Characterizing ARS2 localization and function in differentiating myoblasts

by

Jennifer Christie
B.Sc., University of Victoria, 2009

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

SUPERVISORY COMMITTEE

Characterizing ARS2 localization and function in differentiating myoblasts

by

Jennifer Christie
B.Sc., University of Victoria, 2009

Supervisory Committee

Dr. Perry L. Howard, Department of Biology and Department of Biochemistry & Microbiology
Supervisor

Dr. Robert D. Burke, Department of Biochemistry & Microbiology
Departmental Member

Dr. Leigh Anne Swayne, Division of Medical Sciences
Outside Member

ABSTRACT

Supervisory Committee
Dr. Perry L. Howard, Department of Biology and Department of Biochemistry & Microbiology
Supervisor
Dr. Robert D. Burke, Department of Biochemistry & Microbiology
Departmental Member
Dr. Leigh Anne Swayne, Division of Medical Sciences
Outside Member

ARS2 is a member of the nuclear cap-binding complex (CBC) that is critical for a number of RNA processing pathways. The emerging model is that ARS2 acts as a master regulator of RNAPII transcript maturation by bringing capped RNA substrates together with the appropriate processing machinery. ARS2 is essential for early mammalian development but it remains unclear precisely how ARS2 functions in stem and progenitor cell maintenance and differentiation. The purpose of this study was to answer basic questions about the localization and function of ARS2 in muscle progenitor cells. Here I describe the localization of ARS2 in proliferating myoblasts and post-mitotic differentiating myotubes and show that disruption of ARS2 expression levels by knockdown or overexpression results in impaired myogenic differentiation. I also discovered a new isoform of ARS2 that is localized exclusively in the cytoplasm and found preliminary evidence that ARS2 is required for nonsense-mediated decay (NMD). This study includes the first evidence that an ARS2 isoform is expressed in the cytoplasm and opens the door for the discovery of new ARS2 functions beyond its reported roles in the nucleus.

PREFACE

I have performed all experiments and generated the data and figures presented in this thesis with the following exceptions:

- Antibody validation by immunofluorescence staining of mouse embryos was performed by Dr. Mike Wilson
- Antibody validation by western blotting was performed by Connor O'Sullivan
- Connor O'Sullivan performed BrdU staining of C2C12 myotubes and ARS2 detection in myotube lysates by western blotting. He also generated the corresponding figure.
- The overexpression differentiation assay was performed by former honours student Chris Yeker according to my design and under my supervision
- Co-immunoprecipitation of 3XFLAG-ARS2 and EGFP-FARB was performed by Jake Gambling
- Validation of ARS2 and DROSHA knockdown by western blotting was performed by Connor O'Sullivan
- Detection of the 60 kDa ARS2 band by western blotting was performed by Connor O'Sullivan
- EGFP-Δ227-ARS2 was cloned by Dr. Perry Howard
- Co-immunoprecipitation of 3XFLAG-ARS2 and RNPS1 was performed by Dr. Perry Howard
- The nonsense-mediated decay luciferase assay was performed by former honours student Rachel Wong according to my design and under my supervision

TABLE OF CONTENTS

SUPERVISORY COMMITTEE ii
ABSTRACT iii
PREFACE iv
TABLE OF CONTENTS v
LIST OF TABLES vii
LIST OF FIGURES viii
LIST OF ABBREVIATIONS ix
ACKNOWLEDGMENTS xii
DEDICATION xiii
CHAPTER 1: INTRODUCTION 1
1.1 Stem Cells 1
1.2 Skeletal myogenesis 2
1.3 RNAPII transcript processing 5
1.4 miRNA biogenesis 6
1.5 Replication-dependent histones (RDHs) 8
1.5.1 RDH pre-mRNA 3' processing 9
1.6 Nonsense-mediated decay (NMD) 11
1.7 Overview of Arsenite resistance protein 2 (ARS2) 15
1.7.1 ARS2 conservation and structure 15
1.7.2 ARS2 subcellular localization 17
1.7.3 ARS2 is essential for early mammalian development 17
1.7.4 ARS2 plays an important role in stem and progenitor cell function 17
1.7.5 Plant ARS2 is involved in RNA metabolism 18
1.7.6 ARS2 is required for the processing of RNAPII transcripts 18
1.8 Overview of FLASH 22
1.9 Hypothesis and objectives 23
CHAPTER 2: MATERIALS AND METHODS 25
2.1 Cell culture 25
2.1.1 Cell line passaging 25
2.2 Bacterial strains 25
2.3 Plasmids 25
2.4 Recombinant DNA techniques 28
2.5 DNA sequencing 28
2.6 Transient transfection 28
2.7 Immunocytochemistry of C2C12 myoblasts and myotubes 28
2.8 Differentiation assay 30
2.9 SDS-PAGE and western blotting 30
2.10 Immunoprecipitations 31
2.11 Replication-dependent histone (RDH) reporter assay 31
2.12 Detection and cloning of *Ars2 -X5* 32
2.13 Nonsense-mediated decay luciferase assay 33
2.14 Statistics 33

CHAPTER 3: RESULTS 34
3.1 Localization of ARS2 in proliferating myoblasts and post-mitotic myotubes 34
3.2 Localization of FLASH in proliferating C2C12 myoblasts 39
3.3 ARS2 is required at specific levels for C2C12 myogenic differentiation. 43
3.4 FLASH knockdown by RNAi is sufficient to disrupt FLASH function in N2a cells 49
3.5 C2C12 cells express a truncated C-terminal *Ars2* transcript 53
3.6 ARS2 stimulates a reduction in luciferase activity when tethered to the 3' UTR of a reporter mRNA 57
CHAPTER 4: DISCUSSION AND FUTURE DIRECTIONS 62
4.1 Discussion 62
4.2 Future Directions 67
4.2.1 ARS2 and myogenesis in vivo 67
4.2.2 ARS2 and miRNA expression in myoblasts 68
4.2.3 Further validation of ARS2-X5 69
4.2.4 Future NMD experiments 70
REFERENCES 72
APPENDIX A : ADDITIONAL FIGURES 83
APPENDIX B : KNOCKDOWN DIFFERENTIATION RAW DATA 88
APPENDIX C : OVEREXPRESSION DIFFERENTIATION RAW DATA 94
APPENDIX D : ARS2-X5 SEQUENCE DATA 97

LIST OF TABLES

Table 2-1 shRNA sequences. 26
Table 2-2 FARB and control peptide amino acid sequences. 26
Table 2-3 Primer sequences used to generate λN-RNPS1. 27
Table 2-4 Primer sequences used to detect and clone *Ars2-X5* 33

LIST OF FIGURES

Figure 1-1 Hierarchy of stem cells 1
Figure 1-2 Mammalian skeletal myogenesis 3
Figure 1-3 Examples of miRNA regulation during myogenesis 5
Figure 1-4 The biogenesis of miRNAs 10
Figure 1-6 Model of nonsense-mediated decay (NMD) 14
Figure 1-7 Metazoan ARS2 is a highly conserved protein with multiple protein motifs 16
Figure 1-8 CBCA functions 19
Figure 3-1 ARS2 is expressed in the cytoplasm and nucleus of C2C12 myoblasts 35
Figure 3-2 ARS2 is expressed in differentiating, post-mitotic myotubes 35
Figure 3-3 Exogenously expressed full-length ARS2 is expressed as a 100 kDa protein 38
Figure 3-4 Exogenous full-length ARS2 is expressed predominantly in the nucleus 39
Figure 3-5 Exogenous ARS2 and FLASH co-localize in C2C12 myoblasts 41
Figure 3-6 Endogenous ARS2 and endogenous FLASH do not appear to colocalize in C2C12 myoblasts 42
Figure 3-7 FLASH co-immunoprecipitates with 3XFLAG-ARS2 in C2C12 myoblasts 43
Figure 3-8 Knockdown of ARS2 expression results in decreased myoblast differentiation 45
Figure 3-9 FARB specifically co-immunoprecipitates with 3XFLAG-ARS2 47
Figure 3-10 Overexpression of ARS2 results in decreased myoblast differentiation but expression of FARB has no effect 48
Figure 3-11 Replication-dependent histone (RDH) mRNA processing assay 50
Figure 3-12 ARS2 is required at specific levels for RDH pre-mRNA processing 52
Figure 3-13 ARS2 isoform detection 55
Figure 3-14 ARS2-X5 57
Figure 3-15 RNPS1 co-immunoprecipitates with 3XFLAG-ARS2 59
Figure 3-16 NMD tethering assay 60
Figure 3-17 ARS2 triggers a decrease in luciferase activity when tethered to the 3' UTR of reporter mRNA 61
Figure A-4-1 ARS2 distribution in E3.5 embryos after 24 and 48 hours of culture 84
Figure A-4-2 ARS2 knockdown in C2C12 cells 85
Figure A-4-3 ARS2 is expressed in post-mitotic C2C12 myotubes 86
Figure 4-4 Validation of knockdowns 87

LIST OF ABBREVIATIONS

ARS2	arsenite resistance protein 2
ATCC	American Tissue Culture Collection
BSA	bovine serum albumin
BTZ	barentsz
CAP3	contig assembly program 3rd generation
CBC	cap binding complex
CBP20	cap binding protein 20 kDa
CBP80	cap binding protein 80 kDa
Cdk2	cyclin-dependent kinase 2
cDNA	complementary DNA
CMV	cytomegalovirus
CPSF-100	cleavage and polyadenylation specificity factor 100 kDa
CPSF-73	cleavage and polyadenylation specificity factor 73 kDa
CstF-64	cleavage stimulation factor 64 kDa
Cstf-77	cleavage stimulation factor 77 kDa
DAPI	4',6-diamidino-2-phenylindole
DCL1	DICER-LIKE 1
DM	differentiation medium
DMEM	Dulbecco's modified eagle media
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
DUF	domain of unknown function
EGFP	enhanced green fluorescent protein
eIF4AIII	eukaryotic translation initiation factor 4A-III
EJC	exon junction complex
eRF1	eukaryotic release factor 1
eRF3	eukaryotic release factor 3
ESC	Embryonic stem cell
FARB	FLASH-ARS2 binding
FBS	fetal bovine serum
FLASH	FLICE-associated huge protein
GFP	green fluorescent protein
GM	growth medium
HCC	histone pre-mRNA cleavage complex
HDAC4	histone deacetylase 4
HDE	histone downstream element
HEK293	human embryonic kidney 293 cells
HLB	histone locus body
HRP	horseradish peroxidase

