

8-2016

INTERROGATING DUX4 MRNA 3' END PROCESSING

Natoya J. Peart

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations



Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

Recommended Citation

Peart, Natoya J., "INTERROGATING DUX4 MRNA 3' END PROCESSING" (2016). *The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access)*. 692.

https://digitalcommons.library.tmc.edu/utgsbs_dissertations/692

This Dissertation (PhD) is brought to you for free and open access by the The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digitalcommons@library.tmc.edu.

INTERROGATING DUX4 MRNA 3'END PROCESSING

by

Natoya Janeen Peart, B.S

APPROVED:

Eric J. Wagner, Ph.D.
Advisory Professor

Joseph L. Alcorn, Ph.D.

Swathi Arur, Ph.D.

Thomas A. Cooper, MD.

Anil Sood, MD.

Ambro van Hoof, Ph.D.

APPROVED:

Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

INTERROGATING DUX4 MRNA 3'END PROCESSING

A

DISSERTATION

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
MD Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Natoya Janeen Peart, B.S.

Houston, Texas

August 2016

Dedication

In memory of my mother, Nalda Eugene Peart.

Acknowledgments

This is what the Lord says: “Let not the wise boast of their wisdom or the strong boast of their strength or the rich boast of their riches, but let the one who boasts boast about this: that they have the understanding to know me, that I am the Lord, who exercises kindness, justice and righteousness on earth, for in these I delight,” declares the Lord.

Jeremiah 9:23-24 (NIV)

First, I would like to thank God, for his mercy and guidance as He brought me through this with my sanity intact.

Second, I would like to thank my advisor, Dr. Eric Wagner. It was challenging, I cried a bit, I fumed some, and in the end I am glad it is over. I am glad it is over, but I can go out saying this experience has taught me much. So, thank you Eric: for your patience, and kindness. Thank you for giving me the opportunity to be a part of your lab; your passion for and rigor in science has been instrumental in my growth as a scientist. I am little tougher now, a little more skeptical and lot more excited about science.

Third, I would like to thank my committee, for sitting with me through all these committee meetings. Dr. Joseph Alcorn, Dr. Swathi Arur, Dr. Michael Blackburn, Dr. Thomas Cooper, Dr. Anil Sood, and Dr. Ambro van Hoof, thank you for your advice, and for your constructive criticism.

Fourth, I would like to thank my family. You are all special, and I am grateful to you all for your words of encouragement and prayers. I would in particular like to thank those whose actions were instrumental in allowing me to complete this stage. My Dad, who I have always seen as an anchor, silently supportive, my Dad who sent me timely and encouraging texts, my dad who first taught me to “aim for the sky and not for the coconut tree”, who in his

own way reached out to me when I needed it. Thank you to my sister, Judian for tolerating my moods, and encouraging me to keep going. Thank you also to Natalie and Marie, for taking care of my Dad and my brother, and giving me peace of mind to know that they are well cared for, and for helping me feel at “home” whenever I am home, in spite of all the changes. To my uncle and aunt, Leon and Beverly Cummings, I really appreciate all the care packages you sent, the phone calls just to see how I was doing. I really appreciate and reciprocate the care and the love that I always feel from you. My cousin Christine, for being there for me and listening to me whether it was at 4:00AM in the morning or 9:00PM in the evening. There are innumerable other occasions, that my family has been so helpful during this time.

Fifth, I would like to thank my friends, Patience, Damien, Melissa, Sherine, Nicole, Domila, Loïc, and Drew, in particular, but for so many others whom I haven't mentioned by name. Thank you for commiserating with me, for reminding me that I am not a failure, for reminding me that we all have roadblocks, for reminding me that there isn't just one path, for laughing with me, or consoling me, and for encouraging me to try new things.

Sixth, I would like to thank students, faculty and staff who made up the Biochemistry and Molecular Biology Graduate Program at UTHealth during my time there. In particular, to the blue hallway crew, Rita, Swarna, Nick and Tre', thank you for your enthusiasm and your passion for life and science. For reminding me that it can be fun!

Seventh, I would like to thank the members of the Wagner lab both past and present with whom my tenure in the lab overlapped. You were, in fact, the people I saw more often than my own family, I worked with you, I laughed with you, but most importantly I learned from you. I was surrounded by good and interesting science, but also good and interesting people.

Finally, I would like to thank those agencies that supported my research financially, specifically the Schissler Foundation and the Friends of FSH Research. I would also like to

thank the Graduate School of Biomedical Sciences, the Graduate School Education Committee, and the RNA Society for travel awards to attend conferences to present my work.

INTERROGATING DUX4 MRNA 3'END FORMATION

Natoya Janeen Peart, B.S

Advisory Professor: Eric J. Wagner, Ph.D.

Double Homeobox 4, Dux4, is the leading candidate gene for Facioscapulohumeral Dystrophy (FSHD). FSHD is the third most common muscular dystrophy, and is characterized by progressive muscle weakness primarily in the upper body. In individuals diagnosed with FSHD, Dux4 is inappropriately expressed in somatic cells due to two conditions. The first is hypomethylation of the subtelomeric D4Z4 repeats on chromosome 4. Each D4Z4 repeat on chromosome 4 is 3.3kb in length and contains the open reading frame for Dux4. Hypomethylation of the D4Z4 repeats primarily occurs due to contraction of the repeats from 11-100 (typical numbers in the healthy population) to between 1 and 10 repeats. Concomitant with the hypomethylation of the D4Z4 repeats on chromosome 4 is a single nucleotide polymorphism in the flanking DNA that generates a non-consensus polyadenylation signal (PAS). This PAS allows for the productive transcription of a polyadenylated Dux4 mRNA from the terminal D4Z4 repeat. Dux4 is anemically expressed in patient somatic cells, but contributes to FSHD pathology due to Dux4-dependent cellular reprogramming.

We aim to understand what regulatory elements facilitate the cleavage and polyadenylation (CPA) of the Dux4 mRNA beyond the non-consensus PAS and to determine if inefficient CPA underlies the poor expression of Dux4 in patient cells. We

designed a transcriptional read-through reporter to assay cleavage and polyadenylation in cells and confirm that additional *cis* elements are required for CPA of Dux4 besides the non-consensus PAS. This element is located outside the region where *cis* regulatory elements for CPA are usually present. Moreover, the element which lies downstream of the PAS, is within a degenerate repeat region, called β -satellite DNA. Using the knowledge gained from characterizing Dux4 mRNA 3'end formation, we designed antisense oligonucleotides (ASOs) to impair the production of polyadenylated Dux4. Prior to antagonizing Dux4 CPA, we demonstrate, in proof of principle experiments that ASOs directed toward required CPA regulatory elements can impair gene expression, and may redirect polyadenylation. Finally, the work presented here lays the foundation for us to impair Dux4 CPA in reporter driven assays and patient cells; and to exploit currently available deep sequencing technology to determine the specificity of PAS-directed ASOs.

Table of Contents

Dedication	iii
Acknowledgments	iv
INTERROGATING DUX4 MRNA 3'END FORMATION	vii
Chapter 1 : Eukaryotic RNA Processing	1
What is RNA Processing?.....	2
Nuclear 3' end RNA Processing.....	2
The 3' end formation of non-coding RNA: snRNA 3'end formation is not an integrated tail	4
The 3' end formation of non-coding RNA: telomerase RNA 3'end formation, means to an end.....	8
The 3' end formation of non-coding RNA: A NEAT trick	14
The 3' end formation of Coding RNA: To cleave and polyadenylate	17
The 3' end formation of Coding RNA : To cleave and not polyadenylate	21
Significance: All's well that ends well.	22
Chapter 2 : Introduction to Facioscapulohumeral Dystrophy	24
Overview of Facioscapulohumeral Dystrophy	25
Clinical Features	25
Molecular Features	26
Model System for FSHD Investigation.....	30
Dux4: Leading Candidate Gene of FSHD	31

RNA Processing in FSHD	33
The Dux4-regulated transcriptome	38
Cell Physiology of FSHD and Correlation to Dux4	41
Current State	46
Chapter 3 : Materials and Methods.....	47
Cloning	48
Cell Culture.....	68
Stable Line Development	70
Transfections	71
Small Interfering RNA (siRNA) transfections	71
Antisense Oligonucleotide (ASO) transfections	74
Plasmid transfections	81
Microscopy	81
Cell Fixation and Fluorescence Microscopy.	81
Western Blotting and Immunoprecipitation.....	82
General Cell Lysis	82
Immunoprecipitation of Dux4.....	82
Electrophoresis and Immunoblotting	83
Antibodies	83
Plate Reader Assays	84
Luciferase Assay.....	84

Fluorescence Plate Reader Assay	84
Nucleic Acid Preparation Analysis and Methodology:	85
Total RNA Extraction.....	85
Complementary DNA (cDNA) synthesis	85
Rapid Amplification of Complementary DNA Ends (RACE)	86
RT-PCR	86
qPCR	87
Genomic DNA Extraction	95
Chapter 4 : Development of Tools to Assay mRNA 3'end processing in cells.....	96
Development of Tools to Assay mRNA 3' end processing in cells.....	97
Introduction	97
Gain of Function Reporter	97
Loss of Function Reporter	106
Chapter 5 : Exploring the 3' end processing of Dux4 mRNA.....	111
FSHD, a disease of the dysregulated 3' end processing?	112
β-Satellite Sequences Contribute to Dux4 mRNA 3' End Formation.	115
Downstream Auxiliary Elements aids in cleavage and polyadenylation of Dux4.....	124
Evaluating Dux4 3' end processing “natively”	131
Design and Use of ASO targeting CPA.....	136
Short Antisense Oligonucleotide shows primarily nuclear localization.....	138
Determining the utility of ASOs with use a reporter	144

Redirecting CPA	151
Dux4 in 15Abic cells	154
Conclusions	156
Chapter 6 : Conclusions, Perspectives and Future Directions	160
Downstream Auxiliary Sequences aid cleavage and polyadenylation of Dux4PAS	161
Testing a Model: Are the Dux4PAS auxiliary sequences RNA or DNA?.....	161
The significance of non-canonical <i>cis</i> elements for cleavage and polyadenylation and their interactions with cleavage and polyadenylation <i>trans</i> factors	162
Downstream Auxiliary Sequences aid cleavage and polyadenylation of Dux4PAS, but is there more?	165
Antagonizing Dux4 mRNA 3' end formation	166
Significance	169
References.....	170
VITA.....	199

List of Illustrations

Figure 1.1. snRNA 3' end processing in humans and <i>S. cerevisiae</i>	6
Figure 1.2. Yeast and Human Telomerase RNA 3' end processing	12
Figure 1.3. 3' end processing of NEAT lncRNA transcripts	16
Figure 1.4. The 3' end processing of mRNA in yeast and humans, and human replication dependent histone mRNA..	19
Figure 2.1. Schematic of the D4Z4 Repeats and the FSHD Locus.	28
Figure 2.2 Schematic of Dux4 transcripts.	36
Figure 2.3 Model of Dux4 in FSHD: A Metastable Epiallele.....	44
Figure 3.1. Schematic of transfection protocol with co-transfection of siRNA or ASO with Reporter.....	73
Figure 4.1. PAS-GFP reporter analysis using an SV40 polyadenylation signal (PAS).	100
Figure 4.2 Effect of start codon context on polyadenylation signal (PAS)-GFP reporter analysis.	104
Figure 4.3 pCCHam reporter analysis using an SV40 polyadenylation signal (PAS)..	109
Figure 5.1. Schematic of Chromosome 4qA chromosome.	114
Figure 5.2. Identifying the minimal region necessary to suppress GFP.....	117
Figure 5.3. 3' RACE of Dux4 constructs in PAS-GFP Reporter.	120
Figure 5.4. Elements for Dux4 CPA lie downstream of PAS.....	122
Figure 5.5. Deletion series identifies element in the extreme 3' terminus of the Dux4 CPA construct.	126
Figure 5.6. Point mutation analysis in construct $\Delta 40$ compared with $\Delta 60$	128
Figure 5.7. Effects of mutation in Dux4 Downstream Auxiliary Element(DAE).	130
Figure 5.8. Histogram of terminal Exon 3 of pathogenic Dux4 in dual luciferase reporter. .	133

Figure 5.9. Reporter gene expression as a result of Dux4PAS.....	134
Figure 5.10. Fluorescein labelled Antisense Oligonucleotide transfected into immortalized FSHD patient cells.	140
Figure 5.11. Merged Images of cells transfected with low concentration of Fluorescein labelled ASO.	141
Figure 5.12. Microscope images at HEK293T cells transfected with increasing concentrations of fluorescein labelled ASO.	142
Figure 5.13. Western blot of HEK293T cells transfected using with SV40 PAS directed ASOs using a two hit protocol.....	145
Figure 5.14. Western blot of HEK293T cells transfected using with SV40 PAS directed ASOs.	147
Figure 5.15. Ethidium Bromide stained agarose gels of stable GFP positive clonal lines...	149
Figure 5.16. PAS directed ASO redirect PAS or reduce gene expression.	152
Figure 5.17. Detecting Dux4 Protein in Immortalized FSHD patient lines.	155
Figure 5.18. Model of Dux4 Processing.....	157

List of Tables

Table 3.1. Table of Plasmids	50
Table 3.2. Table of DNA Oligonucleotides for Cloning	51
Table 3.3. Table of Cell Lines	69
Table 3.4. Table of siRNAs	72
Table 3.5. Table of Antisense Oligonucleotides	75
Table 3.6. Table of DNA Oligonucleotides for PCR	88

Chapter 1 : Eukaryotic RNA Processing

Whatever begins, also ends?

-Seneca the Younger

What is RNA Processing?

Ribonucleic Acid (RNA) is one the three major macromolecules important for life. In eukaryotes, genetic material is stored in the deoxyribonucleic acid (DNA) and then transmitted as RNA in response to stimuli, to catalyze reactions, regulate gene expression, influence cell structure, and regulate cell behavior, in part through the translation of the RNA into a protein. There are many types of RNA molecules and they serve important roles in the eukaryotic cell. Here, I will provide background on RNA processing by focusing on RNA transcribed by RNA Polymerase II (RNAPII). Specifically, I will focus on the maturation and metabolism of the RNA accomplished through processing of the 3' end in order to remove it from the template DNA and package it for stability in the nucleus and ultimately cytoplasm.

Nuclear 3' end RNA Processing

RNA polymerase II (RNAPII) is a large multi-subunit and tightly regulated enzyme responsible for producing messenger RNA (mRNA), and several non-coding RNA (ncRNA) including some microRNAs (miRNAs), small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs). Post translational modifications which regulate the function of RNAPII occur on the highly conserved C-terminal domain (CTD) and take the form of various phosphorylation events which facilitate initiation, elongation and termination of transcription (Hampsey, 1998; Jeronimo et al., 2013; Mayfield et al., 2016). For initiation of transcription, RNAPII associates with various factors ranging from the general transcription factors (for initiation of transcription of mRNAs) to specific transcription factors such as small nuclear RNA activating protein complex (for transcription initiation of snRNAs) (Chen and Wagner, 2010; Gupta et al., 2016). Following the initiation of transcription and the synthesis of nascent RNA, all RNA transcripts have a 5' triphosphate. Typically, RNAPII transcripts undergo additional modification to protect the RNA, by addition of an inverted N⁷-

methylguanosine (m⁷G) to the 5' terminus of the RNA (Byszewska et al., 2014), which 'caps' the RNA. Subsequent to capping and during the elongation of the RNA transcript, additional processing occurs co-transcriptionally to yield a mature RNA product. For example, the majority of eukaryotic mRNAs are spliced to remove introns. However, some RNA processing events occur post-transcriptionally depending on the type RNA molecule. For example, the m⁷G cap of several RNAPII substrates, such as spliceosomal snRNAs and snoRNAs, undergo additional methylation to form 2,2,7-trimethyl Guanosine (TMG), which occurs cytoplasmically and in a few instances in the nucleus (Jády et al., 2004; Seto et al., 1999; Webb and Zakian, 2008).

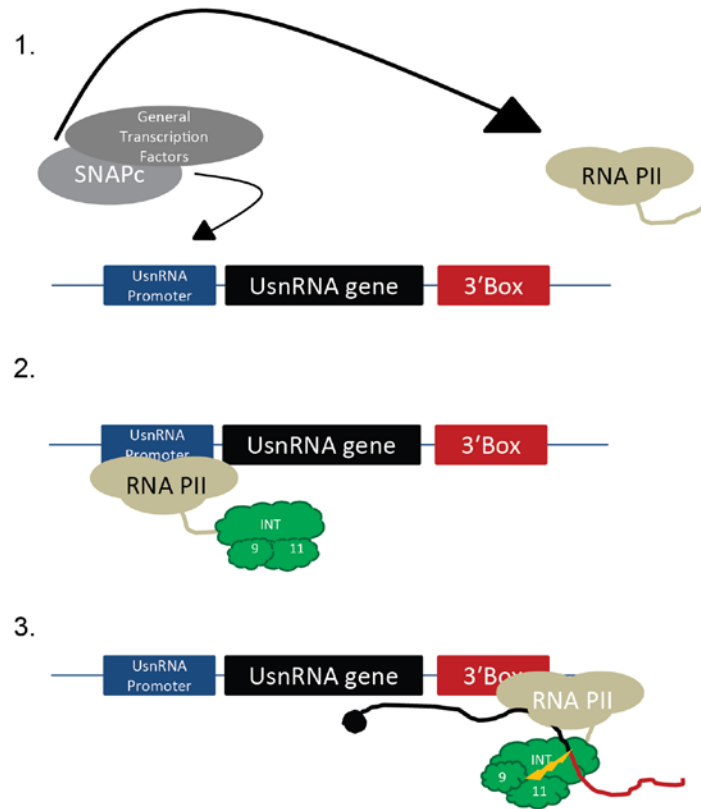
Regardless of their specific downstream function, all RNAPII transcripts must be separated from the polymerase and DNA template and protected to avoid non-regulated degradation. Unsurprisingly, given the diversity of RNAPII transcripts, the process of separating the RNA transcript from the DNA template and RNAPII, as well as protecting the 3' terminus of the RNA transcript (here referred to as 3' end processing) is complex and tightly regulated. The mechanism of 3' end processing can be thought of as a two-step event involving first cleavage of the nascent RNA followed by additional modification to stabilize the processed transcript. The purpose of cleaving RNAPII transcripts is two-fold. First, to release the nascent transcript from the polymerase promoting further modification and second to promote termination of RNAPII through the action of exonucleases (such as Xrn2) on the downstream RNA product (Rosonina et al., 2006). The 3' terminus of the transcripts are further modified to stabilize the transcript through the activity of polyA polymerase (in the case of mRNA) or by exonucleases to generate a stable secondary structure at the 3' end (in the case of histone mRNA and non-coding RNA). The process of 3' end processing is governed by *cis* elements and *trans* acting factors. Moreover, the diversity of RNAPII transcripts means that the regulatory elements of 3' end processing of these transcripts vary tremendously (for reviews see (Legendre and Gautheret, 2003; Peart

et al., 2013; Tian and Graber, 2012; Wilusz and Spector, 2010)) . Below, I provide a brief discussion on the *cis* elements and *trans* factors that govern many of the RNAPII 3' end processing events including coding RNA (both polyadenylated and non-polyadenylated) and non-coding RNA (including snRNA, telomerase RNA, and long non-coding RNA).

The 3' end formation of non-coding RNA: snRNA 3'end formation is not an integrated tail

The majority of the Uridine rich small nuclear RNAs (snRNAs) are transcribed by RNAPII and are ~60-200 nucleotides in length (Peart et al., 2013). These snRNAs are packaged into RNA protein complexes called small nuclear ribonucleoproteins (snRNPs) and they facilitate splicing of pre-mRNAs or promote the 3' end formation of replication dependent histone mRNA. In metazoans, transcription of the snRNAs by RNAPII requires the small nuclear activating protein complex at the promoter (Jawdekar and Henry, 2008; Yoon et al., 1995). This is relevant because the accurate and precise 3' end formation of the RNAPII-transcribed snRNAs is intimately connected to promoter identity as well as the distance between the site of termination and the promoter (Hernandez and Weiner, 1986; Ramamurthy et al., 1996). The 3' end formation of the snRNAs is dependent on recognition of, not only the correct promoter, but also the 3'box downstream of the gene body and specific phosphorylation of the CTD of RNA PII (Figure 1.1A). Together the *cis* regulatory elements allow for the recruitment of a protein complex (Chen and Wagner, 2010; Peart et al., 2013), which is responsible for cleaving the snRNA.

A



B

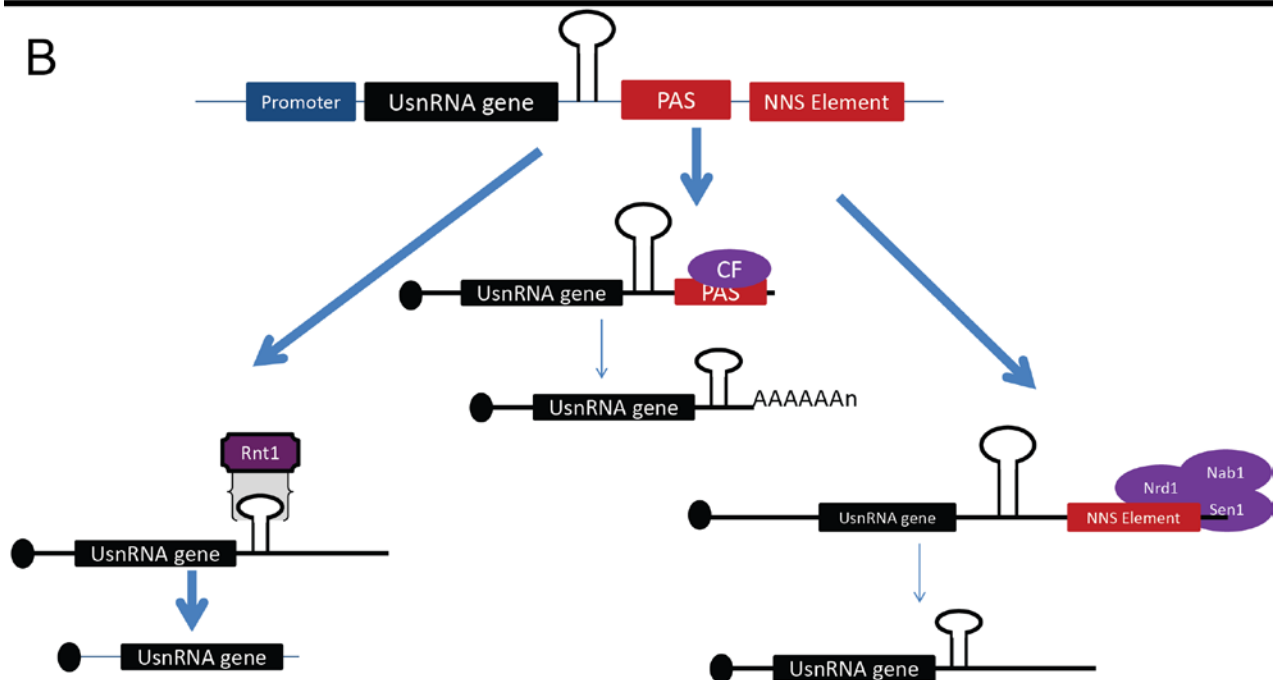


Figure 1.1. snRNA 3' end processing in humans and *S. cerevisiae*. (A). The 3' end formation of the RNAPII transcribed snRNA in humans requires a snRNA specific promoter and conserved downstream element, the 3' box located proximal to the end of the mature snRNA. The snRNA promoter is recognized by the small nuclear activating protein complex (SNAPc), and with the presence of the general transcription proteins recruits RNAPII. Along with the RNAPII, a multi-subunit complex, Integrator, located on the CTD of the polymerase is brought in close proximity to the nascent RNA. The snRNA is cleaved by a heterodimer of subunits 9 and 11 of the Integrator complex. (B). *S. cerevisiae* does not have integrator proteins and instead the snRNAs are either polyadenylated using the cleavage and polyadenylation factors to cleave the RNA and add a poly (A) tail. Another pathway, involves endonucleolytic cleavage of a stem loop structure in the pre-snRNA by the RNase, Rnt1. Subsequently, the snRNA can be further trimmed to generate the final 3' terminus. Alternatively, the snRNA 3' end formation uses the Nrd1/Nab3/Sen1 termination pathway to cause the release of the nascent RNA from the polymerase.

In metazoans, snRNA 3' end formation is accomplished by a large multi-protein complex, termed the Integrator complex (Baillat et al., 2005). The Integrator complex is comprised of ~14 subunits, which play different but as yet unresolved roles in snRNA transcription and termination (Baillat et al., 2005; Chen and Wagner, 2010; Chen et al., 2012). The subunits responsible for cleaving the snRNA are Integrator subunit 9 and Integrator subunit 11, and following the cleavage by these subunits the snRNA is released from the DNA template (Figure 1.1A). Several of the Integrator subunits display reciprocal dependency, in that depletion of one subunit leads to the depletion of another (Albrecht and Wagner, 2012), however the functional relevance of this observation for snRNA 3' end formation is not straightforward. The RNA binding partner within the complex is unknown resulting in a black box for how Integrator positions the RNA for cleavage; and while it is known that the Integrator associates with the CTD of RNAPII, during snRNA transcription, the sequence of events leading to the recruitment of the Integrator complex to facilitate cleavage of the snRNA remains nebulous (Baillat and Wagner, 2015). However, the identity of the protein complex, as well as the *cis* requirements for snRNA 3' end formation is not conserved between metazoans and fungi (Peart et al., 2013).

In certain fungi, such as *Saccharomyces cerevisiae*, the cleavage and subsequent 3' end of the Uridine rich snRNA is accomplished through several distinct processes, which may act as a failsafe to ensure the production of the snRNA (reviewed by (Peart et al., 2013)). The snRNAs in *S. cerevisiae* are processed at the 3' terminus utilizing endonucleolytic cleavage, exonucleolytic trimming and tailing. One pathway, uses endonucleolytic cleavage of *S. cerevisiae* snRNA by Rnt1, a double strand specific RNase III, which cuts the stem loop of pre-snRNA to release a transcript with an unprotected 3' OH. This pre-snRNA is then trimmed by the exonuclease, like the RNA Exosome or Rex1. Alternate pathways use transcription termination by means of the Nrd1/Nab3/Sen1 pathway, or likely involve the cleavage and polyadenylation factor complex (Peart et al., 2013).

U2snRNA in yeast can be polyadenylated utilizing the components of the cleavage and polyadenylation machinery, although some of the factors were dispensable for processing (Abou Elela and Ares, 1998; Morlando et al., 2002). Still, snRNAs in *S. cerevisiae* also use the polyadenylation independent Nrd1/Nab3/Sen1 pathway, which has some functional interconnections with the cleavage and polyadenylation machinery (Porrúa and Libri, 2015). Nrd1 and Nab3 are RNA binding proteins, and Sen1 is a helicase; the three proteins can associate with a phosphorylated RNAPII CTD, and play an important role in termination of RNA PII transcription (Arndt and Reines, 2015). Thus, utilization of this pathway effectively couples transcription termination with RNA 3' end formation. Despite the different complexes required for 3' end formation of snRNA in fungi and vertebrates, the subsequent fate of the snRNAs is similar.

The 3' end formation of non-coding RNA: telomerase RNA 3' end formation, means to an end.

The telomerase RNA is transcribed by RNAPII in vertebrates and fungi and forms the RNA component of the telomerase enzyme, which maintains the length of the telomere. The sequence of the telomerase RNA varies, and only few of the critical secondary structure elements are highly conserved (Rubtsova et al., 2012). Variability in the conservation of several elements, reflects the diversity in how the RNA is processed despite the end result from all species being the formation of the ribonucleoprotein telomerase. Like the snRNAs, the telomerase RNA also harbors a hypermethylated m⁷G in *Homo sapiens*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* (Jády et al., 2004; Seto et al., 1999; Webb and Zakian, 2008). While significant inroads have been made in identifying and characterizing the 3' end formation of several RNAPII transcripts, most notably mRNA which will be discussed later and snRNAs, this compendium of knowledge is less generalized and

congruent for telomerase RNA. The insights that have been gathered over the past few years for the formation of 3' end formation of telomerase RNA (hereafter referred to as TER for simplicity) show great diversity between fungi and vertebrates (Rubtsova et al., 2012).

In *H. sapiens*, and presumably other vertebrates the formation of the 3' terminus of TER is dependent on a highly conserved domains, CR7 and box H/ACA (Figure 1.2A) (Mitchell et al., 1999; Theimer et al., 2007). This domain combination is a unique feature to vertebrate TER (Chen and Greider, 2004). The H/ACA motif is a feature of a class of snoRNAs and the 3' end processing is likely analogous (Balakin et al., 1996; Mitchell et al., 1999), however the precise mechanism is yet to be delineated. While heterologous expression of the human TER in *S. cerevisiae* indicates that its 3' end processing is dependent on several protein factors that mediate processing of yeast H/ACA snRNAs (Dez et al., 2001), the additional requirement of the CR7 domain for 3' processing (Fu and Collins, 2003) of the human TER suggests that novel factors are in play.

In yeast, the 3' terminus of TER is defined after the Sm binding site (Figure 1.2A) (Gunisova et al., 2009). However, the different yeast species process the 3' terminus differently, two case examples, *S. pombe* and *S. cerevisiae* using will be presented.

In *S. cerevisiae* there are also two forms of TER, a polyadenylated and non-polyadenylated form, the former being a minor species. The major non-polyadenylated species ultimately forms the RNA component of the TER. Rather than being processed from the longer polyadenylated species (Chapon et al., 1997), the non-polyadenylated TER appears to be processed independently (Noël et al., 2012). The TER of *S. cerevisiae* has multiple 3' end processing signals, one of which is dependent on the Nrd1/Nab3 termination pathway (Noël et al., 2012). The other appears to utilize the canonical pathway for cleaving and polyadenylation. Polyadenylated TER production is dependent on the Cleavage Factor and PolyA polymerase (Chapon et al., 1997), which are essential factors for cleavage and polyadenylation of mRNA in budding yeast. This observation has created a model where the

polyadenylated form of TER appears to be generated from transcriptional read-through, but what its functional role is or whether it is further processed is unclear. To create the non-polyadenylated form of TER, it is thought that the Nrd1/Nab3 protein complex binds to a terminator sequence in TER and facilitates termination of TER. An intriguing idea posited suggests that presence of non-polyadenylated and polyadenylated forms of TER in several yeast species (although the 3' end formation process is accomplished through different means) is a kind of precautionary measure to ensure the production of the required RNA (Noël et al., 2012). Interestingly, this Nrd1/Nab3 dependence for TER 3' end processing did not show a dependency on the Sen1 helicase. This may be analogous to some of the functional redundancy of snRNA 3' end processing in *S. cerevisiae* (Peart et al., 2013). In both humans and *S. cerevisiae* the precise exonuclease that trims the mature TER or the endonuclease which cleaves the TER from the DNA template after the cleavage site is demarcated by the interacting protein is unclear.

Two models for the formation of mature TER are possible. Non-polyadenylated TER and polyadenylated TER in both humans and *S. cerevisiae* may be differentially processed, utilizing different endonucleases corresponding to the *cis* elements present. Alternatively, the polyadenylated forms of TER may also be a precursor for the mature TER (Chapon et al., 1997). It has led to the speculation that nuclear exosome may play a role in this process, where, in *S. cerevisiae* the Sm site in the TER likely bound by Sm proteins marks the termination site to which the exosome trims the polyadenylated precursor (Figure 1.2B) (Coy et al., 2013).

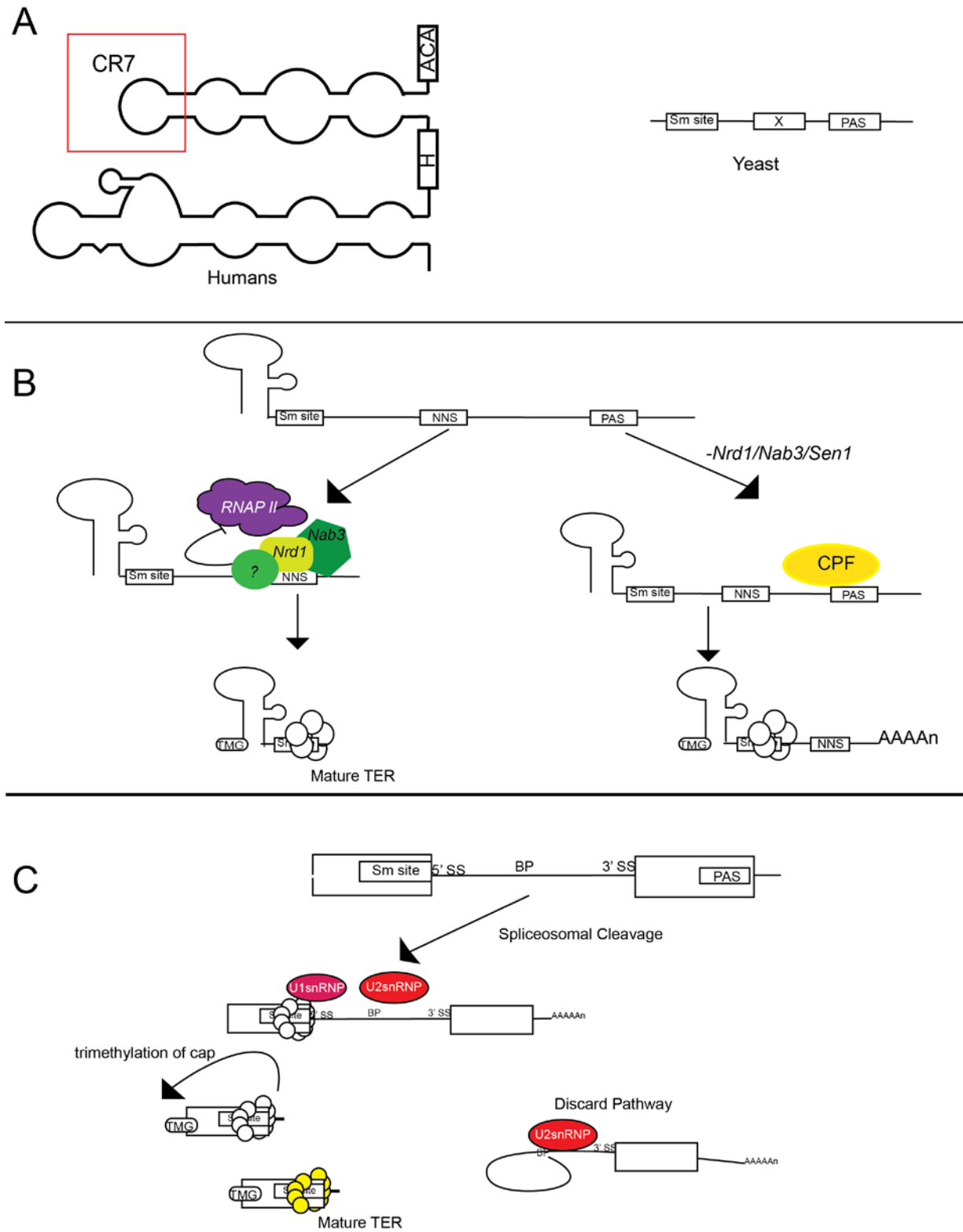


Figure 1.2. Yeast and Human Telomerase RNA 3' end processing. (A). Signals required for Telomerase 3' end formation in humans and yeast. Efficient TER 3' end processing in humans requires a CR7 loop (indicated by red box) and H/ACA box, the human TER is not polyadenylated. Yeast TER may be polyadenylated, however, mature TER incorporated into the telomerase do not necessarily require poly(A) tails. The 3' terminus of yeast TER share a Sm site (posited to act as a boundary element), X indicates additional cis elements used by different yeast genus to generate mature TER. (B). 3' end formation of TER in *S. cerevisiae* uses two distinct pathways to generate mature TER. *S. cerevisiae* uses Nrd1/Nab3 termination complex to terminate transcription and release the nascent TER. The Sm protein bound TER is protected from exonucleolytic degradation. Alternatively, the Cleavage and Polyadenylation (CPF) proteins are recruited to cleave the nascent TER which is subsequently polyadenylated. C. The TER in yeast of the genus *Schizosaccharomyces* contain a suboptimal intron. An incomplete splicing reaction, dependent modulated by stringency of snRNP binding to the splice regulators, 5' or 3' splice site (SS) and branch point (BP). The Sm binding site, bound by Sm proteins, is also required for the generation of mature TER. The first transesterification reaction of splicing is completed, by exon ligation is suppressed and the 3' exon and lariat are discarded. The Sm proteins (white circles) promote the trimethylation of the m7G cap of the TER, and then are replaced by the Lsm proteins (yellow circles).

Like in *S. cerevisiae*, the mature TER of *S. pombe* that forms a subunit of telomerase is the non-polyadenylated form, however in *S. pombe*, the 3' end formation of TER uses the spliceosome, to cleave the mature TER from the polyadenylated precursor (Figure 1.2C) (Box et al., 2008). The spliceosome is a large ribonucleoprotein complex that typically removes introns from pre-mRNAs, by catalyzing transesterification reactions, which excise an intron, and ligate the flanking exons (Matera and Wang, 2014; Smith et al., 2008). An analysis of the polyadenylated TER in *S. pombe* revealed that a small portion had an internal deletion of 56 nucleotides, closer analysis of this deletion revealed that the TER contained an intron. Intriguingly, removal of this intron or substituting a heterologous efficiently spliced intron resulted in a decrease in the mature form TER (Box et al., 2008). The 5' splice site is recognized by the U1snRNP (Box et al., 2008) which is the first step of canonical splicing reactions in eukaryotes (Smith et al., 2008). However, in contrast to the canonical splicing reactions with two transesterification, completion of only the first transesterification reaction is required for generation of mature TER (Box et al., 2008). The 5' splice site in TER overlaps with the Sm binding site and it was subsequently demonstrated that Sm protein binding is critical for spliceosomal cleavage (Tang et al., 2012) and thus Sm protein binding may impair the second transesterification reaction. Subsequent to the spliceosomal cleavage, the Sm proteins are replaced by Lsm proteins which stabilize the mature TER (Tang et al., 2012). Spliceosomal cleavage is conserved in yeast of the genus *Schizosaccharomyces* (Kannan et al., 2015) and possibly in other yeast as splice sites were detected in TER genes in several yeast of the genus *Candida* (Gunisova et al., 2009). A significant contribution to impairment of the complete splicing reaction in the TER is a distortion of the kinetics of splicing due to the suboptimality of the intron ((Kannan et al., 2013, 2015) and reviewed (Peart et al., 2013)).

The 3' end formation of non-coding RNA: A NEAT trick

The nuclear enriched abundant transcripts, NEAT1 and NEAT2 are mammalian conserved polyadenylated long non-coding RNA (lncRNA) (Hutchinson et al., 2007). NEAT1 is a structural lncRNA essential for paraspeckles in the nucleus of mammalian cells (Sunwoo et al., 2009). Paraspeckles are subnuclear compartments suggested to control gene expression by retaining certain RNA molecules in the nucleus. They may also serve as marker for loss of pluripotency (reviewed by (Bond and Fox, 2009; Yamazaki and Hirose, 2015)). NEAT2, also known as MALAT1 for metastasis-associated lung adenocarcinoma transcript 1, is a long lived lncRNA that is frequently associated with several cancers, and plays inconclusive roles in modulating mRNA splicing and influencing gene activation (Gutschner et al., 2013). Both NEAT1 and NEAT2 are polyadenylated, however, they also undergo additional processing to generate a triple helical structure at the 3' terminus which is essential for stabilization ((Brown et al., 2012; Wilusz et al., 2012). NEAT1 is transcribed as two isoforms Men ϵ (a polyadenylated ~3.2kb lncRNA) and Men β (a non-polyadenylated ~23kb lncRNA), while NEAT2 is processed to generate the nuclear retained MALAT1 and a cytoplasmic short-lived mascRNA (Wilusz et al., 2008, 2012). Men β and Malat1 do not have canonical polyadenosine tails, instead they have an encoded run of adenosines at the 3' ends (cleavage of the RNA occurs after the run of encoded adenosines). The polyadenosine tails at the 3' ends are both short, however, the transcripts are surprising stable (Sunwoo et al., 2009; Wilusz et al., 2008). The presence of adenosine stretch in the template DNA and the lack of signals for cleavage and polyadenylation near the 3' ends revealed a novel method of processing the 3' ends of an RNAPII transcript. Illumination of this process was possible due to presence of small non-coding RNA, mascRNA. The mascRNA structurally resembles a transfer RNA as it is predicted to adopt a cloverleaf fold and has a CCA modification at the 3' end. The 5' end of mascRNA corresponds to the 3' end of the Malat1

lncRNA, and production of mascRNA was dependent on Malat1 production. This led to the speculation, which was proven, that the 3' ends of the Malat1 lncRNA is processed by endoribonuclease, RNase P (Wilusz et al., 2008). RNase P cleaves Malat1 and Menε/β downstream after the encoded adenosine stretch to generate an lncRNA with a short tail, upstream of the tRNA like fold. The downstream product of the cleavage in the case of mascRNA is further processed at its 3' terminus by RNase Z, prior to the addition of the CCA (Figure 1.3). It remains unclear if the process occurs co-transcriptionally or post transcriptionally (that is, if the process occurs prior to a canonical cleavage and polyadenylation event or on an already cleaved and polyadenylated transcript) (Sunwoo et al., 2009; Wilusz et al., 2008).

The stability of Menε/β and Malat1 given their short adenosine tails was surprising. Located near the 3' end of the Menε/β and Malat1 RNA there is a conserved A and U rich tract similar to an element called the expression and nuclear retention element (ENE) found in the stable nuclear retained polyadenylated RNAs from Kaposi's sarcoma-associated herpesvirus (Brown et al., 2012). It was further demonstrated by two independent groups that the RNA adopts a triple helical fold due to the presence of the U/A rich motif (Brown et al., 2012; Wilusz et al., 2012).

