (Zarnack et al, 2013). Our previous data has shown that the overexpression of hnRNP A1 leads to an increased association of U2AF2 to sense and antisense *Alu* derived RNA sequences. However, we have also identified that the overexpression of hnRNP A1 does not affect hnRNP C expression or localization (Howard et al, 2018). Collectively, our data suggests that *Alu* elements may serve as RNA regulatory elements that are dependent on the intracellular concentration of splicing factors.

Alu elements are short interspaced nuclear elements (SINEs) that are specific to primate genomes (Salem et al, 2003). Our previous results led to the hypothesis that *Alu* derived RNA elements may contribute to the species specific alternative splicing events we observe in the primate lineage. To address this, we performed iCLIP analysis of U2AF2 on 3 primate cell lines (Figure 3.6). We were able to isolate UV dependent, antibody dependent, and nuclease sensitive RNA-protein complexes to generate libraries which could identify the differences in binding profiles of U2AF2 across the three primate genomes. However, our first sequencing data set revealed that our libraries suffered from high duplication rates (Table 3.1). The high duplication rates could have risen from excessive cycling times during the PCR reaction and a deficient amount of cDNA library input when generating our barcoded libraries. Promisingly,

the sequencing data has shown us that most of the sequenced fragments can be mapped to the prospective genomes of each primate species.

Future directions of this investigation will consist of generating more libraries for each of the primate iCLIP experiments with reduced cycling times and titrating the amount of cDNA input used for each of the PCR reactions. Once we have identified some functional *Alu* derived RNA elements which give rise to species specific alternative splicing events in the different primate genomes, our goal would be to generate splicing gene reporters to validate the specific function and mechanism involved. Following these results, another point of investigation could focus on how *Alu* element mediated splicing transcripts are affected by the ribosome during translation, which can be addressed through Fraq-seq experiments on the same primate cell lines used. Lastly, we are also interested in revealing how the structure and function of proteins translated from *Alu* mediated spicing transcripts are affected.

Bibliography

Ast, G. (2004). How did alternative splicing evolve? *Nature Reviews Genetics,5*(10), 773-782. doi:10.1038/nrg1451

Batzer, M. A., & Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nature Reviews Genetics,3*(5), 370-379. doi:10.1038/nrg798

Conti, L. D., Baralle, M., & Buratti, E. (2012). Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA,4*(1), 49-60. doi:10.1002/wrna.1140

Dever, T. E., & Green, R. (2012). The Elongation, Termination and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harbor Perspectives in Biology*, *4*(7), a013706. http://doi.org/10.1101/cshperspect.a013706

Eperon, I. C., Ireland, D. C., Smith, R. A., Mayeda, A., & Krainer, A. R. (1993). Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *The EMBO Journal*, *12*(9), 3607–3617.

Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science,351*(6274). doi:10.1126/science.aac7247

Fredericks, A. M., Cygan, K. J., Brown, B. A., & Fairbrother, W. G. (2015). RNA-Binding Proteins: Splicing Factors and Disease. *Biomolecules*, *5*(2), 893–909. http://doi.org/10.3390/biom5020893

Green, M. R., & Sambrook, J. (2016). Precipitation of DNA with Ethanol. *Cold Spring Harbor Protocols,*2016(12). doi:10.1101/pdb.prot093377

Jean-Philippe, J., Paz, S., & Caputi, M. (2013). hnRNP A1: The Swiss Army Knife of Gene Expression. *International Journal of Molecular Sciences*, *14*(9), 18999–19024. http://doi.org/10.3390/ijms140918999

Johnson, J. M. et al. (2003). Genome-Wide Survey of Human Alternative Pre-mRNA Splicing with Exon Junction Microarrays. Science, 302(5653), 2141-2144. doi:10.1126/science.1090100

König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., . . . Ule, J. (2010). ICLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural & Molecular Biology,17*(7), 909-915. doi:10.1038/nsmb.1838

Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., & Muñoz, M. J. (2013). Erratum: Alternative splicing: A pivotal step between eukaryotic transcription and translation. *Nature Reviews Molecular Cell Biology,14*(5), 306-306. doi:10.1038/nrm3560

Han, K., Lee, J., Meyer, T. J., Wang, J., Sen, S. K., Srikanta, D., . . . Batzer, M. A. (2005). Alu Recombination-mediated Structural Deletions in the Chimpanzee Genome. *PLoS Genetics,Preprint*(2007). doi:10.1371/journal.pgen.0030184.eor

Häsler, J., & Strub, K. (2006). Alu elements as regulators of gene expression. *Nucleic Acids Research,34*(19), 5491-5497. doi:10.1093/nar/gkl706

Howard, J. M., Lin, H., Kim, G., Draper, J. M., Haeussler, M., Katzman, S., . . . Sanford, J. R. (2017). HNRNPA1 promotes recognition of splice site decoys by U2AF2 in vivo. *Genome Research,28*, 689-698. doi:10.1101/175901

Izquierdo, J. M., Majós, N., Bonnal, S., Martínez, C., Castelo, R., Guigó, R., . . . Valcárcel, J. (2005). Regulation of Fas Alternative Splicing by Antagonistic Effects of TIA-1 and PTB on Exon Definition. *Molecular Cell,19*(4), 475-484. doi:10.1016/j.molcel.2005.06.015

Lee, Y., & Rio, D. C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual Review of Biochemistry,84*(1), 291-323. doi:10.1146/annurev-biochem-060614-034316

Locke, D. P., Hillier, L. W., Warren, W. C., Worley, K. C., Nazareth, L. V., Muzny, D. M., … Wilson, R. K. (2011). Comparative and demographic analysis of orangutan genomes. *Nature*, *469*(7331), 529–533. http://doi.org/10.1038/nature09687

Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol. 2007; 8:209–20. [PubMed: 17318225]

Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology,15*(2), 108-121. doi:10.1038/nrm3742

Pan, Q., Shai, O., Lee, L. J., Frey, B. J.& Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).

Salem, A., Ray, D. A., Xing, J., Callinan, P. A., Myers, J. S., Hedges, D. J., . . . Batzer, M. A. (2003). Alu elements and hominid phylogenetics. *Proceedings of the National Academy of Sciences,100*(22), 12787-12791. doi:10.1073/pnas.2133766100

Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., . . . Fu, X. (2014). Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. *Nature Structural & Molecular Biology,21*(11), 997-1005. doi:10.1038/nsmb.2906

Sorek, R. (2002). Alu-Containing Exons are Alternatively Spliced. *Genome Research,12*(7), 1060-1067. doi:10.1101/gr.229302

Tavanez, J., Madl, T., Kooshapur, H., Sattler, M., & Valcárcel, J. (2012). HnRNP A1 Proofreads 3′ Splice Site Recognition by U2AF. *Molecular Cell,45*(3), 314-329. doi:10.1016/j.molcel.2011.11.033

Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell 2009; 136:701-18; PMID:19239890; http://dx.doi.org/10.1016/j.cell.2009.02.009

Wang, Y., & Wang, Z. (2014). Systematical identification of splicing regulatory *cis*-elements and cognate *trans*-factors. *Methods (San Diego, Calif.)*, *65*(3), 350–358. http://doi.org/10.1016/j.ymeth.2013.08.019

Wu, T., & Fu, X.-D. (2015). Genomic functions of U2AF in constitutive and regulated splicing. *RNA Biology*, *12*(5), 479–485. http://doi.org/10.1080/15476286.2015.1020272

Yin, Y., & Steitz, T. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Nature Reviews Molecular Cell Biology*, 16, 129-143. doi:10.2210/pdb1msw/pdb

Zamore, P. D., & Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(23), 9243–9247.

Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., . . . Ule, J. (2013). Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements. *Cell*,*152*(3), 453-466. doi:10.1016/j.cell.2012.12.023