HYL1	HYPONASTIC LEAVES 1
ICM	inner cell mass
IRES	internal ribosome entry site
kDa	kiloDalton
LB	Luria-Bertani
LSM10	U7 snRNP-specific Sm-like protein LSM10
LSM11	U7 snRNP-specific Sm-like protein LSM11
m7G	7-methylguanosine
MAGOH	mago-nashi
MEF	mouse embryonic fibroblast
MEF2	myocyte enhancer-2
MHC	myosin heavy chain
miRNA	microRNA
MRF	myogenic regulatory factor
MRF4	muscle-specific regulatory factor 4
mRNA	messenger RNA
MYF5	myogenic factor 5
MYF6	myogenic factor 6
MYOD	myoblast determination protein
N2a	Neuro-2a
NELF	negative elongation factor
NES	nuclear export signal
NEXT	nuclear exosome targeting
NLS	nuclear localization signal
NMD	nonsense-mediated decay
NPAT	nuclear protein, ataxia-telangiectasia locus
ORF	open reading frame
PAX3	paired box 3
PAX7	paired box 7
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
pre-miRNA	precursor miRNA
pre-mRNA	precursor mRNA
pri-miRNA	primary miRNA
PTC	premature termination codon
PVDF	polyvinylidene fluoride
RBD	RNA binding domain
RDH	replication-dependent histone
RevChrg	reverse charge
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference

RNAPII	RNA polymerase II
RNPS1	RNA binding protein S1, serine-rich domain
RRM	RNA recognition motif
SCR	scrambled
SDS	sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
shRNA	short hairpin RNA
siRNA	small interfering RNA
SLBP	stem-loop binding protein
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SRRT	serrate, RNA effector molecule
SURF	SMG-1-Upf1-release factor
TBS-T	Tris-buffered saline-Tween 20
UPF	Up-frame shift
UTR	untranslated region
ZFP100	zinc finger protein, 100 kDa

ACKNOWLEDGMENTS

Foremost, I would like to thank my supervisor Dr. Perry Howard for all of his support and guidance. I would also like to thank the members of my committee Dr. Robert Burke and Dr. Leigh Anne Swayne for their valuable input and insights throughout my program. I especially want to thank Connor O'Sullivan for his very patient instruction when I first started and later for being the best co-worker and brainstorming buddy. I would like to thank all the members of the Howard lab past and present: Kevin Yongblah, Marcus Pienaar, Rachel Wong, Jake Gambling, Chris Yeker and Patrick Duke. I also want to thank Dr. Spencer Alford without whom I never would have learned the value of a perfectly aligned figure. To my fellow graduate students, thank you for your friendship and camaraderie. Graduate school is hard enough without the resentment and competition that is prevalent in many program. I am so glad that our department has a positive atmosphere where we choose to support one another instead of tear each other down.

I would like to express my gratitude to all the teaching staff that I have worked with during my time in the department: Barb Currie, Allison Maffey, Melissa Doyle and Janice Keliher. Thank you for making the teaching labs a fun and lively place to teach and learn. I would also like to take this opportunity to thank the BCMB administrative staff, especially Melinda Powell without whom I am sure there would be countless empty forms and missing signatures. Thank you to Marj Wilder for being a great friend and for always being there for a quick chat when I needed a break from science.

Finally, I want to thank my friends and family for their amazing support and love. The completion of my thesis would not have been possible without your encouragement and occasional tough love.

DEDICATION

For my family, who kept me going.

CHAPTER 1: INTRODUCTION

1.1 Stem Cells

The process of mammalian development from a single-cell zygote to a multicellular organism with highly heterogeneous tissues and organs requires that cells become specialized for specific functions. This is made possible by the existence of stem cells. Stem cells are developmentally naïve cells that can differentiate into a variety of cell lineages while retaining the ability to self-renew through mitotic division (Fig. 1-1). In mammals there are two general categories of stem cells: pluripotent embryonic stem cells (ESCs) and multipotent adult stem cells. ESCs are derived from the inner cell mass of blastocyst-stage embryos and give rise to every cell type in an organism including all adult stem cells, committed progenitor cells and terminally differentiated cells. Adult stem cells are found in various tissues throughout the body, they can differentiate into a limited number of cell types and function primarily to regenerate organs and tissues that are injured or damaged.

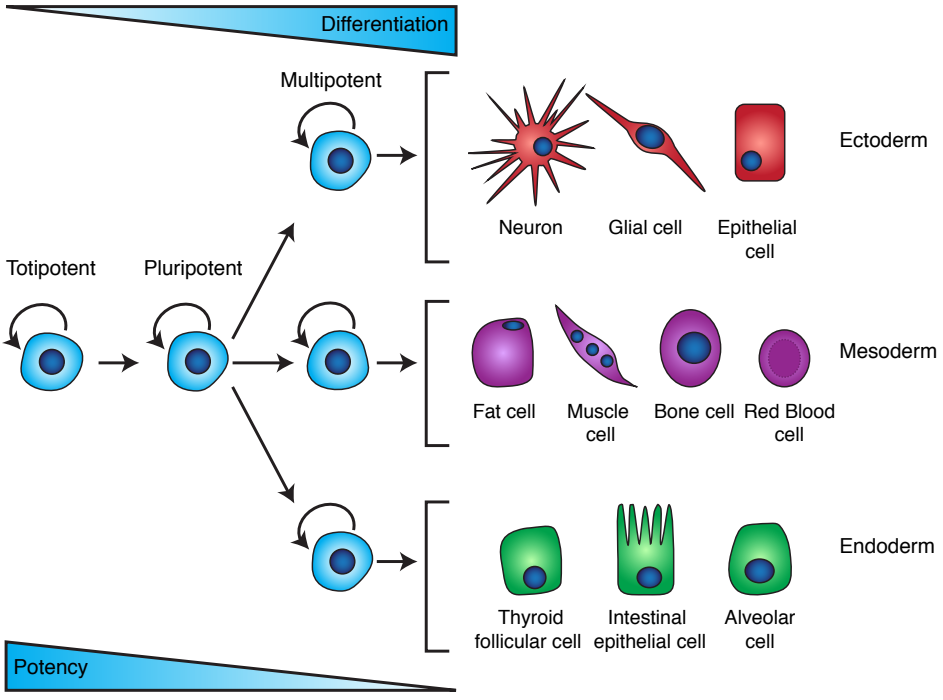

Figure 1-1 Hierarchy of stem cells. A schematic diagram showing the hierarchy of stem cells from totipotent embryonic stem cells to multipotent adult stem cells. Rounded arrows indicate cells that are capable of self-renewal.

The process by which stem and progenitor cells differentiate is not fully understood but it is clear that it involves dynamic interplays between extrinsic signalling, transcriptional, post-transcriptional, translational and epigenetic mechanisms. Further elucidation of cellular differentiation is vital to understanding not only normal development, but also degenerative pathologies and injuries. Greater knowledge about how cells become lineage restricted and progress through differentiation could lead to the development of new regenerative therapies.

In 2008 the Howard and Koop labs at the University of Victoria, discovered that the gene *ARS2* (formally known as *SRRT* in humans) is essential for early mammalian embryogenesis[1]. This work suggested that ARS2 plays an important role in embryonic stem cell maintenance and differentiation. It has recently been determined that ARS2 is a member of the nuclear cap-binding complex (CBC), that binds to 5'-capped RNA, and stimulates multiple RNA processing pathways [2-5]. However, it remains unclear precisely how ARS2 contributes to stem and progenitor cell maintenance and differentiation. The main goal of the Howard lab is to discern how ARS2 deficiencies affect stem and progenitor cell behaviour in muscle tissue.

1.2 Skeletal myogenesis

The formation of skeletal muscle, known as skeletal myogenesis, is a multistep process that involves cell specification, proliferation, differentiation and fusion. During embryonic development, multipotent cells in the mesoderm give rise to committed muscle progenitor cells called myoblasts [6]. Myoblasts undergo extensive proliferation, exit the cell cycle, and differentiate into mononucleated myocytes, which then align and fuse to form nascent myotubes [6]. Nascent myotubes subsequently fuse with additional myocytes and nascent myotubes to form large, multinucleated myotubes that develop into the mature muscle fibres that make up skeletal muscle tissue [6] (Fig. 1-2). In adults, skeletal muscle possesses an extraordinary capacity for regeneration in response to muscle damage. This is mediated by muscle stem cells called satellite cells that are found beneath the basal lamina, adjacent to the plasma membrane of muscle fibres [6-8]. Satellite cells exist in a mitotically quiescent state until they are activated upon muscle injury, at which point they re-enter the cell cycle and give rise to proliferating myoblasts [6,9-13]. Satellite cell derived myoblasts then repair muscle by differentiating and fusing

with existing muscle fibres or with each other in order to form new muscle fibres [6]. Studies reviewed by Charge and Rudnicki [6] have shown that the transcriptional gene regulatory networks that control embryonic muscle development are reiterated during muscle regeneration by satellite cells.