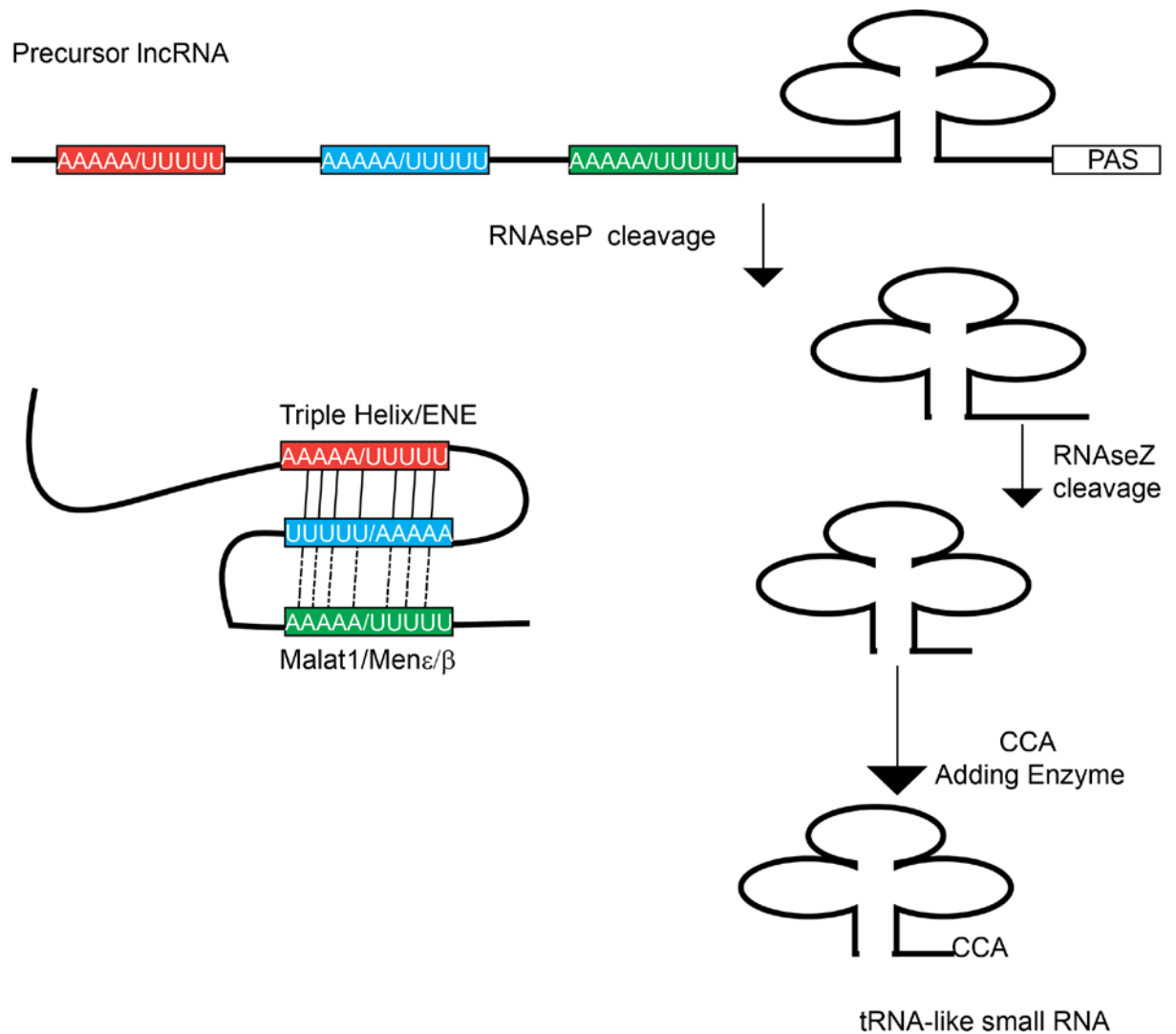


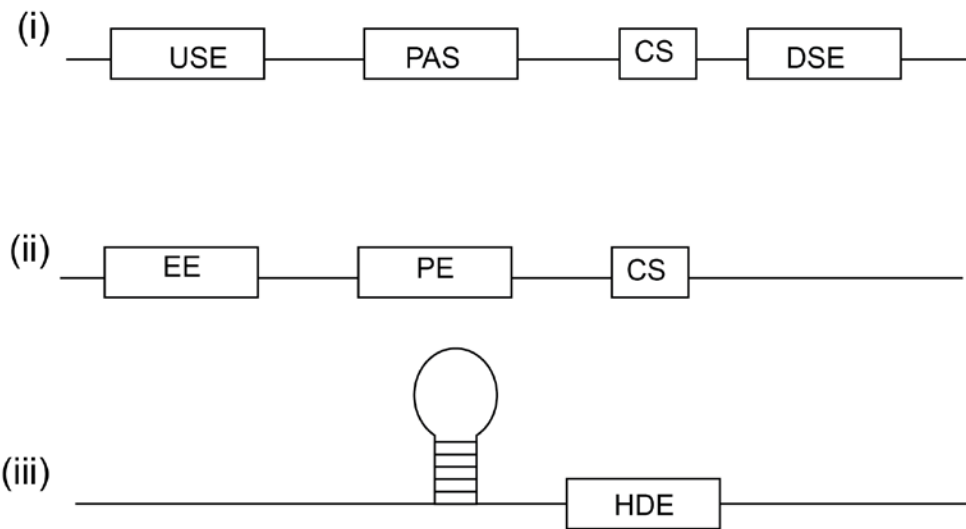
Figure 1.3. 3' end processing of NEAT lncRNA transcripts. The precursor NEAT lncRNA may be polyadenylated. The mature NEAT is generated following cleavage by the endonuclease RNase P between an A/U rich region and a clover like structure. Cleavage generates a small tRNA-like RNA, which is further processed at its 3' end by RNase Z, and stabilized by the addition of a CCA tail by the CCA adding enzyme. The upstream cleavage product is stabilized due to the formation of triple helix structure in the RNA because of the A/U rich stretches near the 3' terminus. Triple helix bonds are Hoogsteen (broken lines between A/U blocks) and Watson crick (solid lines between A/U blocks).

The 3' end formation of Coding RNA: To cleave and polyadenylate

Coding mRNA can be broken down into two categories: those that contain a long adenylated 3' tail and those that lack this feature. In contrast to the majority of mRNAs that are polyA tailed, the 3' end of mature replication dependent histone mRNA is not polyadenylated, rather polyadenylated histone mRNA is aberrant and thought to be produced by inefficient 3' end formation. Remarkably, the process of accomplishing this task is highly complex and tightly regulated. The core machinery responsible for 3' end formation of the coding RNAPII transcripts are the cleavage and polyadenylation specificity factors (CPSF).

In the simplest instance, the cleavage and polyadenylation (CPA) is dependent on the recognition of several *cis* regulatory elements – a polyadenylation signal (PAS), a cleavage site (CS), an upstream sequence element (USE), and a downstream sequence element (DSE). These *cis* elements are recognized and bound by the CPSF complex, the cleavage stimulation factor complex (CstF) and the cleavage factor complex (CF). The PAS is typically AAUAAA, and the DSE is typically U/GU rich in metazoans (Figure 1.4A). However, this general and simple assumption of *cis* element organization seems to be quite rare in mammals (Tian and Graber, 2012). Many mRNAs do not utilize AAUAAA as the PAS, or possess a readily identifiable DSE, or USE, and a few even lack these elements altogether. To add a further level of complexity, several mRNAs possess multiple PASs (Beaudoing et al., 2000; Tian et al., 2005), so the process of selecting a cleavage site must be and is highly regulated. I discuss some of the implications, and processes here.

A.



B.

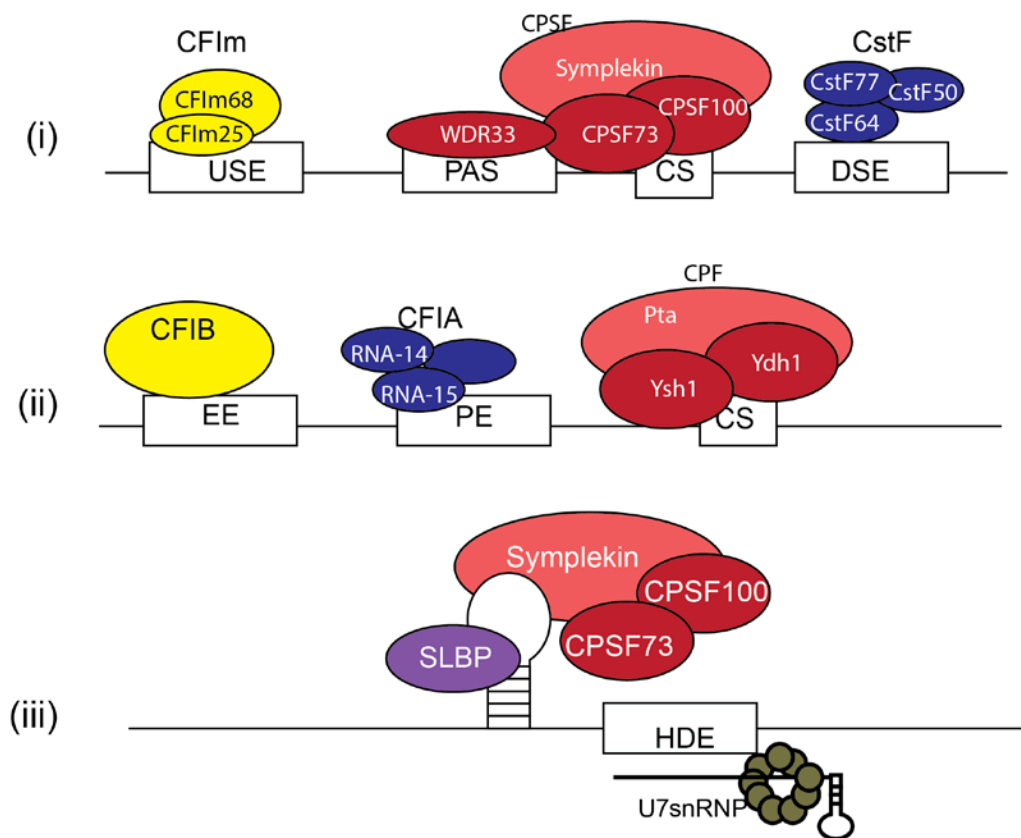


Figure 1.4. The 3' end processing of mRNA in yeast and humans, and human replication dependent histone mRNA. (A) cis regulatory elements for (i) polyadenylated mRNA in humans (ii) polyadenylated mRNA in yeast, and (iii) non-polyadenylated histone mRNA in humans. Canonical elements for polyadenylation of mRNA in humans are an upstream sequence element (USE), a polyadenylation signal (PAS) a cleavage site (CS) and a U-rich downstream sequence element. The elements for CPA in yeast are a U-rich efficiency element (EE), an A rich positioning element (PE) and a cleavage site. The *cis* elements for 3' end formation of the replication dependent histone mRNA in humans are structural and sequence specific. First is the requirement for a stem loop and second is the requirement for histone downstream element (HDE). (B) The *cis* elements of mRNA 3' end formation are recognized by several trans acting factors. Shown are some of the proteins that are shared by all three mRNA processing events. (i) and (iii) human mRNA 3' end processing both use the CPSF proteins 73 and 100 for cleavage supported by the scaffold protein Symplekin. In yeast (ii) the orthologues of these proteins Ysh1, Ydh1 and Pta, respectively, facilitated the cleavage of the mRNA. In yeast the PE is recognized and bound by RNA-15, an orthologue of human CstF64 that binds the DSE in the human pre-mRNA transcript. The PAS in humans is bound by WDR33. (iii). In the replication dependent histone mRNA, the HDE base pairs with the U7 snRNP, and the stem loop binding protein (SLBP) binds the histone stem together recruiting the CPSF proteins.

CPA is reliant on several *cis* elements, key is the hexameric PAS; but how is the hexamer, for example AAUAAA, functionally characterized as a signal for polyadenylation? There is typically an enrichment of uridines in the immediate vicinity of an identifiable hexanucleotide PAS (Legendre and Gautheret, 2003). In addition, the site of cleavage and subsequent tailing is biased toward CA or UA dinucleotide (Derti et al., 2012; Li and Du, 2013; Sheets et al., 1990). The complex of proteins regulating CPA was initially purified in the late 20th century through a series of biochemical purifications and additional proteins have since been shown to play a role in this process (reviewed and references within (Shi and Manley, 2015; Xiang et al., 2014)). The Uridine stretches in the RNA are bound by the CF and CstF complexes, while the hexamer PAS, in the case of AAUAAA, is bound by the CPSF protein WDR33 (Figure 1.3B) (Chan et al., 2014; Schönmeyer et al., 2014). The mRNA is then cleaved by the endonuclease CPSF73 (Dominski, 2010; Mandel et al., 2006), which forms a stable core with CPSF100 and Symplekin ((Xiang et al., 2014). Although there are differences in the stringency of requirement for certain elements or factors depending on the context of the gene expression, these factors work together to facilitate efficient cleavage and polyadenylation the mRNA.

The *trans* acting factors required for cleavage and polyadenylation are highly conserved across fungi and metazoans (Tian and Graber, 2012; Xiang et al., 2014; Yang and Doublé, 2011). In *S. cerevisiae* the *cis* regulatory elements which mediate cleavage and polyadenylation are less defined, but no less critical for CPA. The elements required and sufficient for CPA in *S. cerevisiae* are an efficiency element, a positioning element, a cleavage site, and an enrichment of uridines around the cleavage site (Figure 1.2A) (Guo and Sherman, 1996; Tian and Graber, 2012). While several of the protein machinery are similar, there are differences in the binding specificities which likely reflect the differences in how sites of polyadenylation are determined in metazoans compared to fungi. For example, the A-rich positioning element in yeast which is analogous to the mammalian

polyadenylation signal is recognized by a homologue of the human CstF protein 64. In humans, CstF64 is found downstream of the site of polyadenylation at the U/GU rich DSE (Figure 1.2B). However, both human and yeast *cis* elements for CPA are enriched for uridines (reviewed in (Tian and Graber, 2012; Xiang et al., 2014)).

The process known as alternative cleavage and polyadenylation (APA) describes how one gene uses two or more different PAS, without necessarily changing the coding region to generate different mRNAs. Regulation of polyadenylation site affects the length of the 3' untranslated region (UTR), thus can affect mRNA behavior. As a consequence of APA, mRNA localization, stability and translatability can be changed (Mayr, 2016). These changes can have drastic effects contributing to disease pathologies (Ogorodnikov et al., 2016), nonetheless, APA is not an aberrant occurrence in the cell. Rather, cell type can affect the APA, for example the mRNAs in the brain tend to have longer UTRs, utilizing more distal PAS compared to other cells (Mayr, 2016). Regulation of APA utilizes multiple *trans* acting factors, many of which are part of the core CPA machinery. The mechanisms and signaling cascades that affect the choice of polyadenylation site are still being unveiled, which contributes to understanding of general gene expression regulation, but also introduces new therapeutic targets (Klerk et al., 2012; Masamha et al., 2014, 2016).

The 3' end formation of Coding RNA : To cleave and not polyadenylate

Similar to the other mRNAs in metazoans, the replication dependent histone mRNA depends on the CPSF machinery to remove the nascent RNA from the DNA template. However, of this large multiprotein complex, only a subset of proteins is required. Specifically, the cleavage factor including CPSF73 and 100 and the scaffold Symplekin. As mentioned previously, in higher metazoans such as humans, the histone mRNA is not polyadenylated; in its place, a stable stem loop at the 3' terminus protects the transcript from

decay. In addition to CPSF73 and CPSF100, which cleave the transcript, histone mRNA 3' end formation also requires the U7 snRNP and stem loop binding protein (SLBP). The U7 snRNP and SLBP recognize the *cis* regulatory elements of the histone gene including a histone downstream element (HDE) and stem loop respectively (Figure 1.3A/B) (Dominski and Marzluff, 2007; Marzluff et al., 2008). Additional *trans* acting factor facilitate the recruitment of the CPSF machinery and distinguishes histone mRNA 3' end formation from the canonical mRNA cleavage and polyadenylation (reviewed (Dominski and Marzluff, 2007; Köhn and Hüttelmaier, 2016; Marzluff et al., 2008; Romeo and Schümperli, 2016)).

Significance: All's well that ends well.

The nuclear processing of the 3' end of RNAPII transcripts affects their localization, (whether they are exported, retained or decayed). The 3' end of the RNAPII transcript is subject to regulation to control transcript fate, affecting the behavior and stability. For example, polyadenylated mRNAs are exported to the cytoplasm and once there, removal or shortening of this protective features at the 3' terminus subjects the mRNA to decay, impairs translation of the mRNA into protein, or can lead to sequestration of the mRNA (Weill et al., 2012).

As there are many distinct types of RNAPII transcripts, regulation of the process by which the 3' ends of the transcripts are formed, is critical because it dictates the behavior of the RNA. The redundancies within the pathway, the fail-safes and stopgaps present, demonstrate that appropriate 3' end formation of the transcript is critical for cell homeostasis. The implications of dysregulated RNA 3' end processing is evident in several diseases, dyskeratosis congenita, muscular dystrophy, cancer amongst others (Chen and Greider, 2004; Danckwardt et al., 2008; Ogorodnikov et al., 2016). However, not only is the

understanding of the 3' end formation of RNA important for understanding disease etiology and designing strategies for intervention, but also to contribute to the body of science which governs how we understand how we work. The 3' end formation of RNA can serve as a signal for the termination of transcription for RNAPII, which is important given the myriad of RNAPII transcripts so that the genome of increasingly complex organisms can be partitioned (Kuehner et al., 2011).

Chapter 2 : Introduction to Facioscapulohumeral Dystrophy

Overview of Facioscapulohumeral Dystrophy

Facioscapulohumeral dystrophy (FSHD) is a progressive debilitating muscle disorder. The prevalence of FSHD varies according to the population surveyed; however, since the use of genetic testing the average prevalence is approximately 6.4 per 100000 persons, with the highest incidence and prevalence to date reported in the Dutch population (Deenen et al., 2014). FSHD was first characterized in late 1800s but the molecular mechanism was not realized until the 1990s when the disease was associated with the reduction in D4Z4 microsatellite repeats on chromosome 4 below a threshold number of 10 (healthy individuals contain 10-100 D4Z4 repeats) (van Deutekom et al., 1993; Wijmenga et al., 1992). However, it was not until the last decade that significant inroads have been made in delineating the molecular etiology of the disease (Lemmers et al., 2010a, 2012; Snider et al., 2010).

Clinical Features

The clinical features of FSHD are highly heterogeneous, and range from mild to severe. Severe FSHD may lead to wheelchair dependency, and may involve mental retardation (Tawil and Van Der Maarel, 2006). However, more often individuals with FSHD may present with hearing loss and retinal telangiectasia, the latter occasionally progressing to Coats Disease (Tawil and Van Der Maarel, 2006). Disease severity has been correlated with age, where severe cases of FSHD referred to as infantile onset FSHD lead to progressive muscle strength degeneration and other organ functional diminution. The relative severity of the disease is also often associated with the size of the D4Z4 contractions, with smaller fragments, 1-3 repeats, associated with severe cases (Tawil and Van Der Maarel, 2006). Interestingly, one recent study showed that the severe cases of FSHD associated with smaller D4Z4 fragments were more often observed in patients that

had *de novo* FSHD, which arise without a previous family history of the disease (Nikolic et al., 2016). FSHD patients, in most cases display asymmetric muscle weakness, and facial drooping. FSHD is a rare muscular dystrophy that does not present cardiac involvement, although, there are case studies of FSHD patients presenting with epilepsy (Chen et al., 2013b; Funakoshi et al., 1998; Saito et al., 2007). One of the more prevailing features of FSHD is chronic pain and fatigue (further reading for clinical features of FSHD (Mul et al., 2016; Tawil and Van Der Maarel, 2006; Tawil et al., 2015)). As a consequence of the high clinical variability and the occasionally subtle phenotypes the incidence and prevalence of FSHD is posited to be underestimated (Deenen et al., 2014).

Molecular Features

FSHD is primarily an autosomal dominant disorder, however there is a high incidence of *de novo* mutations engendering disease (Sacconi et al., 2015; Tawil et al., 2014). FSHD is mainly associated with contractions of the D4Z4 repeats on subtelomeric region of chromosome 4 (van Deutekom et al., 1993; Wijmenga et al., 1992). In spite of the similarities of the q arm of chromosomes 4 and 10 (over 98%) (van Geel et al., 2002), early characterization of the FSHD exclusively linked the disease to chromosome 4q, in particular the 4qA allele (de Greef et al., 2009; Lemmers et al., 2004, 2007; van Overveld et al., 2003; Wang et al., 2011). A single D4Z4 repeat is 3.2 kb in length and is thought to contain a single promoter and open reading frame encoding Dux4 (see below) but the lack of a cleavage and polyadenylation signal within the repeat prevents production of a Dux4 mRNA.

Approximately 95% of the cases of FSHD have contracted D4Z4 alleles (Statland and Tawil, 2014). However, patients have been clinically diagnosed with FSHD and bear no contractions of the D4Z4 repeats. Instead, these patients, classified as FSHD2, have reduced repressive methylation marks on their D4Z4 repeats thereby de-repressing this

genomic region similar to what is observed to occur as a consequence of D4Z4 contraction in FSHD1 (van der Maarel et al., 2012; Statland and Tawil, 2014). Concomitant with the repression of the D4Z4 repeats, FSHD is predicated by single nucleotide polymorphisms (SNP) present within the region flanking the D4Z4 repeat on telomeric side that is specific to the 4qA allele (Lemmers et al., 2007, 2010a). The 4qA allele is distinguished, from the 4qB or 4qC allele, because of the presence of a truncated terminal D4Z4 repeat, pLAM, and β -satellite repeat region (Lemmers et al., 2007, 2010b). The 4qA and 4qB alleles are the most predominant within the population, within which the most common haplotypes are A161, A166 B163, and B168 (Lemmers et al., 2010b); however, only the 4qA161 haplotype is associated with FSHD, as it contains a SNP that generate a non-consensus polyadenylation signal (PAS), AUUAAA, in the pLAM region (Lemmers et al., 2010a). The PAS allows for productive transcription of the Dux4 gene from the terminal D4Z4 repeat (Figure 2.1).

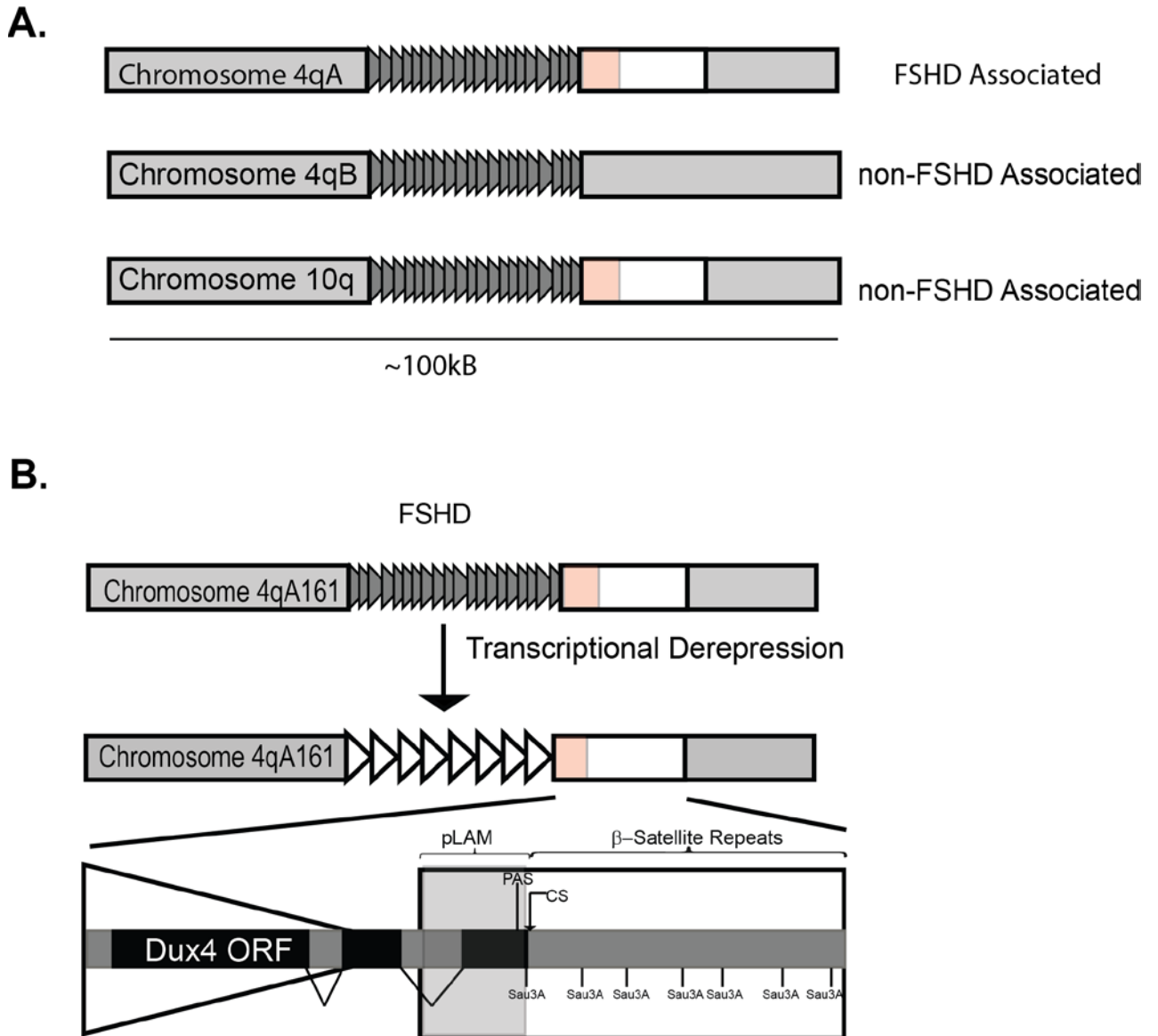


Figure 2.1. Schematic of the D4Z4 Repeats and the FSHD Locus. (A) Chromosome alignment of showing similarity between 4qA, 4qB and 10qA, with only the 4qA allele being associated with FSHD, with 4qB lacking pLAM and β -satellite sequences. D4Z4 repeats represented by gray overlapping triangles representing repressed state, with repeats numbering >10 , pLAM and β -satellite are indicated by red and white boxes respectively. (B) Specific haplotype of FSHD afflicted individual. Transcriptional derepression as a consequence of reduction in number of D4Z4 repeats to 1-10, or mutation in modifier

proteins. The 4qA161 haplotype, contains a SNP generating non-consensus PAS, AUUAAA. The presence of the PAS allows for productive transcription

There is evident, incomplete penetrance of the disease associated with both the SNP generating a non-consensus PAS and the contraction of the D4Z4 repeats (Scionti et al., 2012). For example, there are individuals who display reduction in the number of D4Z4 repeats or hypomethylation of the D4Z4 repeats in the presence of the permissive haplotype who are asymptomatic (Lemmers et al., 2012; Scionti et al., 2012). Moreover, although there is a general correlation between repeat length with disease severity, this is not absolute (Nikolic et al., 2016; Sacconi et al., 2013; Scionti et al., 2012). These observations support the hypothesis that genetic modifiers exist that can affect the D4Z4 locus in *trans*. An example of such a modifier is the gene SMCHD1 (structural maintenance of chromosomes flexible hinge domain containing 1). Individuals with FSHD2 can have the epigenetic state of their D4Z4 repeat maintained in the hypomethylated state due to haploinsufficiency of SMCHD1 (Lemmers et al., 2012). Recently, heterozygous mutations in DNA methyltransferase 3B (DNMT3B) were reported to also derepress the D4Z4 repeats; and are posited to contribute to FSHD2 in the absence of SMCHD1 mutations (van den Boogaard et al., 2016). Intriguingly, the disease displays a strong reliance of the epigenetic state of the D4Z4 repeats, with one group reporting that mutations in the SMCHD1 'modified' the FSHD disease severity when coupled with contracted D4Z4 repeats (Sacconi et al., 2013).

Model System for FSHD Investigation

Currently there is no established animal model for FSHD, although, several animals have been generated to study aspects of FSHD (Jones et al., 2016; Lek et al., 2015). The majority of the shortcomings of the various animal models may be related to the degree of divergence seen between D4Z4 repeats in ape primate and other mammals. Indeed, transgenic mice animals containing the D4Z4 repeats fail to completely recapitulate the disease (Krom et al., 2013). Several candidate genes located in the vicinity of the D4Z4

repeat region have been used to generate animal models of FSHD (Dandapat et al., 2014, 2016; Gabellini et al., 2006; Jones et al., 2016; Mitsuhashi et al., 2012). For example, one of the longstanding candidate genes of FSHD, FRG1, has been used to generate mice that presents some of the features of the disease (Gabellini et al., 2006), however, this required high overexpression of the gene. The level of FRG1 overexpression in FSHD patients is correlatively but not definitively related to the D4Z4 repeat length, and thus the phenotype in the mouse may not be directly related to FSHD. In addition the reproducibility of FRG1 expression in biomarker assays for FSHD is low (Osborne et al., 2007; Rahimov et al., 2012). The variability, sensitivity and, or subtlety of phenotypes in animal models generated for FSHD, underscores the complexity of the disease. The lineage specificity of the D4Z4 repeats, the subtelomeric location of the FSHD locus, the variable expression of several candidate genes (gene present within the D4Z4 repeats – Dux4, or adjacent to the D4Z4 repeats – ANT1, FRG1, and FRG2) as well as the epigenetics all contribute to making the development of an animal model for FSHD technically challenging.

Dux4: Leading Candidate Gene of FSHD

Double Homeobox 4, Dux4, has emerged as the leading candidate gene of FSHD and is the singular, highly conserved open reading frame located in the D4Z4 repeats. Previously, stated to be primate specific, it has been shown that the D4Z4 repeats are found in several placental mammals, however, the topology of the 3.3kb D4Z4 repeat with the Dux4 ORF (with two homeodomains and an activation domain) is nearly perfectly aligned in higher primates, humans, chimpanzees and orangutans (Clapp et al., 2007; Giussani et al., 2012). Despite the high degree of ORF conservation of the Dux4 gene (Clapp et al., 2007), the precise biological role of Dux4 is unclear. Dux4 protein can be detected in muscle biopsies of fetuses with FSHD and the Dux4 mRNA can be detected in the patients with

FSHD but is not typically found in somatic cells of healthy individuals (Ferreboeuf et al., 2014; Snider et al., 2010). Recent studies employing overexpression of Dux4 in mouse derived myoblasts do not replicate the transcriptome changes of Dux4 expression in human cell lines (Sharma et al., 2013); and in FSHD mice models that feature Dux4 expression only ~22% of the genes differentially regulated overlap with human genes (Krom et al., 2013). This is likely due to the Dux4 targets not being present in mice, as Dux4 mainly binds to retroelements and LTR regions (Sharma et al., 2013). Nonetheless, transgenic mice bearing randomly integrated D4Z4 repeats, show a human-like epigenetic topology, in that the locus is typically epigenetically silenced and shortened repeats bear less repressive features, marked with reduced CpG methylation and reduced ratio of H3K9me3:H3K4me2 (Krom et al., 2013). So, while the ORF (in particular the homeodomain) is conserved, suggesting a protein coding function, its presence in the likewise conserved D4Z4 repeats, suggests that the expression of Dux4 is subject to a high degree of epigenetic regulation.

Animal models using Dux4 expression as a driver for pathogenesis have replicated aspects of FSHD pathology but not the full scale of the disease (Lek et al., 2015) (Lek et al., 2015). One mouse model, integrating the Dux4 gene on the X chromosome, showed that the presence of the gene led to increased male lethality and the X chromosome bearing the gene was preferentially silenced in female offspring. Although, there was not observed significant muscle weakness, retinal abnormalities were reported in surviving mice (Dandapat et al., 2014). The retinal abnormalities were also reported in another mouse model (Krom et al., 2013). However, likely due to a dissimilarity in the transcriptome of Dux4 in mice versus humans (Sharma et al., 2013), none of the mouse models fully recapitulate the disease. An alternative approach utilized xenograft models, in which skeletal muscle tissue from FSHD patients is engrafted into mice to reproduce the Dux4 expression profile (Zhang et al., 2014). However, this model is more suited to assaying molecular outcomes due to the presence of the human tissue and localized repair of the xenograft muscle, but,

will not be sufficient for functional studies, such as grip strength which assesses the muscle strength of the mouse. In zebrafish, the ectopic expression of Dux4 led to severe skeletal malformations, which could be rescued by morpholinos reducing Dux4 (Mitsuhashi et al., 2012). In human cell lines it has been shown that Dux4 inhibits myogenesis – differentiation of myoblast to myotubes (Bosnakovski et al., 2008). However, FSHD patients present no difference in the apparent myogenesis and Dux4 expression appears to be enhanced by myogenesis (Block et al., 2013; Tsumagari et al., 2011). Collectively, the efforts to recapitulate aspects of FSHD using Dux4 have been accomplished with mixed success further reflecting the complexity of the disease.

RNA Processing in FSHD

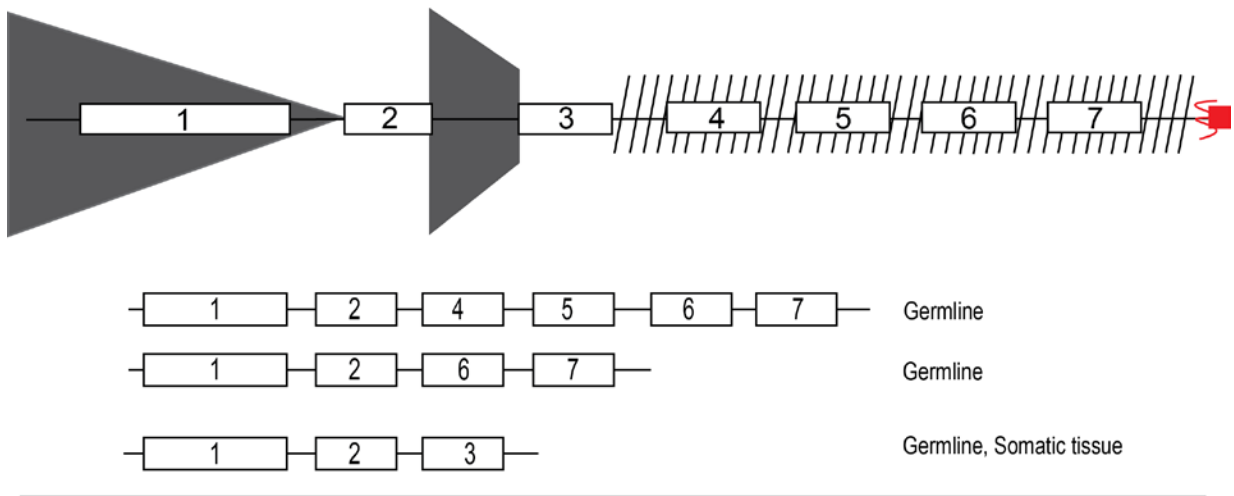
The Dux4 ORF is located entirely in the first exon of the gene, while the 3' untranslated region (UTR) has two introns. In the testis, where Dux4 is found to be expressed in healthy individuals, an alternatively spliced UTR results in usage of a polyadenylation signal in a distant downstream exon, exon 7 (Figure 2.2A). This distal PAS appears to be not active in normal somatic cells and, typically only in the case of FSHD is the Dux4 transcript expressed using a polyadenylation signal AUUAAA found due to a SNP in Exon 3 on chromosome 4, exclusively (Snider et al., 2010). Alternative splicing of Dux4, reportedly generates several isoforms of Dux4 (Figure 2.2 B), including a non-pathogenic Dux4 isoform, called Dux4-s. This splice isoform, is generated from a cryptic splice site in the Dux4 ORF in the first exon, and retains the homeodomain, but does not have the activation domain. The pathogenic Dux4 isoforms, collectively referred to as Dux4-fl retain the entire ORF of Dux4 containing two homeodomains and an activation domain (Snider et al., 2010). There are two annotated RefSeq isoforms of the Dux4 transcript in Genbank: (NM_001306068.1) and variant 2 (NM_001293798.1) that differ due to intron retention in

variant 1 (Figure 2.2 B). The implication of the intron retention will be discussed later in this study, however, it is likely that splicing of the Dux4 mRNA itself is highly regulated. The expression of Dux4 mRNA is low and the full length mRNA is barely detectable (reviewed (Richards et al., 2012)). This could be due to inefficient cleavage and polyadenylation at the non-consensus PAS, rapid mRNA turnover or weak promoter, or due to the epigenetic modifications at the locus. Detection of Dux4 is technically challenging, requiring a high cycle number (typically 50 cycles) nested amplification by PCR.

One possible explanation for the low abundance of Dux4 is that it undergoes splicing downstream of a stop codon. Moreover, the exons are short which lead to the hypothesis that Dux4 mRNA is subject to non-sense mediated decay (NMD) (Feng et al., 2015). Feng and company demonstrated using a mini-gene reporter system that the second intron of Dux4 makes the mRNA susceptible to NMD, and thus results in its decay. It is likely that both variants of Dux4 are NMD targets. The detection of the Dux4-s in patient cells and healthy cells, as well as numerous small transcripts from the D4Z4 repeats of patient cells (Snider et al., 2009) do not support the hypothesis that there is a weak promoter in the D4Z4 repeat. Several studies have shown that there are many different RNA transcripts generated from the D4Z4 repeats (Snider et al., 2009), and while some of these transcripts may aid in increasing overall transcription activity (Cabianca et al., 2012) from the D4Z4 repeats, some have inhibitory effects (Lim et al., 2015). Significant efforts have been devoted to understand the epigenetic regulation of the D4Z4 repeats, as this may contribute to the poor expression of the Dux4. While the locus shows decreased repressive DNA and histone methylation (Hewitt, 2015), it is still unclear if this contributes to Dux4 inefficient expression. Besides, the identification of the non-consensus PAS in the Dux4 transcripts from patient cells (Lemmers et al., 2010a), and a myriad of studies which amplify the polyadenylated Dux4, not much work has been done characterizing the 3' end processing of the Dux4 transcript.

In this thesis work, I explore the 3' processing of Dux4 to determine whether inefficient cleavage and polyadenylation contributes to poor Dux4 expression.

A.



B.

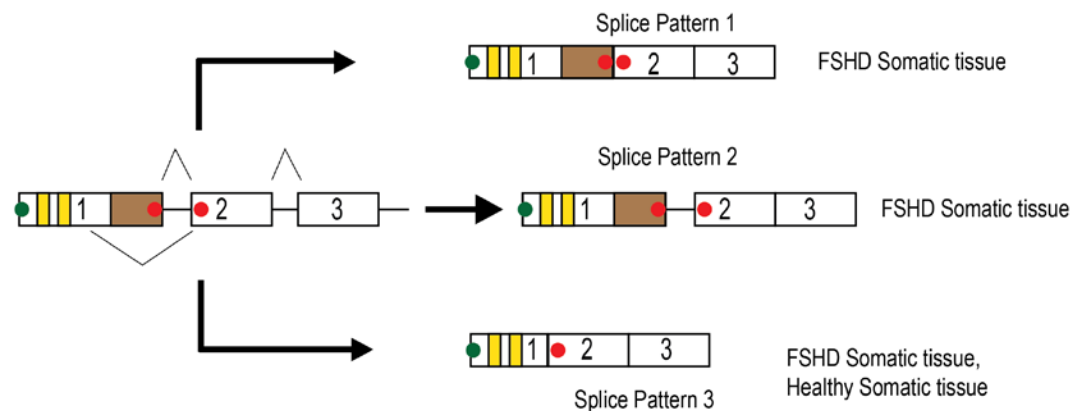


Figure 2.2 Schematic of Dux4 transcripts. (A) Grey triangle represents terminal D4Z4 repeat, grey trapezoid repeats truncated D4Z4 repeat. Exons indicated by white rectangles, hatched diagonal line indicate the degenerate β -satellite repeats and the red square indicates telomeric portion of the chromosome. The Dux4 transcripts are shown indicating the exons detected from germline and somatic cells. Somatic tissue solely produces transcripts with using PAS with exon 3. (B) Dux4 ORF located in exon 1, contains two homeodomains, shown as yellow rectangle, and an activation domain shown as a brown square. Alternative splicing produces three isoforms containing exon 3, Dux4-fl variant 2, is spliced to contain exons 1-3 consecutively (splice pattern 1). Dux4-fl variant 1 retains the first intron (splice pattern 2). Dux4-s, a short isoform produced using a cryptic splice site in the Dux4 ORF,

Dux4-s lacks the activation domain, and can use a stop codon present in exon 2 (splice pattern 3). Start and stop codons are indicated by green and red circles, respectively.

The Dux4-regulated transcriptome

As mentioned, Dux4 contains two homeodomains and an activation domain, and belongs to the family Homeobox proteins, which are typically transcription factors. Consistent with this idea, Dux4 can potently activate transcription at its target genes and preferentially binds and transcriptionally activates multimeric copies of TAATCTAATCA (Choi et al., 2016; Geng et al., 2012; Young et al., 2013; Zhang et al., 2016). The potent transcriptional activation capacity of Dux4 is also evident in its role in a subset of undifferentiated small round cell sarcomas, in which a translocation event fuses the activation domain in the C-terminus of Dux4 to the CIC (human homologue of the *drosophila* Capicua transcriptional repressor) gene and enhances its transcriptional activity; in addition, the fusion gene can transform NIH/3T3 cells (reviewed by (Antonescu, 2014; Haidar et al., 2015)).

Despite this clear activity, the precise biological role of Dux4 is unknown. Overexpression has been shown to be toxic to a variety of cell lines (Block et al., 2013; Bosnakovski et al., 2008; Wallace et al., 2011; Wuebbles et al., 2010) but its presence in germline tissue (Snider et al., 2010) suggests a role in organismal development. Supporting this idea, Dux4 has been shown to activate genes that are typically involved in germline development (Geng et al., 2012). Interestingly, ChIP-Seq data from Dux4 transduced cells shows peaks over the entire body of genes targeted by Dux4, in addition many of the Dux4 binding sites were found within intergenic regions. Further expounding on this observation, it was shown that Dux4 binding was enriched at repetitive elements, such as mammalian apparent long terminal repeat retrotransposons (MaLR) (Geng et al., 2012; Young et al., 2013). The implication of this was that several of these elements became activated as promoters driving the expression of somatically silenced transcripts. While many of the

transcripts detected in Dux4 transduced lines were also found in the testes, the biological role of these changes remains elusive and speculative.

Gene ontology and pathway analyses of Dux4 regulated gene expression changes have been carried out and have shed some light on how Dux4 expression may affect cellular behavior (Bosnakovski et al., 2008; Feng et al., 2015; Geng et al., 2012; Rickard et al., 2015; Young et al., 2013). Using the data generated from genome wide analysis of transcriptome changes, Feng et al reported that there was an increase in the population of mRNA that is predicted to be subject to NMD. The mechanism for this observation is that Dux4 is thought to cause a modest reduction in the NMD regulator UPF1. This report is particularly intriguing because given that Dux4 is regulated by NMD, due to the presence of its second intron, the authors posit a positive feedback loop in which Dux4 may modulate NMD and consequently modulate itself. However, the exact mechanism through which Dux4 regulates NMD remains to be verified. In particular they do not see an overall decrease in a NMD genes, and only show a modest change in UPF1 protein itself. However, others have shown that Dux4 overexpression does lead to changes in relative abundance of splicing proteins (Geng et al 2012). Intriguingly one may speculate that dysregulation of NMD may account for some of the splicing changes observed in patient cells. However, there is likely more at play, consequently it is imperative to understand the splicing dysfunction in FSHD populations compared to control populations.

Toward the goal of further understanding transcriptome changes in FSHD patient cells caused by Dux4 expression, Rickard and colleagues use a Dux4 responsive reporter to enrich for FSHD cells that express Dux4 (Rickard et al., 2015). They and others (Snider et al., 2010; Tassin et al., 2012) show asynchronous expression of Dux4, and demonstrate a diffusion gradient of Dux4 within the syncytia. The asynchronous Dux4 expression, may suggest circadian regulation wherein the gene oscillates between high and low expression, however this has not been tested. Using Dux4 positive cells identified by sorting of reporter

positive cells, a population of cells that would have a transcriptome representative of Dux4 action was enriched (Rickard et al., 2015). It was found that ~20% of the transcriptome is directly changed as a consequence of expression of Dux4. Pathway analysis of the RNA-seq data suggest that one the major pathways affected is RNA metabolism – in particular splicing, surveillance and export. The dysregulation in RNA metabolism is thought to occur through up-regulation of several proteins that belong to either mRNA splicing, surveillance and export pathways. Interestingly amongst these genes are splicing factors SRSF2, STAU1, DDX39B which either skip an exon or retain an intron within their own mRNA transcripts. It must be noted that others have observed that culture conditions likely also affect the degree of Dux4 expression (Block et al., 2013; Pandey et al., 2015) thus the impact of Dux4 on the transcriptome may also be connected to the microenvironment.

To resolve the paradigm of how a low abundance protein like Dux4 that may be expressed asynchronously can have meaningful impact on the transcriptome, a model has been put forward where expression in one nucleus is sufficient to trigger activation of Dux4 target genes in a temporal and spatial manner within the syncytia. The model here posits that Dux4 is stochastically expressed, and it may activate genes in “sentinel” nuclei, or may diffuse across the syncytia and activate genes in other nuclei. Stochastic expression or pulsed expression from sentinel nuclei is typical of myotubes (Newlands et al., 1998). To date, there has been no in-depth study on the half-life of the Dux4 protein, although it is predicted to be unstable and decayed by the proteasome (Tassin et al., 2012). It had previously been demonstrated, in a model interrogating Dux4 expression with respect to telomere shortening (which serves as a proxy for aging due to aging related onset of FSHD symptoms) there is likely 1 in 2000 nuclei expressing Dux4. Upon telomere shortening, this number significantly increased to 1 in 200 nuclei (Stadler et al., 2013). Dux4 mRNA transcribed in one nucleus is exported to the cytoplasm where it is translated and returns to the nuclei closest to its translation site, but it may also diffuse within the syncytia and

activate transcription in distal nuclei. However, the length of time and the distance it diffuses is dependent upon its stability and the proximity of nuclei, thus a careful determination of Dux4 protein stability is necessary to develop this model further.

But questions still remain, what is the permissive amount of Dux4 protein before it becomes toxic to the cells? Overexpression can lead to cell death, and can contribute to deformation effects in animal models. However, what happens when Dux4 is expressed at low level? At what point does the threshold exist? Also, how frequently, and what quantity of RNA is produced in the event of productive transcription of the Dux4? What are the relative stabilities of Dux4 mRNA and protein, and to what extent could the Dux4 mRNA itself play a role in the pathogenicity of the Dux4 mediated disease?

Cell Physiology of FSHD and Correlation to Dux4

There is a decrease in the histone and DNA methylation of the D4Z4 repeats in FSHD patients, which leads to transcriptional derepression. While the skeletal muscle is the primary tissue type affected in FSHD, Dux4 can be detected in the non-skeletal muscle cells in FSHD patients (Snider et al., 2010). The hypomethylation of the D4Z4 repeats in FSHD is attributed to the action of several different factors which will be discussed here.

One longstanding theory underlying the association of hypomethylation of the D4Z4 repeats with FSHD, as well as the subtelomeric localization of the repeats is that disease presents due to dysregulated heterochromatin (van der Maarel et al., 2012). Several lines of evidence support a model in which silenced genomic DNA is being reactivated in somatic cells in FSHD patients.

In a subset of individuals, mutations in *SMCHD1* occurring with 4qA permissive haplotype result in FSHD2, even in the absence of contractions/reduction of the D4Z4 repeats (Lemmers et al., 2012). The monosomy or mutations identified resulted in

haploinsufficiency of SMCHD1 due to the presence of less than 50% of the protein.

Consequently, FSHD2 patients with SMCHD1 mutations, showed reduced occupancy of the protein on the D4Z4 repeats (Lemmers et al., 2012, 2015). *SMCHD1* is an essential gene, and in mice it is embryonic lethal, although in certain genetic backgrounds reduce male lethality (Mould et al., 2013). This apparent gender disparity in embryonic lethality is a consequence of the essential role of SMCHD1 in X-chromosome inactivation. SMCHD1 is recruited at sites with H3K9me3 or H3K9me2 (Brideau et al., 2015), and is potentially a DNA methylase or essential cofactor for acquired and sustained DNA methylation (Blewitt et al., 2008; Gendrel et al., 2012). However, not all FSHD2 patients have haploinsufficiency mutations in SMCHD1, and it has only been recently demonstrated that mutations in DNMT3B also modify the disease in another set of FSHD2 patients, thus there are likely other modifiers. It was previously shown that depletion SUV39H1 specifically reduces H3Kme3 on the D4Z4 repeats (Zeng et al., 2009). Given that SMCHD1 recruitment at the D4Z4 repeats is H3Kme3 dependent (Zeng et al., 2014), mutations in histone methylase or demethylase may also modify disease.

Other work has provided some indication that a feed forward mechanism may be at play to allow for increased derepression of the locus, and thus only a subtle change is necessary to start a destabilizing cascade. An example that supports this model is DBE-T, which is a long non-coding RNA detected in FSHD patients that originates from the D4Z4 repeat (Cabianca et al., 2012). DBE-T is likely polyadenylated because of detection in polyA+ fraction, and thus, is likely an RNA Polymerase II transcript. Fluorescence in situ hybridization (FISH), shows a faintly nuclear localized, chromatin associated DBE-T and it is posited to bind the D4Z4 repeats causing derepression. Evidence that supports this model is that depletion of DBE-T reduces transcription of D4Z4 proximal genes like *Ant1* and *Frg1*, as well as *Dux4*. Notably, the function of the RNA in de-repressing the locus cannot be accomplished in *trans*. The action of DBE-T is likely mediated through its recruitment of

histone-lysine N-methyltransferase, ASHL1. ASHL1, a member of the Trithorax complex was found enriched on non-deleted element (NDE) which lies upstream of the first D4Z4 repeat. In a heterologous host, treated with inhibitors of DNA methylation and histone acetylation (here referred to as enforced transcriptionally permissive conditions), DBE-T associates with ASH-L determined by RNA-IP, moreover knockdown of DBE-T reduces ASHL1 presence at the NDE.

There is potentially a positive feedback effect, wherein, DBE-T production enhances ASHL-1 recruitment at the NDE, which increases H3K36me2 and thus enhances its own transcription. In essence creating a trickle effect wherein occasional transcription, which is enhanced upon deletion of D4Z4 repeats due to more accessibility, allows for more DBE-T to be produced, which in turn recruits ASHL-1 to further derepress the locus.

However questions remain. First, how is DBE-T processed? Without enforcing transcriptional permissive conditions, the most abundant transcripts originating for the region appear discontinuous, therefore is the DBE-T processed to generate smaller RNA species? Second, what is the timing for the recruitment or the sequence of events that occur to permit a feedback loop?

At present, a compelling model of FSHD is one in which the D4Z4 repeat locus on chromosome 4 is a metastable epiallele (Himeda et al., 2014; Jones et al., 2015; Lemmers et al., 2012), where modified expression of genes at the locus contributes to the etiology of the disease (Figure 2.3). Key in this model is that, derepression of the somatically repressed locus in certain chromosomal backgrounds (giving haplotype specificity) allows the production of a homeodomain transcription factor, Dux4.

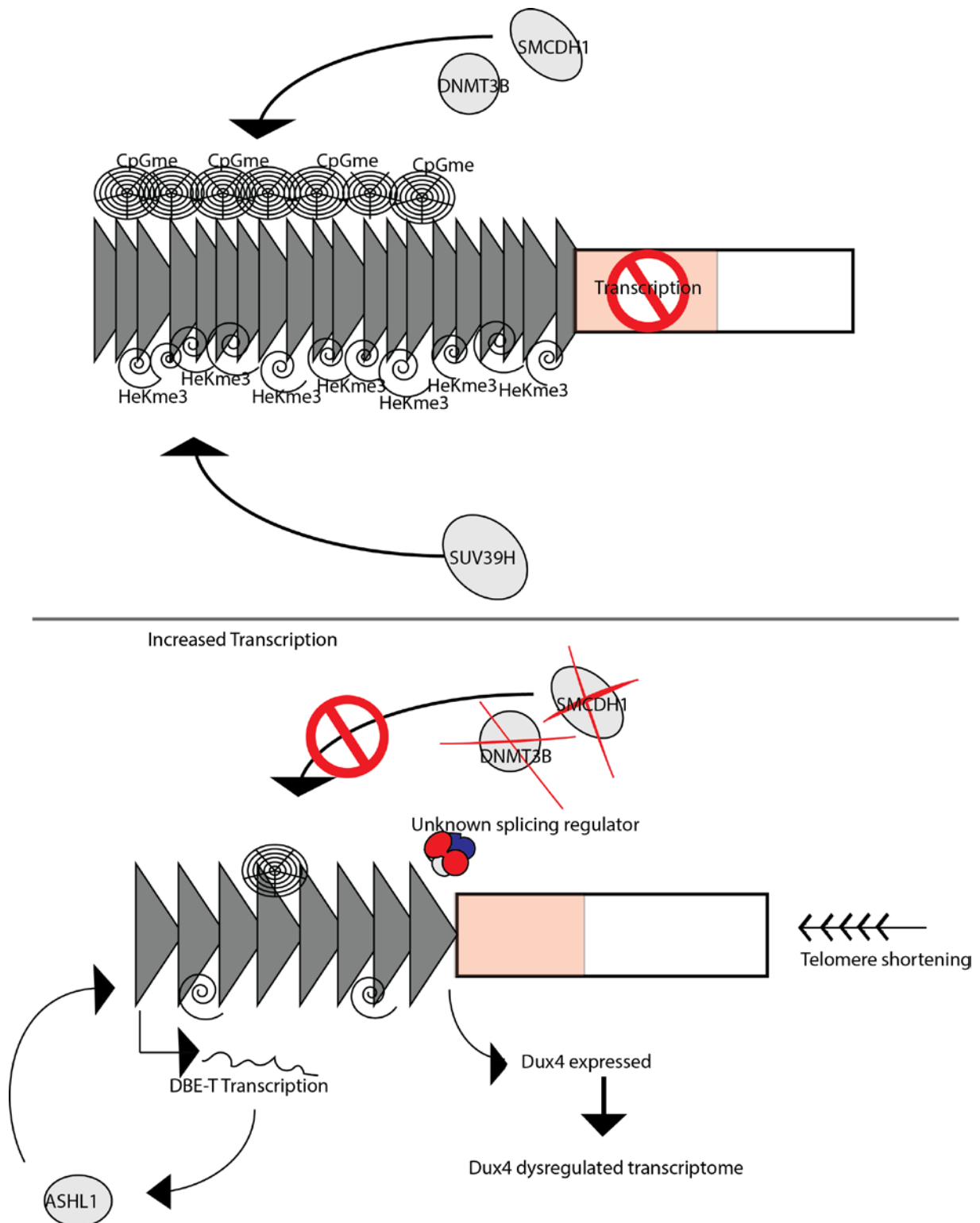


Figure 2.3 Model of Dux4 in FSHD: A Metastable Epiallele. (Top) D4Z4 depicted as heterochromatin region with negligible transcription due to the histone methylation and DNA

methylation by DNMT3B, SMCDH1, SUV39H amongst other factors. (Bottom) Dux4 is derepressed, and the pathogenic splice isoform is enriched due to contractions of the repeats which alleviate some of the repressive heterochromatin signals or mutations which impair the function of DNMT3B or SMCDH1. Transcription of lncRNA DBE-T recruits ASHL1 to further derepress the locus. Dux4 is expressed and activates a transcriptional cascade that includes dysregulated RNA processing – NMD, which stabilizes Dux4-fl mRNA and potentially increases production of the pathogenic protein product.

Current State

Genetic variants, epigenetic modifications and RNA splicing contribute to complex diseases (Li et al., 2016). The transcriptome changes as a consequence of Dux4 expression, the genetic variation modifying disease onset and severity, and the epigenetic contributions to disease pathology, demonstrate that FSHD may be characterized as a complex disease.

The current therapies for treatment of FSHD are physiotherapy, dietary supplementation, corticosteroids and T-cell infiltration (Sacconi et al., 2015; Tawil et al., 2015) and have resulted in variable success rates. The unclear molecular etiology had long impaired therapeutic progress. However, the recent advances into the molecular basis of the disease shows a mosaicism that may not be suitable for a simple panacea and may require varying strategies for treating and, or preventing the disease. For example, recent reports have demonstrated in cell culture models, the utility of RNA therapeutics (Lim et al., 2015; Marsollier et al., 2016), and CRISPR/CAS9 technology (Himeda et al., 2015, 2016) in suppressing Dux4.

The productive transcription of the Dux4 mRNA in FSHD patients is because of a SNP which generates a non-consensus PAS. However, beyond the identification of the PAS, the requirements for efficient 3' end processing of the Dux4 mRNA have not been examined. Here, we extend upon these efforts to investigate and characterize the Dux4 3' end processing signals required for cleavage and polyadenylation. Further, we investigate the use of utility of RNA therapeutics, in particular cleavage and polyadenylation inhibiting antisense oligonucleotides to impair processing at the Dux4 polyadenylation signal.

Chapter 3 : Materials and Methods

Cloning

Plasmids used in the study are described in Table 3.1. In the majority of constructs traditional restriction enzyme cloning was used to generate reporters. Primers and oligonucleotides used for cloning are described in Table 3.2. Inserted DNA was amplified using Pfu polymerase (purified by the Wagner Laboratory), and gel purified using GeneJet Gel Purification System (Thermo Fisher Scientific, Massachusetts USA). The SV40 late polyadenylation signal was amplified using the pGL4.13 (Promega, Wisconsin USA). Putative Dux4 processing signals were amplified from p2loxDux4 that was kindly provided by the Kyba Laboratory (Bosnakovski et al., 2008). Dux4 3' UTRs and cDNA sequences were obtained from ensembl.org and cross referenced with Ref Seq. The Dux4 UTR constructs and the Dux4 cDNA were artificially synthesized by GenScript (New Jersey, USA) and subcloned into the appropriate vector. All clones were sequenced to confirm identity.

In the remaining constructs, annealed oligonucleotide cloning was used and proceeded as follows: unphosphorylated polyacrylamide gel electrophoresis (PAGE) purified oligonucleotides were mixed equimolar at 20 μ M in annealing buffer (100mM Tris pH 7.5, 1M NaCl, 10mM EDTA). The mixture was heated to 95°C and then using a thermocycler the temperature was ramped down 5°C per cycle for three minutes per cycle, to a final temperature of 25°C. 5 μ L of the annealed oligonucleotides was then incubated at 37°C for 30 minutes with T4 Polynucleotide Kinase (New England Biolabs, Massachusetts USA) and 10 μ M ATP (Thermo Fisher Scientific) for phosphorylation. The phosphorylated annealed oligonucleotide was diluted 1:1000 and ligated to the appropriate alkaline phosphatase treated vector. Ligation reactions were carried out using T4 Ligase (purified by the Wagner Laboratory). Ligations were transformed into XL-1 Blue Competent cells (Stratagene, California USA).

In some constructs site-directed mutagenesis was performed to remove cryptic start codons, or introduce defects in the polyadenylation *cis* elements. In brief, to perform site-directed mutagenesis 10-25 ng of template plasmids were used for PCR using oligonucleotides for site directed mutagenesis listed in Table 3.2 in a total volume of 25 μ L using Pfu polymerase. Methylated template DNA was digested with 1 μ L DpnI (Thermo Fisher Scientific) for thirty minutes. Following Dpn I digestion, 1 μ L of the PCR product was transformed in XL1-Blue Competent Cells.

Table 3.1. Table of Plasmids

Name	Source/Reference
pcDNA3.1(+)	Courtesy of the Jayaraman Lab, University of Texas McGovern Medical School
p2loxDux4	Courtesy of the Kyba Lab, Lillehei Heart Institute, University of Minnesota
pLentiDuxBSntGFP	Acquired from the Miller Lab, University of Washington
pcDNA6tr	Invitrogen
pcDNA3	Invitrogen
pdp20	Derived by Dr. Eric J. Wagner from pdp19 vector initially purchased from Ambion
pTZHIVdGless	Suñe Lab, Instituto de Parasitología y Biomedicina “López Neyra”, Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), PTS, Granada, Spain
psiCheck2	Promega
pgl4.13	Promega
pUC19	NEB
pGINT	Addgene
pUC57-Kan-Dux4ORF	Genscript
pUC57-Kan-Dux4UTR1	Genscript
pUC57-Kan-Dux4UTR2	Genscript
pUC57-Kan-Dux4UTR-Int2minus	Genscript
pUC57-Kan-Dux4UTRUnspliced	Genscript

Table 3.2. Table of DNA Oligonucleotides for Cloning

Name	Purpose	Sequence
N060	Annealled Oligo Cloning of Hammerhead ribozyme	GCCGGCGTCCTGGTATCCAATCCT TCGGGATGTACTACCAGCTGATGA GTCCCAAATAGGACGAAACGCCGG A
N061A	Annealled Oligo Cloning of Hammerhead ribozyme	AGCTTCCGGCGTTTCGTCCTATTTG GGACTCATCAGCTGGTAGTACATC CCGAAGGATTGGATACCAGGACGC CGGCTGCA
N078	Cloning CMV Promoter	GCCGAATTCGCGTTGACATTGATTA TTGAC
N079	Cloning CMV Promoter	GGCCGAATTCGAGCTCTGCTTATAT AGACCT
N0D19	Cloning Dux4PAS	GGCCGCGGCCGCTACATATCTCTA CACTGATCACGTAAGTGATGTA
N0D20	Cloning Dux4PAS	GGCCAAGCTTCTTCCGTGAAATTCT GGCTGAATGTCTCC
N0D21	Cloning Dux4PAS	GGCCGCGGCCGCTACCAATTTTCAG GCTTTTTGTACAGGGGATA
N0D22	Cloning Dux4PAS	GGCCAGATCTCTTCCTGGCTAGAC CTGCGC
N0D23	Cloning Dux4PAS	GGCCAGATCTTTCTATAGGATCCAC AGGGAGGG

N0D18B	Cloning Dux4PAS	GGCCAAGCTTTAGACCTGCGCGCA GTGCGCACCCC
N0D24	Cloning Dux4PAS	GGCCAGATCTCGAGTAGACCTGCG CGCAGTGCGCACCCC
N0D25	Cloning Dux4PAS	GGCCAGATCTACATATCTCTACACT GATCAC
N0D26	Cloning Dux4PAS	GGCCAGATCTCGAGCTTCCGTGAA ATTCTGGCTGAATGTCTCC
N0D27	Cloning Dux4PAS	GGCCAGATCTACCAATTTAGGCTT TTTCTACAGGGGATA
N0D28	Cloning Dux4PAS	CATCTCCTGGATGATTACTTCAGAG ATATATTAATGCCC
N0D29	Cloning Dux4PAS	GGGCATTTTAATATATCTCTGAAGT AATCATCCAGGAGATG
N0D30	Cloning Dux4PAS	GTCACAATATCCCCTGTACAAAAAG CCTGAAATTGG
N0D31	Cloning Dux4PAS	CCAATTTAGGCTTTTTGTACAGGG GATATTGTGAC
N0D32	Cloning Dux4PAS	GGCCAGATCTATCGATTGCCTACA CTCTGCCTACAGGAGGC
N0D33	Cloning Dux4PAS	GGCCAGATCTCGAGATCGATTAGA CCTGCGCGCAGTGCGCAC
N0D34	Cloning Dux4PAS	GGCCGCGGCCGCTTTAAGTGATGT AACCATTCTC

N0D35	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTCAGTC TACTATGG
N0D36	Cloning Dux4PAS	GGCCGCGGCCGCTGAGTTCTGAAA CACATCTGC
N0D37	Cloning Dux4PAS	GGCCGCGGCCGCTGCACTGATCAC CGAAGTTATG
N0D34B	Cloning Dux4PAS	GGCCGCGGCCGCTGTAAGTGATGT AACCATTCTC
N0D92	Cloning Dux4PAS	GGCCTCTAGATTAAGTGATGTAACC ATTCTC
N0D93	Cloning Dux4PAS	GGCCTATAGATAGGTTCACTCTACT ATGG
N0D94	Cloning Dux4PAS	GGCCTCTAGAGAGTTCTGAAACAC ATCTGC
N0D95	Cloning Dux4PAS	GGCCTCTAGAGCACTGATCACCGA AGTTATG
N0D96	Cloning Dux4PAS	GGCCTCTAGAACCAATTTCAAGGCTT TTTGTACAGGGGATA
N0D97	Cloning Dux4PAS	GGCCCTCGAGTAGACCTGCGCGCA GTGCGCACCCC
N0D98	Cloning Dux4PAS	GGCC GGATCC TAGACCTGCGCGCAGTGC
N0D106	Cloning Dux4PAS	GGCCGAATTCTAGACCTGCGCGCA GTGCGCAC

N0D107	Cloning Dux4PAS	GGCCCTCGAGACATATCTCTACACT GATCAC
N0D108	Cloning Dux4PAS	GGCCCTCGAG CTATAGGATCCACAGGGAG
N0D109	Cloning Dux4PAS	GGCCCTCGAGGCACTGATCACCGA AGTTATG
N0D111	Cloning Dux4PAS	GGCCGAATTCCTTCCGTGAAATTCT GGCTGAATGTCTCC
N0D120	Cloning Dux4PAS	CATGCGGCCGCTCACGCGCTCTAC ACTGATCACGTAAGTGATG
N0D121	Cloning Dux4PAS	CATGCGGCCGCTACATATAGAGCA ACTGATCACGTAAGTGATG
N0D122	Cloning Dux4PAS	GGCCGCGGCCGCTACTGATCACGT AAGTGATGTAAC
N0D145	Cloning Dux4PAS	CATGCGGCCGCTAAATATCTCTACA CTGAT
N0D125	Cloning Dux4PAS	GGCCGCGGCCGCTCTCTACACTGA TCACGTAAGTG
N0D144	Cloning Dux4PAS	CATGCGGCCGCTCCATATCTCTAC ACTGAT
N0D146	Cloning Dux4PAS	CATGCGGCCGCTACCTATCTCTAC ACTGAT
N0D147	Cloning Dux4PAS	CATGCGGCCGCTACAGATCTCTAC ACTGATCAC

N0D148	Cloning Dux4PAS	CATGCGGCCGCTACATCTCTCTAC ACTGATCAC
N0D149	Cloning Dux4PAS	CATGCGGCCGCTACATAGCTCTAC ACTGATCAC
N0D150	Cloning Dux4PAS	GGCCGCGGCCGCTACATATATCTA CACTGATCACGTAAGT
N0D151	Cloning Dux4PAS	GGCCGCGGCCGCTACATATCGCTA CACTGATCACGTAAGT
N0D152	Cloning Dux4PAS	GGCCGCGGCCGCTACATATCTGTA CACTGATCACGTAAGT
N0D153	Cloning Dux4PAS	GGCCGCGGCCGCTACATATCTCTA CCCTGATCACGTAAGTGAT
N0D154	Cloning Dux4PAS	GGCCGCGGCCGCTACATATCTCTA CAATGATCACGTAAGTGAT
N0D155	Cloning Dux4PAS	GGCCGCGGCCGCTACATATCTCTA CACGGATCACGTAAGTGAT
N0D156	Cloning Dux4PAS	GGCCATCGATTAGACCTGCGCGCA GTGCGCACCCC
N0D157	Cloning Dux4PAS	GGCCATCGATCTTCCGTGAAATTCT GGCTGAATGTCTCC
N0D163	Cloning Dux4PAS	GGCCAAGCTTACATATCTCTACACT GATCACGTAAGTGATGTA
N0D164	Cloning Dux4PAS	GGCCAAGCTTACCAATTTGAGGCTT TTTGTACAGGGGATA

NOD165	Cloning Dux4PAS	GGCCAAGCTTTGCCTACACTCTGC CTACAGGAGGC
NOD36B	Cloning Dux4PAS	GGCCGCGGCCGCTGAGTTCTGAAA CAGATCTGC
NOD167	Cloning Dux4PAS	GGCCAAGCTTAGGGGCTTTGTGAG ATATCTCTG
NOD168	Cloning Dux4PAS	GGCCAAGCTTATTTCCACTGCTCAA ACAGGTGATG
NOD169	Cloning Dux4PAS	GGCCAAGCTTGAGATGTAAAAATTG TCTGGGCTTTGTC
NOD170	Cloning Dux4PAS	GGCCAAGCTTAAGCTCTGCCTACA GGGGCATTG
NOD173	Cloning Dux4PAS	GGCCGCGGCCGCTTCGGTTCAGTC TACTATGGAGTTC
NOD174	Cloning Dux4PAS	GGCCGCGGCCGCTTAGTGGCAGT CTACTATGGAGTTC
NOD175	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTACTTC TACTATGGAGTTC
NOD176	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTACAGGA GACTATGGAGTTC
NOD177	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTACAGTC TCAGATGGAGTTC
NOD178	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTACAGTC TACTCGTGAGTTC

NOD179	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTCAAGTC TACTATGTCTTTC
NOD181	Cloning Dux4PAS	GGCCGCGGCCGCTGCTTGGCAGT CTACTATGGAGTTCTGAAAC
NOD182	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTACTGA GACTATGGAGTTCTGAAAC
NOD183	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTCAAGTC TCAACGTGAGTTCTGAAAC
NOD178B	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTCAAGTC TACTCGTGAGTTC TGAAAC
NOD179B	Cloning Dux4PAS	GGCCGCGGCCGCTTAGGTTCAAGTC TACTATGTCTTTCTGAAAC
NOD189	Cloning Dux4PAS	GGCCAGATCTAGGGGCTTTGTGAG ATATCTCTG
NOD195	Cloning Dux4PAS	GGCCGCGGCCGCTAAATATCTCTA CCAGACGCACGTA
NOD196	Cloning Dux4PAS	GGCCGCGGCCGCTAAATATAGAGC AACTGATCACGTA
NOD197	Cloning Dux4PAS	GGCCGCGGCCGCTCGCGCGCTCT ACACTGATCACGTA
NOD225	Cloning Dux4PAS	GGCCGCGGCCGCTCACGCTCTCTA CACTGATCACGTA
NOD223a	Cloning Dux4PAS	GGCCGCGGCCGCTAAATATCTCTC ACAGGATCACGTA

NOD224a	Cloning Dux4PAS	GGCCGCGGCCGCTAAATAGAGAGA CACTGATCACGTA
NOD226	Cloning Dux4PAS	GGCCGCGGCCGCTGAACTAATCAT CCAGGAGATG
NOD227	Cloning Dux4PAS	GGCCCTCGAGACCTGCGCGCAGT GCGCACC
NOD228	Cloning Dux4PAS	GGCCCTCGAGCTTCCGTGAAATTC TGGC
N0D08	Cloning Dux4PAS Exon3	GGCCAGATCTTTCTATAGGATCCAC AGGGAGGG
N0D17	Cloning Dux4PAS Exon3	GGCCGCGGCCGCTTGCGTACACTC TGCCTACAGGAGGC
N0D01B	Cloning Dux4PAS pLAM	GGCCAAGCTTCGGTCAAAGCATA CCTCTGTCTGTCT
N054	Cloning HAmCherry	GGCCGGTACCATGTACCCATACGA TGTTCTGACTATGCGGGCGTGAG CAAGGGCGAG
N055	Cloning HAmCherry	GGCCGGATCCTTACTTGTACAGCT CGTCCATG
N081	Cloning Random Sequences from pUC19	GGCCAAGCTTGTGATGACGGTGAA AACCTC
N082	Cloning Random Sequences from pUC19	GGCCGGATCCGTACAATCTGCTCT GATGCC
N083	Cloning Random Sequences from pUC19	GGCCAAGCTTGAGAGTGCACCATA TGCGGT

N084	Cloning Random Sequences from pUC19	GGCCGGATCCCAACGTCGTGACTG GGAAAACCCTGGCGTT
N085	Cloning Random Sequences from pUC19	GGCCAAGCTTATGTTGTGCAAAAAA GCGGTTAGCTCCTTC
N086	Cloning Random Sequences from pUC19	GGCCGGATCCGAATGACTTGGTTG AGTACT
N087	Cloning Random Sequences from pUC19	GGCCAAGCTTATGATACCGCGAGA CCCACG
N088	Cloning Random Sequences from pUC19	GGCCGGATCCGGCAACAACGTTGC GCAAAC TATTA ACTGG
N089	Cloning Random Sequences from pUC19	GGCCGGATCCTTTAAAAGTGCTCAT CATTG
N090	Cloning Random Sequences from pUC19	GGCCAAGCTTATGAGTATTCAACAT TTCCGTG
N091	Cloning Random Sequences from pUC19	GGCCAAGCTTGGGTGCCTAATGAG TGAGCT
N092	Cloning Random Sequences from pUC19	GGCCGGATCCGAGGAAGCGGAAG AGCGCCCAATACGCAAA
N081p	Cloning Random Sequences from pUC19	GGCCAGATCTGTGATGACGGTGAA AACCTC
N083p	Cloning Random Sequences from pUC19	GGCCAGATCTGAGAGTGCACCATA TGCGGT
N085p	Cloning Random Sequences from pUC19	GGCCAGATCTATGTTGTGCAAAAAA GCGGTTAGCTCCTTC

N087p	Cloning Random Sequences from pUC19	GGCCAGATCTATGATACCGCGAGA CCCACG
N089p	Cloning Random Sequences from pUC19	GGCCAGATCTTTTAAAAGTGCTCAT CATTG
N091p	Cloning Random Sequences from pUC19	GGCCAGATCTGGGTGCCTAATGAG TGAGCT
N050	Cloning SV40LPAS	GGCCAAGCTTGCCGTGTAATAATTC TAGAGTC
N051	Cloning SV40LPAS	GGCCGGATCCGAAAACCTCCCACA CCTCCCC
N0120	Cloning SV40LPAS	GGCCATCGATGCCGTGTAATAATTC TAGAGT C
N0121	Cloning SV40LPAS	GGCCGGATCCGAAAACCTCCCACA CCTCCCC
N0122	Cloning SV40LPAS	GGCCGGATCCCGATTTTACCACATT TG TAGAGG
N0145	Cloning SV40LPAS	GGCCATCGATGAGTTTGGACAAAC CACAAAC
N0146	Cloning SV40LPAS	GGCCATCGATGCAGCTTATAATGG TTACAAAT
N0147	Cloning SV40LPAS	GGCCAAGCTTCAACAATTGCATTCA TTTTATGTTTC
N099	Cloning SV40LPAS	GGCCAGATCTGAAAACCTCCCACA CCTCCCC

N0121B	Cloning SV40LPAS	GGCCAAGCTTGAAAACCTCCCACA CCTCCCC
N0122B	Cloning SV40LPAS	GGCCAAGCTTCGATTTTACCACATT TGTAGAGG
N080	Cloning SV40LPAS DSE null	GGCCGTCGACATTGTTGTTGTTAAC TTGTTTATTGC
N069	Cloning SV40LPAS DSE null	GGCCGGATCCATTGTTGTTGTTAAC TTGTTTATTGC
N0150A	Site directed mutagenesis create ClaI restriction site in Dux4PAS	CTCCTGGATGATTAGTTCATCGATA TATTAATGCCCCCTCCCT
N0150B	Site directed mutagenesis create ClaI restriction site in Dux4PAS	AGGGAGGGGGCATTTTAATATATC GATGAACTAATCATCCAGGAG
N0149A	Site directed mutagenesis create HindIII restriction site in Dux4PAS	GTGGATCCTATAGAAGATTTGAAGC TTTTGTGTGATGAGTGCAG
N0149B	Site directed mutagenesis create HindIII restriction site in Dux4PAS	CTGCACTCATCACAAAAGCTTCA AATCTTCTATAGGATCCAC
N076	Site directed mutagenesis create SV40L clvnull short oligo	GCAATAACAAGTTAACGGCGGGCG GTTGCATTCATTTTATG
N077	Site directed mutagenesis create SV40L clvnull short oligo	CATAAATGAATGCAACCGCCGCC GTTAACTTGTTTATTGC
N074	Site directed mutagenesis create SV40L PASNull short oligo	GTAACCATTATAAGCTGCGGGAAA CAAGTTAACAAC
N075	Site directed mutagenesis create SV40L PASNull short oligo	GTTGTAACTTGTTTCCCGCAGCTT ATAATGGTTAC

N0D113A	Site directed mutagenesis Dux4PAS	CTCCATAGTAGACTGAACCTATCTC CTGGTTACATCACTTAC
N0D113B	Site directed mutagenesis Dux4PAS	GTAAGTGATGTAACCAGGAGATAG GTTCAGTCTACTATGGAG
N0D114A	Site directed mutagenesis Dux4PAS	GTAGACTGAACCTAGAGAAGTTGG ACATCACTTACGTGATCAG
N0D114B	Site directed mutagenesis Dux4PAS	CTGATCACGTAAGTGATGTCCAAC TCTCTAGGTTTCAGTCTAC
N0D115A	Site directed mutagenesis Dux4PAS	CTGAACCTAGAGAATGGTTCACGA ACTTACGTGATCAGTGTAGAG
N0D115B	Site directed mutagenesis Dux4PAS	CTCTACACTGATCACGTAAGTTCGT GAACCATTCTCTAGGTTTCAG
N0D116A	Site directed mutagenesis Dux4PAS	CTAGAGAATGGTTACATCCAGGCC GTGATCAGTGTAGAGATATG
N0D116B	Site directed mutagenesis Dux4PAS	CATATCTCTACACTGATCACGGCCT GGATGTAACCATTCTCTAG
N0D117A	Site directed mutagenesis Dux4PAS	GAGAATGGTTACATCACTTAATGTC TCAGTGTAGAGATATGTAGC
N0D117B	Site directed mutagenesis Dux4PAS	GCTACATATCTCTACACTGAGACAT TAAGTGATGTAACCATTCTC
N0D117C	Site directed mutagenesis Dux4PAS	GAGAATGGTTACATCACTTATTGTC TCAGTGTAGAGATATGTAGC
N0D117D	Site directed mutagenesis Dux4PAS	GCTACATATCTCTACACTGAGACAA TAAGTGATGTAACCATTCTC

NOD118A	Site directed mutagenesis Dux4PAS	GGTTACATCACTTACGTGAGACTG GTAGAGATATGTAGC
NOD118B	Site directed mutagenesis Dux4PAS	GCTACATATCTCTACCAGTCTCACG TAAGTGATGTAACC
NOD166A	Site directed mutagenesis Dux4PAS	GTTACATCTCCTGGAGGATTACTTC AGAGATATATTTAAATCCCC
NOD166B	Site directed mutagenesis Dux4PAS	GGGGATTTTAAATATATCTCTGAAGT AATCCTCCAGGAGATGTAAC
NOD112C	Site directed mutagenesis Dux4PAS	CATTACTTCAGAGATATATTTAAATC CCCCCTCCCTGTG
NOD112D	Site directed mutagenesis Dux4PAS	CACAGGGAGGGGGGATTTAAATAT ATCTCTGAAGTAATG
NOD180A	Site directed mutagenesis Dux4PAS	GATCATTACTTCAGAGATATAGGGA AATCCCCCCTCCCTGTG
NOD180B	Site directed mutagenesis Dux4PAS	CACAGGGAGGGGGGATTTCCCTAT ATCTCTGAAGTAATGATC
NOD190A	Site directed mutagenesis Dux4PAS	CAGAACTCCATAGTAGACTGCCAA GCGAGAATCGTTACATCTACGTGAT C
NOD190B	Site directed mutagenesis Dux4PAS	GATCACGTAGATGTAACGATTCTCG CTTGGCAGTCTACTATGGAGTTCTG
NOD191A	Site directed mutagenesis Dux4PAS	CTCCATAGTAGACTGAACCTATCTC CGCGTTACATCACTTACGTGATCAG

NOD191B	Site directed mutagenesis Dux4PAS	CTGATCACGTAAGTGATGTAACGC GGAGATAGGTTTCAGTCTACTATGG AG
NOD192A	Site directed mutagenesis Dux4PAS	CCATAGTAGACTGAACCTAGAGAAT ATCTACATCACTTACGTGATCAGTG
NOD192B	Site directed mutagenesis Dux4PAS	CACTGATCACGTAAGTGATGTAGAT ATTCTCTAGGTTTCAGTCTACTATGG
NOD192C	Site directed mutagenesis Dux4PAS	CCATAGTAGACTGAACCTAGAGAAT ATCGCAATCACTTACGTGATCAGTG
NOD192D	Site directed mutagenesis Dux4PAS	CACTGATCACGTAAGTGATTGCGAT ATTCTCTAGGTTTCAGTCTACTATGG
NOD193A	Site directed mutagenesis Dux4PAS	GTAGACTGAACCTAGAGAATCGTTA CCGACAGTACGTGATCAGTGTAGA G
NOD193B	Site directed mutagenesis Dux4PAS	CTCTACACTGATCACGTAAGTGTCGG TAACGATTCTCTAGGTTTCAGTCTAC
NOD194A	Site directed mutagenesis Dux4PAS	CCTAGAGAATCGTTACATCACTGCA GTGATCAGTGTAGAGATATTTTG
NOD194B	Site directed mutagenesis Dux4PAS	CAAAATATCTCTACACTGATCACTG CAGTGATGTAACGATTCTCTAGG
NOD194C	Site directed mutagenesis Dux4PAS	CCTAGAGAATCGTTACATCACTGCA TATATCAGTGTAGAGATATTTTG
NOD194D	Site directed mutagenesis Dux4PAS	CAAAATATCTCTACACTGATATATG CAGTGATGTAACGATTCTCTAGG

NOD214A	Site directed mutagenesis Dux4PAS	CAGATCTGTTTCAGAACTCACGCTT AGACTGAACCTAGAGAATC
NOD214B	Site directed mutagenesis Dux4PAS	GATTCTCTAGGTTTCAGTCTAAGCGT GAGTTCTGAAACAGATCTG
NOD215A	Site directed mutagenesis Dux4PAS	CTGTTTCAGAACTCCATAGGCTCGT GAACCTAGAGAATCGTTAC
NOD215B	Site directed mutagenesis Dux4PAS	GTAACGATTCTCTAGGTTTCACGAGC CTATGGAGTTCTGAAACAG
NOD216A	Site directed mutagenesis Dux4PAS	CAGAACTCCATAGTAGACGTCCACT AGAGAATCGTTACATCAC
NOD216B	Site directed mutagenesis Dux4PAS	GTGATGTAACGATTCTCTAGTGGAC GTCTACTATGGAGTTCTG
NOD217A	Site directed mutagenesis Dux4PAS	CTCCATAGTAGACTGAACAGCTCG AATCGTTACATCACTTACGTG
NOD217B	Site directed mutagenesis Dux4PAS	CACGTAAGTGATGTAACGATTCTGA GCTGTTTCAGTCTACTATGGAG
NOD218A	Site directed mutagenesis Dux4PAS	CTCCATAGTAGACTGAACCTAGATC CTAGTTACATCACTTACGTG
NOD218B	Site directed mutagenesis Dux4PAS	CACGTAAGTGATGTAAGTAGGATCT AGGTTTCAGTCTACTATGGAG
NOD219A	Site directed mutagenesis Dux4PAS	GTAGACTGAACCTAGAGAATCTGG CAATCACTTACGTGATCAGTG
NOD219B	Site directed mutagenesis Dux4PAS	CACTGATCACGTAAGTGATTGCCA GATTCTCTAGGTTTCAGTCTAC

NOD220A	Site directed mutagenesis Dux4PAS	GAACCTAGAGAATCGTTACCGACAT TACGTGATCAGTGTAGAG
NOD220B	Site directed mutagenesis Dux4PAS	CTCTACACTGATCACGTAATGTCCG TAACGATTCTCTAGGTTT
NOD221A	Site directed mutagenesis Dux4PAS	CTAGAGAATCGTTACATCACGGCAT TGATCAGTGTAGAGATATC
NOD221B	Site directed mutagenesis Dux4PAS	GATATCTCTACACTGATCAATGCCG TGATGTAACGATTCTCTAG
NOD222A	Site directed mutagenesis Dux4PAS	GAATCGTTACATCACTTACGGTCGA AGTGTAAGATATCTAGCG
NOD222B	Site directed mutagenesis Dux4PAS	CGCTAGATATCTCTACACTTCGACC GTAAGTGATGTAACGATT
N066	Site-directed mutagenesis remove start codon in SV40LPAS	CCGCTTCGAGCAGACGTGATAAGA TACATTGATGAGTTTGG
N067A	Site-directed mutagenesis remove start codon in SV40LPAS	CCAAACTCATCAATGTATCTTATCA CGTCTGCTCGAAGCGG
N064	Site-directed mutagenesis remove stop codon in SV40LPAS	CCACAACTAGAATGCAGGGAAAAA AATGCTTTATTTGTG
N065A	Site-directed mutagenesis remove stop codon in SV40LPAS	CACAAATAAAGCATTTTTTTCCCTG CATTCTAGTTGTGG
N093	Site-directed mutagenesis remove stop codon in SV40LPAS	GGACAAACCACAACTAGAGTGCAG GGA AAAAAATGC
N094	Site-directed mutagenesis remove stop codon in SV40LPAS	GCATTTTTTTCCCTGCACTCTAGTT GTGGTTTGTCC

N095	Site-directed mutagenesis remove stop codon in SV40LPAS	TGCATTCATTTTGTGTTTCAGGTTC AGGGGGAGGTG
N096	Site-directed mutagenesis remove stop codon in SV40LPAS	CACCTCCCCCTGAACCTGAAACAC AAAATGAATGCA

Cell Culture

All cell lines used in this study are described in Table 3.3. Patient cell lines were obtained from University of Rochester and University of Massachusetts Wellstone Center for FSHD. Patient cells were supplied as de-identified lines by the aforementioned institutes. HeLa, HEK293T and RD cells were cultured in Dulbecco's modified Eagle medium (DMEM) (Lonza, Maryland USA) containing L-Glutamine and Sodium pyruvate and supplemented with 1% penicillin-streptomycin (Thermo Fisher Scientific) and 10% fetal bovine serum (Hyclone, Utah USA). Hereafter, referred to as DMEM Complete. Cells were passaged 1:10 and grown in 5% CO₂.