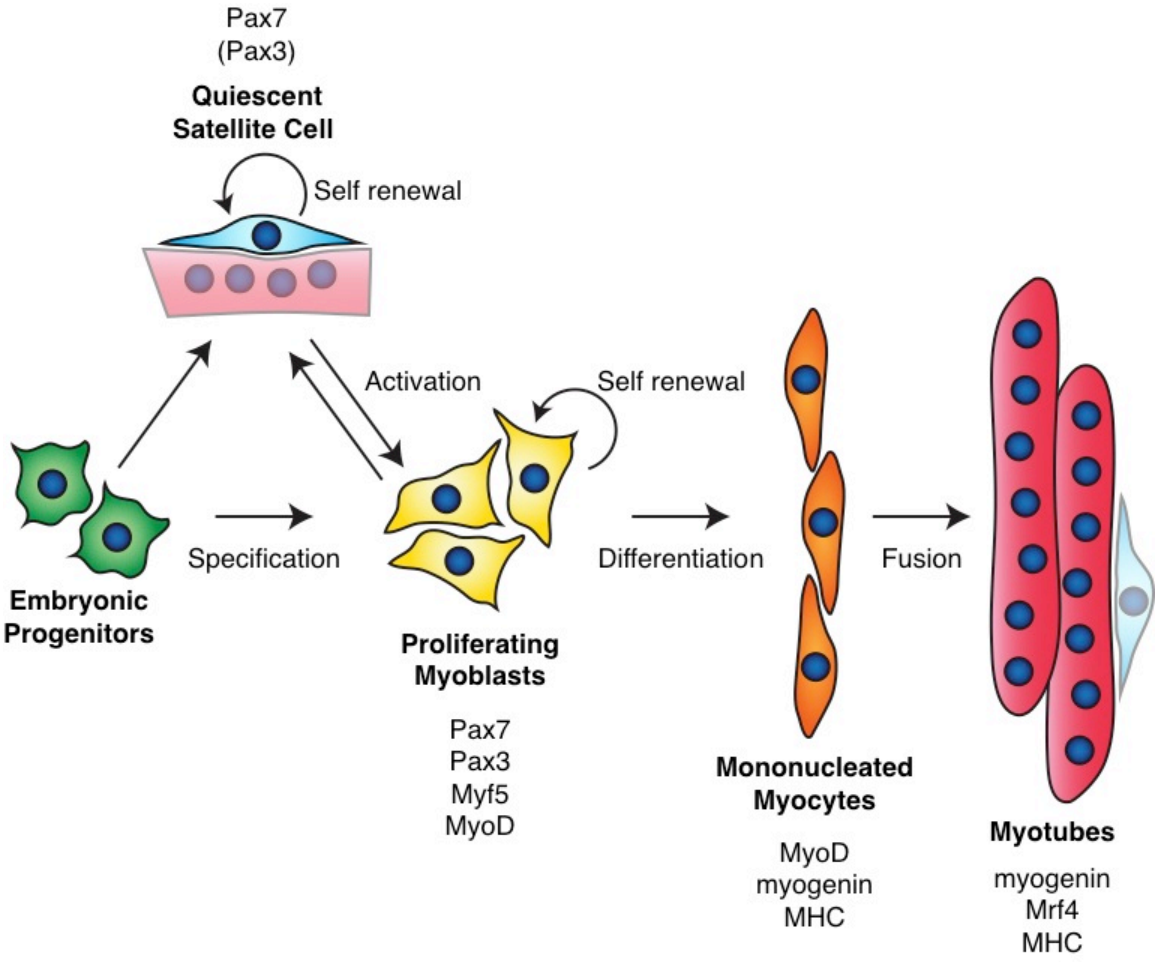

Figure 1-2 Mammalian skeletal myogenesis. Embryonic progenitor cells give rise to committed myoblasts and satellite cells. Satellite cells remain quiescent until they are activated by signals from muscle injury at which point they start to proliferate and differentiate into myoblasts. After several rounds of cell division myoblasts exit the cell cycle and give rise to myocytes, which fuse with each other to form multinucleated myotubes. Markers expressed in each cell type are summarized.

Skeletal muscle development is regulated by a number of transcription factors that activate and suppress the gene expression programs required for muscle progenitor cells to become terminally differentiated myotubes. The principal regulators of myogenesis are

four muscle-specific transcription factors called the myogenic regulatory factors (MRFs): myogenic factor 5 (MYF5), myoblast determination protein (MYOD), muscle-specific regulatory factor 4 (MRF4; also known as MYF6), and myogenin [6]. The MRFs were originally identified by their ability to transform non-muscle cell types, such as fibroblasts, into fusion-competent cells that can differentiate into mature myotubes [14-19]. The current consensus in the field is that MYF5, MYOD, and MRF4 act in specification of progenitor cells to the myogenic lineage, while myogenin is required for the terminal differentiation of committed myoblasts [20]. The transcriptional network regulating myogenesis also includes the paired-box transcription factors PAX3 and PAX7, which promote myoblast specification and are downregulated during differentiation [20-23] and the myocyte enhancer-2 (MEF2) family of MADS-box transcription factors, which act in coordination with MRFs to activate skeletal muscle gene expression [20,24,25].

The ability of muscle progenitor cells to either remain cycling or initiate differentiation requires regulation of the balance between pro-maintenance factors that inhibit differentiation and pro-differentiation factors that promote differentiation [26,27]. There is growing evidence that microRNAs (miRNAs) are key players in the mechanisms that maintain this balance. miRNAs are a class of short non-coding RNA molecules that suppress the expression of target genes at a post-transcriptional level through inhibition of protein translation or degradation of the mRNA transcript. Early studies revealed that many miRNAs are expressed specifically in skeletal muscle tissue, suggesting that miRNAs play an important role in modulating gene expression during myogenesis [28-30]. Global loss of miRNA in muscle, via a conditional knockout of *DICER*, results in decreased skeletal muscle mass, increased apoptosis of myogenic cells, abnormal muscle fiber morphology and premature death [31]. Several specific roles for miRNAs in myogenesis have been reported (Fig. 1-3). For example, myogenic miRNAs miR-1, miR-27, miR-206 and miR-486 promote myoblast differentiation by downregulating the pro-maintenance factors PAX3 and PAX7, thereby suppressing the progenitor state and allowing myoblast differentiation to occur [32-36]. In addition to relieving inhibitory effects on terminal differentiation, miRNAs also play a role in establishing the muscle-specific gene expression profile. In muscle progenitor cells, histone deacetylase 4

(HDAC4) binds to MEF2 resulting in repression of MEF2-dependent muscle gene expression [37]. Upon myoblast differentiation, the expression of HDAC4-targeting miRNAs miR-1, miR-206 and miR-29 increases, and thereby, HDAC4 expression decreases allowing MEF2 to activate transcription of muscle-specific genes [26,38-43]. The regulation of myogenesis by miRNAs is one example of a multitude of post/co-transcriptional gene regulatory mechanisms that rely on RNA processing.

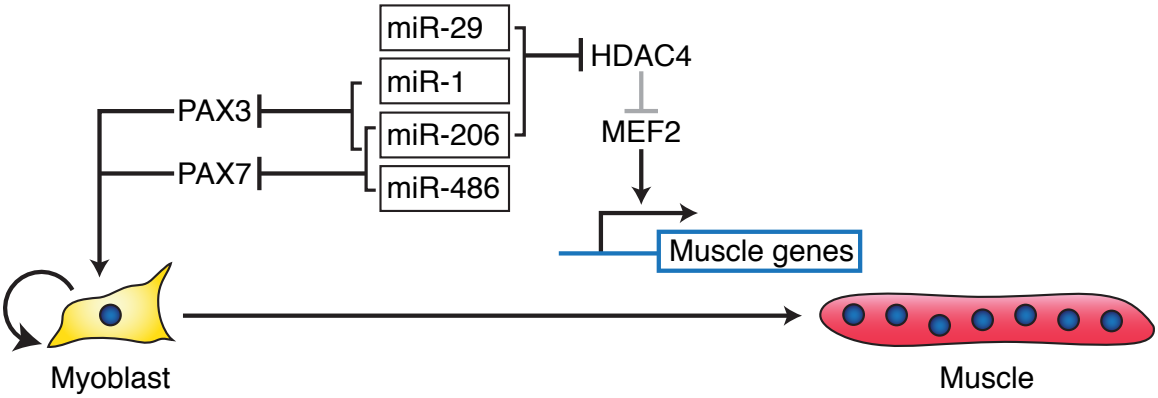

Figure 1-3 Examples of miRNA regulation during myogenesis. Expression of myogenic miRNAs increases during myogenesis. These miRNAs promote myoblast differentiation by downregulating genes that promote the progenitor state and genes that inhibit myogenic differentiation. miR-1, miR-486 and miR-206 repress pro-myoblast factors PAX3 and PAX7, thereby promoting myoblast differentiation. miR-29, miR-1 and miR-486 repress HDAC4, relieving the inhibition on MEF2 and allowing MEF2 to activate transcription of muscle-specific genes. Adapted from **[44]**.

1.3 RNAPII transcript processing

Traditionally, the majority of gene expression regulation was believed to occur at the level of transcription initiation, however, over the last two decades it has become increasingly clear that co/post-transcriptional gene regulation is far more important than originally thought. At the heart of co/post-transcriptional regulatory pathways is RNA processing. Eukaryotic RNAPII produces a diversity of primary RNA transcripts that undergo processing by specific multi-protein complexes in order to become mature active species. RNAPII transcripts include at least 4 different classes of RNA: mRNA, miRNA, small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA). The first step in the processing of all RNAPII transcripts is the exclusive, co-transcriptional addition of a 7-

methylguanosine (m7G) cap to the 5' end of nascent transcripts [45,46]. The m7G cap is then bound by a multi-protein complex called the nuclear cap-binding complex (CBC) that includes the proteins CBP20 and CBP80 [47]. CBP20 binds directly to the m7G cap while CPB80 induces a conformational change in CPB20 that ensures high-affinity cap binding [48]. The CBC functions to increase mRNA stability and provide a platform for interaction with downstream RNA processing machinery [49]. The CBC has been implicated in several RNA processing events including pre-mRNA splicing [47], mRNA 3' end formation [50], RNA nuclear export [51-53], miRNA processing [4,54,55] and nonsense-mediated decay [56,57]. The process by which RNAPII transcripts are recognized for differential processing is not fully understood. The fact that the CBC is involved in such a wide variety of RNA processing pathways suggests that it may play a central role in discriminating transcripts and recruiting the appropriate processing machinery. ARS2 was recently identified as a member of the CBC and implicated in the processing of a number of different RNA classes, indicating that ARS2 may be a master regulator of RNAPII transcript processing [2-5,55]. My thesis focuses on ARS2's role in 3 RNA processing pathways: miRNA biogenesis, replication-dependent histone (RDH) pre-mRNA 3' end processing and nonsense-mediated decay (NMD), which I will briefly describe in the following sections.

1.4 miRNA biogenesis

miRNAs are a class of short (21-22 nucleotide [nt]), non-coding RNA molecules that critically regulate gene expression at a post-transcriptional level. They can be transcribed individually, as a cluster of miRNAs or within an intron of a protein-coding gene. The majority of miRNAs are transcribed in the nucleus by RNAPII as part of larger, stem-loop containing transcripts known as primary miRNAs (pri-miRNA). These transcripts usually consist of a ~33 nt, double-stranded stem flanked by a loop on one end and single stranded segments on the other [58].