Immortalized myoblast cells obtained from the Wellstone Center for FSHD were culture in LHCN , Medium (4:1 DMEM: Medium 199, supplemented with 20% fetal bovine serum, 30 mg/L ZnSO₄ (Sigma, Missouri USA), 1.4 mg/mL Vitamin B₁₂ (Sigma, Missouri USA), 55µg/mL dexamethasone (Sigma, Missouri USA), 2.5 µg/mL hepatocyte growth factor (Sigma), 25 µg/mL basic fibroblast growth factor (Promega, Wisconsin, USA) and 1% antibiotic/antimycotic (Thermo Fisher Scientific, Massachusetts USA). For differentiation, cells at ≥90% confluence were fed 4:1 DMEM:Medium 199 supplemented with 2% horse serum (Hyclone, Utah USA) or 15% Knock-Out Serum Replacer (Thermo Fisher Scientific, Massachusetts USA), 2 mM L-glutamine (Thermo Fisher Scientific, Massachusetts USA), 1% antibiotics/antimycotics, 20 mM HEPES (Sigma, Missouri USA) and 1 mM sodium pyruvate (Thermo Fisher Scientific, Massachusetts USA). Immortalized human myoblast cells were grown on plates treated with 0.1% gelatin.

Table 3.3. Table of Cell Lines

Cell Line	Supplier	Descriptor	Reference
HeLa	ATCC	Immortalized human cervical cancer cell line	HeLa (ATCC® CCL-2™)
HEK293T	ATCC	Immortalized human embryonic kidney cells containing the SV40 large T antigen	293T (ATCC® CRL-3216™)
RD	ATCC	Immortalized human rhabdomyosarcoma striated muscle cell line	RD (ATCC® CCL-136™)
15Abic Ct#24	University of Massachusetts Wellstone Program	Immortalized human skeletal muscle cell line: myoblasts. FSHD line derived from proband with ~8 D4Z4 repeats.	(Homma et al., 2012; Rahimov et al., 2012; Stadler et al., 2011)
15Vbic CT#9	University of Massachusetts Wellstone Program	Immortalized human skeletal muscle cell line: myoblasts. Control line derived from unaffected sibling of 15Abic line. Normal D4Z4 repeat length >11.	(Homma et al., 2012; Rahimov et al., 2012; Stadler et al., 2011)

Stable Line Development

A kill curve was performed on HEK293T cells and the optimal dose for selecting resistance within 6 days using blasticidin (Thermo Fisher Scientific, Massachusetts USA) was 10 mg/mL. HEK293T cells were plated at 6×10^5 cells per well in 6 well dish, and transfected the following day with the reporter plasmid and selection plasmid (10:1 ratio) using Polyethylenimine (PEI). After a period of 24 hrs after transfection, visual confirmation of the expression of the selection plasmid (mCherry) and the reporter plasmid (GFP) was done. The medium was then removed from the cells and replaced with DMEM complete containing blasticidin at 10 $\mu\text{g/mL}$. Cells were allowed to grow in selection in six-well dish for a week with medium replenished every other day. After day six, cells were moved to 10 cm^2 dish, while still maintaining selection. Cells were allowed to grow to generate sufficient plates for freezing down and cell sorting. Within two days of moving to 10 cm^2 dish, cell selection was maintained using a lower dose of Blasticidin (5 $\mu\text{g/mL}$) with each passage in DMEM complete. Cells plated for transfection with ASOs or siRNAs were seeded in DMEM complete without blasticidin. Polyclonal cells were sorted at the University of Texas Medical Branch Flow Cytometry and Cell Sorting Core Facility for GFP positive cells. To select clonal cells, cells were seeded at a density of 0.5 cell/ well in 96 well dish in the absence of blasticidin. After sufficient growth, indicated by a change in the color of the medium clones were moved to a 24-well dish and supplemented with 5 $\mu\text{g/mL}$ blasticidin. The clonal identity was determined by sequencing of nested PCR products amplified from the genomic DNA of the clones.

Transfections

All cells were transiently transfected with Lipofectamine 2000 (Thermo Fisher Scientific, Massachusetts USA) or with Polyethylenimine (PEI) (Sigma, Missouri USA). Cells were transfected at 70-80% confluence.

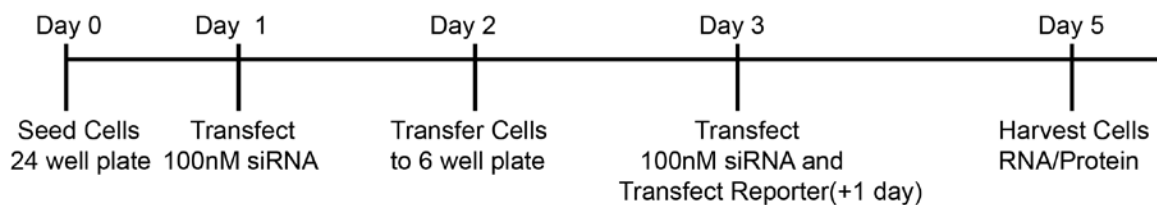
Small Interfering RNA (siRNA) transfections

All siRNAs used in this study are listed in Table 3.4, and were purchased from Sigma-Aldrich. A two hit protocol (Albrecht and Wagner, 2012; Wagner and Garcia-Blanco, 2002) was used for transfection of siRNAs (Figure 3.1A). Briefly, cells were seeded the day before transfection to ensure 80% confluence on day of transfection in a 24-well plate. A total of 3 μ L of Lipofectamine 2000 was incubated with 50 μ L of OptiMEM (Thermo Fisher Scientific) for seven minutes at room temperature. After seven minutes elapsed, the Lipofectamine/OptiMEM mixture was mixed with an equal volume of OptiMEM containing 3 μ L of 20 μ M siRNA. The siRNA:lipid complexes were allowed to form during a 25 minute incubation at room temperature. After which, the complexes were added to cells. Final concentration of siRNA on cells 100 nM. The medium was changed four hours after transfection. Twenty-four hours after transfection, the cells were trypsinized and moved to a six-well plate. Following an additional twenty-four hours of growth, cells were once more transfected with siRNA as described earlier for final concentration of 100 nM siRNA. In some instances at the second transfection, siRNA was co-transfected with DNA plasmids that were typically 100 ng of reporter plasmid with 50ng transfection control plasmid. Cells were lysed for protein or RNA 24-48 hrs after second transfection.

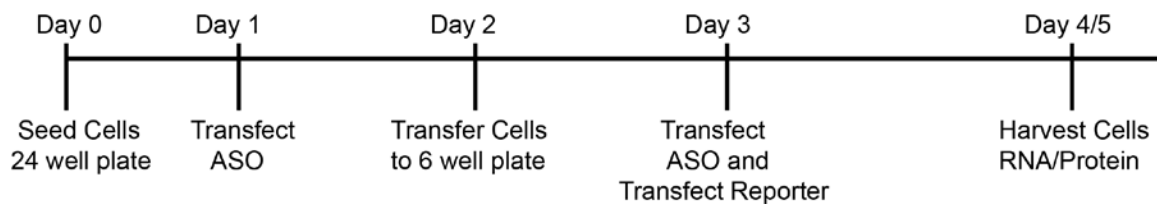
Table 3.4. Table of siRNAs

Name	Sequence	Reference
C2	GGUCCGGCUCCCCCAAUG[dT][dT]	(Wagner et al., 2005)
XRN1	GUAACUGAACUUUCUCGAA[dT][dT]	Sigma-Aldrich #1
Exosc4	GACCGUAAGUCCUGAGA[dT][dT]	Sigma-Aldrich #1
siGFP	CAAGCUGACCCUGAAGUUC[dT][dT]	(Tschuch et al., 2008)
Dicer1	CAUUGAUCCUGUCAUGGAU[dT][dT]	Sigma-Aldrich #1
CPSF160	GCUUUAAGAAGGUCCCUCA[dT][dT]	Sigma-Aldrich #1
UPF1	GAGAU AUGCCUGCGGUACA[dT][dT]	Sigma-Aldrich #1
CstF64	GGCUUUAGUCCCGGGCAGA[dT][dT]	Sigma-Aldrich #1

A.



B.



C.

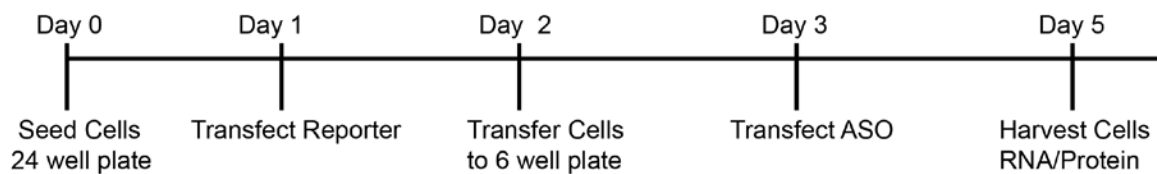


Figure 3.1. Schematic of transfection protocol with co-transfection of siRNA or ASO with Reporter. A and B are two hit protocols, C is a single hit protocol for transfections

Antisense Oligonucleotide (ASO) transfections

All ASO are listed in Table 3.5, and were purchased from Sigma-Aldrich with either a phosphorothioate or phosphodiester backbone. ASOs were transfected with a two-hit protocol as described above for siRNA transfection. Lipofectamine 2000 volume was typically kept constant at 3 μ L and ASO volume maintained below 5 μ L, and if necessary diluted in ultrapure water for the required concentration. Modifications to the ASO transfections include only a single transfection of ASO in 24 well plated cells, with cells lysed for RNA or protein or visualized for fluorescence 24-120 hrs (120 hrs post transfection used for myoblast transfected ASOs which were induced to differentiate). Another modification of the ASO transfection protocol (Figure 3.1B/C) occurs when reporter and transfection control plasmid are first transfected into cells plated in a 24-well dish, described below. Plasmid transfected cells are moved to a 6-well dish and transfected with ASO at the appropriate concentration.

Table 3.5. Table of Antisense Oligonucleotides

Name	Sequence	Bac kbone	Targ et	Refe rence
NPASO -1	[mG][mU][mU][mA][mA][mC][mU][mU][mG][mU][mU][mU][mA][mU] [mU][mG][mC][mA]	PS, PO	SV40PAS	This study
NPASO -2	[mG][mC][mA][mA][mU][mU][mA][mG][mU][mA][mA][mA][mU][mU][mC]	PS, PO	ASO Control Short	This study
NPASO-2F	[Flc][mG][mC][mA][mA][mU][mU][mA][mG][mU][mA][mA][mA][mU][mU][mC]	PS	ASO Control Short '5Fluorecein Tag	This study
NPASO -3	[mU][mU][mG][mC][mA][mA][mU][mU][mA][mG][mU][mA][mA][mA][mU][mU][mC][mA]	PS, PO	ASO Control Long	This study

NPASO -4	[mG][mC][mA][mU][mU][mU][mU][mA][mA][mU][mA][mU][mA][mU][mC]	PS	Dux4PAS at PAS	This study
NPASO -5	[mG][mG][mA][mU][mU][mU][mU][mA][mA][mU][mA][mU][mA][mU][mC]	PS	Dux4PAS at PAS single mismatch	This study
NPASO -6	[mG][mU][mU][mA][mA][mC][dT][dT][dG][dT][dT][dT][dA][dT][mU][mG][mC][mA][mG][mC]	PS	SV40PAS	This study
NPASO -7	[mG][mU][mA][mU][mC][mU][dA][dC][dT][dG][dG][dT][dT][dC][mG][mU][mA][mU][mA][mU]	PS	ASO Control Long	This study
NPASO - 8	GTTAAC TTGTTTATTGCAGC	PS	SV40PAS	This study
NPASO - 9	GTATCTACTGGTTCGTATAT	PS	ASO Control Long	This study
NPASO - 10	[mA][mG][mG][mG][mG][mG][dC][dA][dT][dT][dT][dT][dA][dA][mU][mA][mU][mA][mU][mC]	PS	Dux4PAS at PAS	This study

NPASO - 12	[mC][mU][mA][mG][mG][mU][mU][mC][mA][mG][mU][mC][mU][mA] [mC]	PS	Dux4PAS at ~+160	This study
NPASO - 13	[mA][mC][mA][mU][mA][mU][mC][mU][mC][mU][mA][mC][mA][mC][mU][mG]	PS	Dux4PAS at~+200	This study
NPASO - 14	[mU][mC][mU][mA][dG][dG][dT][dT][dC][dA][dG][dT][mC][mU][mA][mC][mU]	PS	Dux4PAS at~+160	This study
NPASO - 15	[mG][mC][mA][mA][mA][mU][mC][mU][mU][mC][mU][mA][mU][mA][mG]	PS	Dux4PAS at Cleavage site	This study
NPASO - 16	[mC][mU][mA][mU][mA][mG][mG][mA][mU][mC][mC][mA][mC][mA][mG]	PS	Dux4PAS Cleavage site	This study
NPASO - 21	[mA][mG][mG][mG][mG][dG][dC][dA][dT][dT][dT][dT][dA][dA][dT][m A][mU][mA][mU][mC]	PS	Dux4PAS at PAS	This study
NPASO - 22	[mA][mG][mG][mG][mG][mG][mC][mA][mU][mU][mU][mU][mA][mA] [mU][mA][mU][mA][mU][mC]	PS, PO	Dux4PAS at PAS	This study

NPASO - 23	[mU][mG][mC][mG][mC][mG][mC][mA][mG][mG][mU][mC][mU][mA] [mG][mC][mC][mA][mG][mG]	PS, PO	Dux4 Exon 3 Splice Acceptor	This study
NPASO - 24	[mU][mG][mU][mA][mA][mC][mC][mA][mU][mU][mC][mU][mC][mU] [mA][mG][mG][mU][mU][mC]	PS, PO	Dux4PAS ~160	This study
NPASO - 25	[mC][mC][mU][mA][mA][mG][mU][mG][mA][mU][mG][mU][mA][mA] [mC][mC][mA][mU][mU][mC]	PS, PO	Dux4PAS ~160	This study
NPASO - 26	[mC][mU][mA][mC][mA][mC][mU][mG][mA][mU][mC][mA][mC][mC][mU][mA][mA][mG][mU][mG]	PS, PO	Dux4PAS ~200	This study
NPASO - 27	[mC][mG][mA][mG][mA][mA][mU][mU][mU][mU][mA][mA][mC][mA][mU][mA][mU][mC][mU][mC]	PS, PO	Dux4PAS ~200	This study
NPASO- Dicer1A1	[mG][mC][mU][mG][mA][dC][dC][dT][dT][dT][dT][dT][dG][dC][dT][m U][mC][mU][mC][mA]	PS	Dicer Gapmer1 5- 10-5 Positive Control	(Lim et al., 2015)

NPASO-Dicer1B1	[mA][mG][mG][mA][mG][dG][dA][dA][dG][dC][dC][dA][dA][dT][dT][mC][mA][mC][mA][mG]	PS	Dicer Gapmer1 5-10-5 Positive Control	(Lim et al., 2015)
NPASO-Dicer1C1	[mA][mG][mA][mC][mG][dA][dT][dA][dA][dC][dT][dT][dT][dA][dT][mU][mG][mG][mA][mG]	PS	Dicer Gapmer PAS Targeting	This study
NPASO-Dicer1C2	[mA][mG][mA][mC][mG][mA][mU][mA][mA][mC][mU][mU][mU][mA][mu][mU][mG][mG][mA][mG]	PS	Dicer PAS Targeting non-gapmer	This study
NPASO - 28	[mU][mA][mU][mC][mU][mC][mU][mA][mC][mA][mC][mU][mG][mA][mU][mC]	PS	Dux4PAS ~200	This study
NPASO - 29	[mG][mA][mU][mC][mA][mC][mC][mU][mA][mA][mG][mU][mG][mA][mU][mG]	PS	Dux4PAS ~200	This study

NPASO - 30	[mC][mA][mU][mC][mU][mG][mC][mA][mC][mU][mG][mA][mU][mC] [mA][mC][mC][mG]	PS	Dux4PAS ~120	This study
NPASO- INTS4-1	[mU][mC][mA][mA][mC][dT][dT][dT][dA][dT][dT][dG][dT][dG][dG][m A][mC][mA][mG][mG]	PS	Integrator4 PAS	This study
NPASO- INTS4-2	[mU][mC][mA][mA][mC][mU][mU][mU][mA][mU][mU][mG][mU][mG] [mG][mA][mC][mA][mG][mG]	PS	Integrator4 PAS	This study
NPASO - 4+22	[mG][mG][mC][mA][mU][mU][mU][mU][mA][mA][mU][mA][mU][mA][mU][mC][mU][mC][mU][mG][mA][mA][mC]	PS, PO	Dux4PAS at PAS	This study
NPASO - 15+16	[mG][mC][mA][mA][mA][mU][mC][mU][mU][mC][mU][mA][mU][mA][mG][mG][mA][mU][mC][mC][mA][mC][mA]	PS, PO	Dux4PAS at Cleavage site	This study
NPASO - 22MS	[mA][mA][mG][mU][mG][mG][mA][mA][mA][mU][mG][mU][mG][mU] [mU][mA][mU][mC][mU][mC]	PO	Dux4PAS at PAS with mismatch	This study

Plasmid transfections

To prepare DNA for transfection, 0.1 -1.0 µg of DNA was mixed with 50µL of OptiMEM. 0.5-2µL Lipofectamine 2000 was mixed with 50µL of OptiMEM and incubated seven minutes, after which the DNA diluted in OptiMEM was mixed with Lipofectamine Reagent diluted in OptiMEM. Lipid DNA complexes were allowed to form during a 25-minute incubation at room temperature. The lipid:DNA complexes were then added to cells.

Microscopy

Cell Fixation and Fluorescence Microscopy.

15Abic CT#24 cells were plated at 25000cells/cm². Cells were transfected with fluorescein tagged ASO (ASO-2F) at concentrations 2.5 µM, 2 µM, 1.5 µM, 1 µM, 500 nM and 100 nM using Lipofectamine 2000. 24 hrs after transfection cells were shifted to differentiation medium (DM). Day 3 after switch to DM cells were washed once with phosphate buffered saline (PBS), and then fixed with 4% paraformaldehyde (Electron Microscopy Sciences, Pennsylvania USA) for 10 minutes. Cells were washed once with PBS and nuclei were stained with a solution of 1:10000 4',6-diamidino-2-phenylindole (DAPI) in PBS for 10 minutes at 37 °C. After DAPI staining, cells were washed twice with PBS and visualized with the Leica DM IL LED Fluorescence Microscope.

Alternatively, cells were visualized without DAPI staining, and bright field/phase contrast and GFP fluorescence images were taken of live cells.

Western Blotting and Immunoprecipitation

General Cell Lysis

HEK293T cells were lysed in low salt lysis buffer (150mM NaCl, 100mM Tris pH 7.5, 1%NP-40). Patient cell lines and HeLa cells were lysed in high salt lysis buffer (500mM NaCl, 100mM Tris pH 7.5, 1%NP-40). Patient cell lines and RD cells were lysed in RIPA buffer (50mM Tris pH 8, 150mM NaCl, 1% NP-40, 0.5% Sodium Deoxycholate, and 0.1% sodium dodecyl sulfate (SDS)). All cells were washed once with PBS and then incubated with rocking at 4°C for 15 minutes with appropriate lysis buffer. Cell lysis was completed with one round of freeze-thaw. Protein concentration of either low- or high-Salt lysed cells was quantified using the Bradford assay and 10µg-20µg of cell lysate was loaded and resolved using either a 10% or 12.5 %SDS-PAGE. Protein samples were prepared in 4X SDS loading buffer (200mM Tris pH 6.8, 0.8%SDS, 20% glycerol, 400mM Dithiothreitol) and boiled for 5mins at 95°C prior to loading on SDS-PAGE.

Immunoprecipitation of Dux4

Protein G Dynabeads (Thermo Fisher Scientific) was washed twice with Phosphate Buffered Saline with Tween 20 (PBS-T), and once with RIPA buffer. Beads were then bound to 10mg of mAB raised to Dux4 in PBS-T, by rotating 10 minutes at room temperature, followed by 1hr at 4°C. After antibody adsorption to the beads, the supernatant was removed and beads were resuspended in 100µL of PBS-T. Immortalized human myoblasts and myotubes grown on 35mm dishes were washed once with PBS and were then lysed with 500µL of RIPA buffer. 400µL of RIPA lysate was added to the 20µL of mAB Dux4 labelled beads. The slurry was rotated overnight at 4°C. The supernatant was removed and cells were washed once with RIPA buffer, and twice with PBS-T. To elute the protein bound to the beads, after removal of PBS-T at final wash, beads were resuspended in 0.1M

Glycine pH 2.7. To which, equal volume of 2X Laemilli buffer (Biorad) was added. After boiling the sample for 5 minutes, the sample was resolved using SDS-PAGE.

Electrophoresis and Immunoblotting

For all SDS-PAGE experiments, lysates were initially resolved using 90V through the stacking gel, and then the voltage was increase to 140V and the gel was run until the 17kDa marker (PageRuler Plus Prestained Marker, Thermo Fisher Scientific) ran off the gel. The proteins were transferred to PVDF membrane. The membrane was blocked 20 minutes in 5% milk PBS-T, then incubated with primary antibody diluted in 5% milk PBS-T for 1hr with rocking at room temperature or overnight with rocking at 4°C. After three, 10-minute washes with 5% milk PBS-T, the membrane was incubated 1hr with 1:5000 dilution of secondary antibody in 5% milk PBS-T. Membrane was washed thrice for 10 minutes each, then rinsed once with PBS. The membrane was then incubated for 5 minutes with 4-IBPA ECL. Chemiluminescent signal was visualized by exposing film in autoradiograph cassette or with the Bio-Rad ChemiDoc Touch Imaging System. Western blots were quantified using Image J software (imagej.nih.gov/ij/) or Image Lab (Biorad).

Antibodies

Primary antibodies used in this project were: mouse α GFP, JL-8 (632380, Clontech, California USA), mouse α Cherry, 1C51 (ab125096, Abcam, Massachusetts, USA), mouse α -GAPDH (AM4300, Thermo Fisher Scientific), HRP conjugated α HA (A190-107P, Bethyl, Texas, USA), HRP Conjugated α Flag-M2 (A8592, Sigma), mouse α -Flag-M2 (A9469, Sigma), rabbit α -Exosc4 (A303-774A, Bethyl), mouse α -MHC (MAB4470, R&D Systems, Minnesota, USA), mouse α -Dux4, 9A12 (MABD116, Millipore, California USA), rabbit α Dux4, E55 (ab124699, Abcam), rabbit α -Tubulin (ab15246, Abcam), and α -myc. Secondary antibodies were Horse Radish Peroxidase conjugated α mouse (715-053-150,

Jackson Immuno Research, Pennsylvania, USA), and Horse Radish Peroxidase conjugated α rabbit (NA934V, GE Healthcare, Buckinghamshire, UK)

Plate Reader Assays

Luciferase Assay

HEK293T cells or RD cells were seeded in a 24 well plate at 1.25×10^5 and 3.5×10^4 cells/well respectively. The following day, the cells were transfected with 100 ng of psiCheck2 reporter plasmid using Lipofectamine 2000 as described above. 48 hrs after transfection cells were lysed in passive lysis buffer (Promega) for 15 minutes at room temperature. Firefly and Renilla luciferase activities were determined using the Dual Luciferase Reporter Assay System (Promega) in the Enspire 2300 Multilabel Reader. Renilla luciferase activity was normalized to Firefly luciferase activity.

Fluorescence Plate Reader Assay

HeLa cells were seeded in a 24 well plate at 8.5×10^4 cells/well. HEK293T cells were seeded in a 96 well plate at 1.3×10^4 cells/well. The following days, cells were transfected with 500ng or 100ng of reporter plasmid, with 100 or 20 ng of transfection control plasmid using Lipofectamine 2000 as described above. Fluorescence intensity was measure at 24 or 48 hrs post-transfection in the Enspire 2300 Multilabel Reader or the Tecan Infinite 200 Pro Plate Reader. Excitation and emission wavelengths of eGFP was selected as 488/509 and the excitation and emission wavelengths of mCherry selected was 575/610.

Nucleic Acid Preparation Analysis and Methodology:

Total RNA Extraction

RNA was extracted from cells using TriZol Reagent (Thermo Fisher Scientific). Briefly, cells were incubated with rocking for 15 minutes at room temperature with TriZol. Chloroform was added to the TriZol reagent and the mixture was vortexed 10 seconds. The samples were incubated for 10 minutes on ice, then centrifuged at 4°C for 20 minutes at 15,000rpm. The aqueous layer was mixed with isopropanol and 10 µg glycogen (AM9510, Thermo Fisher Scientific), and incubated overnight at -20°C or 30 minutes at -80°C. The samples were then centrifuged at 15,000 rpm for 20 minutes at 4°C. The isopropanol was removed without disturbing the RNA pellet, and the pellet was washed once with 75% ethanol. The samples were centrifuged at 4°C for 20 minutes at 15000rpm and then the ethanol was removed. After drying, the pellet was resuspended in 30µL of DEPC treated water (Thermo Fisher Scientific).

Complementary DNA (cDNA) synthesis

RNA for RT-PCR and qPCR were DNase treated as follows. Briefly, 3 – 10ug of RNA was treated with Turbo DNase (Thermo Fisher Scientific) for 30 minutes at 37°C. After the 30 minutes elapsed the DNase was inactivated with the addition of Turbo DNase inactivation buffer, and the mixture was incubated 10 minutes at 25°C with agitation. Following the incubation, the mixture was briefly centrifuged to pellet the DNase. 1 µg of DNase treated RNA was used for cDNA synthesis with M-MLV Reverse Transcriptase (Thermo Fisher Scientific). Reverse transcription reaction was allowed to proceed for 1hr at 37°C, followed by incubation at 95°C for 5 minutes. In the event, where RNA from myoblasts was being prepared to assay Dux4 mRNA expression, 2 – 3 µg of DNase treated RNA was used for cDNA synthesis with Superscript III Reverse Transcriptase (Thermo

Fisher Scientific). cDNA synthesis for RT PCR and qPCR were primed with OligodT₁₂₋₁₈, OligodT₂₀ (Thermo Fisher Scientific), or OligodT₁₈ (New England Biolabs). cDNA for 3' RACE was primed with SP6 OligodT or T7 OligodT.

Rapid Amplification of Complementary DNA Ends (RACE)

Primers used for 3'RACE are listed in Table 3.6. The polyadenylated cDNA pool was subjected to two consecutive rounds of PCR at 20 cycles each to amplify the 3' RACE product. 1µL of cDNA was used as template for the 1st cycle of PCR, with Pfu polymerase, primed with appropriate forward primer and SP6 OligodT or T7 OligodT, as reverse primer in a 50µL reaction. The cycling conditions are 95°C for 3 minutes, 20 cycles of 95°C for 30s, 55°C for 30s, 72°C for 30s, and final elongation at 72°C for 5minutes. After the 1st PCR cycle, 5 µL of PCR product was run on 1-3% agarose gel. 2 µL of the product from the 1st PCR cycle, was used as template for a second round of PCR, using cycling conditions as above. After 2nd round of PCR, 5 uL of the PCR product was run on 1-3% agarose gel. To determine the site of polyadenylation, 20 µL of the final PCR product was run on 1% agarose gel and bands were excised and purified and submitted for sequencing (Genewiz, NJ USA or Lonestar Labs, TX USA with a nested primer. Alternatively, 4 µL of the PCR product was ligated to the Zero Blunt Topo PCR Vector (Thermo Fisher Scientific) according to manufacturer's recommendation. XL1-Blue competent cells were transformed with 1 µL of the ligation reaction and plated on Luria Broth-Kanamycin Agar plates. Amplified clones that contained inserts as determined by EcoRI restriction digest screen were submitted for sequencing with M13F and M13R primers.

RT-PCR

Primers used for RT-PCR are listed in Table 3.6. 2 µL of cDNA was used as template for PCR with Pfu polymerase. The final PCR product was run on 2-3% agarose gel.

PCR products were ligated and transformed as described above, and clones amplified were submitted for sequencing as above.

qPCR

Primers used for qPCR are listed in Table 3.6. 2 μ L of the undiluted cDNA was used in the reaction with SYBR Green master mix (KAPA Biosystems or Biorad) and appropriate primers. Data were acquired using the Stratagene Agilent MX3000P or Biorad CFX Connect, and calculated using $\Delta\Delta$ CT method.

Table 3.6. Table of DNA Oligonucleotides for PCR

Name	Sequence	Purpose
SP6 3'RACE R1	GGCCGGATTTAGGTGACACTATAGTTTTTTTTTTTTTT TTTTTTTTTTTTT	3'RACE Reverse primer
SP6 3'RACE R2	CCGGATTTAGGTGACACTATAG	3'RACE Reverse primer
TSS F1	CCCACTGCTTACTGGCTTATCG	3'RACE Transcription Start Site PASGFP Reporter
TSS F2	CGAAATTAATACGACTCAC	3'RACE Transcription Start Site PASGFP Reporter
TSS F3	GACCCAAGCTGGCTAGCG	3'RACE Transcription Start Site PASGFP Reporter
mCHE-ORF F1	CTTCAAGTGGGAGCGCGTG	3'RACE, qRT-PCR: mCherry ORF
mCHE-ORF F2	GGCGAGTTCATCTACAAGGTG	3'RACE, qRT-PCR: mCherry ORF
mCHE-ORF F3	GCCTCCTCCGAGCGGATG	3'RACE pCCHAM Reporter mCherry ORF
mCHE-ORF F4	GTCAACATCAAGCTGGACATC	3'RACE pCCHAM Reporter mCherry ORF
mCHE-R1	GCTTCAGCCTCTGCTTGATCTC	qRT-PCR: mCherry ORF

T7 3'RACE R1	GGCCGGTAATACGACTCACTATAGTTTTTTTTTTTTT TTTTTTTTTTTTT	3'RACE Reverse primer
T7 R2	GGCCGGTAATACGACTCACTATAG	3'RACE Reverse primer
NRO-G4A	TGAGCAAGGGCGAGGAGCTGTT	qRT-PCR, RTPCR: GFP ORF, Forward
NRO-G4B	AAGCACTGCACGCCGTAGGTCA	qRT-PCR, RTPCR: GFP ORF, Reverse
NRO-G5A	GGGCATCGACTTCAAGGAGGAC	qRT-PCR, RTPCR: GFP ORF, Forward
NRO-G5B	ACGCTGCCGTCCTCGATGTT	qRT-PCR, RTPCR: GFP ORF, Reverse
NRO-NeoF	ATTCGGCTATGACTGGGCAC	qRT-PCR: Neomycin Resistance Gene ORF
NRO-NeoR	CCAATAGCAGCCAGTCCCTT	qRT-PCR: Neomycin Resistance Gene ORF
eGFP RACE 3	GGATCACTCTCGGCATGG	3'RACE: GFP ORF
eGFP RACE 2	GATCACATGGTCCTGCTGGAG	3'RACE: GFP ORF
eGFP RACE 1	GACAAGCAGAAGAACGGCATC	3'RACE: GFP ORF
NeoF2	TCGTGCTTTACGGTATCGCC	RTPCR: Neomycin Resistance Gene ORF
NeoR2	GGCGAAGAACTCCAGCATGA	RTPCR: Neomycin Resistance Gene ORF
N0143	GGCAAGCTGACCCTGAAGTT	qRT-PCR; GFP ORF, Forward

N0144	GGTCTTGTAGTTGCCGTCGT	qRT-PCR; GFP ORF, Reverse
GAPDH F	CAG GAG GCA TTG CTG ATG AT	qRT-PCR, RTPCR
GAPDH R	GAA GGC TGG GGC TCA TTT	qRT-PCR, RTPCR
B-GUSF	GAAAATATGTGGTTGGAGAGCTCATT	qRT-PCR, RTPCR
B-GUSR	CCGAGTGAAGATCCCCTTTTTA	qRT-PCR, RTPCR
18S F	CAGCCACCCGAGATTGAGCA	qRT-PCR
18S R	TAGTAGCGACGGGCGGTGTG	qRT-PCR
ROCK1P1 F	ACACTCTACCACTTTCCTGCCA	qRT-PCR, RTPCR
ROCK1P1 R	TGTGGCACTTAACATGGCGTCT	qRT-PCR, RTPCR
N0D64	GTGAGCAAGGGCGAGGAGC	qRT-PCR: GFP ORF, Forward
N0D67	GGTCAGCTTGCCGTAGGTGG	qRT-PCR: GFP ORF, Reverse
N0129	GGACTTGAAGAAGTCGTGCTGC	qRT-PCR: GFP w/intron Exon-Exon Junction, Forward
N0130	GAG GGC GAT GCC ACC TAC	qRT-PCR: GFP ORF, Reverse
N0157	CATGGTAACGCTGCCTCCAG	RTPCR: Luciferase, Forward

N0158	GGCGATATGAGCCATTCCCG	RTPCR: Luciferase, Reverse
N0159	TCGAGCTGCTGAACCTTCCA	RTPCR: Luciferase, Forward
N0160	CGATCACGTCCACGACACTC	RTPCR: Luciferase, Reverse
N0161	TAGACGGCCTACCCTCTCCT	RTPCR: Luciferase, Forward
N0162	CCAGGGTCGGACTCGATGAA	RTPCR: Luciferase, Reverse
N0163	CCTGGGCTACCTGATTTGCG	RTPCR: Luciferase, Forward
N0164	GAAGCTGAACAGGGTTGGCA	RTPCR: Luciferase, Reverse
N0165	GCCGAGCTGGAGTCTATCCT	RTPCR: Luciferase, Forward
N0166	TCATGGTCTTGCCGTGTTCC	RTPCR: Luciferase, Reverse
N0167	TTCACCGATGCCCACATTGA	RTPCR: Luciferase, Forward
N0168	GTTCTCAGAGCACACCACGA	RTPCR: Luciferase, Reverse
N0169	CAGCGACGATCTGCCTAAGA	RTPCR: Luciferase, Forward
N0170	TCCAACGCTATTGTCGAGGG	RTPCR: Luciferase, Forward
N0171	ATCAAGAGCTTCGTGGAGCG	RTPCR: Luciferase, Forward
N0172	ACCGAGTTCGTGAAGGTGAAG	RTPCR: Luciferase, Forward

NO173	ACGCTCCAGATGAAATGGGT	RTPCR: Luciferase, Forward
NO174	CGTGCTGAAGAACGAGCAGT	RTPCR: Luciferase, Forward
WO851	GTTTACGTCGCCGTCCAGCTC	qRT-PCR; GFP ORF, Reverse
N0D103(RACE)	AGGCGCAACCTCTCCTAGAAAC	3' RACE: Dux4 Exon 1 C-terminus
N0D104(RACE)	GAAGCACCCCTCAGCGAGGAA	3' RACE: Dux4 Exon 1 C-terminus
N0D105(RACE)	GGCTCTGCTGGAGGAGCTTTAG	3' RACE: Dux4 Exon 1 C-terminus
NOD199	CCAGGAGATGTA ACTCTAATCCAGGTTTGC	RTPCR: Dux4 Exon 3, Reverse
NOD200	GCTGGAAGCACCCCTCAGCGAGGAA	RTPCR: Dux4 Exon 1 C-terminus, Forward
NOD201	GAGCTCCTGGCGAGCCCGGAGTTTCTG	RTPCR: Dux4 Exon 1 C-terminus, Forward
NOD202	GGCCCGGTGAGAGACTCCACAC	RTPCR: Dux4 Exon 2, Forward
NOD203	GGCCCGGTGAGAGACTCCACA	RTPCR: Dux4 Exon 2, Forward
NOD204	GCGCACCCCGGCTGACGTGCAA	RTPCR: Dux4 Exon 3, Forward
NOD205	GTA ACTCTAATCCAGGTTTGCCTAGACAGC	RTPCR: Dux4 Exon 3, Reverse
NOD206	CCCCGAGCCAAAGCGAGGCCCTGCGAGCCT	RTPCR: Dux4 Exon 1 N-terminus, Forward
NOD207	CGGCCCTGGCCCGGGAGACGCGGCCCGC	RTPCR: Dux4 Exon 1 N-terminus, Forward

NOD208A	CCTGGTCTGCACTCCCCT	qRT-PCR, RTPCR: Dux4 Exon1 C-terminus, Forward
NOD208B	CTAAAGCTCCTCCAGCAGAGCC	qRT-PCR, RTPCR: Dux4 Exon1 C-terminus, Reverse
NOD208C	GAGCCCGGTATTCTTCCTCG	qRT-PCR, RTPCR: Dux4 Exon1 C-terminus, Reverse
NOD208Anest	GGCGCAACCTCTCCTAGAAA	RTPCR: Dux4 Exon 1 C-terminus, Forward
NOD208Bnest	GAGCCCGGTATTCTTCCTCG	RTPCR: Dux4 Exon 1 C-terminus, Reverse
NOD209A	GCTTTCGTGAGCCAGGCAGCG	RTPCR: Dux4 Exon 1 C-terminus, Forward
NOD209B	CTTGAGCGGGCCCAGGCTGTG	RTPCR: Dux4 Exon 1 C-terminus, Reverse
NOD210A	TCCCAGGGGAGTCCGTG	RTPCR: Dux4 Exon 1 C-terminus, Forward
NOD210B	TTTCTAGGAGAGGTTGCGCC	RTPCR: Dux4 Exon 1 C-terminus, Reverse
NOD211A	CTGGTCTGCACTCCCCTG	RTPCR: Dux4 Exon 1 C-terminus, Forward
NOD211B	CGTCCTAAAGCTCCTCCAGC	RTPCR: Dux4 Exon 1 C-terminus, Reverse
NOD212	GTCTAGGCCCGGTGAGAGAC	RTPCR: Dux4 Exon 2, Forward
NOD213	ATCCACAGGGAGGGGGCATTTTAATATATC	RTPCR: Dux4 Exon 3, Reverse

NOD229	GCTGGTACCTGGGCCG	RTPCR: Dux4 Exon 2, Reverse
NOD230	CTAGGCCCGGTGAGAGACT	RTPCR: Dux4 Exon 2, Forward
NOD231	GGTTTGCCTAGACAGCGTCG	RTPCR: Dux4 Exon 3, Reverse
NOD232	CGTAGCCAGCCAGGTGTTC	RTPCR: Dux4 Intron 1, Reverse
NOD233	AAGGCAGGAATCCCAGGC	RTPCR: Dux4 Intron 1, Reverse
NOD236	GAAAGGCAGTTCTCCGCGG	RTPCR: Dux4 Exon 2, Reverse
Dicer1 Common F	CTCATTATGACTTGCTATGTCGCCTTG	qRT-PCR
Dicer1 Common R	CACAATCTCACATGGCTGAGAAG	qRT-PCR
Dicer1 Distal F	TGCTTTCCGCAGTCCTAACTATG	qRT-PCR
Dicer1 Distal R	AATGCCACAGACAAAAATGACC	qRT-PCR
PRAMEF1	Biorad, PrimePCR SYBR Green: qHsaCED0057477	qRT-PCR, primer pair
TRIM43	Biorad, PrimePCR SYBR Green: qHsaCID0038709	qRT-PCR, primer pair
TRIM48	Biorad, PrimePCR SYBR Green: qHsaCID0022430	qRT-PCR, primer pair
TRIM49	Biorad, PrimePCR SYBR Green: qHsaCED0046590	qRT-PCR, primer pair
ZSCAN4	Biorad, PrimePCR SYBR Green: qHsaCID0036861	qRT-PCR, primer pair

Genomic DNA Extraction

Genomic DNA was extracted from cells using the Rapid Genomic DNA Extraction (RGDE) method (Ali et al., 2008).

Chapter 4 : Development of Tools to Assay mRNA 3'end processing in cells

This chapter is based in part upon Peart N, Wagner EJ. 2016. Gain-of-function reporters for analysis of mRNA 3' end formation: Design and Optimization. *BioTechniques* 60:137-140. doi 10.2144/000114390

© 2009 BioTechniques. Used by Permission

Development of Tools to Assay mRNA 3' end processing in cells.

Introduction

The formation of the 3' end of mRNA is a complex and highly regulated process that is essential for the generation of mature mRNA (Chan et al., 2011). In eukaryotic cells, the process involves the recognition of *cis* sequence elements by several *trans*-acting protein factors (Mandel et al., 2008). A key *cis* element essential for this process is the hexanucleotide polyadenylation signal (PAS). The recognition of the PAS, typically AAUAAA, is aided by the presence of other *cis* sequence elements to stimulate cleavage and subsequent polyadenylation of the mRNA. These other sequences may include a loosely defined upstream sequence element (USE), a G/U-rich or G-rich downstream sequence element (DSE), and the actual cleavage site itself (Tian and Graber, 2012). The combination of these *cis* regulatory elements creates a biosynthetic context that determines whether an mRNA will be efficiently processed (Hu et al., 2005; Tian and Graber, 2012; Wilusz et al., 1990). Perturbation of this context is evident in the etiologies of a variety of human diseases (Danckwardt et al., 2008).

Gain of Function Reporter

Cell-based reporters can be used to analyze mutations in disease-causing genes for their effect on mRNA 3' end formation, as well as to understand the biological mechanism of 3' end formation. Reporter systems have been successfully used to study the requirements of various RNA 3' end formation, including the Nrd1/Nab3/Sen2 complex in *Saccharomyces cerevisiae* transcription termination (Steinmetz et al., 2001), the *cis* and *trans* requirements of histone mRNA 3' end formation (Yang et al., 2009), the role of the Integrator complex in snRNA 3' end formation (Albrecht and Wagner, 2012; Chen et al., 2013a), and more recently to investigate effects of the nuclear cap binding complex on stimulation of 3' end

processing of several RNA families (Hallais et al., 2013). In the case of snRNA or histone pre-mRNA processing the *cis*-acting signals are compact making reporter design straightforward, however, the processing signals required for cleavage and polyadenylation are much more diverse and cover larger ranges (Hu et al., 2005). Here, we describe the construction of an effective, gain-of-function reporter system to analyze pre-mRNA processing and provide an assessment of the reporter's capabilities and limitations.

To demonstrate the design and utility of a GFP expression-based assay to monitor mRNA 3' end processing, we generated a transcriptional read-through reporter (Figure 4.1A). This reporter (termed PAS-GFP) is constructed in the pcDNA3.1 (+) vector (Thermo Fisher Scientific) where sequences to be analyzed are cloned upstream of a GFP open reading frame (ORF), which itself is followed by a bovine growth hormone PAS. The expectation is that if RNAPII encounters a functional cleavage and polyadenylation signal prior to transcription of GFP-encoding mRNA, no fluorescence will be observed. As a positive control, we tested the late SV40 PAS, which has been well-characterized for cleavage and polyadenylation efficiency (Zarkower and Wickens, 1988). Into the PAS-GFP reporter, we inserted a 233-nt-long sequence that included the SV40 PAS, cleavage site, and a DSE. Importantly, we also created two negative control reporters where random sequences devoid of defined cleavage and polyadenylation elements were inserted upstream of the GFP ORF (Figure 4.1A). All reporter constructs as well as an empty PAS-GFP were each transiently transfected into HeLa or HEK293T cells and both protein and total RNA were isolated for analysis 48 hrs post-transfection (as described in (Albrecht and Wagner, 2012; Chen et al., 2013a)). An additional plasmid encoding HA-tagged mCherry was co-transfected to control for transfection efficiency. We observed that only the insertion of the SV40 PAS into the reporter prevented the expression of the GFP protein (Figure 4.1B). Quantification of the GFP expression using ImageJ analysis correlated with qRT-PCR measurement of GFP mRNA (Figure 4.1C/D). It is noteworthy that both random sequences

slightly reduced the expression of the GFP reporter compared to the empty PAS-GFP vector. This is not surprising as some level of translation inhibition would be expected from placement of a ~220nt insertion upstream of the GFP start codon. Along the same lines, it is important to consider that not all random sequences are “inert” and may contain cryptic regulatory elements therefore it is recommended to test multiple negative control sequences.

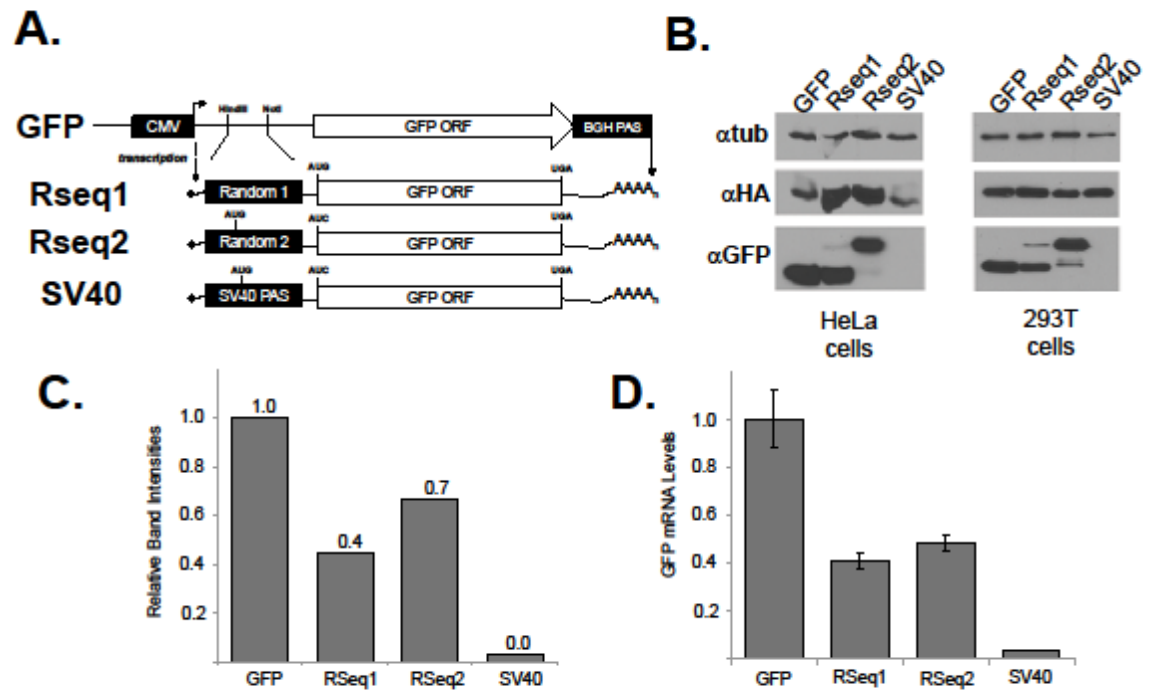


Figure 4.1. PAS-GFP reporter analysis using an SV40 polyadenylation signal (PAS). (A) Schematic of the transcriptional read-through reporter designed to study mRNA 3'-end formation. The GFP open reading frame (ORF) without a native AUG start codon is placed downstream of either the SV40 PAS (SV40-GFP) or random sequences derived from the pUC vector (RSeq-GFP). (B) Western blot of lysates from HeLa cells (left) and HEK293T cells (right) transfected with the reporters shown in panel A. The Western blots were probed for GFP protein using α GFP. HAMCherry protein expressed from a co-transfected HAMCherry plasmid used to normalize transfection was probed with α HA, and tubulin used as a loading control was probed with α -tub. These controls were used in all blots throughout this study. (C) Quantification of the GFP signal in the Western blot of the HeLa lysates shown in panel B (left) was performed using Image J. (D) qRT-PCR analysis of RNA

isolated from the HeLa cells transfected as those analyzed by Western blotting in panel B (left). Error bars show standard deviation from three independent experiments.

To demonstrate sensitivity of the reporter to mutations of the *cis* regulatory element of 3' end processing, we introduced a three-nucleotide mutation (AAUAAA to GGGAAA) of the PAS (PAS null [PN]), and deletion of the DSE (DSE null [DN]) or a mutation of the cleavage site (cleavage null [CN]) (Figure 4.2A) The most significant read-through was observed after mutation of the PAS and to a lesser extent the DSE (Figure 4.2B). It was not unexpected that the cleavage site mutation did not result in significant read-through as this is thought to be the least critical element, as alternative cleavage sites can be utilized if one is mutated (Tian and Graber, 2012). Importantly, we observed strong correlation between GFP protein expression and GFP mRNA levels demonstrating that GFP expression is reflective of mRNA production (Figure 4.2C/D). Finally, 3' RACE and sequencing confirmed that the SV40 PAS located upstream of GFP or the BGH PAS located downstream of GFP represent the only two cleavage and polyadenylation events as predicted (Figure 4.2E). Our data show that while monitoring GFP expression is representative of transcriptional read-through, we did observe slightly greater sensitivity measuring read-through using qRT-PCR analysis, where we measured the GFP mRNA levels normalized to the neomycin resistance mRNA encoded elsewhere within the PAS-GFP plasmid.

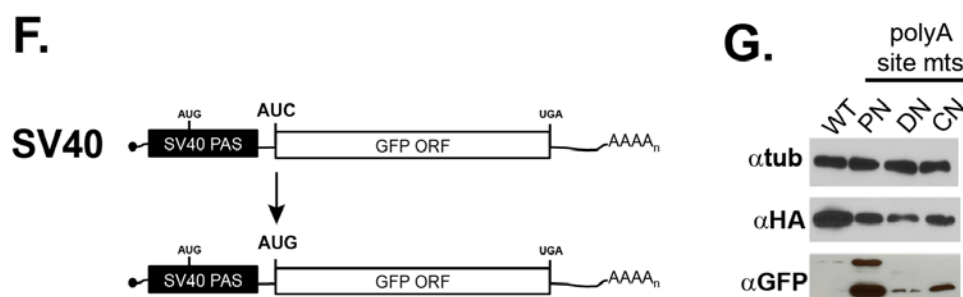
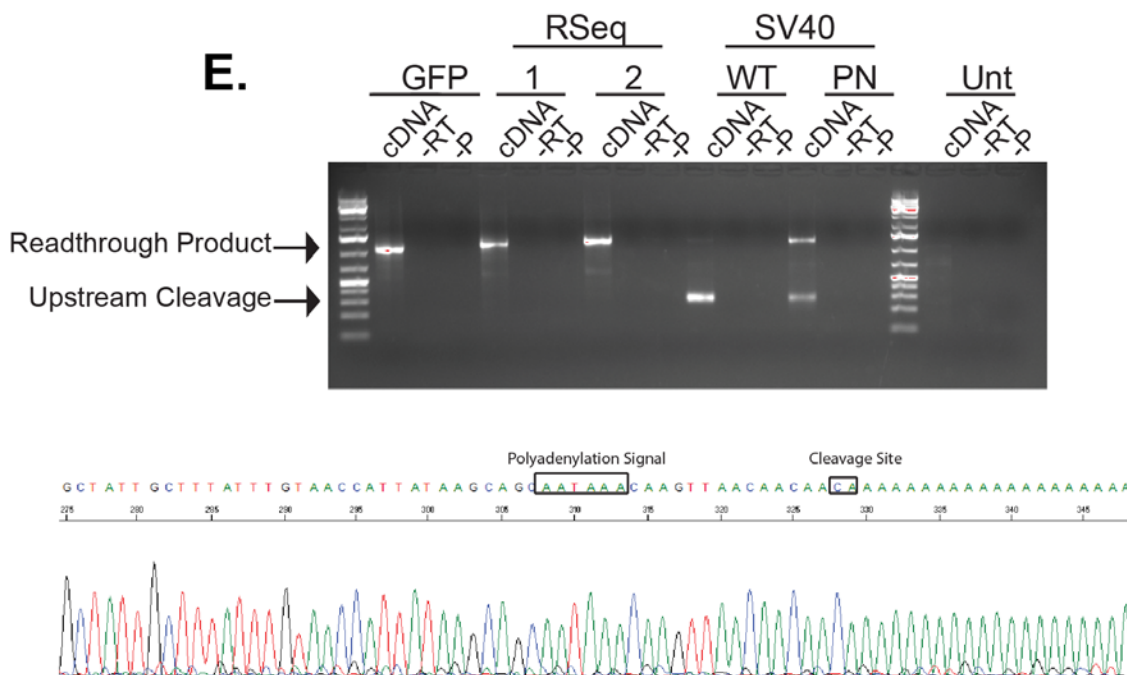
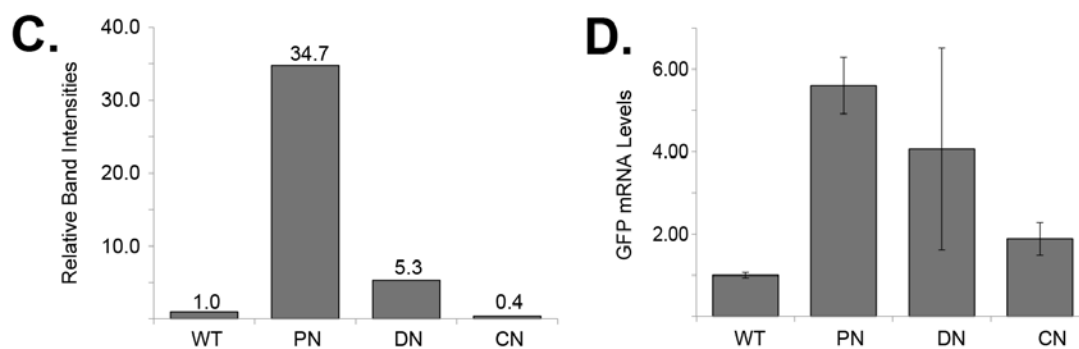
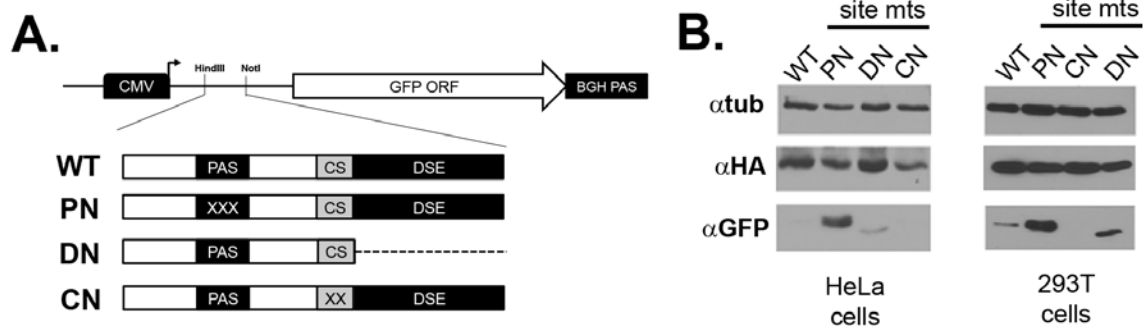



Figure 4.2 Effect of start codon context on polyadenylation signal (PAS)-GFP reporter analysis. (A) Schematic of SV40 mutant GFP reporter constructs with mutations of the PAS from AUAAAA to GGGAAA (PN), deletion of the downstream sequence element (DSE) (DN), and mutation of the cleavage site from CA to CG (CN). (B) Western blot of lysates from HeLa cells (left) and HEK293T cells (right) transfected with the reporters shown in panel A. (C) Quantification of the GFP signal in the Western blot of the HeLa lysates shown in panel B (left) performed using Image J. (D) qRT-PCR analysis of RNA isolated from the HeLa cells transfected with the reporters shown in panel A. Error bars show standard deviation from three independent experiments. (E) Ethidium bromide–stained agarose gel (top) showing products of 3′ rapid amplification of cDNA ends (RACE) of RNA isolated from HeLa cells transfected with the reporters shown in panel A. Sequence chromatogram (bottom) obtained by 3′ RACE showing that the SV40 PAS in the reporter uses the cleavage site previously reported in the literature. (F) Schematic of the transcriptional read-through reporter designed to study mRNA 3′-end formation. The GFP open reading frame (ORF) with a native AUG start codon was placed downstream of the SV40 PAS constructs. (G) Western blot of lysates from HeLa cells transfected with reporter constructs containing the SV40 PASs shown in panel A placed into the reporter plasmid with the native AUG start codon of GFP shown in panel F.

One possible concern for a transcriptional read-through reporter with large inserts upstream is the location and context of the start codon. The constructs shown in (Figure 4.2A) all have the optimized GFP start codon mutated and rely on the start codon fortuitously located within the SV40 polyadenylation signal. We redesigned the PAS-GFP reporter, such that the EGFP ORF cloned downstream of NotI maintains its native AUG (Figure 4.2F). The same SV40PAS and its mutants were then tested in this new context to determine if the presence of an optimal codon would alter the results. While we did not detect any difference in the effects of the SV40 mutants, however, we did observe preferential usage of the native AUG of the EGFP by Western Blot analysis (Figure 4.2G) despite the presence of an upstream start codon. This is most likely because the NotI site flanking the endogenous EGFP start codon is GC rich similar to a Kozak sequence.

Results from recent global analyses of altered RNA 3' end formation events have uncovered a process that is highly dynamic and subject to regulation ((Gruber et al., 2014; Shepard et al., 2011; Xia et al., 2014), and reviewed in (Shi and Manley, 2015)). However, these approaches must be followed up using specific reagents to provide mechanistic understanding on how RNA processing takes place. Typical reporters used by others to investigate mRNA 3' end formation (Gehring et al., 2001; Mayr and Bartel, 2009) place the 3' end processing elements downstream of a heterologous ORF (e.g. luciferase or lobin) and then either the protein or mRNA expression is measured. These reporters have proven to be robust and the primary strength of this design compared to the transcriptional read-through reporter described in this study is that 3' end processing occurs in a more native context, which is downstream of an ORF. However, the potential limitation of reporters with 3' end-processing signals placed downstream lies in the interpretation of loss-of-function events. Mutations that are introduced into these reporters that result in failure to produce the reporter product may be confounded by other factors such as instability of the RNA transcribed due altered 3' UTR content. This concern is mitigated by the design of our

transcriptional read-through reporters. However, the PAS-GFP reporter described here is also not without limitation. As previously mentioned, cryptic regulatory elements in the 5' UTR may impair reporter expression. We addressed one of these limitations by exploring the importance of the start codon context (Figure 4.2). Additional considerations when designing a transcriptional read-through reporter are the presence of upstream ORFs (uORFs), secondary structures limiting start codon recognition, and the potential use of a non-AUG start codon (Araujo et al., 2012; Ivanov et al., 2011). These factors may affect translation of the downstream reporter. These constraints may also place an upper limit on the size of the construct investigated.

Nonetheless, the customizable, gain-of-function PAS-GFP reporter described here offers an immunological and visual output to allow for a case-by-case analysis of mRNA 3' end formation. Moreover, this reporter is readily adaptable for more complicated model systems including *in vivo* expression for tissue-specific analysis of mRNA 3' end formation and genome-wide CRISPR screening to identify novel regulatory factors.

Loss of Function Reporter

While we have alluded to the weakness within loss of function reporters for assaying mRNA 3' end processing in cells, these reporters are still a viable tool for assaying 3' end formation if used appropriately, and the data are interpreted with an understanding of the reporter limitations. Here, we describe a loss of function reporter that addresses two limitations of traditional loss-of-function reporters that assay case specific mRNA 3' end formation in cells. The first limitation is the dependence upon radioactivity by means of northern blot or cell lysis to monitor enzymatic activity of the reporter gene. Here, we use the fluorescent mCherry as reporter. mCherry is a stable monomer with high brightness (Shaner et al., 2004) compared to other fluorescent proteins. We can detect the reporter expression using fluorescence intensity or by immunoblotting. The reporter is also amenable to analysis

of the mRNA by qPCR. The second limitation for assaying 3'end formation using loss-of-function reporters lies in the failure to cleave and polyadenylate at the required downstream PAS. Instead, the polymerase may skip over what it perceives as weak PAS and the reporter may be subject to unintentional alternative cleavage and polyadenylation (APA). To address this limitation downstream of the multiple cloning site into which the interrogated PAS is cloned we placed the self -cleaving ribozyme *Schistosoma Mansoni* Hammerhead (Martick et al., 2008). This reporter, termed pCCHAM (Figure 4.3A) is constructed in the pUC19 vector. The CMV promoter was amplified and cloned in the EcoRI restriction site of the pUC19 vector. We then cloned using annealed oligonucleotide, as described above, the *S. Mansoni* Hammerhead, between PstI and HindIII. HA-tagged mCherry was amplified and cloned into the KpnI and BamHI. Polyadenylation signals interrogated with this reporter were cloned between BamHI and Sall. As a proof of principle, to investigate the utility of the reporter, we cloned the late SV40PAS as described above. We also used site-directed mutagenesis to generate PAS Null, Cleavage Null and DSE null SV40 constructs as described previously in this study for the PAS-GFP Reporter (Figure 4.3A).

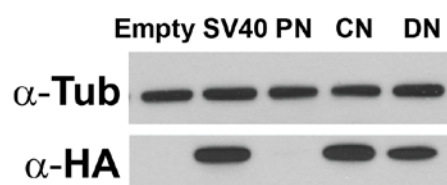
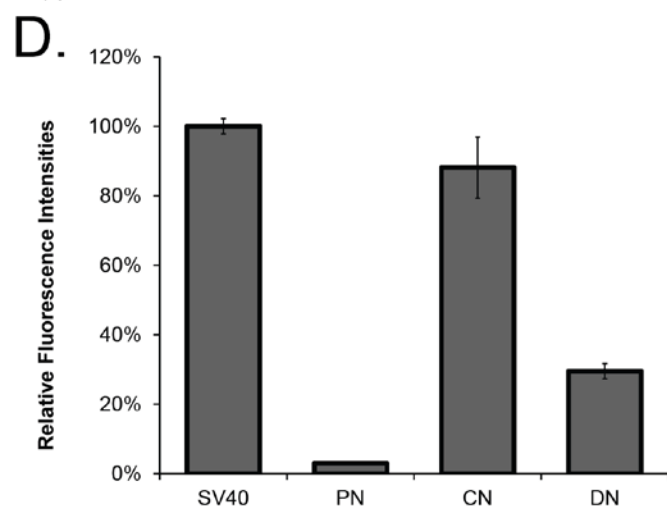
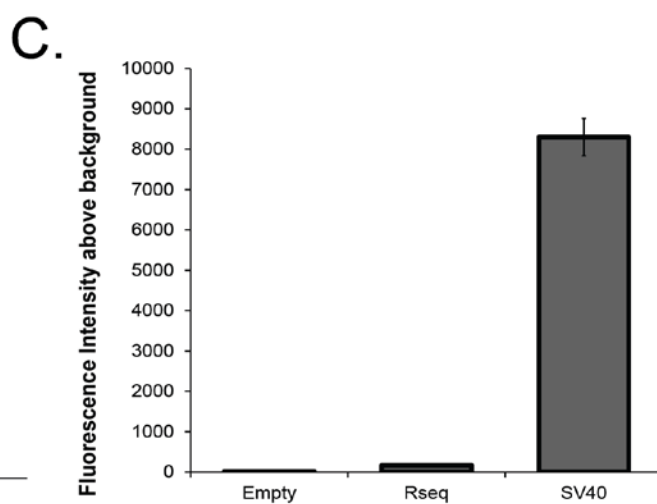
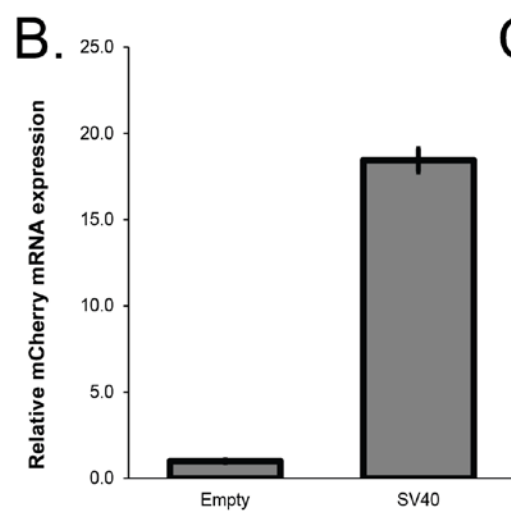
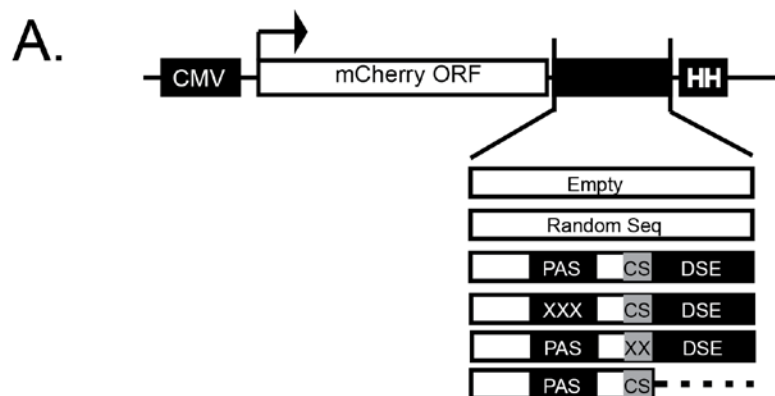


Figure 4.3 pCCHam reporter analysis using an SV40 polyadenylation signal (PAS). A. Schematic of the loss of function pCCHAM reporter designed to study mRNA 3'-end formation. The mCherry open reading frame (ORF) is cloned in the pUC Vector, after a cytomegalovirus (CMV) promoter, and the self-cleaving Hammerhead ribozyme (HH) is placed in the 3' terminus of the multiple cloning site of the vector. Random Sequence (Random Seq or RSeq) derived from elsewhere in the pUC vector or SV40PAS sequences are cloned between the stop codon of mCherry and the Hammerhead ribozyme. (B) qRT-PCR analysis of RNA isolated from the HeLa cells transfected with pCCHAM SV40 or pCCHAM without PAS (empty). Error bars show standard deviation from two independent experiments. (C) Fluorescent plate reader analysis of HeLa cells transfected with pCCHAM SV40, pCCHAM Empty, pCCHAM RSeq. Fluorescence intensity above background was measured relative to untransfected cells, and normalized for transfection using GFP which was co-transfected. (D) Fluorescent plate reader analysis (Top) of HeLa cells transfected with the reporters shown in panel A. Percentage fluorescence intensity of relative to pCCHAMSV40, and normalized for transfection using co-transfected GFP. Western blot (Bottom) of lysates from HeLa cells. The Western blots were probed for HAmCherry protein expressed with α -HA, and tubulin used as a loading control was probed with α -tub.

The reporter was transiently transfected into HeLa cells and protein and RNA were extracted from the cells 48 hrs post transfection. Prior to harvesting RNA or protein from the cells, fluorescence intensity was determined for each construct. We demonstrate that we do not detect reporter expression when there is no PAS, while the SV40 PAS placed downstream of mCherry allows for robust reporter expression (Figure 4.3B/C). Furthermore, we observe a reduction in the expression of the mCherry, when the hexanucleotide PAS is mutated (PN), and the DSE is removed (Figure 4.3D). The cleavage site (CN) mutant likely does not show significant reduction due in part to the promiscuity in the cleavage site selection (Pauws et al., 2001; Tian and Graber, 2012). On the other hand, in our PAS-GFP reporter we still obtained a cleavage product using the CG dinucleotide as a cleavage site, despite its suboptimality (Chen et al., 1995; Furger et al., 1998; Jin et al., 2005). Therefore, these reporters may not be sensitive enough to measure the efficacy of CPA by impairing the cleavage site, or further supports other data (Chen et al., 1995; Furger et al., 1998; Jin et al., 2005) that favor promiscuity in the cleavage site selection.

Nevertheless, this reporter still possess the limitation that reporter output, is subject to UTR destabilization unrelated to CPA, which can obfuscate the results. However, like similar reporters this can be addressed by assaying the mRNA levels in conjunction with the protein output. Nonetheless, it can be useful as a complement to the PAS-GFP reporter to interrogate CPA in a more native context, provided validation in the PAS-GFP reporter supports the interrogated element as a PAS.

Chapter 5 : Exploring the 3' end processing of Dux4 mRNA

FSHD, a disease of the dysregulated 3' end processing?

The importance of the context of the *cis* regulatory elements is evident in disease pathology. Several human diseases can be attributed to dysregulation of 3' end processing of mRNA. Two common human diseases attributed to changes in *cis* regulatory elements that result in disease include β -thalassemia and thrombophilia. Loss-of-function mutation within the PAS of the β -globin gene decreases the efficiency of cleavage and polyadenylation and thereby mRNA production, which consequently impairs accumulation of the β -globin protein and leads to the development of β -thalassemia (Orkin et al., 1985). In contrast, gain-of-function mutations due to CG to CA mutation within the prothrombin gene creates an optimal cleavage site and increases the production of prothrombin mRNA by enhancing cleavage and polyadenylation of the pre-mRNA. This event greatly increases the predisposition toward thromboembolic disorders (Danckwardt et al., 2008; Gehring et al., 2001).

In addition to these well-characterized human pathologies, several other lesser-understood diseases are also associated with changes in the efficiency of 3' end processing of mRNA. Recently, it has been demonstrated that Facioscapulohumeral Dystrophy (FSHD) can also be linked to single nucleotide polymorphisms that generate *cis* elements mediating 3' end formation that affect the CPA of a pathogenic mRNA (Lemmers et al., 2010a).

Individuals with FSHD inappropriately express Dux4. The expression of Dux4 in FSHD muscle cells is not only contingent on D4Z4 derepression, but also requires the presence of a stable collection of single nucleotide polymorphisms (SNPs), or haplotype, in a region of DNA downstream of the D4Z4 repeats. In the general population, there are two equally common haplotypes of chromosome 4, namely A and B. Haplotype A, with which FSHD is associated, is distinguished from haplotype B by the presence of a segment of DNA called pLAM and a 6.2kb degenerate β -satellite repeat immediately downstream of the

D4Z4 repeat (van Geel et al., 2002; Lemmers et al., 2004) (Figure 5.1). Concomitant with the transcriptional depression at the FSHD locus, a SNP in the pLAM creates a functional but non-canonical PAS (AUUAAA) that allows for the production of the Dux4 mRNA (Lemmers et al., 2010a, 2012; Snider et al., 2010). Dux4, a homeodomain gene, is not normally expressed in somatic cells but is robustly expressed in testes and is thus posited to play a role in development (Dixit et al., 2007; Snider et al., 2010; Young et al., 2013). In germline cells, Dux4 is alternatively spliced and uses a PAS distal to that which may be created by SNPs in pLAM (Snider et al., 2010). The inappropriate expression of Dux4 is believed to contribute to muscle damage caused by various mechanisms including defects in myogenesis via induction of apoptosis, immune response stimulation, and inappropriate activation of germline transcription (van der Maarel et al., 2012).

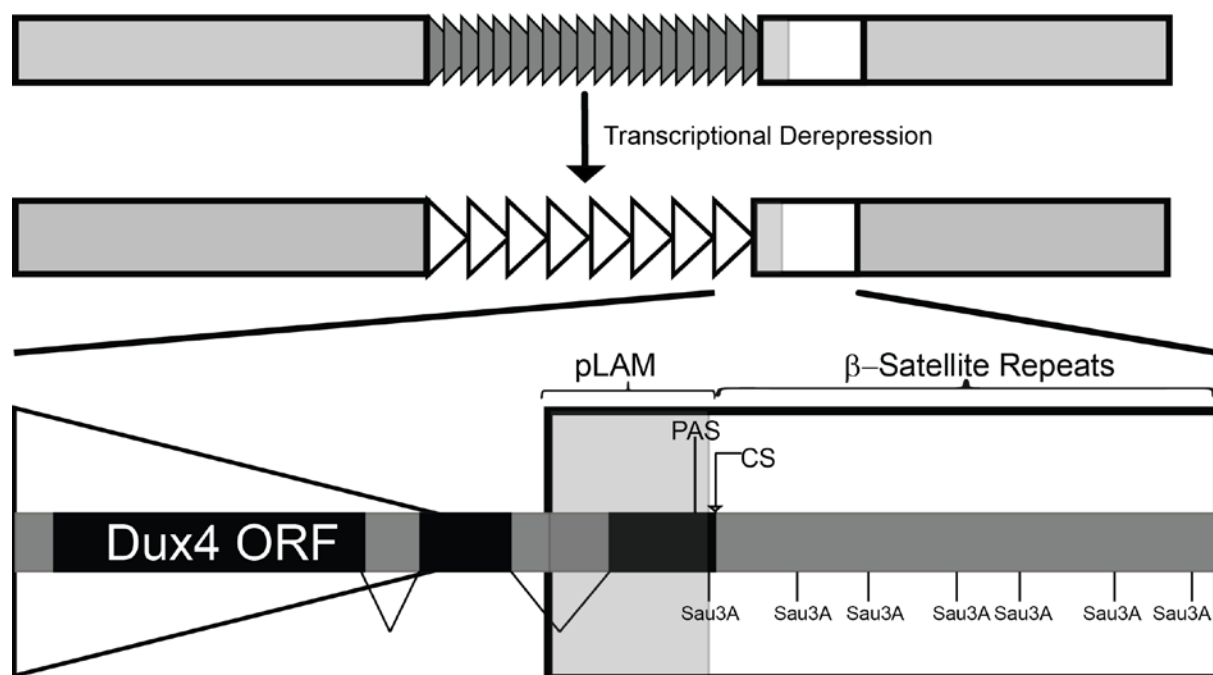


Figure 5.1. Schematic of Chromosome 4qA chromosome. Chromosome 4qA depicted with hypermethylated transcriptionally silenced D4Z4 repeats (gray triangles) upstream of pLAM and the β -satellite repeat region. Reduction in the number of repeats leads to hypomethylation and transcriptional derepression. The image on the bottom magnifies the terminal D4Z4 repeat and downstream sequences, and shows the Dux4 ORF in the D4Z4 repeats. The downstream sequences specific to the haplotype A, shows pLAM with a SNP creating a PAS and the β -satellite repeats defined by recurring sites for the restriction enzyme Sau3A.

In patients with FSHD, the productive transcription of Dux4 mRNA is facilitated by SNPs within certain haplotypes on chromosome 4 that supply a PAS. However, beyond the non-canonical PAS, the *cis* sequence elements that define the context for 3' end formation of Dux4 are undefined. The purpose of this study was to identify *cis* sequence elements that aid in the efficiency of cleavage and polyadenylation of Dux4 mRNA. We anticipated that the Dux4 PAS would be weak because of its non-canonical nature and that it would require additional regulatory elements to mediate 3' end processing. Here, we use a transcriptional read-through reporter (PAS-GFP) described previously in Chapter 4 to interrogate the *cis* requirements of Dux4 cleavage and polyadenylation. With use of this tool, we identified a previously uncharacterized element that is required to mediate Dux4 mRNA 3' end formation. Surprisingly, this element, although potent in enhancing processing, is located more than 100 nucleotides downstream of the PAS and is present within the β -satellite region. This finding underscores the complexity of Dux4 expression and uncovers a potentially new therapeutic target to inhibit Dux4 expression.

β -Satellite Sequences Contribute to Dux4 mRNA 3' End Formation.

The pLAM sequence, which contains the SNP creating the PAS for Dux4, is ~240 base pairs and includes a portion of the terminal intron as well as the terminal exon 3 (Figure 5.1). To test the relative efficiency of this sequence to act as a CPA element, we cloned pLAM into the PAS-GFP reporter (described in Chapter 4) and transfected this construct into HeLa cells and HEK293T (Figure 5.2A). Surprisingly, despite a functioning, albeit non-consensus, PAS present within pLAM, we observed a significant amount of GFP expression relative to the SV40-GFP reporter. This could indicate that (1) the Dux4 PAS is significantly weaker than SV40, (2) other sequence elements are required for efficient CPA, (3) Neither HEK293T nor HeLa cells can recapitulate Dux4 CPA to the extent of muscle tissues, or (4) the small portion of the terminal intron that was included in the pLAM-GFP

reporter somehow disrupts recognition of the Dux4 PAS (Figure 5.2B/C, lane 2). To address this last point, we also generated a PAS-GFP reporter containing only exon 3 upstream of GFP. We observed that removal of the partial intron resulted in a slight decrease in GFP expression; however, GFP protein was still readily detectable (Figure 5.2B/C, lane 3).

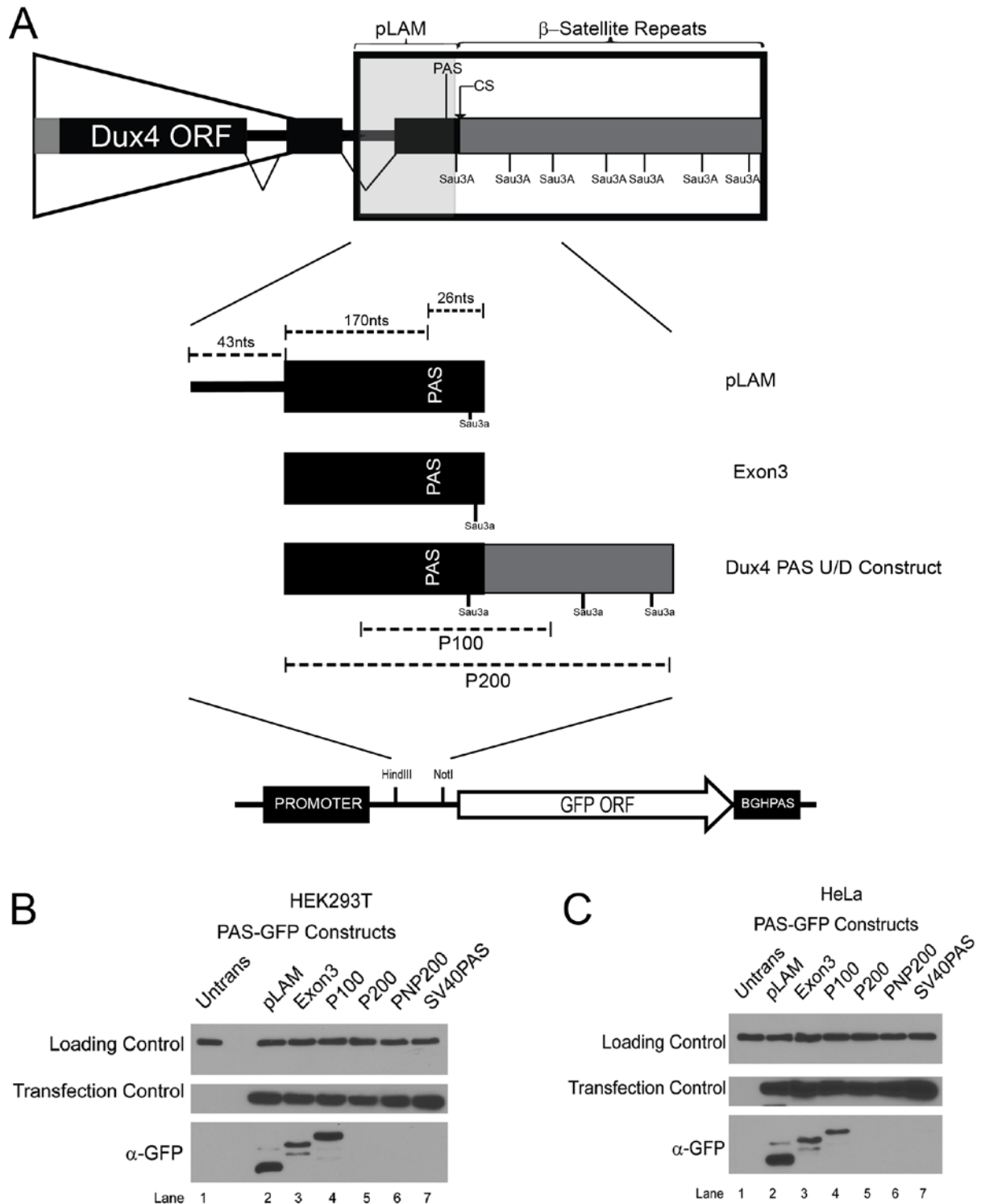


Figure 5.2. Identifying the minimal region necessary to suppress GFP. (A). Schematic of the terminal D4Z4 repeat pLAM and β-satellite to show the Dux4 constructs cloned upstream of

the GFP ORF. (B and C). Western blots of cell lysates for HEK293T (left) and HeLa cells (right). Loading control was α -tubulin, transfection control was α HA to recognize co-transfected HAmCherry.

We then generated two more Dux4 reporters into which we inserted genomic fragments containing increasing amounts of DNA flanking the pLAM PAS to include either ~100 nucleotides on either side of the PAS or ~200 nucleotides (Figure 5.2A). We observed a total suppression of GFP expression only when ~200 nucleotides of sequence flanking on either side of the pLAM PAS was included in the PAS-GFP reporter (Figure 5.2B/C, lanes 3-4). Finally, we conducted 3' RACE on RNA isolated from the pLAM-GFP, Exon3-GFP, P100-GFP or P200-GFP transfected cells. We observed that the primary product was due to transcriptional read-through for the pLAM, Exon 3 and P100 sequence; however, when the P200 sequence was used, we observed a small upstream cleavage product (Figure 5.3). Sequencing of this smaller fragment confirmed that the Dux4 PAS is used, although the cleavage site mapped to 1-2 nucleotides upstream of the annotated sequence.

On the basis of the observation that additional sequences located on either side of the Dux4 PAS are required for its efficient CPA, we sought to increase the resolution of our analysis by using deletions. To first determine whether *cis* elements lie upstream or downstream of the Dux4 PAS, we created two additional reporter constructs. The Δ D-GFP reporter lacked 100 nucleotides from the 3' terminus of P200, thereby removing the second β -satellite repeat, whereas the Δ U-GFP reporter retained only 100 nucleotides upstream of the PAS (Figure 5.4A). In agreement with results presented in Figure 5.2, we observed GFP expression from the P100-GFP reporter but not from the P200-GFP reporter; moreover, we noted GFP expression from the Δ D-GFP reporter but not from the Δ U-GFP reporter (Figure 5.4B/C). These results demonstrate that sequences that were removed in the Δ U-GFP reporter are dispensable for Dux4 CPA, whereas sequences within the second β -satellite repeat are required for Dux4 CPA. Altogether, these results indicate that sequences downstream of the Dux4 PAS are required for its efficient use.

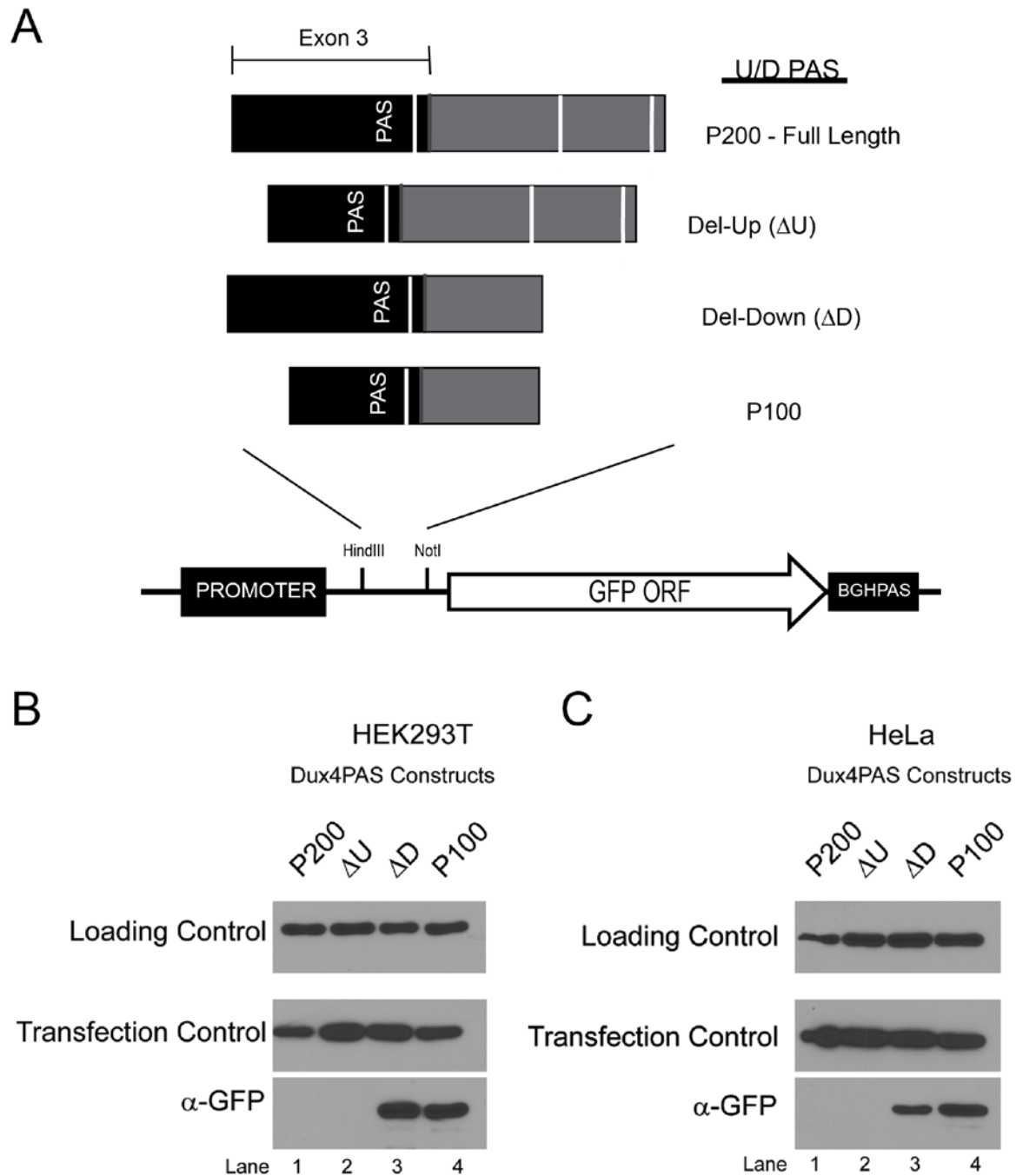


Figure 5.4. Elements for Dux4 CPA lie downstream of PAS. (A). Schematic of Dux4 CPA constructs cloned upstream of GFP. (B and C). Western blot of lysates from HEK293T and

HeLa cells transfected with the constructs in (A). Loading control and transfection control are α -tubulin and α HA respectively.

Downstream Auxiliary Elements aids in cleavage and polyadenylation of Dux4.

To further delineate the location of the downstream element, we performed deletion mutagenesis, in 20-nucleotide increments, of the terminal 100 nucleotides of the P200 sequence (henceforth called full-length, FL) (Figure 5.5A). By Western blot analysis, we observed that we gained GFP expression when the last 40 nucleotides were removed from the FL Dux4 CPA construct (Figure 5.5B). Following up with a fluorescence plate reader assay to increase sensitivity, we observe that there is complete loss of GFP fluorescence only in the FL construct, suggesting that an element lies within the last ~20 nucleotides. This discrepancy may be as a result of the amount of protein loaded, and/or the sensitivity of the GFP antibody used.