Following transcription, nascent pri-miRNAs are processed by two sequential cleavage reactions [58] (Fig. 1-4). The first processing step excises the double stranded RNA stem-loop from the rest of the transcript to create a pre-miRNA. This step is catalyzed in the nucleus by the Microprocessor, a complex that contains RNase III-type enzyme DROSHA and RNA-binding protein DGCR8 (also known as PASHA) [59,60].

miRNAs that are expressed as introns in protein coding genes are processed independently of the Microprocessor [61,62]. Splicing of the host gene pre-mRNA releases the miRNA containing intron, which mimics a Microprocessor-produced pre-miRNA and goes on to enter the conventional miRNA processing pathway. Following release from larger transcripts, pre-miRNAs are then transported from the nucleus to the cytoplasm by Exportin-5 [63]. Once in the cytoplasm a member of the RNase III family known as DICER catalyzes the second processing step [64-66]. DICER binds target double-stranded RNA (dsRNA) via its PAZ domain. Once bound the pre-miRNA substrate lies across DICER's surface until it reaches the active site located between DICER's two RNase III domains. DICER cleavage then excises the terminal loop of the pre-miRNA to produce 21-22 nt dsRNA duplex. The distance from the dsRNA-binding site in the PAZ domain to the active site is equivalent to ~21-22 nt and this dictates the length of the dsRNA duplex [67]

The existence of the RNA duplex in the cytoplasm is fleeting. Shortly after cleavage by DICER, the duplex is unwound and one strand is loaded onto an argonaute/GW182-containing RNA-induced silencing complex (RISC) [58]. Determination of which strand is selected for loading is based on thermodynamic stability of the duplex ends. The strand whose 5' end exists at the duplex terminus with weaker base-pairing is preferentially retained and the other strand is degraded [68]. The single-stranded mature miRNA then acts as a guide and specifically directs the silencing complex to target mRNA [58].

RISC downregulates target genes by two different mechanisms: 1) target mRNA degradation or 2) translational repression. The primary mechanism determinant is the degree of complementarity between the miRNA and the target transcript. In animals, most miRNA target sequences exist as multiple copies in the 3' untranslated region (UTR) of target mRNAs. Nucleotides 2-8 of the miRNA bind to the target sequence with textbook Watson-Crick base-pairing and are collectively referred to as the seed region. Outside the seed region there are often mismatches and loops as a result of imperfect complementarity. If there is sufficient complementarity between the miRNA and the target mRNA, the mRNA will be cleaved by RISC and if there is not sufficient complementarity, translation of the target mRNA will be blocked. As a general rule, a

single miRNA can have multiple target genes and a target gene can be repressed by multiple miRNAs. This functional redundancy can create complex regulatory networks [58].

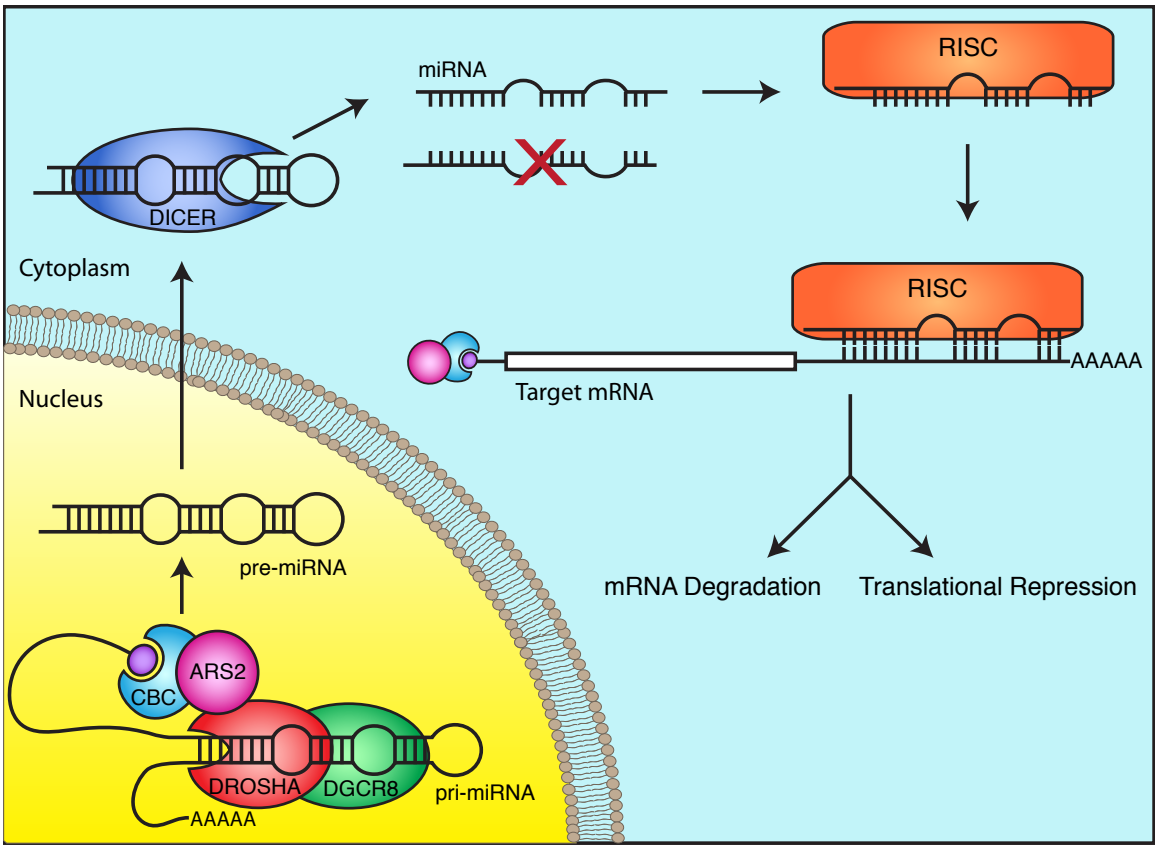

Figure 1-4 The biogenesis of miRNAs. The majority of miRNAs are transcribed by RNAPII as part of stem-loop containing pri-miRNA transcripts. A pri-miRNA is processed into pre-miRNA in the nucleus by the Microprocessor, a complex made up of RNase III-type enzyme DROSHA and double-stranded RNA-binding protein DGCR8. ARS2 and CBC contribute to this step of miRNA biogenesis by increasing the stability of pri-miRNA transcripts and facilitating their delivery to the Microprocessor. The pre-miRNA is then transported from the nucleus to the cytoplasm by Exportin-5, where it is processed by the RNase III enzyme DICER into a 21-22 nt double-stranded duplex. The RNA duplex is unwound and the 'passenger' strand is discarded while the 'guide' strand (mature miRNA) is loaded onto an RNA-induced silencing complex (RISC). The mature miRNA guides RISC to the 3' UTR of target mRNAs. Expression of target mRNAs is then suppressed by degradation of the mRNA or repression of protein translation.

1.5 Replication-dependent histones (RDHs)

In eukaryotic cells, chromosomal DNA is compacted into a dynamic nucleoprotein structure known as chromatin. The bulk of the protein found in chromatin

consists of the nucleosome core histones H2A, H2B, H3 and H4 and the linker histone H1. These histones are known as replication-dependent histones (RDHs) and are encoded by a family of cell cycle regulated genes that are clustered at two major loci in mammalian genomes [69]. The expression of RDHs is highly cell cycle regulated with RDH protein accumulating rapidly during S-phase, which ensures proper packaging of newly synthesized DNA prior to cell division. The accumulation of RDH protein is the result of an increase in RDH mRNA concentration. RDH mRNA levels are regulated through modulation of histone gene transcription, pre-mRNA processing efficiency and mRNA stability [70]. For example, transcriptional activation of RDH genes is coupled to the cell cycle by a cyclin E/Cdk2 substrate called NPAT; a nuclear protein that localizes to histone gene clusters and activates RDH transcription [71]. Phosphorylation of NPAT by cyclin E/Cdk2 during S-phase enhances NPAT-mediated RDH transcriptional activation [71]. RDH mRNAs are expressed at high levels at the onset of S-phase and persist at high levels until they are destroyed at the conclusion of DNA synthesis [70].

1.5.1 RDH pre-mRNA 3' processing

RDH mRNAs are unique in that they are the only metazoan mRNAs that do not have a poly(A) tail and instead end in a conserved 3' stem-loop. They are intronless and undergo a single 3' endonucleolytic cleavage event to become mature mRNAs. In contrast, replication-independent or variant histone pre-mRNAs have introns and undergo splicing and 3' cleavage/polyadenylation, as is typical for all other mRNA transcripts. The molecular machinery that carries out RDH pre-mRNA 3' end processing includes proteins that are unique to this processing event as well as proteins that overlap with the polyadenylation machinery [70]. RDH transcription and pre-mRNA processing take place at histone locus bodies (HLBs), nuclear bodies associated with RDH gene clusters that contain factors required for RDH pre-mRNA processing [72].

There are two highly conserved *cis*-acting elements in the 3' UTR of RDH pre-mRNAs (Fig. 1-5A). The first is a stem-loop containing sequence located 25-60 nt downstream of the open reading frame and the second is a purine-rich sequence known as the histone downstream element (HDE) found an additional 15-20 nt downstream [70]. During processing the stem-loop is bound by stem-loop binding protein (SLBP) [73] and the HDE is recognized by the U7 small nuclear ribonucleoprotein (snRNP) [74] (Fig. 1-

5B). SLBP and U7 snRNP position the pre-mRNA for cleavage and recruit a cleavage complex that contains at least CPSF-73, CPSF-100 and symplekin [75,76].

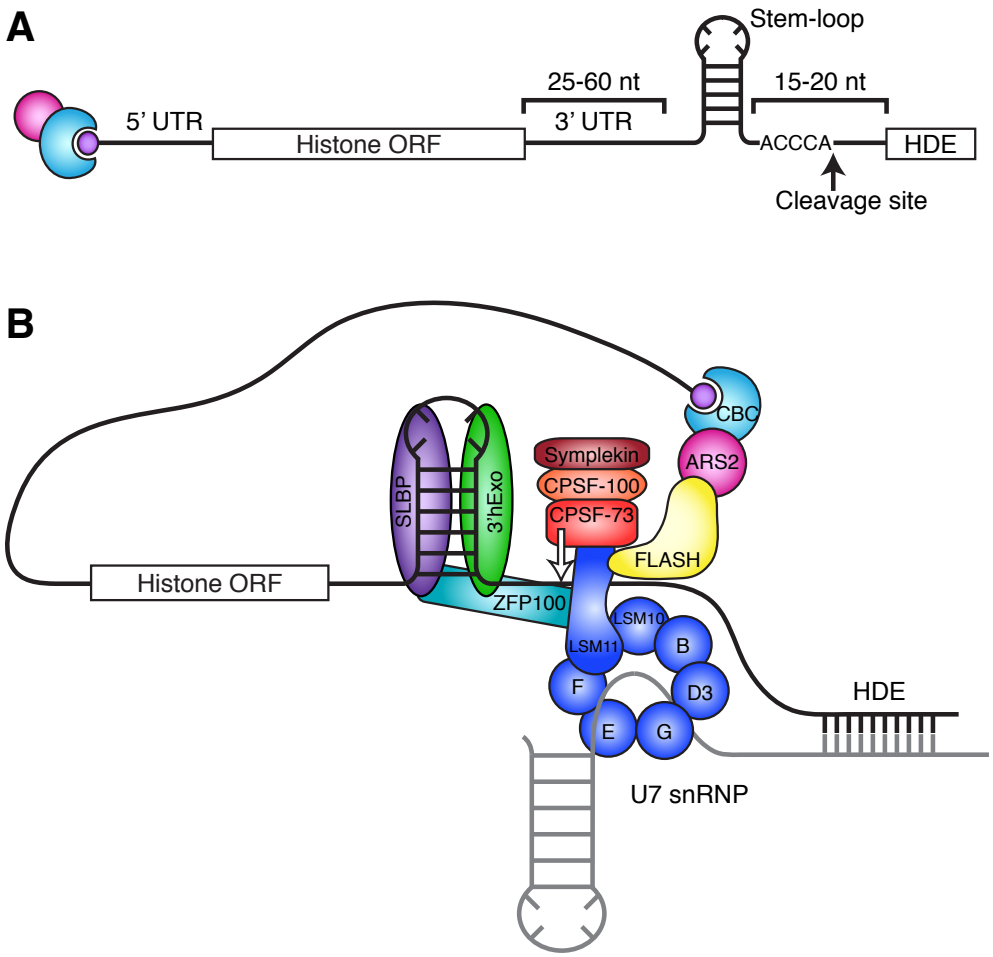

Figure 1-5 Structure and processing of RDH pre-mRNA. (A) The structure of a metazoan RDH pre-mRNA showing the conserved stem-loop and histone downstream element (HDE). **(B)** Proteins involved in replication-dependent histone (RDH) pre-mRNA 3' processing. RDH pre-mRNAs (black line) contain a conserved stem-loop sequence that binds stem-loop binding protein (SLBP) followed by the HDE, which base-pairs with U7 snRNA (grey line). A cleavage complex containing at least CPSF73, CPSF100 and symplekin is recruited to cleave the pre-mRNA. An arrow indicates the site where CPS73-mediated cleavage occurs. The U7 snRNA is a component of the U7 snRNP, which contains a heptameric ring of five Sm proteins and two U7 snRNP-specific Sm-like protiens: LSM11 and LSM10. LSM11 interacts with the large nuclear protein FLASH, which also associates with ARS2 and the CBC. LSM11 also contacts a 100 kDa zinc finger protein (ZFP100), which bridges the U7 snRNP to the SLBP-stem-loop complex.

The stem-loop of RDH pre-mRNAs has a 6 base pair (bp) stem with a rigorously conserved G-C base pair at the second position and a 4 nt loop that is typically

pyrimidine-rich [77]. A recent crystal structure has revealed that SLBP binds to the 5' flanking sequence, the 5' arm of the stem, and the loop of the stem-loop via an RNA-binding domain (RBD) [78]. The U7 snRNP is recruited through base-pairing between the 5' end of U7 snRNA and the HDE [74]. The U7 snRNP is similar to spliceosomal snRNPs in that it has a stem-loop containing snRNA and a heptameric ring of Sm proteins. Five members of the heptameric ring in U7 snRNP overlap with spliceosomal snRNPs (Sm proteins B, D3, E, F and G) while the remaining two are replaced by Sm-like proteins LSM10 and LSM11 [79,80]. At 40 kDa, LSM11 is more than twice the size of a prototypical Sm protein. The C-terminal region of LSM11 contains two Sm motifs and is sufficient for assembly on the U7 snRNA, while the extended N-terminal region is necessary for RDH pre-mRNA 3' end processing [80]. The N-terminal segment interacts with the zinc-finger protein ZFP100, which in turn interacts with SLBP, forming a bridge between SLBP and U7 snRNP [80,81]. In addition, the N-terminal region of LSM11 interacts with FLICE-associated huge protein (FLASH) and together these two proteins form a platform that recruits the endoribonuclease CPSF-73 and several other factors that are shared with the canonical cleavage/polyadenylation machinery [82]. As in 3' end processing of polyadenylated mRNAs, CPSF-73 catalyzes the 3' cleavage of RDH pre-mRNAs [83,84]. Following cleavage by CPSF-73, mature RDH mRNAs are exported from the nucleus for translation.

1.6 Nonsense-mediated decay (NMD)

Nonsense-mediated decay (NMD) is a translation-dependent surveillance mechanism that recognizes and promotes the degradation of transcripts containing premature termination codons (PTCs) during the pioneer round of translation [85]. Initially, NMD was identified as a quality control mechanism that allows the cell to dispose of defective transcripts with PTCs before potentially harmful truncated proteins can accumulate [85]. Gene expression profiling in yeast [86,87], *Drosophila* [88] and human cells [89-91] has since revealed that NMD also regulates the level of thousands of physiologic transcripts, indicating that NMD has important regulatory roles beyond mRNA quality control. There is evidence for NMD-mediated gene regulation in processes such as oxidative stress response [92,93], nutrient homeostasis [89,94], axon guidance [95] and telomere maintenance [96]. In addition, variation in NMD efficiency

has been reported in different strains of *Sacchoaromyces cerevisiae* [97] and HeLa cells [98].

The primary facilitators of eukaryotic NMD are the Up-frame shift (UPF) proteins UPF1, UPF2 and UPF3 [99]. Loss of the UPF proteins results in increased abundance of PTC containing mRNAs [100-102]. Also, tethering of any of the human UPF proteins to the 3' UTR of a β-globin mRNA reporter triggers degradation of the reporter transcript by NMD [103]. Additional proteins that are required for NMD include the SMG factors that regulate a cycle of UPF1 phosphorylation and dephosphorylation, multiple components of the exon junction complex (EJC), and translation release factors eRF1 and eRF3 [99].

Early models of NMD proposed the existence of a protein mark unique to prematurely terminating mRNAs that labels them as targets and initiates NMD [104,105]. NMD typically targets mRNAs with a PTC followed by a downstream exon-exon junction [99]. The fact that the position of a PTC relative to a downstream exon-exon junction determines whether or not a transcript elicits NMD was the first indication that NMD is splicing dependent and suggested that the multi-protein EJC might serve as the protein mark in mammalian cells [106,107]. EJCs are deposited onto mRNA 20-24 nt upstream of exon-exon junctions during pre-mRNA splicing and stay bound until they are displaced by the first translating ribosome [106,108]. The EJC contains an invariable protein core that consists of the Y14-MAGOH heterodimer, BTZ, and eIF4AIII plus a dynamic complement of peripherally associated proteins that is context-dependent [109,110].

Current models of NMD suggest that if a ribosome stops at a PTC and begins to undergo premature translation termination before a downstream EJC is displaced, a surveillance complex is able to recognize the downstream EJC and initiate NMD. This model is supported by the fact that mammalian NMD typically targets mRNAs with a PTC 50-55 nt upstream of an exon-exon junction [99]. Normal termination codons are usually found in the last exon or within 50-55 nt of any downstream introns and therefore generally do not elicit NMD. Data showing that UPF1 interacts with eRF3, a release factor involved in translation termination, supports the involvement of a surveillance complex at the site of premature translation termination [111]. It is now known that

during translation termination at a PTC, UPF1 and its regulating factors SMG1, SMG8 and SMG9 are recruited to the termination complex through interaction with eRF3 to form the SMG1-UPF1-release factor (SURF) complex [112,113] (Fig. 1-6A). In parallel with SURF formation, UPF2 is recruited to the EJC through interaction with UPF3, which is already present on the EJC following binding during mRNA splicing [107,114,115]. When a ribosome stops at a PTC, the SURF complex interacts with the downstream EJC core through interaction between UPF1 and UPF2-UPF3 [116] (Fig. 1.6B). This interaction triggers phosphorylation of UPF1 by SMG1 and allows UPF1 to interact with SMG6 and SMG5-bound SMG7 [117] (Fig. 1.6C). The mRNA target is then degraded by one of two pathways: (1) SMG6-mediated endonucleolytic cleavage followed by rapid decay of the cleavage products [118,119] or (2) SMG5-SMG7-mediated decapping and deadenylation followed by degradation of the body of the transcript [120,121].