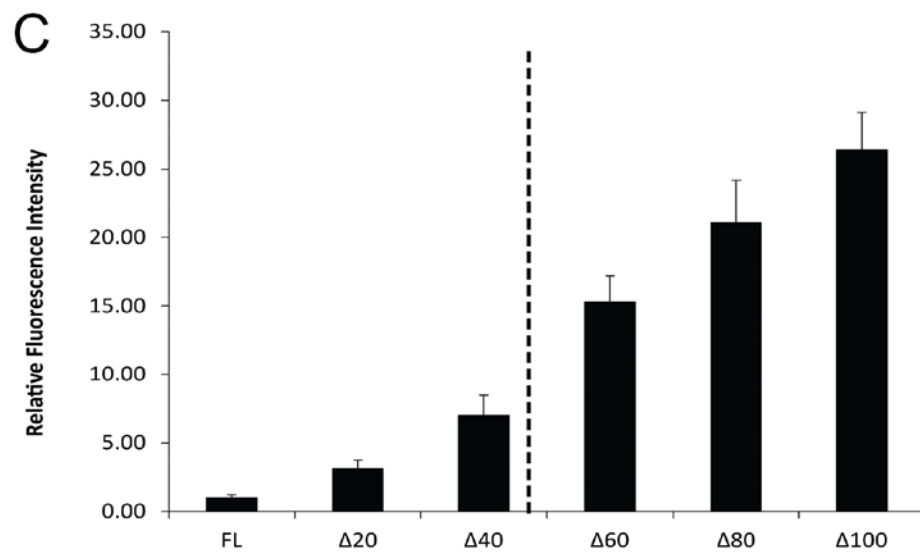
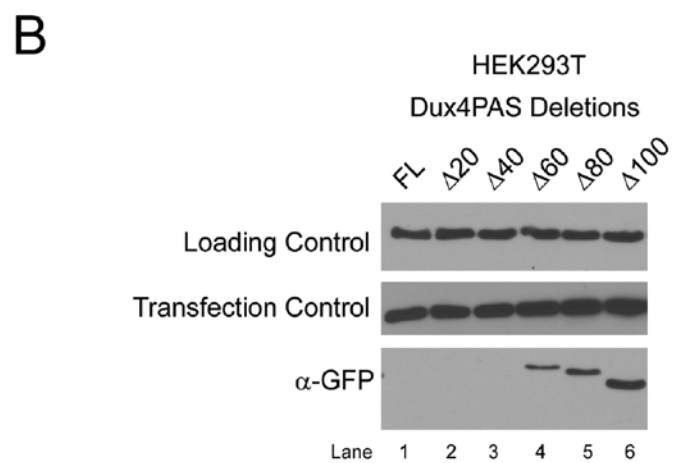
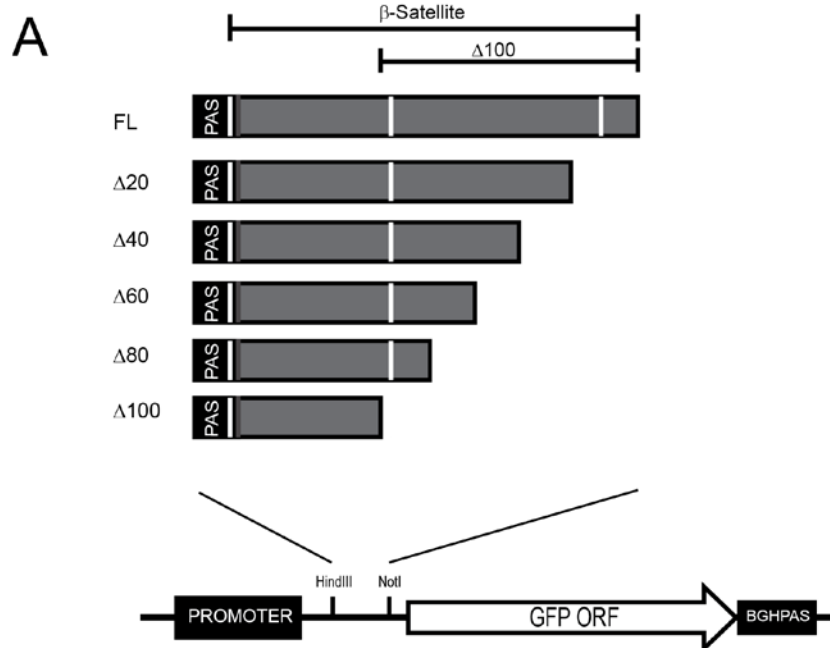


Figure 5.5. Deletion series identifies element in the extreme 3' terminus of the Dux4 CPA construct. (A). Schematic of deletion constructs of Dux4 FL cloned upstream of GFP (B). Western blot of lysates from HEK293T cells transfected with reporter containing Dux4 PAS constructs shown in A. Loading control and transfection control are α -tubulin and α HA respectively. (C). Quantification of data from plate reader assay of GFP fluorescence normalized to co-transfected HA tagged mCherry. Hatched vertical line indicates where the sensitivity of the Western blot is diminished. Error bars calculated from standard deviation of biological triplicates.

To achieve even greater resolution of the regulatory element, we compared the constructs $\Delta 40$ and $\Delta 60$, and further made transversion mutations in six-nucleotide clusters within the region that the $\Delta 40$ possess and the $\Delta 60$ lacks (Figure 5.6, left). We observed that there is an increase GFP expression, when mutations of the last six nucleotides are made, which reproduced using a fluorescence plate reader (data not shown).

We further assessed the validity of our data, by making transversion mutations in the context of our full-length Dux4 CPA (Figure 5.7, left). Due to the weak signal we quantified the expression of the GFP from three different experiments, and our analysis of the western blots (Figure 5.7, right) re-identifies the *cis* regulatory element (mutant C) identified in the context of the $\Delta 40$ mutants in (Figure 5.6). In addition, we identify near the 3' end of the full length Dux4 CPA another element (mutant K), wherein mutation of this sequence resulted in increased GFP expression. In the second instance we mutated AGAGA to CTCTC, while the first element identified in both point mutation analyses share an overlap of the nucleotide AAC mutated to CCA. Mutated sequences that increase GFP expression were over 150 - 130 nucleotides downstream of the Dux4 polyadenylation signal and cleavage site, respectively. Here we identify downstream sequences that are required to suppress GFP expression which we correlate with facilitating cleavage and polyadenylation at the Dux4 PAS. Nonetheless, we acknowledge that confirmation of their importance for 3' end processing requires the analysis of the RNA levels. A limitation of the PAS-GFP reporter is a consequence of the immunoblotting or fluorescent readout for the analysis of CPA events. To ensure that the reporter protein is produced, it is imperative that the upstream PAS: (1) is cloned in frame, (2) contains no in frame stop codons or start codons if the native reporter start codon is used, and (3) contains no start codons out of frame with the reporter protein ORF. All these consideration as well as those discussed in Chapter 3, limit the scope of the PAS-GFP reporter while site-directed mutagenesis to remove these elements may interrupt or create alternative *cis* regulatory elements.

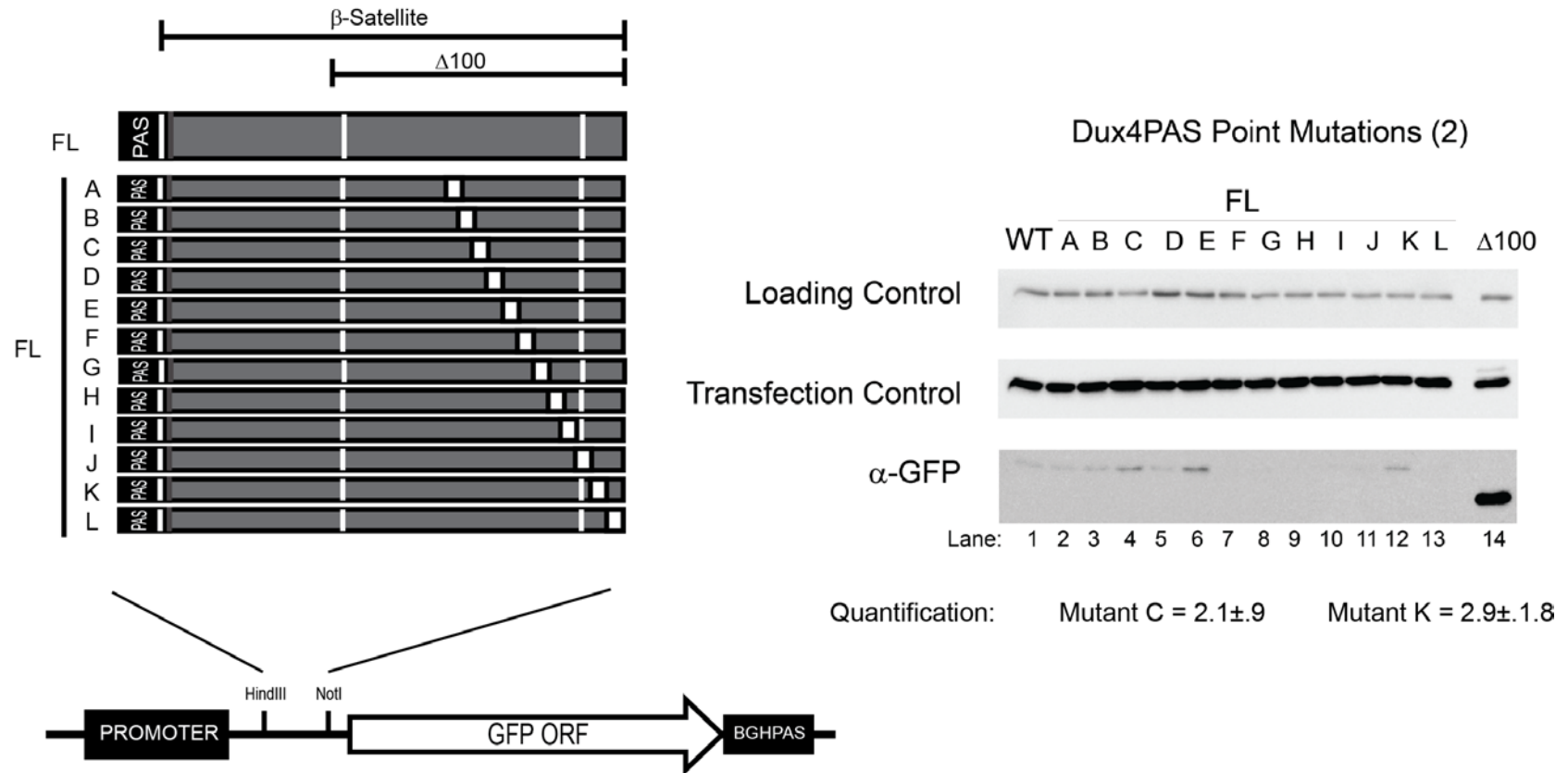


Figure 5.7. Effects of mutation in Dux4 Downstream Auxiliary Element (DAE). (Left) Schematic of the Dux4PAS constructs depicting transversion mutations of five nucleotides made in the DAE of Dux4. (Right) Representative western blot of Dux4PAS constructs depicted on left. Quantification of western blots from 3 experiments performed using Biorad Image Lab software, and mutants with reproducible increases in GFP expression shown. Loading control and transfection control are α -tubulin and α HA respectively.

Evaluating Dux4 3' end processing “natively”

To further assay Dux4 processing we placed the Dux4 P200 construct downstream of the mCherry ORF. Surprisingly, we did not detect robust mCherry expression, although we observed a 3' RACE product from the P200 construct (not shown). To determine if the UTR is destabilizing we used the dual luciferase reporter system and placed the terminal exon of pathogenic Dux4, Exon 3, into the psiCheck 2 vector (Figure 5.8). However, we did not see any decrease in luciferase expression. Consequently, we switched models, and moved the 3' UTR into a splicing capable reporter (Figure 5.9A). As a consequence of frequent PCR errors and recombination events due to the repetitive nature of the Dux4 sequence, we opted to further modify the region cloned downstream of the GFP ORF such that we retained 100 nucleotides upstream of the Dux4 polyadenylation signal and 500 nucleotides downstream of the PAS, as we previously showed that elements downstream of the PAS were critical for 3' end processing (Figure 5.2). When the Dux4 CPA elements were placed downstream of an intron containing GFP, we observed robust protein expression, similar to SV40. Moreover, the strength of the GFP expression was directly related to presence of downstream auxiliary elements of Dux4 (Figure 5.9B). 3'RACE of the reporter containing solely the Dux4CPA produced a polyadenylated product cleaved as reported in the literature and as observed with our PAS-GFP reporter. However, we also observed that there was another band present which when sequenced showed a cleavage product due to utilization of a PAS (not identified) in the vicinity of the GFP stop codon. This product would correspond to GFP transcript with a short UTR (Figure 5.9C). Due to the lack of any destabilizing elements in the Dux4 UTR cloned as determined by luciferase assay, we speculate that the upstream splicing aided in cleavage and polyadenylation with the Dux4 PAS. Alternately, the usage of an alternate polyadenylation signal in the reporter containing the full Dux4 CPA, may result in a robust protein production not dependent on an

mRNA cleaved and polyadenylated at the Dux4 CPA element. However, this is unlikely, as we observe that substituting the 100 nucleotides upstream of the Dux4 polyadenylation signal with the minimal SV40PAS upstream sequence elements, also produced robust GFP expression and we detect only a single band in the 3' RACE (Figure 5.9C). Moreover, sequencing of the PCR product showed that the cleavage and polyadenylation event occurred at the same location as used by the full intact Dux4PAS. In contrast, creating a reciprocal construct, in which the downstream sequence elements (DSE) of SV40PAS are placed after the Dux4 polyadenylation signal, we observed that the primary cleavage and polyadenylation event occurs downstream of the SV40 DSE, suggesting the use of another unidentified polyadenylation signal (Figure 5.9).

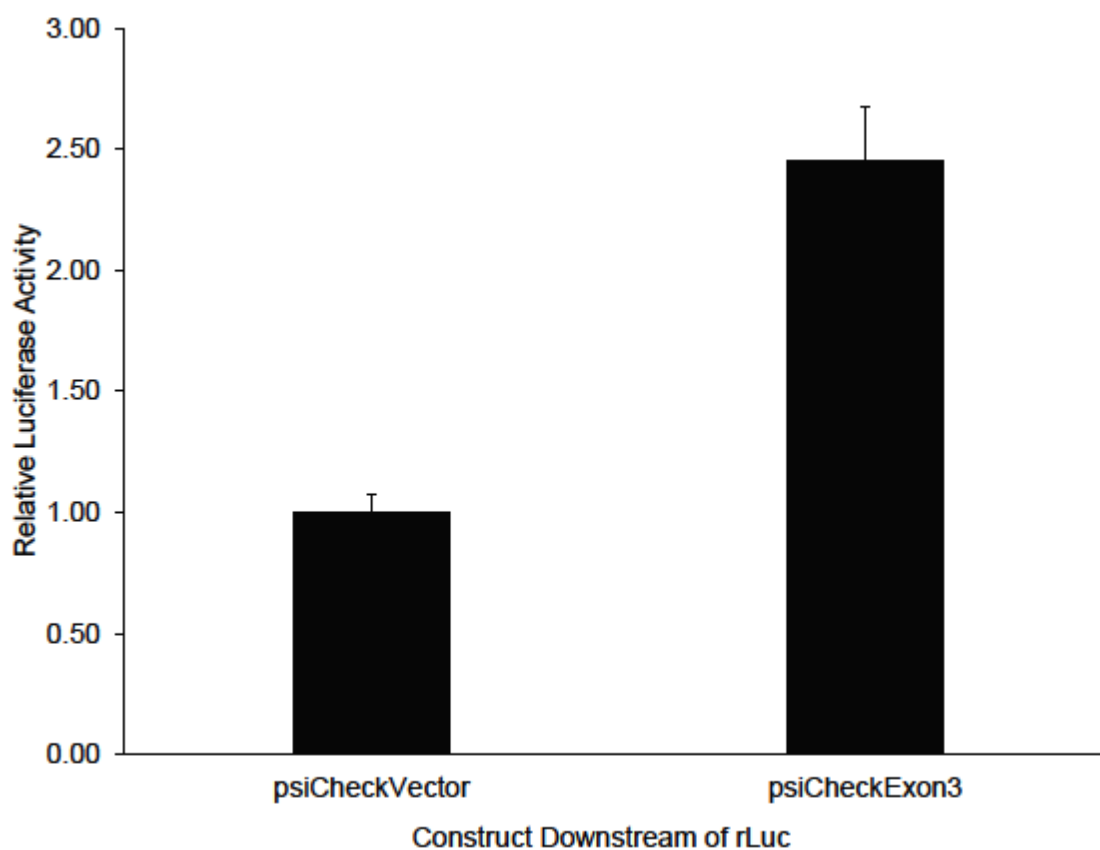


Figure 5.8. Histogram of terminal Exon 3 of pathogenic Dux4 in dual luciferase reporter. Exon 3 cloned downstream of Renilla luciferase and transfected into HEK293T cells. Renilla luciferase expression was normalized to Firefly luciferase. Error bars shown depict standard deviation of biological triplicates.

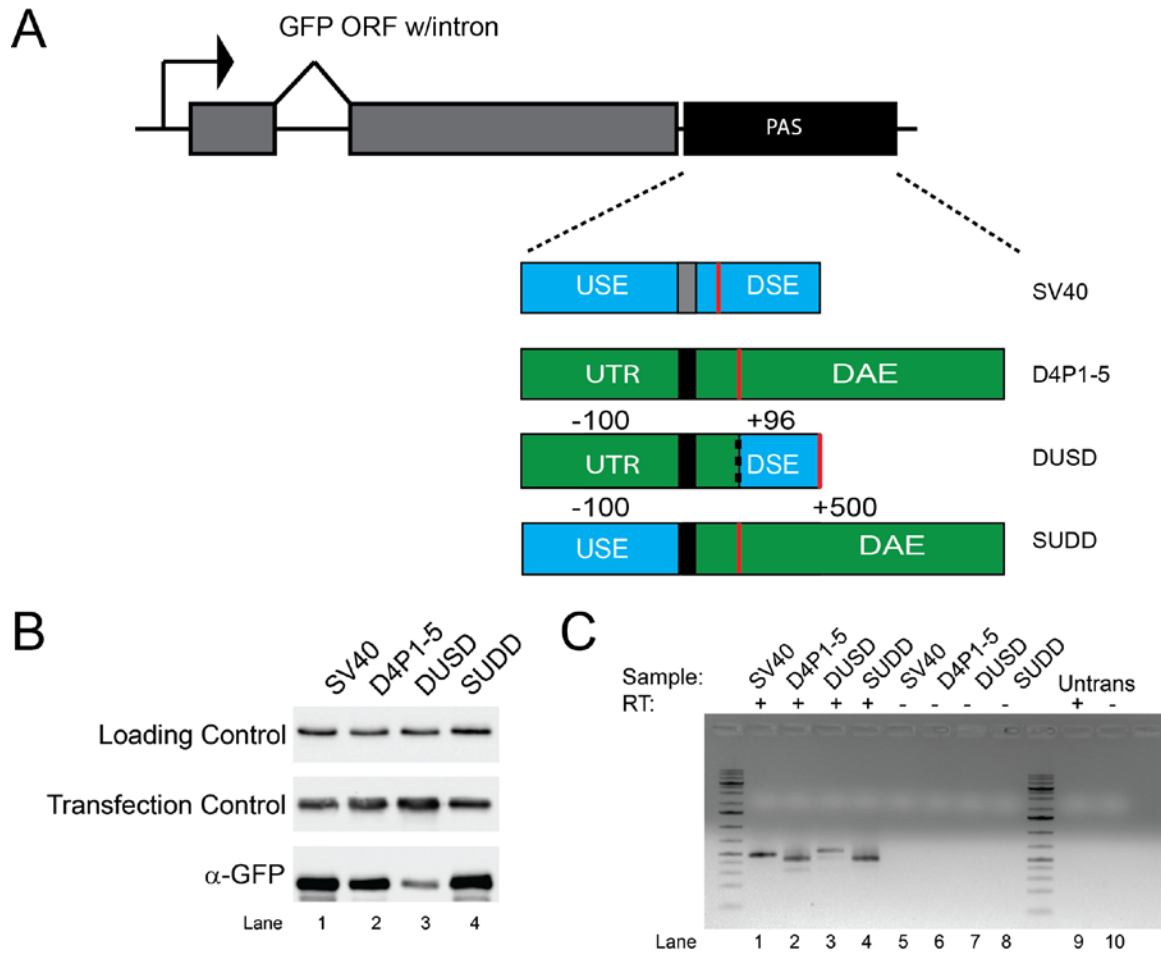


Figure 5.9. Reporter gene expression as a result of Dux4PAS. (A). Schematic of reporter showing intron-containing GFP, cloned downstream are the SV40PAS, Dux4PAS and chimeric constructs which retain the Dux4 hexanucleotide polyadenylation signal, but have either sequence upstream or downstream of the polyadenylation signal substituted for the complementary elements in the SV40PAS. Blue boxes represent SV40 sequences and green boxes represent Dux4 sequences. In pictogram of SV40 gray rectangle denotes polyadenylation signal and red line indicates cleavage site. In the pictogram of the Dux4 construct, black rectangle indicates the polyadenylation signal, and red line shows the cleavage site used. The chimeric construct DUSD contains 100 nucleotides of Dux4 UTR sequences upstream (green box) of the Dux4 PAS (black rectangle), and extends to the

cleavage site (dotted line). The 96 nucleotide SV40DSE is immediately downstream of the cleavage site. Red line denotes the cleavage site used. The chimeric construct SUDD contains 100 nucleotides of SV40 upstream sequence elements (blue bar), the Dux4PAS (black rectangle) and the Dux4 DAE (green bar). The cleavage site used is indicated by the red line. (B). Representative western blot HEK293T cells transfected with constructs shown in A. Transfection control depicts co-transfected myc tagged mCherry. Loading control and transfection control are α -GAPDH and α -myc respectively. (C). Ethidium bromide stained 2% agarose gel showing 3' RACE products obtained from the HEK293T cells transfected with constructs depicted in A. RT denotes reverse transcriptase, and + or -, represents the presence or absence of reverse transcriptase in the cDNA synthesis. Untransfected HEK293T cells were also included in the 3'RACE.

Design and Use of ASO targeting CPA

The utility of RNA as both a target and a modality to influence gene expression is rapidly becoming mainstream. This has been spurred by recent advances in nucleic acid based technology such as clustered regularly-interspaced short palindromic repeats (CRISPR), and improvement in commercial availability of chemically modified nucleic acids (Haussecker, 2016). One of the broadest methods to target RNA and modify gene expression uses antisense mechanisms (reviewed (Kole et al., 2012; Potaczek et al., 2016)). This is exemplified by the increasingly common place use of RNA interference (RNAi) and antisense oligonucleotides (ASOs). The simplest definition of an ASO would describe a short polymer of the nucleotides A, T, U, C or G as a ribose or deoxyribose sugar, which is used to bind RNA or DNA molecules through base pairing. ASOs are typically modified to increase affinity, stability and accessibility. Increases in the stability, that is resistance to nucleases, can be accomplished by modifying the backbone chemistry; for example, using phosphorothioate bonds in lieu of a phosphodiester bonds for polymer conjugation. However, while a phosphorothioate backbone increases ASO stability, it also reduces the affinity of the ASO for the target and introduces chirality (Koziolekiewicz et al., 1995), and indeed alternative backbones have been investigated to achieve increased nuclease resistance without loss of affinity (Freier and Altmann, 1997). Additional modifications, when an RNA template ASO is in use, include substituting the reactive 2' hydroxyl group with a more stabilizing or neutral group. With such substitutions, the loss of binding affinity caused by using a phosphorothioate backbone can be rescued (and further improved). Likewise, in the context of a phosphodiester backbone, substitution of the 2' hydroxyl group can increase nuclease resistance of ASOs (Monia et al., 1993, 1996). Common substitutions are 2'-O-methyl or 2'-O-fluoro. Alternatively, nucleoside analogues

such as peptide nucleic acids (PNA), locked nucleic acids (LNAs), phosphorodiamidate morpholino oligomers (PMOs) may be used.

Antisense strategies, can be exploited in many ways for example, RNAi is a routinely used antisense strategy and employs small RNAs (short interfering RNAs, microRNAs) incorporated in the RNA induced silencing complex (RISC) machinery to decrease translation of target mRNA and degrade the RNA (reviewed by (Iwakawa and Tomari, 2015; Jonas and Izaurralde, 2015)). In the case of ASOs, the chemistry and design affects the behavior when it interacts with the target substrate. One such application ASOs is to degrade the RNA and downregulate gene expression. This method takes advantage of the nuclease, RNase H, which cleaves RNA molecules in RNA:DNA hybrids (reviewed (Bennett and Swayze, 2010; Crooke, 2004)). However, not all ASOs are meant to degrade RNA. Fully modified antisense oligonucleotides which do not induce RNase H action ASOs can be and have been used to repurpose the molecule by affecting how it is processed (reviewed (Bennett and Swayze, 2010; Kole et al., 2012)). The best studied application in modulating RNA processing, is the use of ASOs to change splicing events. In this instance, the ASOs are designed to bind RNA and sterically impair protein:RNA interactions, such as those between splicing factors and exonic splicing silencer or enhancers, thus promoting exon inclusion or skipping. The application of ASOs in redirecting splicing has been widely explored, however, there are other RNA processing events which can be interrogated by splicing (Kole et al., 2012; Sazani and Kole, 2003).

Cleavage and polyadenylation of mRNA is a tenable target for application of ASOs, and others have shown that using CPA targeting ASO can redirect polyadenylation signal choice or impair CPA altogether (Marsollier et al., 2016; Vickers et al., 2001). Moreover, they have been used to elucidate molecular mechanisms behind RNAPII termination events (Zhang et al., 2015). Here, we design steric blocking ASOs targeting CPA elements of endogenous genes and reproduce a redirection of polyadenylation signal choice. In addition,

we use steric blocking and RNase H sensitive (gapmer) ASOs to interrogate the elements critical for 3' end formation of Dux4 mRNA in a reporter system, and to downregulate the endogenous expression of the gene.

Short Antisense Oligonucleotide shows primarily nuclear localization

We designed a short, 15 nucleotide ASO, with a fluorescein moiety at the 5' end, ASO-2F (Table 3.5) and transfected this nucleotide into RD cells, HEK293T, HeLa, and an immortalized FSHD patient myoblast cell line, 15Abic CT#24. HEK293T cells were transfected with ASO 2F at concentration ranging from 100nM to 2.5 μ M, and cells were visualized by microscopy twenty four hours after transfections. The immortalized patient myoblast cell line was transfected at the same concentrations of ASO. Twenty four hours after transfection the cells were shifted to differentiation medium, and cells were ultimately visualized by microscopy five days after inducing differentiation (see Chapter 3, methods).

While there was minimal toxicity to HEK293T cells to even the highest dose of ASO used, we did observe that the 15Abic cells were less resilient. The toxicity of the higher doses (>500nM) of ASO-2F to 15Abic cells is likely to account for the decreased myotube formation observed (Figure 5.10). Furthermore, we observed that in addition to the toxicity, the localization of ASO-2F in 15Abic cells was primarily cytoplasmic at lower concentrations, however, the ASO was enriched in the nucleus with increasing concentrations. In contrast, ASO-2F was primarily nuclear (Figure 5.11) at low concentrations in HeLa, HEK293T and RD cells. However, as the ASO concentration increased, the localization became more diffuse and punctate (Figure 5.12). Differences in ASO behavior dependent on the cell line has been previously reported (Crooke, 2004). Another consideration, that after six days the fluorescein group on the ASO may have been hydrolyzed and the observed fluorescent signals were the consequence of free fluorescein. However, we do observe that there

appears to be diffuse localization of the ASOs in the cells even after changing the medium at 4 hrs post transfection.

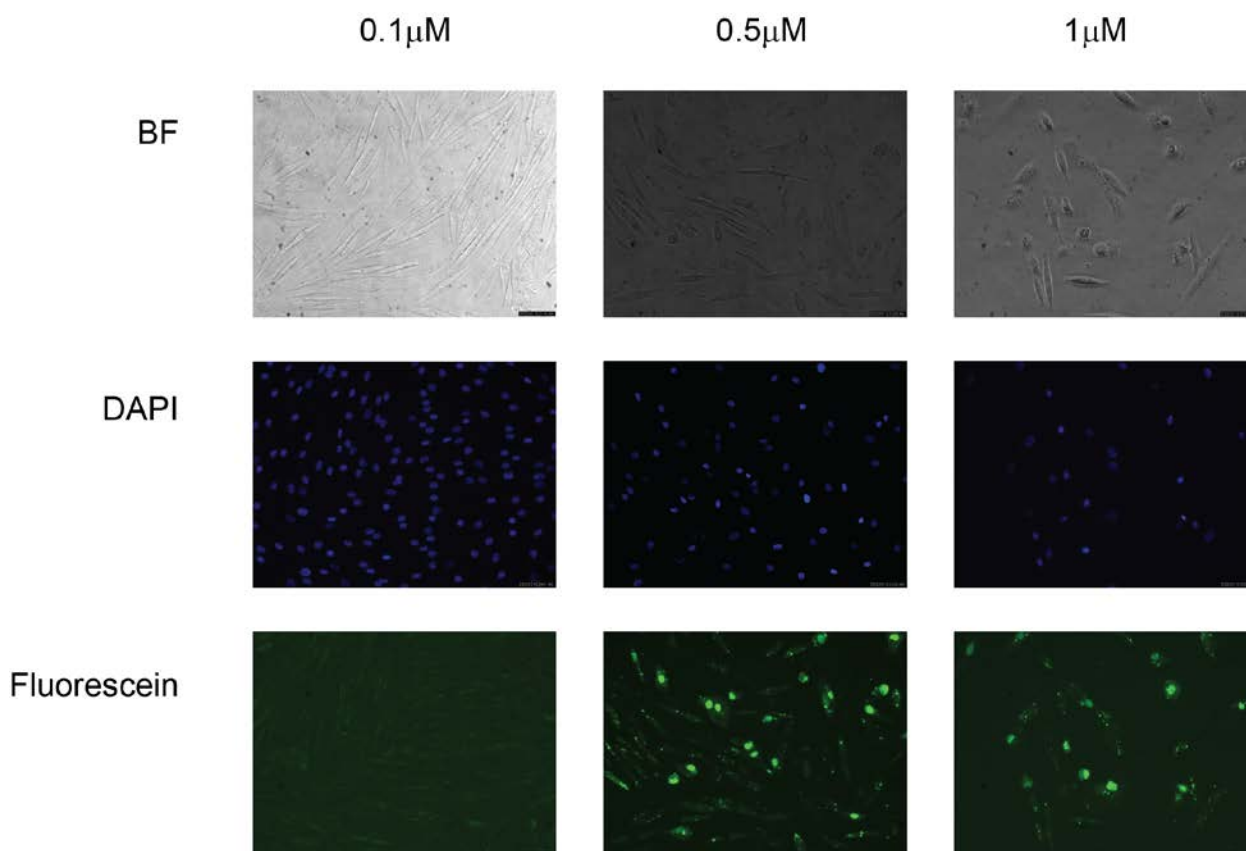


Figure 5.10. Fluorescein labelled Antisense Oligonucleotide transfected into immortalized FSHD patient cells. Images of 20X magnification of 15Abic CT#24 cells, five days after initiating differentiation, six days after transfection with fluorescent ASO. Phase contrast images (BF), DAPI and green fluorescence images shown, for cells transfected with increasing concentrations of fluorescent ASO. Toxicity observable at higher concentrations (1 μ M) where there is an increase in the balled cells as well as decreased cell density. Transfection of lower concentrations (0.1 and 0.500 μ M) had less toxicity and cell morphologies were consistent with healthy growing cells upon verge of differentiation, although we do not observe robust myotube formation (multinucleated cells). Increasing concentration of ASO resulted in more diffuse localization of ASO, as well as the appearance of punctate spots. Exposure times were kept constant.

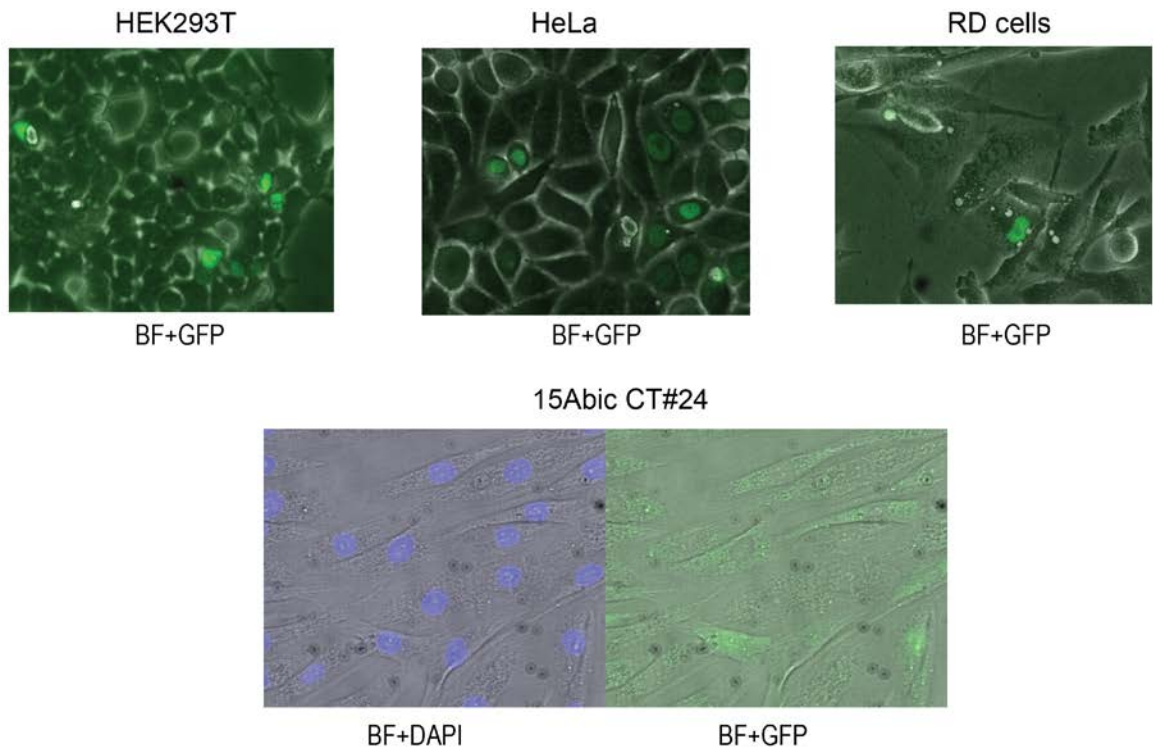


Figure 5.11. Merged Images of cells transfected with low concentration of Fluorescein labelled ASO. (Top) Merged bright field (BF) and green channel images (GFP) of HeLa, HEK293T and RD cells transfected with 100nM fluorescein labelled ASO at 10X magnification. Localization of the ASO is primarily nuclear. (Bottom) Merged images (bright field with DAPI and bright field with GFP) of immortalized FSHD patient cells transfected with 100nM fluorescein labelled ASO at 10X magnification. Localization heterogeneous, majority of cells show cytoplasmic localization, but some cells show both nuclear and cytoplasmic localization.

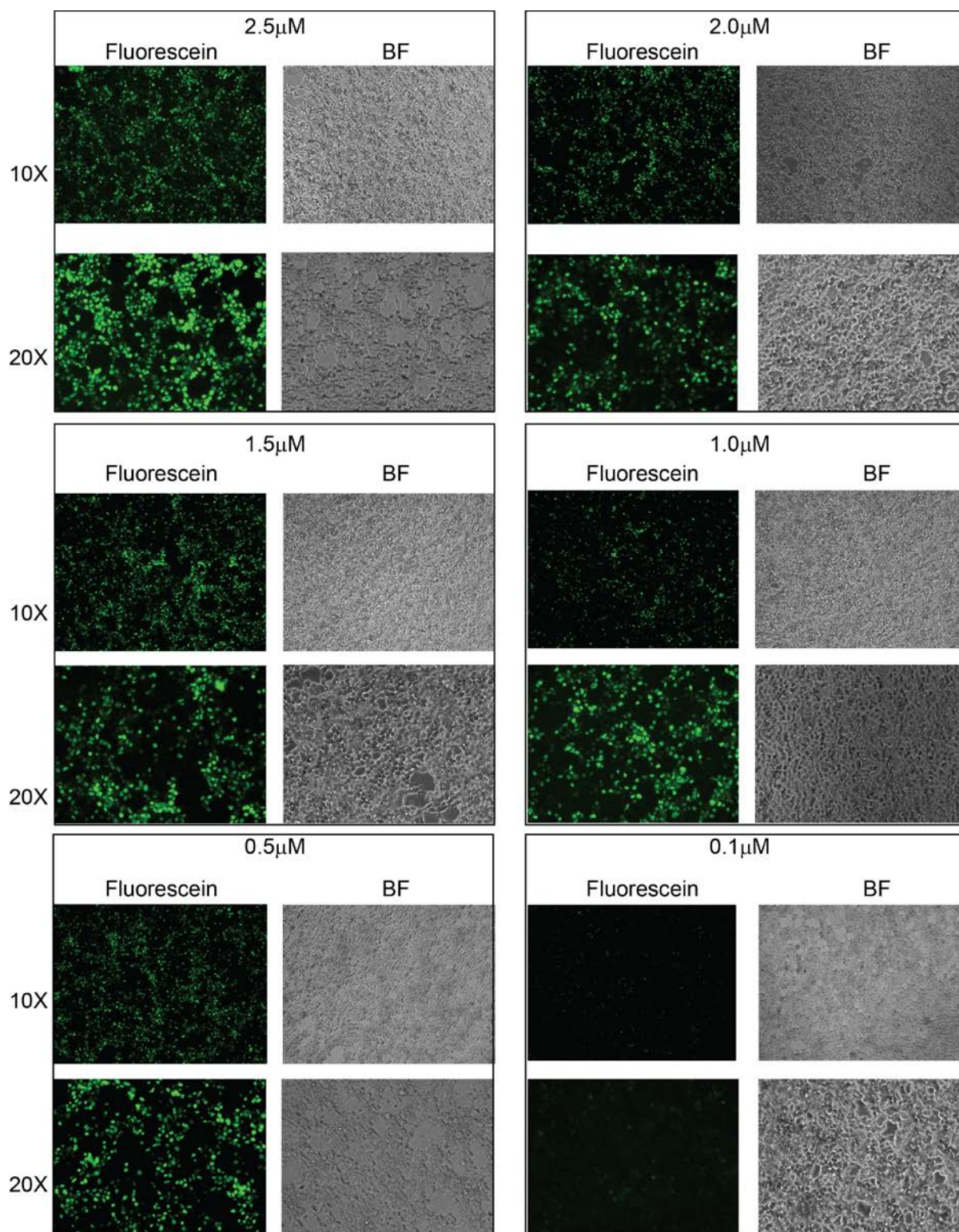


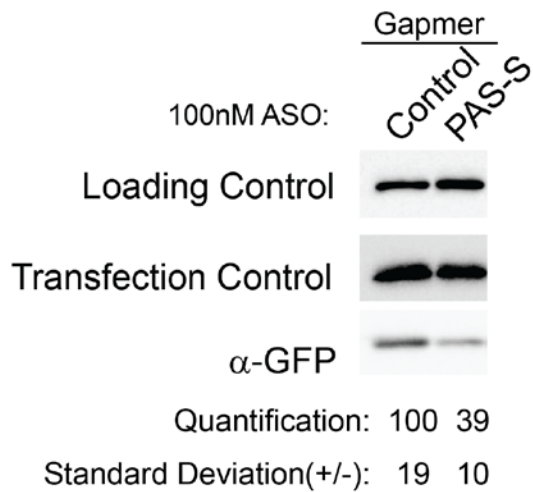
Figure 5.12. Microscope images at HEK293T cells transfected with increasing concentrations of fluorescein labelled ASO. Bright field (BF) images show little cellular

toxicity, in the green channel (fluorescein) increasing concentration of ASO in more visibly fluorescent cells and more diffuse localization of ASO in the cell. Exposure times were kept constant.

Determining the utility of ASOs with use a reporter

As a proof of principle and to investigate the utility of CPA targeting ASOs to decrease protein expression we designed ASOs targeting the SV40 polyadenylation signal. We generated a reporter in which the SV40PAS is placed downstream of an intron containing GFP ORF. The HEK293T cells were first transfected with antisense oligonucleotides, and later transfected with reporter (see Chapter 3 Methods). After 48 hrs, the cells were harvested for protein. Western blot analysis and quantification demonstrates that the different approaches have different levels of efficiency, and also that the chemistry employed has different levels of antagonistic effects for protein expression. Using this protocol we observed that both steric and gapmer ASOs resulted in ~70% decrease in protein expression, however, the decrease was modest and not significant with the steric blocking ASO (Figure 5.13).

A



B

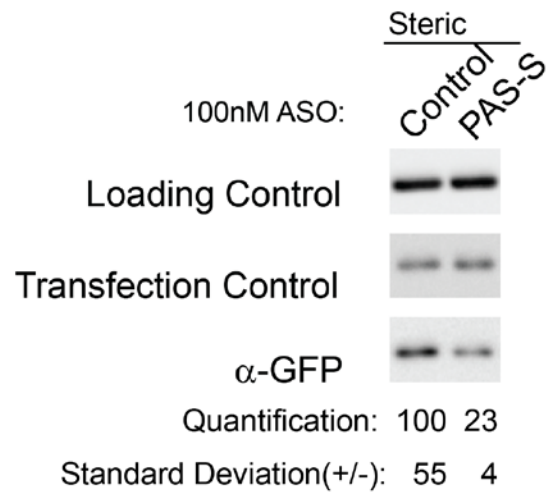


Figure 5.13. Western blot of HEK293T cells transfected using with SV40 PAS directed ASOs using a two hit protocol. Cells were first treated with ASOs, subsequently cells were simultaneously transfected with ASOs plasmid containing SV40PAS downstream GFP. Control ASO is a scrambled version of the SV40PAS directed ASO. Loading control and transfection control are α -GAPDH and α -Cherry respectively. Quantification performed using Bio-Rad Image Lab on western blots from biological triplicates. Error depicts standard deviation of biological triplicates. (A). Western blot of HEK293T cells transfected using 2hit protocol with 100nM SV40PAS directed gapmer ASO. (B). Western blot of HEK293T cells transfected using 2hit protocol with 100nM SV40PAS directed steric ASO.