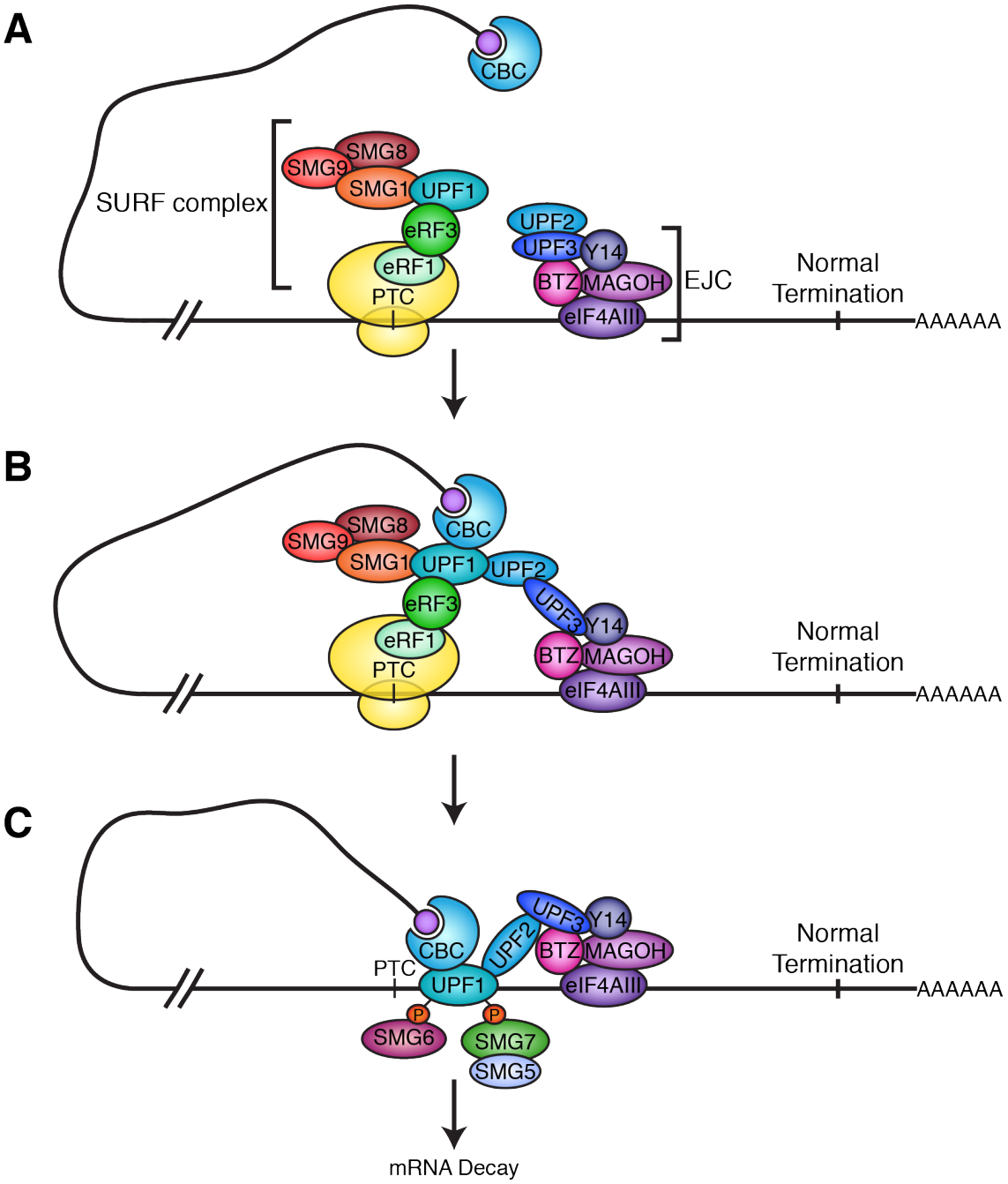

Figure 1-6 Model of nonsense-mediated decay (NMD). (A) The exon junction complex (EJC) is deposited onto mRNA 20-24 nt upstream of exon-exon junctions during pre-mRNA splicing. UPF3 binds to the EJC core during splicing and UPF2 joins following mRNA export to the cytoplasm. UPF1 and its regulators (SMG1, SMG8 and SMG9) are recruited to the translation termination complex through interaction with release factor eRF3, forming the SURF (SMG1-UPF1-release factor) complex. **(B)** When a ribosome stops at a premature termination codon (PTC) with a downstream EJC, UPF1 interacts with UPF2-UPF3. **(C)** Interaction between the SURF complex and the EJC triggers phosphorylation of UPF1 by SMG1. UPF1 phosphorylation allows it to interact SMG6 and SMG5-bound SMG7. The mRNA NMD target is degraded by: 1) SMG6-medieated

cleavage followed by rapid decay of cleavage products or 2) SMG5-SMG7-mediated decapping and/or deadenylation followed by degradation of the body of the transcript. Adapted from [99].

1.7 Overview of Arsenite resistance protein 2 (ARS2)

1.7.1 ARS2 conservation and structure

Arsenite resistance protein 2 (ARS2) is a highly conserved protein that has ≥98% amino acid identity within mammals[1]. The *Ars2* gene (formally *SRRT* in humans) has a single ortholog in all eukaryotic organisms with the exception of budding yeast [1]. Little is known about the structure of metazoan ARS2 but a crystal structure of the plant ortholog SERRATE has been reported. X-ray crystallography showed that the SERRATE core assumes a 'walking-man-like' topology with an N-terminal domain of unknown function (DUF3546) as the leading leg, a novel Mid domain as the body and the C-terminal zinc finger (ZnF) domain as the lagging leg [122]. Both the N- and C-terminal tails of SERRATE are predicted to be unstructured [122]. While the overall amino acid identity between *Arabidopsis* SERRATE and mouse ARS2 is moderately low (28%), bioinformatic analysis by Connor O'Sullivan in our lab revealed that the DUF3546, Mid and ZnF domains are conserved (O'Sullivan; in submission). Based on O'Sullivan's homology modeling, metazoan ARS2 contains 4 domains: an N-terminal DUF3546, a Mid domain, a C-terminal ZnF domain and an additional RNA recognition motif (RRM) -containing domain that is absent in plants. Metazoan ARS2 also contains a putative nuclear localization signal (NLS), a putative nuclear export signal (NES), an arginine-rich region and multiple PXXP proline-rich motifs[1] (Fig. 1-7A).

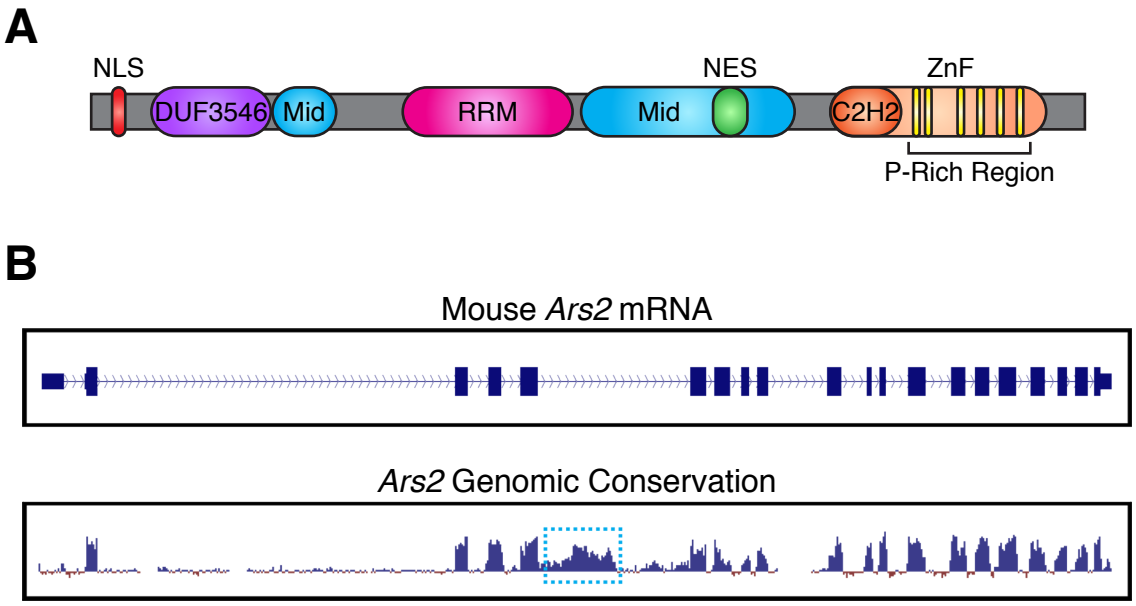

Figure 1-7 Metazoan ARS2 is a highly conserved protein with multiple protein motifs. (A) The protein motif structure of metazoan ARS2. DUF3546 – domain of unknown function 3546, RRM – RNA recognition motif, NLS – nuclear localization signal, NES – nuclear export signal, ZnF – zinc finger. Yellow lines indicate PXXP motifs. **(B)** Mouse *Ars2* mRNA and genomic conservation of the *Ars2* gene (formally *SRRT*) between mouse and 39 additional placental mammal species. The alignments were generated using mulitz and other tools in the UCSC/Penn State Bioinformatics comparative genomics alignment pipeline and evolutionary conservation was measured using *phyloP*. This data can be found at positions chr5:137,295,704 to 137,307,674 of the mouse December 2011 asssembly on the UCSC genome browser website http://genome.ucsc.edu/ [123]. Dotted box indicates region in intron 5 with high conservation.

Wilson *et al.* found that *Ars2* mRNA is expressed in all major mouse and human tissues with the highest levels found in mouse testes, heart, and liver and human heart and skeletal muscle[1]. Northern blots showed that the predominant *Ars2* transcript size is ~3.0 kb, which correlates with the full-length mouse cDNA sequence (GenBank: BC066831.1). With a 3' probe, additional weak bands were visible at ~2 kb and 1 kb. It is possible that these bands represent *Ars2* isoforms produced from internal promoters. The existence of alternative *Ars2* isoforms is further supported by the fact that shorter C-terminal *Ars2* transcripts have been identified in mouse, human, Chinese hamster and *Drosophila*[1]. An unexpectedly high degree of genomic conservation in intron 5 of the *Ars2* gene suggests that it may contain functional elements such as internal promoters (Fig. 1-7B).

1.7.2 ARS2 subcellular localization

Our lab found that 3XFLAG-ARS2 expressed in BOSC kidney cells is localized predominantly in the nucleus with some expression in the cytoplasm [1]. This is consistent with a study from Craig Thompson's group that reported a similar localization pattern for V5-tagged ARS2 in 3T3 MEFs [4]. This report also found that ARS2 shuttles between the cytoplasm and the nucleus in heterokaryon shuttling experiments [4]. Although these data suggest that ARS2 is predominantly a nuclear protein, the subcellular localization of endogenous ARS2 has not yet been reported.

1.7.3 ARS2 is essential for early mammalian development

ARS2 is essential for early mammalian embryogenesis. *Ars2*-null mouse embryos are embryonic lethal and die shortly after implantation between embryonic day 3.5-5.5. *In vitro* culturing of *Ars2*-null blastocysts revealed that embryos can differentiate into trophectoderm and inner cell mass but cannot progress beyond the blastocyst stage of development [1]. This is consistent with *Ars2* deletion in zebrafish [124,125] and *Drosophila* [126], which also result in developmental lethality.