Two limitations of the two-hit protocol used above are pretreatment of the ASOs, as well as co-transfection with the reporter. It is conceivable in the latter instance, that co-transfecting the ASO with the reporter may be biased toward either the plasmid DNA or the ASO. Another possibility is that there is segregation of plasmid and ASO upon transfection such that cells which are transfected with ASOs do not necessarily take up the plasmid DNA, and vice versa. The first caveat in which the cells are pre-treated with ASO prior to the addition of the reporter, may influence the data in that cells do not have a pre-existing pool of the target substrate, so the data would have a limited range of applicability. Similar protocols have been employed by others in which cells are induced to produce the gene of interest subsequent to addition of the antisense (Bennett et al., 1992; Chiang et al., 1991; Vickers et al., 2001). To address these limitations, we first transfected the reporter, and following transfection of the reporter we treated the cells once or twice with antisense oligonucleotides (see Chapter 3 Methods). To deplete the preexisting pool as well as decrease subsequent production we transfected a higher concentration of both steric and gapmer ASO. Once more we observe a ~70% reduction in GFP protein expression with a PAS directed gapmer ASO (Figure 5.14). Although, we do see a reduction in the GFP protein expression with PAS directed steric ASO, there was high degree of variability which resulted in a loss of significance. A pertinent consideration in this approach would be the consideration of the half-life of the reporter protein (estimated to be 26 hrs) (Corish and Tyler-Smith, 1999) and mRNA, however, cells are harvested for protein 48 hrs post transfection which should provide sufficient time to observe an effect.

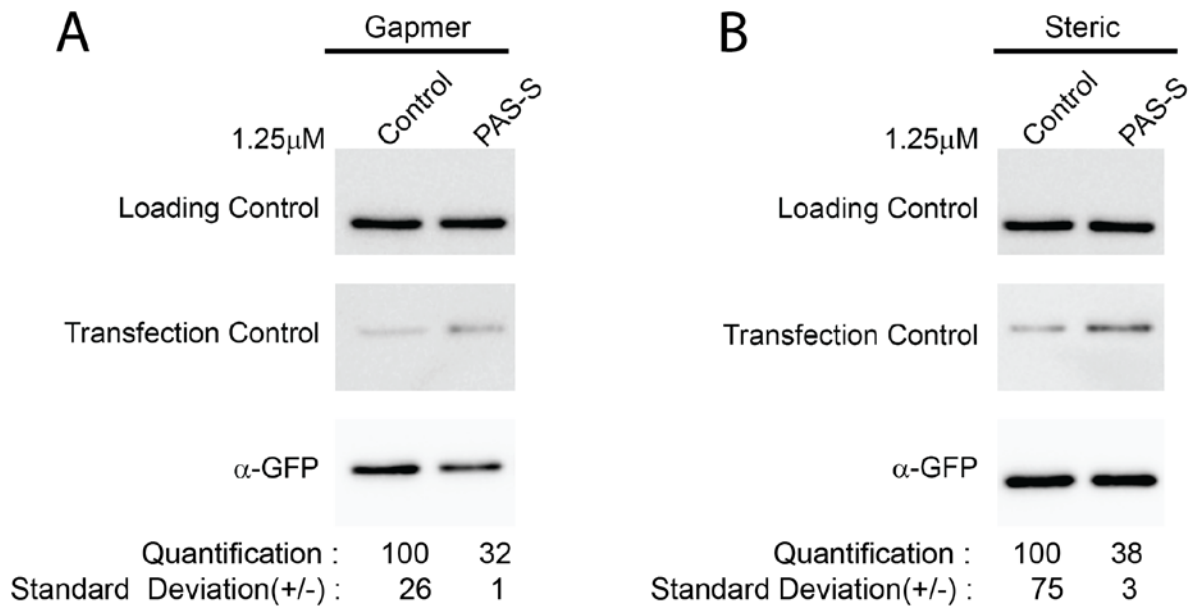


Figure 5.14. Western blot of HEK293T cells transfected using with SV40 PAS directed ASOs. Cells transiently transfected with plasmid containing SV40PAS downstream GFP, and subsequently transfected with 1.25 μ M of ASO. Control ASO is a scrambled version of the SV40PAS directed ASO. Loading control and transfection control are α -GAPDH and α -myc respectively. Quantification performed using Biorad Image Lab on western blots from biological triplicates. Error depicts standard deviation of biological triplicates. (A). Western blot of HEK293T cells transfected with 1.25 μ M SV40PAS directed gapmer ASO. (B). Western blot of HEK293T cells transfected using 2hit protocol with 1.25 μ M SV40PAS directed steric ASO.

Having cloned the Dux4PAS and DAE downstream of the intron-containing GFP, we attempted to use CPA targeting ASOs to decrease the reporter protein expression. In our pilot experiments we first transfected HEK293T cells with the reporter containing the Dux4PAS, and subsequently transfected the cells with increasing concentrations of ASO, targeting the PAS and DAE. Using gapmer ASOs, we were not able to achieve reproducible depletion in GFP expression when cells were transfected with the PAS directed ASO compared to the control treated cells (not shown). Likewise, changes in the GFP expression in cells transfected with steric ASO, were highly variable. One cause for concern is the usage of alternate PAS in the reporter when Dux4 CPA elements are placed downstream (Figure 5.9) this could prevent the detection of any changes in the protein expression, due to inducing alternative cleavage polyadenylation. As such, we feel that further optimization is necessary. One method to address this concern, would be the development of a stable clonal line with a reporter using the Dux4PAS. However, care would have to be taken to ensure that GFP expression detected is as a consequence of Dux4PAS usage, and not due random integration into the genome supplying alternative polyadenylation signals. Using an analogous reporter in which we have the chimeric SV40UTR/Dux4PASDAE (Figure 5.9) placed downstream of the intron-containing GFP we generated clonal stable cell lines. Although amplification of genomic DNA suggested that all the elements Dux4CPA downstream of the intron containing GFP were present as desired (Figure 5.15) we observed the difficulty in identifying clones that solely use the Dux4 CPA (Figure 5.15).

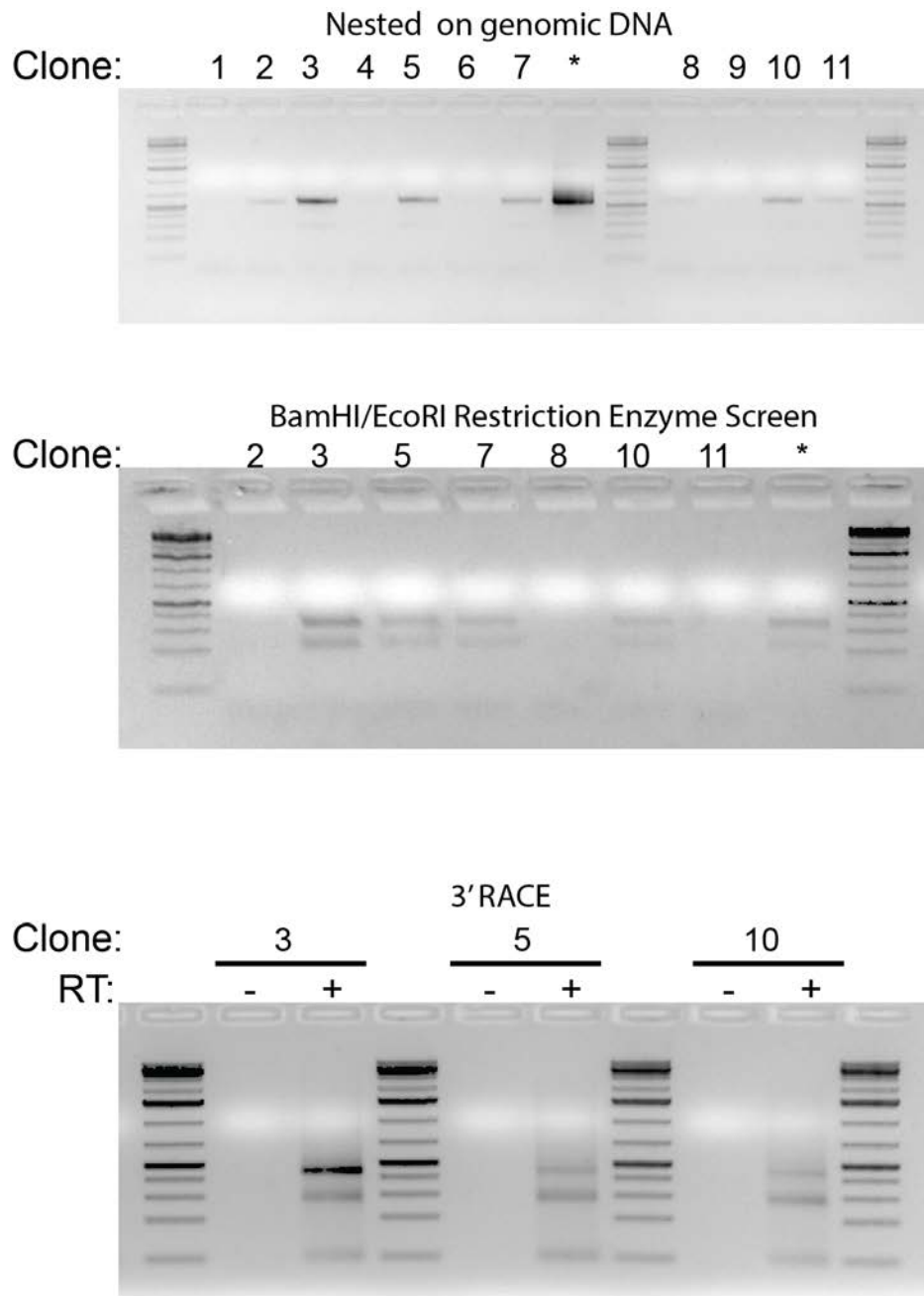


Figure 5.15. Ethidium Bromide stained agarose gels of stable GFP positive clonal lines. (Top) 2% agarose gel with products of nested PCR amplification of genomic DNA to identify GFP positive clones that had the GFP ORF and Dux4PAS and CPA elements downstream. PCR products were excised and purified, then submitted for sequencing and restriction digest screen. The * (positive control) indicates the parental plasmid used to generate the

stable cell line amplified by PCR. (Middle) Clones which were positive in the nested PCR amplification, were digested with BamHI and EcoRI and resolved on 3% agarose gel. The * indicates the control positive control PCR product digested also with BamHI and EcoRI. (Bottom) 3' RACE on clonal lines which contained GFP upstream of the Dux4PAS.

Redirecting CPA

We next turned our attention to an endogenous gene in HeLa cells. Dicer has multiple polyadenylation signals (Masamha et al., 2014). We designed both steric and gapmer ASOs targeting the distal PAS of Dicer. A double transfection of 100nM dicer siRNA using the standard two hit protocol (see Chapter 3 Methods) was able to give robust knockdown ~80% compared to a non-targeting siRNA (Figure 5.16). Comparatively, we performed a single transfection of both steric and gapmer ASOs targeting Dicer1 at a final concentration of 100nM, and observe that a PAS directed gapmer ASO produced a ~40-50% decrease in the Dicer1 mRNA levels, when measured using a primer within the ORF and a primer which detects Dicer1 transcripts using a distal polyadenylation signal. In contrast, using the PAS directed steric blocking ASO, we did not detect a significant decrease in the overall Dicer1 mRNA levels, and instead, we observed ~ 60% decrease in usage of distal polyadenylation signal. We interpret the data presented here, as redirection of the polyadenylation signal choice by a PAS directed steric ASO. The availability of more than one polyadenylation signal likely, limits the extent of knockdown that can be achieved with the steric blocker. In contrast, the gapmer ASO allows for degradation of the mRNA due to the action of RNase H, regardless of whether or not there is an alternative polyadenylation signal. In addition, the gapmer ASO is not restricted to working during the process of CPA, but can also lead to degradation of the processed mRNA in both the cytoplasm and nucleus. Increasing the concentration of gapmer ASO or following a two hit protocol similar to the siRNA transfection may increase the level of knockdown. Although the depletion would be most observable in transcripts using the distal polyadenylation signal.

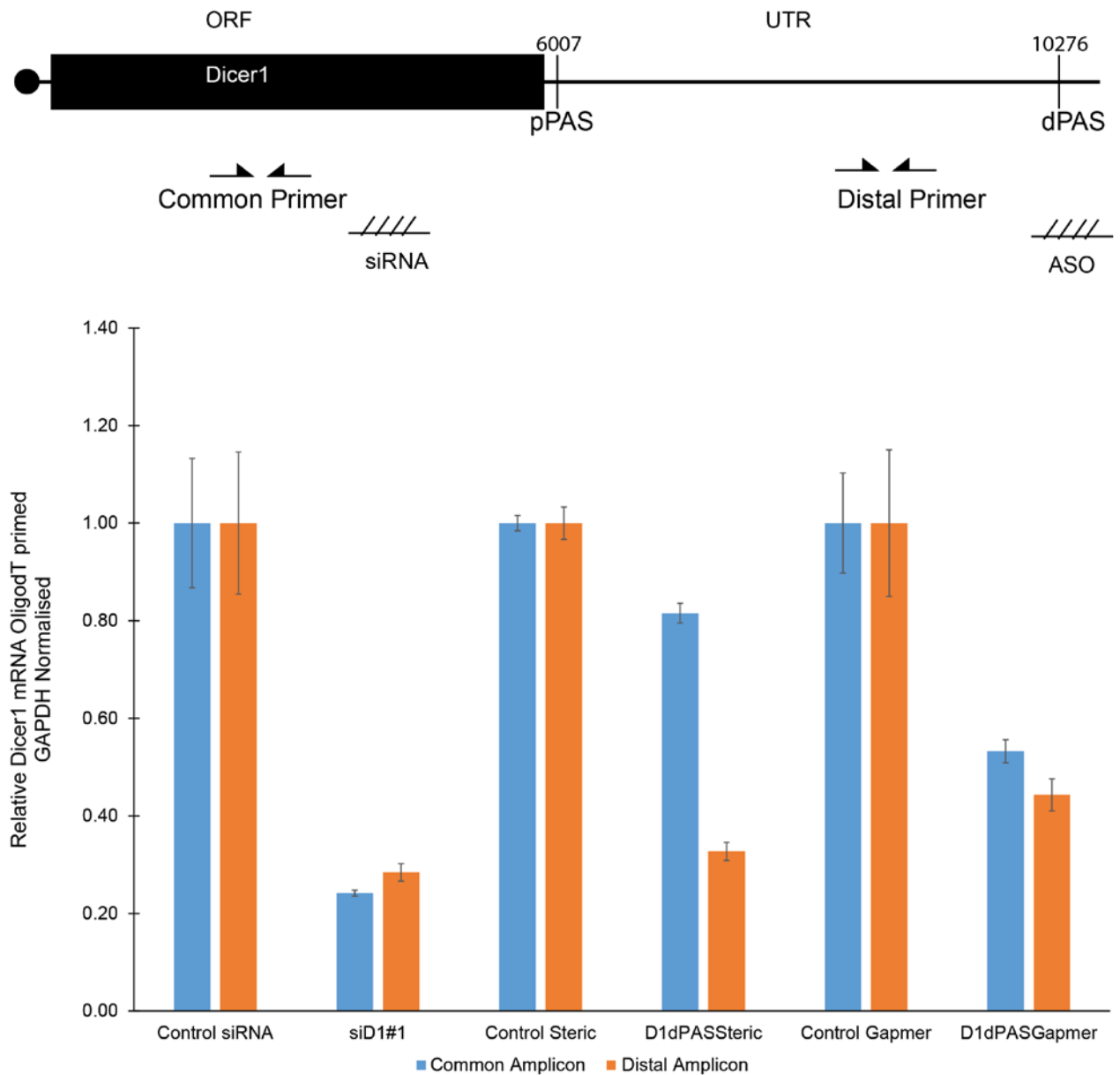


Figure 5.16. PAS directed ASO redirect PAS or reduce gene expression. (Top) Schematic of Dicer1 open reading frame (ORF) and untranslated region (UTR), head to head arrows indicate the location of the primers used to amplify Dicer. Vertical bars in the UTR denote position of the proximal (pPAS) and distal (dPAS) polyadenylation signal. Horizontal lines with diagonal hatches indicate location of siRNA and PAS-Directed ASO. (Bottom) qRT-PCR analysis of HeLa cells transfected with control or Dicer1 siRNA (siD1#1), and both

steric and gapmer ASOs directed toward the distal polyadenylation signal of Dicer1. Control ASOs are derived from a scramble of the SV40PAS directed ASO. Error bars calculated from standard deviation of biological triplicates. Distal and proximal polyadenylation signal indicated are as previously described by Masamha et al. 2014. Primers to amplify Dicer1 common and distal previously described by Masamha et al. 2014.

Dux4 in 15Abic cells

One of our primary goals was to decrease the Dux4 protein expression. Dux4 is typically poorly expressed, and detection is difficult. Nonetheless, we attempted to detect Dux4 protein in the 15Abic immortalized cell line. We immunoprecipitated with a Dux4 antibody (9A12) that recognizes the Dux4 (but is reported to cross hybridize with other Dux4-like proteins) and then subsequently performed a western blot to detect Dux4 using an antibody which uniquely recognizes the Dux4 C-terminus (Geng et al., 2011). Consistent with previous reports that Dux4 is induced upon differentiation (Block et al., 2013), we observe a strong signal at ~55kDa (likely Dux4) (Figure 5.17A), due to the predicted size of the Dux4 protein being 45-52kDa. However, contrary to previous reports (Pandey et al., 2015), substitution of knockout serum replacer (KOSR) in place of horse serum (HS) in the differentiation medium, did not enrich or stimulate for Dux4 protein detection (Figure 5.17B). Knockout serum replacer is posited to enhance myotube formation, which would consequently promote Dux4 expression (Block et al., 2013). However, it is notable that, at the time of lysis, both the horse serum and knockout serum replacer differentiated cells had roughly the same amount of multinucleated cells (not shown).

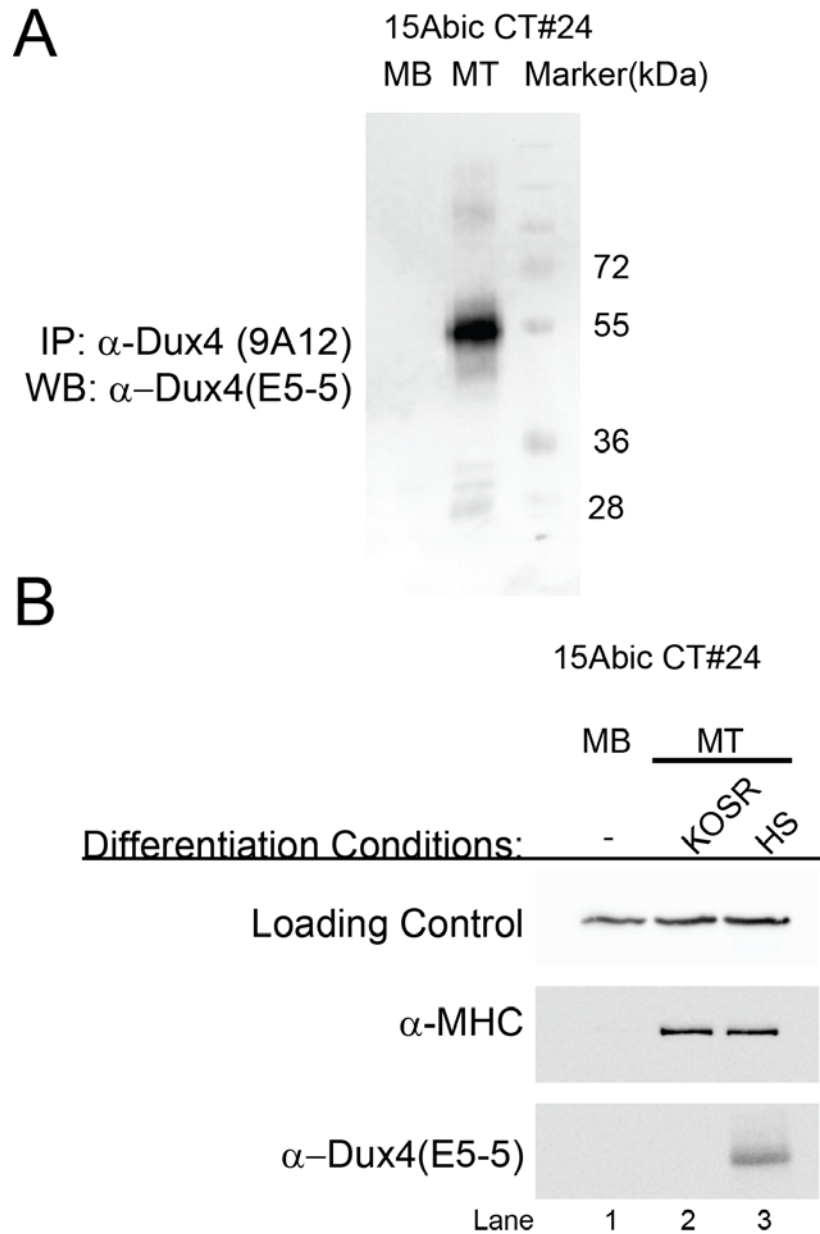


Figure 5.17. Detecting Dux4 Protein in Immortalized FSHD patient lines. (A). Western blot probed with α -Dux4 (E5-5) of 5Abic myoblasts (MB) and myotubes (MT) immunoprecipitated with α -Dux4(9A12). Protein marker shown. (B). Western blot of 15Abic myoblasts (MB) and myotubes (MT). Myoblasts were differentiated over 5 days in either 20% Knockout Serum Replacer (KOSR) or 2% Horse Serum. Blots probed for differentiation marker myosin heavy chain (α -MHC) and Dux4 protein with α Dux4 (E5-5). Loading control is α -GAPDH.

Conclusions

Here, I identified that additional *cis* regulatory elements, besides the non-consensus PAS are required for cleavage and polyadenylation of Dux4. Previous bioinformatics analyses posit that RNA elements critical for cleavage and polyadenylation optimally lie within 100 nucleotides of cleavage site and are typically Uridine rich (Hu et al. 2005; Legendre and Gautheret 2003). The location of these *cis* regulatory elements for Dux4 cleavage and polyadenylation are outside the optimal 100 nucleotide window downstream of both the polyadenylation signal and the cleavage site. Moreover, this element lies within a degenerate repeat regions, which leads us to form two hypotheses to model Dux4 cleavage and polyadenylation. The *cis* elements identified are not enriched for uridines, and the degeneracy of the repeats impairs the identification of similar motifs, although the commonality between the regions identified to increase GFP expression seems to be biased toward purines.

In the context of our reporter, we identify a DAE that appears to enhance Dux4 mRNA 3' end processing from the non-consensus PAS. Thus, we present the following models for how the DAE identified enhances CPA of Dux4 (Figure 5.18). We propose that the DAE facilitates efficient cleavage and polyadenylation by one of two potential mechanisms. The first involves the recognition of the DAE within the nascent pre-mRNA by a yet-to-be identified RNA binding factor that would enhance the recognition of the PAS by the CPSF machinery.

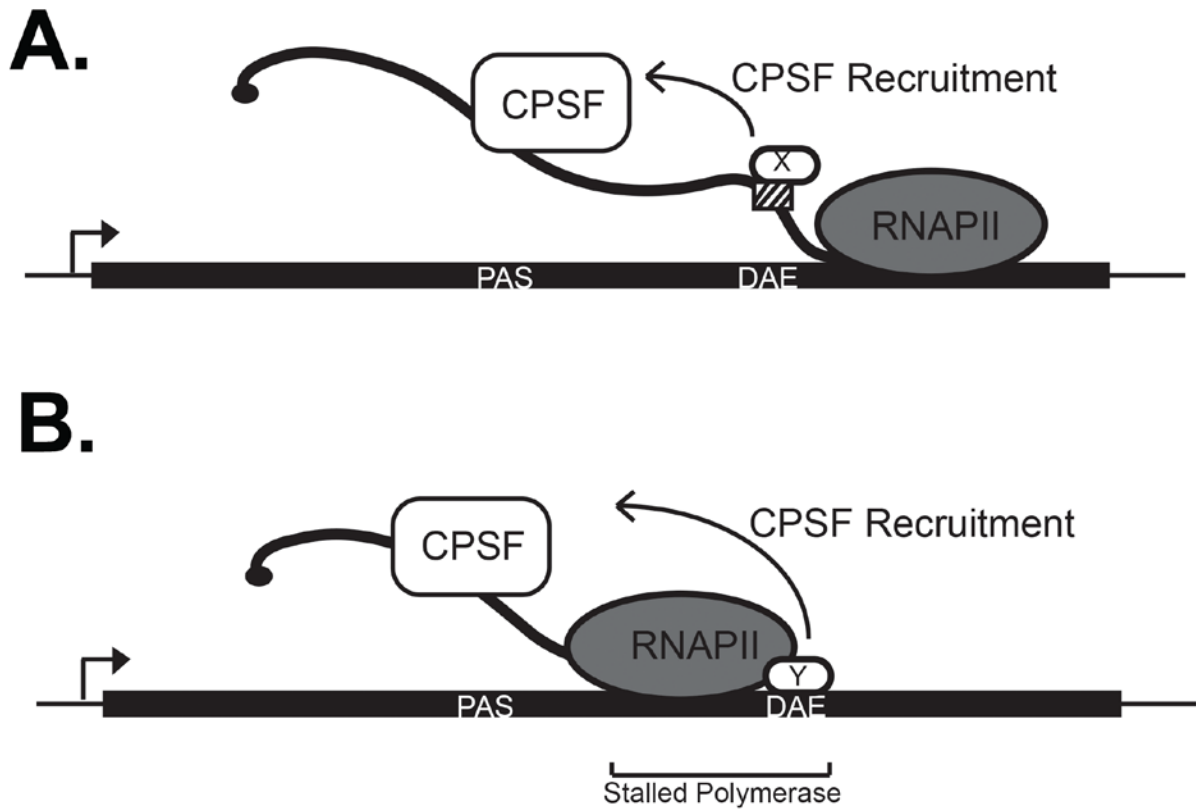


Figure 5.18. Model of Dux4 Processing. (A.) RNA binding proteins interacting with the downstream auxiliary element (DAE) identified may facilitate Dux4 cleavage and polyadenylation by aiding in the recruitment of CPSF machinery to the non-consensus PAS. Alternatively, (B). The DAE may be a DNA element that induces RNA pauses due either to sequence identity of protein blockage. The stalled polymerase kinetically favors cleavage and polyadenylation at the non-consensus Dux4 PAS.

An alternative explanation, is that the DAE aids in the 3' end formation of Dux4 mRNA by pausing (or terminating) of the RNA polymerase. Polymerase pausing is well defined in terms of its role in transcription initiation, particularly for stress response genes (Kwak and Lis 2013), however, less is known about the role of RNAPII pausing to facilitate mRNA maturation. One of the best understood pausing elements, which can influence mRNA 3'end processing, is the MAZ sequence (Yonaha and Proudfoot 1999, 2000). It has been shown that placement of MAZ elements downstream of the of poorly defined exons such as α -tropomyosin exon 3 or Fibroblast Growth Factor Receptor II exon IIIb, increases their inclusion presumably through providing additional time for splice site recognition (Roberts et al. 1998; Robson-Dixon and Garcia-Blanco 2004).

Upon cloning the Dux4 CPA downstream of mCherry we disappointingly did not observe robust expression. However, as we found increasing the β -satellite repeats sequences downstream increased reporter expression. Significantly, we also observed that introduction of splicing upstream of the Dux4CPA increased our detection of the Dux4 mRNA. However, due to the presence of additional cleavage products in the 3'RACE we cannot exclude that the Dux4CPA elements may result in a lower production of mRNA compared to the upstream PAS.

Finally, using the knowledge acquired we designed Dux4 CPA targeting ASOs to prevent protein expression. We undertook proof of principle experiments and demonstrate that CPA targeting ASOs can decrease gene expression or redirect cleavage and polyadenylation. Unfortunately, we have not yet been able to directly impair gene expression using Dux4PAS targeting ASOs in either our reporter system or in patient cell lines. In the immortalized patient cell line we observed a wide range of toxicity in response to several of the ASOs tested in 15Abic cells. In our reporter system, we do not get reproducible decrease in reporter protein expression. Analysis of the RNA would illuminate whether we are observing alternative polyadenylation in the reporter system. Further, in the patient cell

lines optimization of the ASO delivery, with due consideration for the ASO chemistry that we have accessible, is required.

Chapter 6 : Conclusions, Perspectives and Future Directions

Downstream Auxiliary Sequences aid cleavage and polyadenylation of Dux4PAS

Using a transcriptional read-through reporter, we confirm that additional *cis* elements are required for cleavage and polyadenylation (CPA) of Dux4 besides the non-consensus polyadenylation signal (PAS). Moreover, this element is not located within a 100 nucleotides of either the Dux4 PAS, or the cleavage site. Interestingly, the downstream auxiliary element lies within a degenerate repeat region, called β -satellite DNA. Moreover using the RegRNA2.0 tool (<http://regrna2.mbc.nctu.edu.tw/index.html>) (Chang et al., 2013), we do not observe an enrichment for any known RNA motifs associated with cleavage and polyadenylation within this region.

Testing a Model: Are the Dux4PAS auxiliary sequences RNA or DNA?

In Chapter 6, we propose two models describing how the downstream auxiliary sequence elements (DAE) may aid in CPA of Dux4. While the models are not necessarily mutually exclusive, we can interrogate the downstream sequences to examine the extent to which either DNA or RNA elements facilitates CPA of the Dux4PAS. To interrogate the model that the DAE is a DNA element we may use two complementary methods: Nuclear Run-On assay and transcriptional pausing assay. Previously, we attempted to use a non-radioactive Nuclear Run-On assay to interrogate Dux4 PAS sequence elements for transcriptional pausing in PAS-GFP reporter transfected cells, however, our available cell lines proved refractory to labelling with 5-Bromouridine. The Nuclear Run On/Off Assay (Li and Chaikof, 2002; Smale, 2009), provides a cell based method to assay nascent and active transcription on genes. However, we can also use the assay to comparatively demonstrate the absence of transcription in the regions that are anticipated to not have robust transcriptional activity due to termination events (Skourti-Stathaki et al., 2011), such

as cleavage and polyadenylation of the mRNA and the subsequent removal of the RNA Polymerase II (RNAPII) from the template. However, this assay does not prove pausing, and as such the complementary cell-free transcriptional pausing assay, such as the G-Less Cassette Assay (Carey et al., 2010) is also useful to investigate this hypothesis of transcriptional pausing.

To address the hypothesis that the DAE functions primarily as an RNA element, antisense oligonucleotides (ASO) may serve as a useful tool. The utility of ASOs for therapeutic intervention is slowly growing (Kole et al., 2012; Potaczek et al., 2016), however, we can also use ASOs in cell-free RNase protection assays and cleavage and polyadenylation assays to interrogate the relevance of the RNA DAE for processing. Steric impairment of DAE:Protein interaction (Zhang et al., 2015) or removal of the DAE by RNase H activating ASO may reveal whether the DAE in the pre-mRNA is critical for cleavage and polyadenylation.

The significance of non-canonical *cis* elements for cleavage and polyadenylation and their interactions with cleavage and polyadenylation *trans* factors

The size and complexity of protein factors required to cleave the mRNA transcript from transcribing RNAPII and DNA template underscores the importance of *cis* regulatory elements to facilitate accurate and correct cleavage and subsequent polyadenylation ((Shi et al., 2009) and reviewed in part (Chan et al., 2011; Wahle and Rügsegger, 1999; Yang and Doublé, 2011)). It has been shown that WDR33 and CPSF30, as part of the cleavage and polyadenylation specificity factor (CPSF) recognize and bind to the consensus PAS, AAUAAA (Chan et al., 2014; Schönemann et al., 2014). WDR33 appears to preferentially associate with Uridines immediately downstream of the PAS, while the CPSF30 shows a preference toward the AU PAS (Schönemann et al., 2014; Shimberg et al., 2016). However,

in the case of non-consensus PAS, the tolerance for the divergence from the AAUAAA and low Uridine enrichment is unclear (Chan et al., 2014; Schönemann et al., 2014; Shimberg et al., 2016). Interestingly, although AUUAAA is fairly frequently used as a PAS, (~17% compared to >50% for the AAUAAA, and <20% for other variants) (Tian and Graber, 2012), WDR33 enrichment at AUUAAA PAS was low compared to the AAUAAA and other variants (Schönemann et al., 2014). In the case of pathogenic expression of Dux4, the PAS used is AUUAAA; and although there is a short stretch of containing two small Uridine tracks (<4 nucleotides) and intermittent presence of UG, immediately downstream of the cleavage site, there is no apparent enrichment elsewhere. Additionally, using our PAS-GFP reporter, we demonstrate that a cleavage product with Dux4PAS is detected when we place more than 100 nucleotides downstream of the PAS, the additional sequences do not show an enrichment for Uridine stretches.

Cleavage and polyadenylation assays suggests that cleavage after a non-adenosine nucleotides is poor (Sheets et al., 1990). In the literature the reported cleavage site of Dux4 is AG (Lemmers et al., 2010a; Snider et al., 2009). The RNA template within this region reads AGA and so could be cleaved at either the G or A. In our hands, we reproducibly detect a cleavage product within the TATA region a single nucleotide upstream of the GA. However, an important observation here is that the cleavage site is close to the upper limit downstream of the PAS; cleavage and polyadenylation typically occurs within 30 nucleotides of the PAS (Chan et al., 2011). While there is a known micro-heterogeneity in cleavage site, the significance is unclear (Chan et al., 2011). However, we believe that an additional consideration should be whether there is a relevance between PAS consensus (that is the sequence of the PAS as well as the presence and strength of USE and DSE) and the distance of cleavage site from the PAS. Mutation of the cleavage site does not significantly impair processing, because the endonuclease, CPSF73, can simply cleave the RNA at an alternate location, these observations were reported in the context of a RNA substrates

modelled on the late SV40 polyadenylation signal or L3 adenovirus-2 polyadenylation signal that bare the consensus, PAS, AAUAAA (this study and (Mandel et al., 2006; Sheets et al., 1990)).

We propose that the Dux4 PAS, with its paucity of notable motifs for CPA, would be a good substrate to interrogate the factors that bind to a non-consensus PAS. By developing a stable cell line with the gain of function reporter (such as the PAS-GFP) using the Dux4PAS, a large scale CRISPR/Cas9 screen (Shalem et al., 2014) can be then employed to identify genes which cause of loss of cleavage and polyadenylation of Dux4. Identified genes can then be investigated using RNA immunoprecipitation or RNA binding assays to determine interaction with Dux4PAS. Likely such a screen would provide a lot of candidates; thus to narrow the investigation, combining unbiased RNA immunoprecipitation and mass spectrometry to identify proteins interacting with the Dux4PAS with the aforementioned CRISPR/Cas9 screen to identify genes that result in loss of cleavage and polyadenylation of Dux4 would be ideal.

Chromosome 10 contains D4Z4 repeats, and pLAM and is nearly identical to the D4Z4 repeats and pLAM on chromosome 4, however FSHD is exclusively associated with the hypomethylation of D4Z4 repeats on chromosome 4 (van Geel et al., 2002; Lemmers et al., 2010b; van Overveld et al., 2003; Wijmenga et al., 1992). Non-pathogenic single nucleotide polymorphisms (SNPs) generate ATTTAA or ATCAAA, which are not recognized as PAS on chromosome 10. Interestingly, placing the pathogenic SNP (generating ATTAAA) in the context of a non-permissive chromosome 10 supported cleavage and polyadenylation, albeit inefficiently (Lemmers et al., 2010a). Nonetheless, we posit that the context of the SNP that generates the PAS is critical for CPA of Dux4. It is imperative to consider what additional SNPs are present that allow for the AUUAAA to be used as a PAS for Dux4, and also to what extent does the degeneracy of the β -satellite repeats affect the CPA? While overwhelming evidence supports the hypothesis that poor expression of Dux4 expression is

a consequence of dysregulated epigenetic silencing (resulting in a leaky and stochastic event), it is also possible that cleavage and polyadenylation at this non-consensus PAS is impaired due to the degeneracy of the downstream β -satellite repeats. There have been dissenting data presented that discounted the importance of the pathogenic SNP, there exist individuals who fit the molecular criteria for FSHD, but who do not present with disease (Ricci et al., 2013; Scionti et al., 2012). The molecular criteria being contracted D4Z4 repeats on chromosomal haplotype 4qA161. As such we recommend additional sequencing of patient DNA, with due consideration to the sequence identity and arrangement of the β -satellite to determine if there is a correlation between severity of disease, disease onset, or disease presence or absence (asymptomatic FSHD). This may indicate differential efficiency in the CPA of Dux4, and thus production of the somatically toxic gene.

Downstream Auxiliary Sequences aid cleavage and polyadenylation of Dux4PAS, but is there more?

Although, we identify downstream sequences critical for production of a cleavage product using the PAS-GFP reporter we failed to detect robust mRNA and protein expression when the identified Dux4 CPA is placed downstream of the a reporter such as mCherry. This suggests that there are additional requirements beyond the PAS and the downstream auxiliary sequences. This is likely, because although the P200 construct suppresses detection of the reporter protein in the PAS-GFP, we still detect an upstream read-through product. However, it is unlikely that this read-through product is abundant as it was detected only in the second round of PCR in the 3' RACE. Moreover, the read-through transcript, unlike the short upstream cleavage transcript, contains a stop codon and as such is not susceptible non-stop decay (reviewed by (Klauer and van Hoof, 2012)) which would reduce abundance. Further supporting low abundance of the read-through transcript in the

P200 constructs, the P100, Exon 3 and pLAM constructs all produced a detectable read-through transcripts in the first round of PCR.

While cloning the Dux4 PAS sequences required for CPA and reporter protein expression we made an interesting observation: splicing upstream enhanced reporter expression from the Dux4PAS. The observation that splicing affects CPA is not novel (Kaida, 2016), however, when considering the architecture of the Dux4 gene, the observation is intriguing. The Dux4 pre-mRNA has two introns in the untranslated region (UTR). It had been previously demonstrated that the second intron causes the transcript to be subject to nonsense mediated decay (Feng et al., 2015). However, the relevance of the first intron which is retained in Dux4 isoform, *Dux4FI-1*, had not been investigated. As previously mentioned (see Chapter 2), a cryptic splice site within the first exon reportedly generates a short isoform, Dux4-s which if translated would lack the C-terminal domain, and may act as a dominant negative (Geng et al., 2012). Additionally, it has been observed by others during an attempt to generate a transgenic Dux4 mouse containing the terminal D4Z4 repeat and abutting pLAM that there was frequent missplicing (Ansseau et al., 2015). Upon analysis, we observe that the 5' splice site after Dux4 exon 1 deviates 50% from the consensus sequence bound by the U1snRNP. This poor splice site is likely in competition with alternative sequences that have higher complementarity with the U1snRNP, and consequently make better substrates for exon ligation, and we further posit that missplicing, should it happen in patient cells, may contribute to the poor detection/expression of the Dux4 mRNA.

Antagonizing Dux4 mRNA 3' end formation

One of the primary goals of our investigation is to use Dux4PAS directed ASOs to downregulate gene expression. One consideration for using ASOs is the accessibility of the

target sequence. RNA molecules are not naked single stranded moieties in the cell, instead, they adopt intricate secondary and tertiary structures by forming intra- and inter- molecular bonds. Alternatively, or simultaneously RNA may be bound by proteins and packed into ribonucleoproteins (RNPs). The result of this propensity to adopt secondary, tertiary and quaternary structural conformations is that sequence elements targeted by antisense oligonucleotides may be inaccessible. Consequently, the application and utility of ASO is dependent on the stability and affinity of the interaction between RNA targets and non-ASO molecules (proteins, RNA etc.). In addition, ASO utility and application is dependent on the kinetics of the interactions. There are various strategies to limit these restrictions, but, for use in therapeutics some of these options may be inapplicable. For example, increasing the concentrations of ASO, increases the probability of ASO:RNA interactions, but may prove toxic to patients. In another example, pretreating with ASOs prior to the activation of the targeted event increases the likelihood of modulating the gene expression, however, this would be a preventative measure rather than corrective.

Apart from consideration of accessibility to target by ASO, we should also consider the implication of escape of targeted RNA from ASO. Here we use diphtheria toxin as an example, wherein low expression of the diphtheria toxin (purportedly a single molecule) is enough to cause toxicity (Murphy, 1996). By extension, we must address if this observation is pertinent to Dux4, and to what extent must we reduce Dux4 for there to be a beneficial effect. Dux4 expression is toxic to a wide variety of cells, with the possible exception of germ cells (given the robust detection of Dux4 in the testicular tissue) (Kowaljow et al., 2007; Snider et al., 2010). Given the stochastic and low expression of Dux4, we must ask how many molecules of Dux4 are required to trigger Dux4-dependent pathologies. Although, Dux4 mice do not recapitulate the full scope of FSHD, it has been shown that even leaky expression of the protein can be lethal, and that surviving mice preferentially had the Dux4 gene silenced (Dandapat et al., 2014).

Recently, others have demonstrated that Dux4PAS directed antisense oligonucleotides can impair Dux4 expression in patient cells (Marsollier et al., 2016). However, a global assessment of the effects of PAS directed ASOs, has not yet been performed. The rationale for global assessment are (1) ASOs are generally very stable, and stay in the cells for prolonged period and thus may have off-target effects (Bennett and Swayze, 2010; Crooke, 2004). ASOs have also been shown to activate innate immunity pathways (Agrawal and Kandimalla, 2004; Burel et al., 2012; Crooke, 2004; Senn et al., 2005; Watts and Corey, 2012).

Successfully modulating gene expression with ASOs is empirical in both the clinic and laboratory. Considerations for when using ASOs must include dosage frequency and concentration and chemical modifications on the ASO to decrease toxicity while maintaining efficacy (Crooke, 2004) amongst other thing. Yet another consideration, it that the clinical response to use of antisense oligonucleotide therapies may be independent and distinct from the response observed in the laboratory setting (reviewed by(Watts and Corey, 2012)). As previously mentioned, ASOs can activate immune response pathways, due to triggering interferon and extracellular signal–regulated kinases (ERKs) signaling (Burel et al., 2012; Senn et al., 2005) and thus the utility of ASOs in mitigating the debilitating effects of one disease is cancelled by the induction of acute inflammation. Furthermore, ASO delivery may result in inappropriate organ targeting, and often hepatotoxicity (Burel et al., 2016; Watts and Corey, 2012). There is increased commercial availability of a variety of chemical modifications for RNA or DNA based antisense oligonucleotides as well as the availability of nucleoside analogues. Together with new modalities of ASO delivery the off target effects in the clinical setting may be further reduced (Watts and Corey, 2012).

Using a reporter system where the Dux4 CPA is placed downstream of an intron-containing GFP, we have not been able to reproducibly deplete Dux4. 3' RACE of the reporter shows that there is another cleavage event upstream of the Dux4PAS which would

still generate GFP protein. In contrast, the SV40PAS only shows one cleavage event. Others have shown that steric blocking ASOs can be used to redirect polyadenylation signal choice and cleavage site (Marsollier et al., 2016; Vickers et al., 2001), and we reproduce this effect. As such it is possible, that Dux4PAS-directed ASOs may redirect cleavage and polyadenylation. However, the irreproducible decrease in reporter expression was observed for both steric and gapmer ASO.

Continued investigation of the use of Dux4PAS directed ASOs is required, not only to reproduce decrease in the gene expression, but also to address specificity. This would be broadly applicable to any PAS-directed ASO therapeutic especially with the consideration of causing alternative cleavage and polyadenylation.

Significance

Here we study the non-consensus PAS, AUUAAA, of Dux4 and interrogate the adjacent sequences to identify *cis* elements that enhance cleavage and polyadenylation from the Dux4PAS. The identification of downstream sequences in a degenerate repeat (regardless of the distance from the PAS) is intriguing because of the implication that a lack of consensus elements would allow for great variability in how this element mechanistically aids cleavage and polyadenylation.

References

- Abou Elela, S., and Ares, M. (1998). Depletion of yeast RNase III blocks correct U2 3' end formation and results in polyadenylated but functional U2 snRNA. *EMBO J.* 17, 3738–3746.
- Agrawal, S., and Kandimalla, E.R. (2004). Antisense and siRNA as agonists of Toll-like receptors. *Nat. Biotechnol.* 22, 1533–1537.
- Albrecht, T.R., and Wagner, E.J. (2012). snRNA 3' End Formation Requires Heterodimeric Association of Integrator Subunits. *Mol. Cell. Biol.* 32, 1112–1123.
- Ali, S.M., Mahnaz, S., and Mahmood, T. (2008). Rapid genomic DNA extraction (RGDE). *Forensic Sci. Int. Genet. Suppl. Ser.* 1, 63–65.
- Ansseau, E., Domire, J.S., Wallace, L.M., Eidahl, J.O., Guckes, S.M., Giesige, C.R., Pyne, N.K., Belayew, A., and Harper, S.Q. (2015). Aberrant Splicing in Transgenes Containing Introns, Exons, and V5 Epitopes: Lessons from Developing an FSHD Mouse Model Expressing a D4Z4 Repeat with Flanking Genomic Sequences. *PLoS ONE* 10, e0118813.
- Antonescu, C. (2014). Round cell sarcomas beyond Ewing: emerging entities. *Histopathology* 64, 26–37.
- Araujo, P.R., Yoon, K., Ko, D., Smith, A.D., Qiao, M., Suresh, U., Burns, S.C., Penalva, L.O.F., Araujo, P.R., Yoon, K., et al. (2012). Before It Gets Started: Regulating Translation at the 5' UTR. *Int. J. Genomics Int. J. Genomics* 2012, 2012, e475731.
- Arndt, K.M., and Reines, D. (2015). Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu. Rev. Biochem.* 84, 381–404.
- Baillat, D., and Wagner, E.J. (2015). Integrator: surprisingly diverse functions in gene expression. *Trends Biochem. Sci.* 40, 257–264.

Baillat, D., Hakimi, M.-A., Nääär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 123, 265–276.

Balakin, A.G., Smith, L., and Fournier, M.J. (1996). The RNA World of the Nucleolus: Two Major Families of Small RNAs Defined by Different Box Elements with Related Functions. *Cell* 86, 823–834.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.-M., and Gautheret, D. (2000). Patterns of Variant Polyadenylation Signal Usage in Human Genes. *Genome Res.* 10, 1001–1010.

Bennett, C.F., and Swayze, E.E. (2010). RNA Targeting Therapeutics: Molecular Mechanisms of Antisense Oligonucleotides as a Therapeutic Platform. *Annu. Rev. Pharmacol. Toxicol.* 50, 259–293.

Bennett, C.F., Chiang, M.Y., Chan, H., Shoemaker, J.E., and Mirabelli, C.K. (1992). Cationic lipids enhance cellular uptake and activity of phosphorothioate antisense oligonucleotides. *Mol. Pharmacol.* 41, 1023–1033.

Blewitt, M.E., Gendrel, A.-V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* 40, 663–669.

Block, G.J., Narayanan, D., Amell, A.M., Petek, L.M., Davidson, K.C., Bird, T.D., Tawil, R., Moon, R.T., and Miller, D.G. (2013). Wnt/ β -catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum. Mol. Genet.* 22, 4661–4672.

Bond, C.S., and Fox, A.H. (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* 186, 637–644.

van den Boogaard, M.L., Lemmers, R.J.L.F., Balog, J., Wohlgemuth, M., Auranen, M., Mitsuhashi, S., van der Vliet, P.J., Straasheijm, K.R., van den Akker, R.F.P., Kriek, M., et al. (2016). Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. *Am. J. Hum. Genet.* 98, 1020–1029.

Bosnakovski, D., Xu, Z., Ji Gang, E., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppée, F., et al. (2008). An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J.* 27, 2766–2779.

Box, J.A., Bunch, J.T., Tang, W., and Baumann, P. (2008). Spliceosomal cleavage generates the 3' end of telomerase RNA. *Nature* 456, 910–914.

Brideau, N.J., Coker, H., Gendrel, A.-V., Siebert, C.A., Bezstarosti, K., Demmers, J., Poot, R.A., Nesterova, T.B., and Brockdorff, N. (2015). Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. *Mol. Cell. Biol.* 35, 4053–4068.

Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2012). Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MEN β noncoding RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19202–19207.

Burel, S.A., Machemer, T., Ragone, F.L., Kato, H., Cauntay, P., Greenlee, S., Salim, A., Gaarde, W.A., Hung, G., Peralta, R., et al. (2012). Unique O-methoxyethyl ribose-DNA chimeric oligonucleotide induces an atypical melanoma differentiation-associated gene 5-dependent induction of type I interferon response. *J. Pharmacol. Exp. Ther.* 342, 150–162.

Burel, S.A., Hart, C.E., Cauntay, P., Hsiao, J., Machemer, T., Katz, M., Watt, A., Bui, H.-H., Younis, H., Sabripour, M., et al. (2016). Hepatotoxicity of high affinity gapmer antisense oligonucleotides is mediated by RNase H1 dependent promiscuous reduction of very long pre-mRNA transcripts. *Nucleic Acids Res.* *44*, 2093–2109.

Byszewska, M., Śmietański, M., Purta, E., and Bujnicki, J.M. (2014). RNA methyltransferases involved in 5' cap biosynthesis. *RNA Biol.* *11*, 1597–1607.

Cabianca, D.S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., and Gabellini, D. (2012). A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* *149*, 819–831.

Carey, M.F., Peterson, C.L., and Smale, S.T. (2010). G-less cassette in vitro transcription using HeLa cell nuclear extracts. *Cold Spring Harb. Protoc.* *2010*, pdb.prot5387.

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

Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., Ule, J., Manley, J.L., and Shi, Y. (2014). CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev.* *28*, 2370–2380.

Chang, T.-H., Huang, H.-Y., Hsu, J.B.-K., Weng, S.-L., Horng, J.-T., and Huang, H.-D. (2013). An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics* *14*, S4.

Chapon, C., Cech, T.R., and Zaug, A.J. (1997). Polyadenylation of telomerase RNA in budding yeast. *RNA N. Y. N* *3*, 1337–1351.

- Chen, J., and Wagner, E.J. (2010). snRNA 3' end formation: the dawn of the Integrator complex. *Biochem. Soc. Trans.* 38, 1082–1087.
- Chen, J.-L., and Greider, C.W. (2004). Telomerase RNA structure and function: implications for dyskeratosis congenita. *Trends Biochem. Sci.* 29, 183–192.
- Chen, F., MacDonald, C.C., and Wilusz, J. (1995). Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res.* 23, 2614–2620.
- Chen, J., Ezzeddine, N., Waltenspiel, B., Albrecht, T.R., Warren, W.D., Marzluff, W.F., and Wagner, E.J. (2012). An RNAi screen identifies additional members of the Drosophila Integrator complex and a requirement for cyclin C/Cdk8 in snRNA 3'-end formation. *RNA N. Y. N* 18, 2148–2156.
- Chen, J., Waltenspiel, B., Warren, W.D., and Wagner, E.J. (2013a). Functional analysis of the integrator subunit 12 identifies a microdomain that mediates activation of the Drosophila integrator complex. *J. Biol. Chem.* 288, 4867–4877.
- Chen, T.-H., Lai, Y.-H., Lee, P.-L., Hsu, J.-H., Goto, K., Hayashi, Y.K., Nishino, I., Lin, C.-W., Shih, H.-H., Huang, C.-C., et al. (2013b). Infantile facioscapulohumeral muscular dystrophy revisited: Expansion of clinical phenotypes in patients with a very short EcoRI fragment. *Neuromuscul. Disord. NMD* 23, 298–305.
- Chiang, M.Y., Chan, H., Zounes, M.A., Freier, S.M., Lima, W.F., and Bennett, C.F. (1991). Antisense oligonucleotides inhibit intercellular adhesion molecule 1 expression by two distinct mechanisms. *J. Biol. Chem.* 266, 18162–18171.

Choi, S.H., Gearhart, M.D., Cui, Z., Bosnakovski, D., Kim, M., Schennum, N., and Kyba, M. (2016). DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.*

Clapp, J., Mitchell, L.M., Bolland, D.J., Fantes, J., Corcoran, A.E., Scotting, P.J., Armour, J.A.L., and Hewitt, J.E. (2007). Evolutionary Conservation of a Coding Function for D4Z4, the Tandem DNA Repeat Mutated in Facioscapulohumeral Muscular Dystrophy. *Am. J. Hum. Genet.* 81, 264–279.

Corish, P., and Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Eng.* 12, 1035–1040.

Coy, S., Volanakis, A., Shah, S., and Vasiljeva, L. (2013). The Sm complex is required for the processing of non-coding RNAs by the exosome. *PloS One* 8, e65606.

Crooke, S.T. (2004). Progress in antisense technology. *Annu. Rev. Med.* 55, 61–95.

Danckwardt, S., Hentze, M.W., and Kulozik, A.E. (2008). 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* 27, 482–498.

Dandapat, A., Bosnakovski, D., Hartweck, L.M., Arpke, R.W., Baltgalvis, K.A., Vang, D., Baik, J., Darabi, R., Perlingeiro, R.R., Hamra, F.K., et al. (2014). Dominant lethal pathologies in male mice engineered to contain an X-linked DUX4 transgene. *Cell Rep.* 8, 1484–1496.

Dandapat, A., Perrin, B.J., Cabelka, C., Razzoli, M., Ervasti, J.M., Bartolomucci, A., Lowe, D.A., and Kyba, M. (2016). High Frequency Hearing Loss and Hyperactivity in DUX4 Transgenic Mice. *PloS One* 11, e0151467.

Deenen, J.C.W., Arnts, H., van der Maarel, S.M., Padberg, G.W., Verschuuren, J.J.G.M., Bakker, E., Weinreich, S.S., Verbeek, A.L.M., and van Engelen, B.G.M. (2014). Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* 83, 1056–1059.

Derti, A., Garrett-Engle, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., and Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Res.* 22, 1173–1183.

van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., and Frants, R.R. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037–2042.

Dez, C., Henras, A., Faucon, B., Lafontaine, D.L.J., Caizergues-Ferrer, M., and Henry, Y. (2001). Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res.* 29, 598–603.

Dixit, M., Anseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Mattéotti, C., Acker, A.M. van, Leo, O., et al. (2007). DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci.* 104, 18157–18162.

Dominski, Z. (2010). The hunt for the 3' endonuclease. *Wiley Interdiscip. Rev. RNA* 1, 325–340.

Dominski, Z., and Marzluff, W.F. (2007). Formation of the 3' end of histone mRNA: Getting closer to the end. *Gene* 396, 373–390.

Feng, Q., Snider, L., Jagannathan, S., Tawil, R., van der Maarel, S.M., Tapscott, S.J., and Bradley, R.K. (2015). A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *eLife* 4.

Ferreboeuf, M., Mariot, V., Bessi res, B., Vasiljevic, A., Atti -Bitach, T., Collardeau, S., Morere, J., Roche, S., Magdinier, F., Robin-Ducellier, J., et al. (2014). DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. *Hum. Mol. Genet.* 23, 171–181.

Freier, S.M., and Altmann, K.H. (1997). The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.* 25, 4429–4443.

Fu, D., and Collins, K. (2003). Distinct Biogenesis Pathways for Human Telomerase RNA and H/ACA Small Nucleolar RNAs. *Mol. Cell* 11, 1361–1372.

Funakoshi, M., Goto, K., and Arahata, K. (1998). Epilepsy and mental retardation in a subset of early onset 4q35-facioscapulohumeral muscular dystrophy. *Neurology* 50, 1791–1794.

Furger, A., Schaller, A., and Sch mperli, D. (1998). Functional importance of conserved nucleotides at the histone RNA 3' processing site. *RNA* 4, 246–256.

Gabellini, D., D'Antona, G., Moggio, M., Prella, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R., et al. (2006). Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* 439, 973–977.

van Geel, M., Dickson, M.C., Beck, A.F., Bolland, D.J., Frants, R.R., van der Maarel, S.M., de Jong, P.J., and Hewitt, J.E. (2002). Genomic Analysis of Human Chromosome 10q and 4q Telomeres Suggests a Common Origin. *Genomics* 79, 210–217.

Gehring, N.H., Frede, U., Neu-Yilik, G., Hundsdoerfer, P., Vetter, B., Hentze, M.W., and Kulozik, A.E. (2001). Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat. Genet.* 28, 389–392.

Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* 23, 265–279.

Geng, L.N., Tyler, A.E., and Tapscott, S.J. (2011). Immunodetection of Human Double Homeobox 4. *Hybridoma* 30, 125–130.

Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38–51.

Giussani, M., Cardone, M.F., Bodega, B., Ginelli, E., and Meneveri, R. (2012). Evolutionary history of linked D4Z4 and Beta satellite clusters at the FSHD locus (4q35). *Genomics*.

de Greef, J.C., Lemmers, R.J.L.F., van Engelen, B.G.M., Sacconi, S., Venance, S.L., Frants, R.R., Tawil, R., and van der Maarel, S.M. (2009). Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* 30, 1449–1459.

Gruber, A.R., Martin, G., Müller, P., Schmidt, A., Gruber, A.J., Gumienny, R., Mittal, N., Jayachandran, R., Pieters, J., Keller, W., et al. (2014). Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat. Commun.* 5, 5465.

Gunisova, S., Elboher, E., Nosek, J., Gorkovoy, V., Brown, Y., Lucier, J.-F., Laterreur, N., Wellinger, R.J., Tzfati, Y., and Tomaska, L. (2009). Identification and comparative analysis of telomerase RNAs from *Candida* species reveal conservation of functional elements. *RNA* 15, 546–559.

Guo, Z., and Sherman, F. (1996). 3'-end-forming signals of yeast mRNA. *Trends Biochem. Sci.* 21, 477–481.

Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S., and Berger, I. (2016). Zooming in on Transcription Preinitiation. *J. Mol. Biol.*

Gutschner, T., Hämmerle, M., and Diederichs, S. (2013). MALAT1 — a paradigm for long noncoding RNA function in cancer. *J. Mol. Med.* 91, 791–801.

Haidar, A., Arekapudi, S., DeMattia, F., Abu-Isa, E., and Kraut, M. (2015). High-Grade Undifferentiated Small Round Cell Sarcoma with t(4;19)(q35;q13.1) CIC-DUX4 Fusion: Emerging Entities of Soft Tissue Tumors with Unique Histopathologic Features – A Case Report and Literature Review. *Am. J. Case Rep.* 16, 87–94.

Hallais, M., Pontvianne, F., Andersen, P.R., Clerici, M., Lener, D., Benbahouche, N.E.H., Gostan, T., Vandermoere, F., Robert, M.-C., Cusack, S., et al. (2013). CBC–ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat. Struct. Mol. Biol.* 20, 1358–1366.

Hampsey, M. (1998). Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. *Microbiol. Mol. Biol. Rev.* 62, 465–503.

Haussecker, D. (2016). Stacking up CRISPR against RNAi for therapeutic gene inhibition. *FEBS J.*

Hernandez, N., and Weiner, A.M. (1986). Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell* 47, 249–258.

Hewitt, J.E. (2015). Loss of epigenetic silencing of the DUX4 transcription factor gene in facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* 24, R17-23.

Himeda, C.L., Jones, T.I., and Jones, P.L. (2014). Facioscapulohumeral Muscular Dystrophy As a Model for Epigenetic Regulation and Disease. *Antioxid. Redox Signal.* 22, 1463–1482.

Himeda, C.L., Jones, T.I., and Jones, P.L. (2015). CRISPR/dCas9-mediated transcriptional inhibition ameliorates the epigenetic dysregulation at D4Z4 and represses DUX4-fl in FSH muscular dystrophy. *Mol. Ther. J. Am. Soc. Gene Ther.*

Himeda, C.L., Jones, T.I., and Jones, P.L. (2016). Scalpel or Straitjacket: CRISPR/Cas9 Approaches for Muscular Dystrophies. *Trends Pharmacol. Sci.* 37, 249–251.

Homma, S., Chen, J.C.J., Rahimov, F., Beermann, M.L., Hanger, K., Bibat, G.M., Wagner, K.R., Kunkel, L.M., Emerson, C.P., and Miller, J.B. (2012). A unique library of myogenic cells from facioscapulohumeral muscular dystrophy subjects and unaffected relatives: family, disease and cell function. *Eur. J. Hum. Genet. EJHG* 20, 404–410.

Hu, J., Lutz, C.S., Wilusz, J., and Tian, B. (2005). Bioinformatic Identification of Candidate Cis-Regulatory Elements Involved in Human mRNA Polyadenylation. *RNA* 11, 1485–1493.

- Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* 8, 39.
- Ivanov, I.P., Firth, A.E., Michel, A.M., Atkins, J.F., and Baranov, P.V. (2011). Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res.* 39, 4220–4234.
- Iwakawa, H., and Tomari, Y. (2015). The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol.* 25, 651–665.
- Jády, B.E., Bertrand, E., and Kiss, T. (2004). Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body–specific localization signal. *J. Cell Biol.* 164, 647–652.
- Jawdekar, G.W., and Henry, R.W. (2008). Transcriptional regulation of human small nuclear RNA genes. *Biochim. Biophys. Acta* 1779, 295–305.
- Jeronimo, C., Bataille, A.R., and Robert, F. (2013). The Writers, Readers, and Functions of the RNA Polymerase II C-Terminal Domain Code. *Chem. Rev.* 113, 8491–8522.
- Jin, Y., Chen, Y., and Gong, Z. (2005). Determination and analysis of the pre-mRNA cleavage sites in Arabidopsis. *RNA Biol.* 2, 83–85.
- Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* 16, 421–433.
- Jones, T.I., King, O.D., Himeda, C.L., Homma, S., Chen, J.C.J., Beermann, M.L., Yan, C., Emerson, C.P., Miller, J.B., Wagner, K.R., et al. (2015). Individual epigenetic status of the

pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. *Clin. Epigenetics* 7, 37.

Jones, T.I., Parilla, M., and Jones, P.L. (2016). Transgenic *Drosophila* for Investigating DUX4 and FRG1 , Two Genes Associated with Facioscapulohumeral Muscular Dystrophy (FSHD). *PLOS ONE* 11, e0150938.

Kaida, D. (2016). The reciprocal regulation between splicing and 3'-end processing. *Wiley Interdiscip. Rev. RNA*.

Kannan, R., Hartnett, S., Voelker, R.B., Berglund, J.A., Staley, J.P., and Baumann, P. (2013). Intronic sequence elements impede exon ligation and trigger a discard pathway that yields functional telomerase RNA in fission yeast. *Genes Dev.* 27, 627–638.

Kannan, R., Helston, R.M., Dannebaum, R.O., and Baumann, P. (2015). Diverse mechanisms for spliceosome-mediated 3' end processing of telomerase RNA. *Nat. Commun.* 6, 6104.

Klauser, A.A., and van Hoof, A. (2012). Degradation of mRNAs that lack a stop codon: A decade of nonstop progress. *Wiley Interdiscip. Rev. RNA* 3, 649–660.

Klerk, E. de, Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O., Trollet, C., Dickson, G., Dunnen, J.T. den, Maarel, S.M. van der, Raz, V., et al. (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res.* 40, 9089–9101.

Köhn, M., and Hüttelmaier, S. (2016). Non-coding RNAs, the cutting edge of histone messages. *RNA Biol.* 13, 367–372.

Kole, R., Krainer, A.R., and Altman, S. (2012). RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug Discov.* 11, 125–140.

Kowaljew, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Mattéotti, C., Arias, C., Corona, E.D., Nuñez, N.G., Leo, O., et al. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.* *17*, 611–623.

Koziolkiewicz, M., Krakowiak, A., Kwinkowski, M., Boczkowska, M., and Stec, W.J. (1995). Stereodifferentiation--the effect of P chirality of oligo(nucleoside phosphorothioates) on the activity of bacterial RNase H. *Nucleic Acids Res.* *23*, 5000–5005.

Krom, Y.D., Thijssen, P.E., Young, J.M., den Hamer, B., Balog, J., Yao, Z., Maves, L., Snider, L., Knopp, P., Zammit, P.S., et al. (2013). Intrinsic Epigenetic Regulation of the D4Z4 Macrosatellite Repeat in a Transgenic Mouse Model for FSHD. *PLoS Genet* *9*, e1003415.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* *12*, 283–294.

Legendre, M., and Gautheret, D. (2003). Sequence determinants in human polyadenylation site selection. *BMC Genomics* *4*, 7.

Lek, A., Rahimov, F., Jones, P.L., and Kunkel, L.M. (2015). Emerging preclinical animal models for FSHD. *Trends Mol. Med.* *21*, 295–306.

Lemmers, R.J.F.L., Wohlgemuth, M., Frants, R.R., Padberg, G.W., Morava, E., and van der Maarel, S.M. (2004). Contractions of D4Z4 on 4qB Subtelomeres Do Not Cause Facioscapulohumeral Muscular Dystrophy. *Am. J. Hum. Genet.* *75*, 1124–1130.

Lemmers, R.J.L.F., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M.M., de Knijff, P., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2007).

Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* 81, 884–894.

Lemmers, R.J.L.F., van der Vliet, P.J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010a). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650–1653.

Lemmers, R.J.L.F., van der Vliet, P.J., van der Gaag, K.J., Zuniga, S., Frants, R.R., de Knijff, P., and van der Maarel, S.M. (2010b). Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. *Am. J. Hum. Genet.* 86, 364–377.

Lemmers, R.J.L.F., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W.E., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.*

Lemmers, R.J.L.F., van den Boogaard, M.L., van der Vliet, P.J., Donlin-Smith, C.M., Nations, S.P., Ruivenkamp, C.A.L., Heard, P., Bakker, B., Tapscott, S., Cody, J.D., et al. (2015). Hemizyosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. *Hum. Mutat.* 36, 679–683.

Li, L., and Chaikof, E.L. (2002). Quantitative nuclear run-off transcription assay. *BioTechniques* 33, 1016–1017.

Li, X.-Q., and Du, D. (2013). RNA Polyadenylation Sites on the Genomes of Microorganisms, Animals, and Plants. *PLOS ONE* 8, e79511.

Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science* 352, 600–604.

Lim, J.-W., Snider, L., Yao, Z., Tawil, R., Van Der Maarel, S.M., Rigo, F., Bennett, C.F., Filippova, G.N., and Tapscott, S.J. (2015). DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FSHD. *Hum. Mol. Genet.* 24, 4817–4828.

van der Maarel, S.M., Miller, D.G., Tawil, R., Filippova, G.N., and Tapscott, S.J. (2012). Facioscapulohumeral muscular dystrophy: consequences of chromatin relaxation. *Curr. Opin. Neurol.* 25, 614–620.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end processing endonuclease. *Nature* 444.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell. Mol. Life Sci.* 65, 1099–1122.

Marsollier, A.-C., Ciszewski, L., Mariot, V., Popplewell, L., Voit, T., Dickson, G., and Dumonceaux, J. (2016). Antisense targeting of 3' end elements involved in DUX4 mRNA processing is an efficient therapeutic strategy for Facioscapulohumeral Dystrophy: a new gene silencing approach. *Hum. Mol. Genet.*

Martick, M., Horan, L.H., Noller, H.F., and Scott, W.G. (2008). A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature* 454, 899–902.

- Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* 9, 843–854.
- Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.-B., Li, W., and Wagner, E.J. (2014). CFI_{m25} links alternative polyadenylation to glioblastoma tumour suppression. *Nature* 510, 412–416.
- Masamha, C.P., Xia, Z., Peart, N., Collum, S., Li, W., Wagner, E.J., and Shyu, A.-B. (2016). CFI_{m25} regulates glutaminase alternative terminal exon definition to modulate miR-23 function. *RNA* 22, 830–838.
- Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* 15, 108–121.
- Mayfield, J.E., Burkholder, N.T., and Zhang, Y.J. (2016). Dephosphorylating eukaryotic RNA polymerase II. *Biochim. Biophys. Acta BBA - Proteins Proteomics* 1864, 372–387.
- Mayr, C. (2016). Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol.* 26, 227–237.
- Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138, 673.
- Mitchell, J.R., Cheng, J., and Collins, K. (1999). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell. Biol.* 19, 567–576.
- Mitsuhashi, H., Mitsuhashi, S., Lynn-Jones, T., Kawahara, G., and Kunkel, L.M. (2012). Expression of DUX4 in zebrafish development recapitulates Facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.*

Monia, B.P., Lesnik, E.A., Gonzalez, C., Lima, W.F., McGee, D., Guinosso, C.J., Kawasaki, A.M., Cook, P.D., and Freier, S.M. (1993). Evaluation of 2'-modified oligonucleotides containing 2'-deoxy gaps as antisense inhibitors of gene expression. *J. Biol. Chem.* 268, 14514–14522.

Monia, B.P., Johnston, J.F., Sasmor, H., and Cummins, L.L. (1996). Nuclease Resistance and Antisense Activity of Modified Oligonucleotides Targeted to Ha-ras. *J. Biol. Chem.* 271, 14533–14540.

Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W., and Bozzoni, I. (2002). Functional Analysis of Yeast snoRNA and snRNA 3'-End Formation Mediated by Uncoupling of Cleavage and Polyadenylation. *Mol. Cell. Biol.* 22, 1379–1389.

Mould, A.W., Pang, Z., Pakusch, M., Tonks, I.D., Stark, M., Carrie, D., Mukhopadhyay, P., Seidel, A., Ellis, J.J., Deakin, J., et al. (2013). Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin* 6, 19.

Mul, K., Lassche, S., Voermans, N.C., Padberg, G.W., Horlings, C.G., and van Engelen, B.G. (2016). What's in a name? The clinical features of facioscapulohumeral muscular dystrophy. *Pract. Neurol.*

Murphy, J.R. (1996). *Corynebacterium Diphtheriae*. In *Medical Microbiology*, S. Baron, ed. (Galveston (TX): University of Texas Medical Branch at Galveston), p.

Newlands, S., Levitt, L.K., Robinson, C.S., Karpf, A.B.C., Hodgson, V.R.M., Wade, R.P., and Hardeman, E.C. (1998). Transcription occurs in pulses in muscle fibers. *Genes Dev.* 12, 2748–2758.

Nikolic, A., Ricci, G., Sera, F., Bucci, E., Govi, M., Mele, F., Rossi, M., Ruggiero, L., Vercelli, L., Ravaglia, S., et al. (2016). Clinical expression of facioscapulohumeral muscular dystrophy in carriers of 1-3 D4Z4 reduced alleles: experience of the FSHD Italian National Registry. *BMJ Open* 6, e007798.

Noël, J.-F., Larose, S., Elela, S.A., and Wellinger, R.J. (2012). Budding yeast telomerase RNA transcription termination is dictated by the Nrd1/Nab3 non-coding RNA termination pathway. *Nucleic Acids Res.* 40, 5625–5636.

Ogorodnikov, A., Kargapolova, Y., and Danckwardt, S. (2016). Processing and transcriptome expansion at the mRNA 3' end in health and disease: finding the right end. *Pflüg. Arch. - Eur. J. Physiol.* 1–20.

Orkin, S.H., Cheng, T.C., Antonarakis, S.E., and Kazazian, H.H.J. (1985). Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *EMBO J.* 4, 453–456.

Osborne, R.J., Welle, S., Venance, S.L., Thornton, C.A., and Tawil, R. (2007). Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* 68, 569–577.

van Overveld, P.G.M., Lemmers, R.J.F.L., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.-J.B., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* 35, 315–317.

Pandey, S.N., Khawaja, H., and Chen, Y.-W. (2015). Culture Conditions Affect Expression of DUX4 in FSHD Myoblasts. *Mol. Basel Switz.* 20, 8304–8315.

Pauws, E., van Kampen, A.H.C., van de Graaf, S.A.R., de Vijlder, J.J.M., and Ris-Stalpers, C. (2001). Heterogeneity in polyadenylation cleavage sites in mammalian mRNA sequences: implications for SAGE analysis. *Nucleic Acids Res.* 29, 1690–1694.

Peart, N., Sataluri, A., Baillat, D., and Wagner, E.J. (2013). Non-mRNA 3' end formation: how the other half lives. *Wiley Interdiscip. Rev. RNA* 4, 491–506.

Porrua, O., and Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* 16, 190–202.

Potaczek, D.P., Garn, H., Unger, S.D., and Renz, H. (2016). Antisense molecules: A new class of drugs. *J. Allergy Clin. Immunol.* 137, 1334–1346.

Rahimov, F., King, O.D., Leung, D.G., Bibat, G.M., Emerson, C.P., Kunkel, L.M., and Wagner, K.R. (2012). Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. *Proc. Natl. Acad. Sci.* 109, 16234–16239.

Ramamurthy, L., Ingledue, T.C., Pilch, D.R., Kay, B.K., and Marzluff, W.F. (1996). Increasing the distance between the snRNA promoter and the 3' box decreases the efficiency of snRNA 3'-end formation. *Nucleic Acids Res.* 24, 4525–4534.

Ricci, G., Scionti, I., Sera, F., Govi, M., D'Amico, R., Frambolli, I., Mele, F., Filosto, M., Vercelli, L., Ruggiero, L., et al. (2013). Large scale genotype-phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. *Brain J. Neurol.* 136, 3408–3417.

Richards, M., Coppée, F., Thomas, N., Belayew, A., and Upadhyaya, M. (2012). Facioscapulohumeral muscular dystrophy (FSHD): an enigma unravelled? *Hum. Genet.* 131, 325–340.

Rickard, A.M., Petek, L.M., and Miller, D.G. (2015). Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol. Genet.*

Romeo, V., and Schümperli, D. (2016). Cycling in the nucleus: regulation of RNA 3' processing and nuclear organization of replication-dependent histone genes. *Curr. Opin. Cell Biol.* 40, 23–31.

Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: breaking up is hard to do. *Genes Dev.* 20, 1050–1056.

Rubtsova, M.P., Vasilkova, D.P., Malyavko, A.N., Naraikina, Y.V., Zvereva, M.I., and Dontsova, O.A. (2012). Telomere lengthening and other functions of telomerase. *Acta Naturae* 4, 44–61.

Sacconi, S., Lemmers, R.J.L.F., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *Am. J. Hum. Genet.* 93, 744–751.

Sacconi, S., Salviati, L., and Desnuelle, C. (2015). Facioscapulohumeral muscular dystrophy. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* 1852, 607–614.

Saito, Y., Miyashita, S., Yokoyama, A., Komaki, H., Seki, A., Maegaki, Y., and Ohno, K. (2007). Facioscapulohumeral muscular dystrophy with severe mental retardation and epilepsy. *Brain Dev.* 29, 231–233.

Sazani, P., and Kole, R. (2003). Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing. *J. Clin. Invest.* 112, 481–486.

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev.* 28, 2381–2393.

Scionti, I., Greco, F., Ricci, G., Govi, M., Arashiro, P., Vercelli, L., Berardinelli, A., Angelini, C., Antonini, G., Cao, M., et al. (2012). Large-Scale Population Analysis Challenges the Current Criteria for the Molecular Diagnosis of Fascioscapulohumeral Muscular Dystrophy. *Am. J. Hum. Genet.* 90, 628–635.

Senn, J.J., Burel, S., and Henry, S.P. (2005). Non-CpG-Containing Antisense 2'-Methoxyethyl Oligonucleotides Activate a Proinflammatory Response Independent of Toll-Like Receptor 9 or Myeloid Differentiation Factor 88. *J. Pharmacol. Exp. Ther.* 314, 972–979.

Seto, A.G., Zaug, A.J., Sobel, S.G., Wolin, S.L., and Cech, T.R. (1999). *Saccharomyces cerevisiae* telomerase is an Sm small nuclear ribonucleoprotein particle. *Nature* 401, 177–180.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* 343, 84–87.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* 22, 1567–1572.

Sharma, V., Harafuji, N., Belayew, A., and Chen, Y.-W. (2013). DUX4 differentially regulates transcriptomes of human rhabdomyosarcoma and mouse C2C12 cells. *PloS One* 8, e64691.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* 18, 5799–5805.

Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA N. Y. N* 17, 761–772.

Shi, Y., and Manley, J.L. (2015). The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 29, 889–897.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 33, 365–376.

Shimberg, G.D., Michalek, J.L., Oluyadi, A.A., Rodrigues, A.V., Zucconi, B.E., Neu, H.M., Ghosh, S., Sureschandra, K., Wilson, G.M., Stemmler, T.L., et al. (2016). Cleavage and polyadenylation specificity factor 30: An RNA-binding zinc-finger protein with an unexpected 2Fe–2S cluster. *Proc. Natl. Acad. Sci.* 113, 4700–4705.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol. Cell* 42, 794–805.

Smale, S.T. (2009). Nuclear run-on assay. *Cold Spring Harb. Protoc.* 2009, pdb.prot5329.

Smith, D.J., Query, C.C., and Konarska, M.M. (2008). “Nought may endure but mutability”: spliceosome dynamics and the regulation of splicing. *Mol. Cell* 30, 657–666.

Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J.L.F., Winokur, S.T., Tawil, R., et al. (2009). RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.* 18, 2414–2430.

Snider, L., Geng, L.N., Lemmers, R.J.L.F., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.* 6, e1001181.

Stadler, G., Chen, J.C., Wagner, K., Robin, J.D., Shay, J.W., Emerson, C.P., and Wright, W.E. (2011). Establishment of clonal myogenic cell lines from severely affected dystrophic muscles - CDK4 maintains the myogenic population. *Skelet. Muscle* 1, 12.

Stadler, G., Rahimov, F., King, O.D., Chen, J.C.J., Robin, J.D., Wagner, K.R., Shay, J.W., Emerson, C.P., Jr., and Wright, W.E. (2013). Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* 20, 671–678.

Statland, J., and Tawil, R. (2014). Facioscapulohumeral Muscular Dystrophy. *Neurol. Clin.* 32, 721–728.

Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413, 327–331.

Sunwoo, H., Dinger, M.E., Wilusz, J.E., Amaral, P.P., Mattick, J.S., and Spector, D.L. (2009). MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* 19, 347–359.

- Tang, W., Kannan, R., Blanchette, M., and Baumann, P. (2012). Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* 484, 260–264.
- Tassin, A., Laoudj-Chenivesse, D., Vanderplanck, C., Barro, M., Charron, S., Anseau, E., Chen, Y.-W., Mercier, J., Coppée, F., and Belayew, A. (2012). DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? *J. Cell. Mol. Med.*
- Tawil, R., and Van Der Maarel, S.M. (2006). Facioscapulohumeral muscular dystrophy. *Muscle Nerve* 34, 1–15.
- Tawil, R., Shaw, D.W., van der Maarel, S.M., and Tapscott, S.J. (2014). Clinical trial preparedness in facioscapulohumeral dystrophy: Outcome measures and patient access: 8–9 April 2013, Leiden, The Netherlands. *Neuromuscul. Disord.* 24, 79–85.
- Tawil, R., Kissel, J.T., Heatwole, C., Pandya, S., Gronseth, G., and Benatar, M. (2015). Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the Practice Issues Review Panel of the American Association of Neuromuscular & Electrodiagnostic Medicine. *Neurology* 85, 357–364.
- Theimer, C.A., Jády, B.E., Chim, N., Richard, P., Breece, K.E., Kiss, T., and Feigon, J. (2007). Structural and Functional Characterization of Human Telomerase RNA Processing and Cajal Body Localization Signals. *Mol. Cell* 27, 869–881.
- Tian, B., and Graber, J.H. (2012). Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA* 3, 385–396.

- Tian, B., Hu, J., Zhang, H., and Lutz, C.S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* 33, 201–212.
- Tschuch, C., Schulz, A., Pscherer, A., Werft, W., Benner, A., Hotz-Wagenblatt, A., Barrionuevo, L.S., Lichter, P., and Mertens, D. (2008). Off-target effects of siRNA specific for GFP. *BMC Mol. Biol.* 9, 60.
- Tsumagari, K., Chang, S.-C., Lacey, M., Baribault, C., Chittur, S.V., Sowden, J., Tawil, R., Crawford, G.E., and Ehrlich, M. (2011). Gene expression during normal and FSHD myogenesis. *BMC Med. Genomics* 4, 67.
- Vickers, T.A., Wyatt, J.R., Burckin, T., Bennett, C.F., and Freier, S.M. (2001). Fully modified 2' MOE oligonucleotides redirect polyadenylation. *Nucleic Acids Res.* 29, 1293–1299.
- Wagner, E.J., and Garcia-Blanco, M.A. (2002). RNAi-Mediated PTB Depletion Leads to Enhanced Exon Definition. *Mol. Cell* 10, 943–949.
- Wagner, E.J., Berkow, A., and Marzluff, W.F. (2005). Expression of an RNAi-resistant SLBP restores proper S-phase progression. *Biochem. Soc. Trans.* 33, 471–473.
- Wahle, E., and Rügsegger, U. (1999). 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.* 23, 277–295.
- Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J., and Harper, S.Q. (2011). DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann. Neurol.* 69, 540–552.
- Wang, Z.-Q., Wang, N., van der Maarel, S., Murong, S.-X., and Wu, Z.-Y. (2011). Distinguishing the 4qA and 4qB variants is essential for the diagnosis of

facioscapulohumeral muscular dystrophy in the Chinese population. *Eur. J. Hum. Genet.* EJHG 19, 64–69.

Watts, J.K., and Corey, D.R. (2012). Gene silencing by siRNAs and antisense oligonucleotides in the laboratory and the clinic. *J. Pathol.* 226, 365–379.

Webb, C.J., and Zakian, V.A. (2008). Identification and characterization of the *Schizosaccharomyces pombe* TER1 telomerase RNA. *Nat. Struct. Mol. Biol.* 15, 34–42.

Weill, L., Belloc, E., Bava, F.-A., and Méndez, R. (2012). Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat. Struct. Mol. Biol.* 19, 577–585.

Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., and Williamson, R. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* 2, 26–30.

Wilusz, J.E., and Spector, D.L. (2010). An unexpected ending: noncanonical 3' end processing mechanisms. *RNA N. Y. N* 16, 259–266.

Wilusz, J., Shenk, T., Takagaki, Y., and Manley, J.L. (1990). A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. *Mol. Cell. Biol.* 10, 1244–1248.

Wilusz, J.E., Freier, S.M., and Spector, D.L. (2008). 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135, 919–932.

Wilusz, J.E., JnBaptiste, C.K., Lu, L.Y., Kuhn, C.-D., Joshua-Tor, L., and Sharp, P.A. (2012). A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev.* 26, 2392–2407.

- Wuebbles, R.D., Long, S.W., Hanel, M.L., and Jones, P.L. (2010). Testing the effects of FSHD candidate gene expression in vertebrate muscle development. *Int. J. Clin. Exp. Pathol.* 3, 386–400.
- Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., and Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat. Commun.* 5, 5274.
- Xiang, K., Tong, L., and Manley, J.L. (2014). Delineating the Structural Blueprint of the Pre-mRNA 3'-End Processing Machinery. *Mol. Cell. Biol.* 34, 1894–1910.
- Yamazaki, T., and Hirose, T. (2015). The building process of the functional paraspeckle with long non-coding RNAs. *Front. Biosci. Elite Ed.* 7, 1–41.
- Yang, Q., and Doubl  , S. (2011). Structural biology of poly(A) site definition. *Wiley Interdiscip. Rev. RNA* 2, 732–747.
- Yang, X.-C., Burch, B.D., Yan, Y., Marzluff, W.F., and Dominski, Z. (2009). FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs. *Mol. Cell* 36, 267–278.
- Yoon, J.B., Murphy, S., Bai, L., Wang, Z., and Roeder, R.G. (1995). Proximal sequence element-binding transcription factor (PTF) is a multisubunit complex required for transcription of both RNA polymerase II- and RNA polymerase III-dependent small nuclear RNA genes. *Mol. Cell. Biol.* 15, 2019–2027.
- Young, J.M., Whiddon, J.L., Yao, Z., Kasinathan, B., Snider, L., Geng, L.N., Balog, J., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2013). DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet.* 9, e1003947.

Zarkower, D., and Wickens, M. (1988). A functionally redundant downstream sequence in SV40 late pre-mRNA is required for mRNA 3'-end formation and for assembly of a precleavage complex in vitro. *J. Biol. Chem.* 263, 5780–5788.

Zeng, W., de Greef, J.C., Chen, Y.-Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., et al. (2009). Specific Loss of Histone H3 Lysine 9 Trimethylation and HP1 γ /Cohesin Binding at D4Z4 Repeats Is Associated with Facioscapulohumeral Dystrophy (FSHD). *PLoS Genet* 5, e1000559.

Zeng, W., Chen, Y.-Y., Newkirk, D.A., Wu, B., Balog, J., Kong, X., Ball, A.R., Zanotti, S., Tawil, R., Hashimoto, N., et al. (2014). Genetic and epigenetic characteristics of FSHD-associated 4q and 10q D4Z4 that are distinct from non-4q/10q D4Z4 homologs. *Hum. Mutat.* 35, 998–1010.

Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol. Cell* 59, 437–448.

Zhang, Y., King, O.D., Rahimov, F., Jones, T.I., Ward, C.W., Kerr, J.P., Liu, N., Emerson, C.P., Jr, Kunkel, L.M., Partridge, T.A., et al. (2014). Human skeletal muscle xenograft as a new preclinical model for muscle disorders. *Hum. Mol. Genet.*

Zhang, Y., Lee, J.K., Toso, E.A., Lee, J.S., Choi, S.H., Slattery, M., Aihara, H., and Kyba, M. (2016). DNA-binding sequence specificity of DUX4. *Skelet. Muscle* 6.

VITA

Natoya Janeen Peart was born in St. Catherine, Jamaica on April 18, 1988, the daughter of Reverend James Michael Anthony Peart and Nalda Eugene Peart. After completing her secondary education at Manning's School in the parish of Westmoreland, Jamaica, she entered Yale University, in New Haven, Connecticut. She received the degree of Bachelor of Science in Molecular Biophysics and Biochemistry from Yale in May 2010. In August of 2010, she entered the University Of Texas Graduate School Of Biomedical Sciences at Houston. She completed her dissertation under the supervision of Dr. Eric J. Wagner.

Permanent Address:

426 Cane Avenue

Shrewsbury Housing Scheme, Petersfield

Westmoreland

Jamaica, W.I.