1.7.4 ARS2 plays an important role in stem and progenitor cell function

Following the discovery that *Ars2* deletion is embryonic lethal in mice, a number of reports emerged implicating *Ars2* in stem and progenitor cell function. Craig Thompson's lab showed that loss of *Ars2* in adult mice leads to increased apoptosis in hematopoietic organs and decreased bone marrow cellularity, suggesting that the ARS2 protein is required for proper hematopoiesis [4]. In another report, Eric Lai's lab found that ARS2-deficient neural stem cells (NSCs) fail to self-renew and have decreased multipotency [127]. Our lab in collaboration with Bob Chow's lab has shown that loss of ARS2 function in mouse retinal explants disrupts retinal progenitor cell differentiation resulting in increased production of photoreceptors (Nickerson, in preparation). We have also found that ARS2 is required for proliferation of the C2C12 muscle progenitor cell line (O'Sullivan, in submission). ARS2-deficient C2C12 myoblasts accumulate in S-phase and demonstrate delayed cell cycle progression. In addition, overexpression of ARS2 causes a potent dominant-negative effect resulting in a cell cycle defect more severe than that seen in ARS2-depleted cells, with C2C12 cells expressing high levels of

CMV-driven EGFP-ARS2 arresting in early S-phase (O'Sullivan, in submission). Together this data clearly establishes that ARS2 is essential for stem and progenitor cell maintenance and differentiation.

1.7.5 Plant ARS2 is involved in RNA metabolism

The first clues about the molecular roles of ARS2 came from studies of SERRATE (gene name *se*), the ARS2 ortholog in plants. Loss of SERRATE is embryonic lethal and hypomorphic mutations result in a variety of developmental defects such as serrated/hyponastic leaves and reduced fertility [128,129]. A number of these phenotypes have been linked to aberrant gene regulation by miRNA indicating that SERRATE is involved in miRNA metabolism [128]. It is now known that SERRATE interacts with both pri-miRNA transcripts and essential components of miRNA biogenesis – the RNaseIII-like enzyme DICER-LIKE 1 (DCL1) and the double-stranded RNA-binding protein HYPONASTIC LEAVES 1 (HYL1) [122,130]. SERRATE plays a vital role in miRNA biogenesis by increasing the accuracy and efficiency with which DCL-1 processes pri-miRNA to mature miRNA [122,130]. A recent study found that SERRATE interacts with both CBP80 and CBP20 suggesting that SERRATE may function in cooperation with CBC [131]. Further evidence of this is provided by the fact that SERRATE influences alternative splicing and pre-mRNA splicing in a similar manner as CBC [54,131]. Altogether these data suggest that SERRATE functions as a scaffold by bringing together RNA substrates with the appropriate RNA processing machinery. There is growing evidence that ARS2 plays a similar role in animals. A number of recent reports have shown that ARS2 is a member of CBC and that it is critical for the metabolism of several different classes of RNAPII transcribed RNA [2-4,55].

1.7.6 ARS2 is required for the processing of RNAPII transcripts

All RNAPII transcripts undergo co-transcriptional addition of a 7-methylguanosine (m7G) cap, which is subsequently bound by the heterodimeric CBC. It has recently been established that ARS2 interacts directly with the assembled CBC to from a tertiary complex known as CBCA [2-4]. ARS2 and the CBC bind a similar set of capped RNAs and have been implicated in a number of the same RNA processing pathways including RNA nuclear export, pre-mRNA splicing, RNA exosomal

degradation, miRNA biogenesis and RNA 3' end processing [2-4,47,50-55,132] (Fig. 1-8).

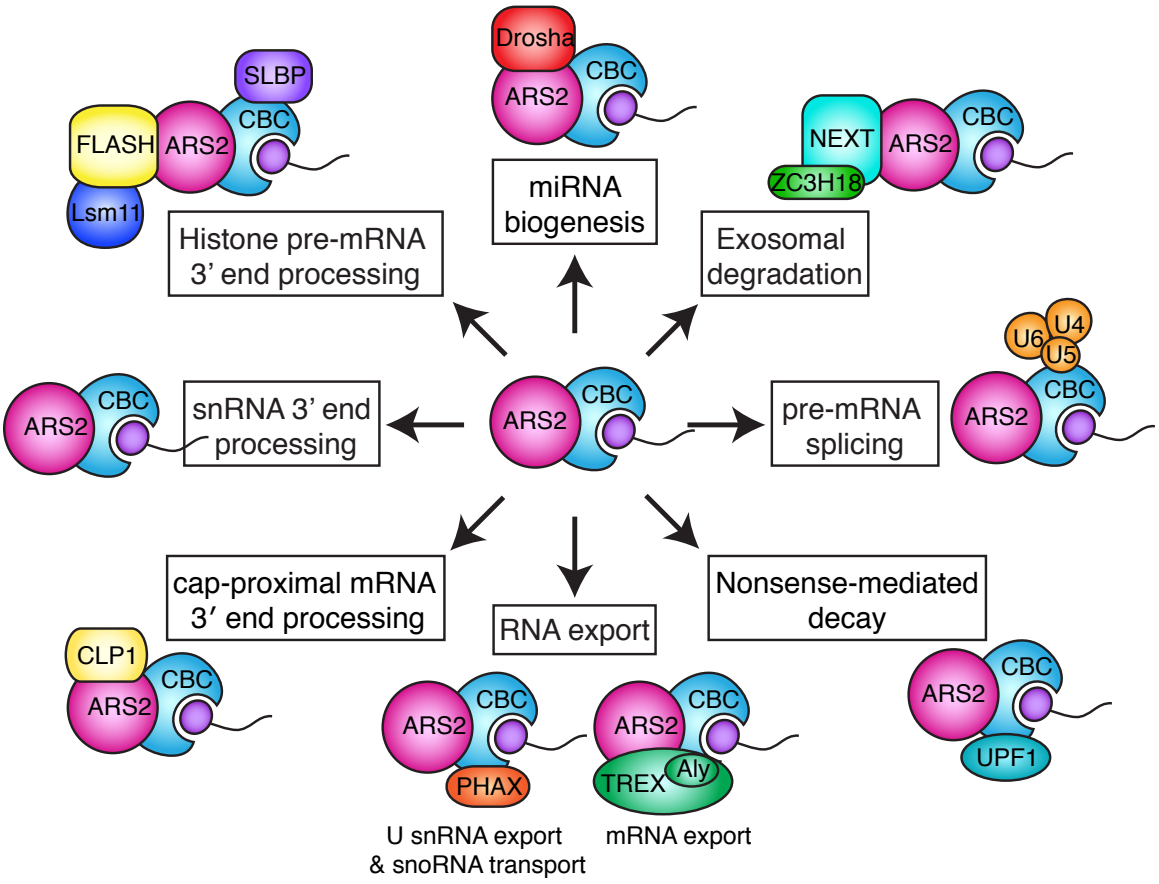

Figure 1-8 CBCA functions. The CBC-ARS2 (CBCA) complex has been implicated in many RNA processing events through its interaction with a wide variety of factors mediating RNA metabolism.

ARS2 promotes efficient nuclear export of polyadenylated mRNAs. This was demonstrated by an accumulation of polyadenylated RNAs in the nuclei of ARS2-depleted cells [133]. ARS2 associates with the transcription export (TREX) complex, a multiprotein complex that facilitates mRNA export and is recruited to mRNA during splicing [53,133-135]. TREX recruitment is mediated by interaction between CBP80 and the TREX component Aly [53]. Recruitment of TREX releases mRNAs from discrete foci known as nuclear speckle domains that are enriched with splicing machinery [136]. The majority of accumulated polyadenylated RNAs in the nuclei of ARS2-depleted cells

co-localize with the nuclear speckle domain marker SC35 suggesting that ARS2 contributes to mRNA export by promoting TREX recruitment [133].

It is likely that ARS2 is also involved in the nuclear export of snRNAs and intranuclear transport of snoRNAs. It has been previously shown that nuclear export of snRNAs requires CBC, the small RNA adaptor PHAX and the nuclear export receptor CRM1-RanGTP [137]. PHAX interacts directly with snRNA, CBC, and CRM1-RanGTP thereby bridging the gap between snRNA cargo and their export machinery [137]. CBC, PHAX and CRM1 are also involved in intranuclear transport of snoRNAs [138]. ARS2 allosterically enhances the binding of PHAX to CBC *in vivo* suggesting that ARS2 is involved in snRNA export and snoRNA intranuclear transport [2].

ARS2 has also been identified in RNA-protein complexes enriched for splicesomes in human cells but it is currently unknown if metazoan ARS2 functions in pre-mRNA splicing [139,140]. It is clear, however, that mammalian CBC interacts with the U4/U6.U5 tri-snRNP and that depletion of CBC from HeLa cells results in inhibition of pre-mRNA splicing and impaired spliceosome assembly [47,141,142]. In addition, the fact that mutations of *Arabidopsis CBP80*, *CBP20* and the plant *Ars2* ortholog *SERRATE* all lead to an accumulation of partially and alternatively spliced transcripts further supports a role for metazoan ARS2 as a splicing regulator [54,131].

ARS2 is part of an association of proteins that functionally links CBC to the nuclear RNA exosome, a multi-subunit complex that is engaged in RNA degradation, processing, and surveillance [3]. In order to participate in a diverse array of pathways with a multitude of substrates, the exosome operates with several cofactors that regulate its enzymatic activity and RNA-binding adaptors that mediate substrate recruitment. In humans, one complex that facilitates substrate recruitment is the nuclear exosome targeting (NEXT) trimer, which consists of hMTR4, RBM7 and ZCCHC8 [132]. Recently, Andersen *et al.* proposed that NEXT interacts with CBP80, CBP20, ARS2 and the uncharacterized zinc finger protein ZC3H18 to form a subcomplex that has been dubbed the NEXT-containing CBC subcomplex (CBCN) [3]. It has been demonstrated that CBCN associates with known exosome targets and that depletion of CBCN components ARS2, CBP80 or ZCCHC8 results in increased levels of these targets [3].

This data indicates that ARS2, in the context of CBCN, participates in recruiting a subset of substrates to the nuclear RNA exosome.

In 2009, accompanying reports published in Cell revealed that metazoan ARS2 and CBC are required for processing pri-miRNA into mature miRNA [4,55]. In the absence of ARS2 or CBP80, there is a loss of miRNA-mediated repression, a decrease in pri-miRNA stability and a corresponding decrease in mature miRNA [4,55]. ARS2 interacts with the microprocessor protein DROSHA and thereby provides a link between CBC and miRNA processing machinery [4]. ARS2 and CBC contribute to miRNA biogenesis by increasing the stability of pri-miRNA transcripts and facilitating their delivery to the Microprocessor [4].

Yet another RNA processing pathway that ARS2 is involved in is RDH mRNA 3' end processing. In a 2012 report, Gruber *et al.* observed that loss of ARS2 is accompanied by a decrease in RDH mRNAs and that ARS2 and CBC physically associate with RDH transcripts [5]. This is consistent with a previous report that found a reduction in total RDH mRNA following ARS2 knockdown [143]. Precisely how ARS2 executes its role in RDH pre-mRNA processing is not fully understood. ARS2 likely acts through its association with CBC, which has been shown to regulate RDH 3' end-processing through interaction with the negative elongation factor (NELF) complex [144]. However, ARS2 may also act through its interaction with FLASH, a nuclear protein that binds the U7 snRNP and is essential for RDH pre-mRNA 3' cleavage and S-phase progression [143]. Disruption of the interaction between ARS2 and FLASH is correlated with a reduction in total RDH mRNA levels and deficient S-phase progression but it is not known if the interaction is necessary for RDH pre-mRNA 3' cleavage [143]. Recently, ARS2 has been implicated in the 3' end-processing of two additional classes of small RNAPII transcripts: snRNA and snoRNA [2]. Depletion of ARS2, CBP80 or CBP20 in HeLa cells results in increased read-through of U4 snRNA- and U3 snoRNA-based reporters, indicating that ARS2 is necessary for the 3' end-processing of at least a subset of snRNAs and snoRNAs [2]. ARS2 also promotes 3' end-processing of mRNAs with cap-proximal polyadenylation sites further suggesting that ARS2, in the context of CBCA, is a key regulator of RNAPII transcript 3' end-processing [2].

Another level of RNA metabolism that ARS2 may play a role in is NMD, a surveillance mechanism that prevents expression of PTC-containing mRNAs. Evidence that CBC mediates NMD suggests that ARS2 may be involved as part of CBCA. For example, NMD is partially inhibited in CBP80-deficient mammalian cells [56]. The current understanding is that CBC contributes to NMD by interacting with UPF1 via CBP80 and thereby promoting the assembly of NMD machinery on target mRNAs [56,57]. It is also possible that ARS2 influences NMD through interaction with the RNA-binding protein RNPS1. The association between ARS2 and RNPS1 was first detected by two separate high-throughput protein interaction studies in human cells [145,146]. RNPS1 is a nucleocytoplasmic shuttling protein that contains a single RRM, an N-terminal serine-rich domain, and a C-terminal arginine/serine/proline-rich domain that includes an active NLS [147-149]. RNPS1 was first identified as a general splicing activator that is associated with active spliceosomes and stimulates both constitutive and alternative splicing [148]. It has since been reported that RNPS1 is also a peripheral member of the EJC that can trigger NMD when it is tethered to the 3' UTR of a β-globin mRNA reporter [150]. In addition, it has been shown that exogenously expressed FLAG-RNPS1 co-immunoprecipitates with each of the three UPF proteins in HEK293 lysates [150]. Furthermore, one study reported a functional link between RNPS1 cellular concentration and NMD efficiency [98]. Altogether, the data described above provides evidence that ARS2 may function as a master regulator of RNAPII transcript maturation by bringing the CBC and capped RNA transcripts together with the appropriate processing machinery.

1.8 Overview of FLASH

As previously mentioned, ARS2 interacts with multiple proteins that are important for RDH pre-mRNA processing, including CBP80, CBP20 and FLASH [5,143]. FLASH is a 220 kDa nuclear protein that was originally described as a pro-apoptotic factor involved in the activation of caspase-8 in response to Fas ligand binding [151]. The molecular mechanism by which FLASH contributes to apoptosis activation is poorly understood and FLASH's role in this process has been questioned [152,153]. Recently, study of FLASH has focused primarily on its role in regulating RDH expression. FLASH is expressed in a cell-cycle dependent manner with levels peaking

during S-phase, it is required for normal S-phase progression and it is localized in HLBs [154]. Depletion of FLASH by RNA interference results in loss of both RDH mRNA and protein, indicating that FLASH is involved in RDH expression [154]. An anti-FLASH antibody generated against the N-terminal region of human FLASH reduces cleavage of a histone H2A pre-mRNA reporter in mouse nuclear extracts [155]. Addition of bacterially expressed FLASH completely rescues RDH pre-mRNA processing indicating that FLASH is necessary for 3' end cleavage of RDH pre-mRNAs [155]. Disruption of the interaction between FLASH and ARS2 in KB cells results in deficient S-phase progression, inefficient localization of FLASH to HLBs and reduced levels of RDH mRNAs suggesting that the interaction may be necessary for RDH transcript processing [143].

1.9 Hypothesis and objectives

Despite recent growth in our knowledge of this previously uncharacterized protein, it remains unclear precisely how ARS2 functions in stem and progenitor cell maintenance and differentiation. The guiding hypothesis of my thesis work was that ARS2 contributes to progenitor cell maintenance and differentiation by promoting miRNA biogenesis and/or RDH pre-mRNA processing. To address this I used the myogenic progenitor cell line C2C12. My first objective was to determine the localization of ARS2 in undifferentiated, proliferating myoblasts and differentiating, post-mitotic myotubes. The discovery of abundant ARS2 expression in post-mitotic myotubes led me to my second objective – to determine if ARS2 is required for myogenic differentiation. The data from this portion of my work suggested that the interaction between ARS2 and FLASH is not necessary for progenitor cell differentiation. In response I set out to ensure that the degree of FLASH knockdown achieved in my experiments was sufficient to disrupt FLASH function. In addition, I observed a discrepancy in the localization of ARS2 signal between full-length, epitope-tagged ARS2 and endogenous ARS2. This suggested that there might be multiple ARS2 isoforms with distinct subcellular localizations therefore my fourth objective was to determine if an ARS2 isoform is expressed in C2C12 myoblasts. My work on this objective uncovered evidence of a cytoplasmic ARS2 isoform, raising the issue of what ARS2 is doing in the cytoplasm since all its reported functions take place in the nucleus. Data showing that

ARS2 interacts with RNPS1, a protein involved in NMD, led me to my fifth objective – to determine if ARS2 plays a role in NMD.

CHAPTER 2: MATERIALS AND METHODS

2.1 Cell culture

C2C12 and HeLa cell lines were purchased from American Tissue Culture Collection (ATCC). Neuro-2a (N2a) cells were received courtesy of Dr. Leigh Anne Swayne and were originally purchased from ATCC. All cell lines were cultured in growth medium (GM) consisting of DMEM (HyClone™) supplemented with 10% fetal bovine serum (FBS) (HyClone™) and 1% penicillin-streptomycin (Gibco®) at 37°C + 5% CO_2. Fusion of C2C12 myoblasts was induced by culturing in differentiation medium (DM) made up of DMEM + 2% horse serum (Fisher).

2.1.1 Cell line passaging

All cell lines were passaged or harvested for use in experiments when they reached 80-90% confluence. Cells were washed once with Ca^{2+}, Mg^{2+} and phenol red-free phosphate buffered saline (PBS) (HyClone) and then trypsinized for 3-5 minutes with 0.05% Trypsin-EDTA (Gibco®) at 37°C. Trypsin was quenched with GM and cells were centrifuged for 5 minutes at 350 × *g* to remove residual enzyme. The cells were then resuspended in GM and plated at a density of 3.5×10^4 cells/cm^2.

2.2 Bacterial strains

Escherichia coli DH5α (Invitrogen) was used to perform all molecular cloning procedures and cultured in Luria-Bertani (LB) broth supplemented with either 100 μg/mL ampicillin or 50 μg/mL kanamycin.

2.3 Plasmids

The mammalian expression vectors ARS2-Myc-His and 3XFLAG-ARS2 were cloned previously as described by Wilson *et al.* [1]. The full-length *Ars2* cDNA was subcloned from ARS2-Myc-His into pEGFP-C1 (Clontech) using the EcoRI and BamHI restriction sites. HA-FLASH and H2B-GFP expression plasmids were generous gifts from Dr. Gerry Melino and Dr. Tony Pawson respectively. EGFP-ARS2-X5 was produced from EGFP-ARS2 by Dr. Perry Howard.

All shRNAs were purchased from the HuSH™ shRNA collection at Origene in the vector pGFP-V-RS. The vector contains an shRNA cassette driven by the human U6

promoter and a turboGFP element driven by a CMV promoter. Integration of turboGFP allows for ready verification of transfection efficiency. The sequences of each shRNA used are listed in Table 2-1.

Table 2-1 shRNA sequences.

Target	shRNA sequence (5' to 3')
Ars2	CAGGCTGAGAATGACAGTTCCAACGATGA
Drosha	GTTCATTGAGCGGAAATACAGACAAGAGT
FLASH	AACATTGTGCCAATAATGTCTGGTCACGT
LSM11	GGATTACCAGCAGGTATTCACTCGGCACA
Non-targeting	GCACTACCAGAGCTAACTCAGATAGTACT

EGFP-FARB, EGFP-SCR and EGFP-RevChrg were produced by annealed oligo cloning into the EcoRI and BamHI sites of pEGFP-C1. The sequence for RevChrg was designed by replacing each charged amino acid in FARB with one of the opposite charge. The sequence for SCR was obtained by using a random sequence generator to scramble the FARB sequence. Amino acid sequences of FARB, RevChrg and SCR are listed in Table 2-2.

Table 2-2 FARB and control peptide amino acid sequences.

Peptide	Amino acid sequence
FARB	PKKKRKVGGDELEEGEIRSDDE
RevChrg	PKKKRKVGGRKLRKGRIESTTK
SCR	PKKKRKVGGEGEIESELDEDDR

For the replication-dependent histone (RDH) pre-mRNA 3' cleavage reporter (H2A-Reporter) a partial histone H2A coding sequence corresponding to the first 67 amino acids of H2A was synthesized upstream of an RDH 3' processing signal consisting of the conserved stem-loop and histone downstream element (HDE). For the control construct (H2A- Control) the partial H2A coding sequence was synthesized without the 3' processing signal. Both constructs were synthesized by Bio Basic Canada Inc. and subsequently subcloned upstream of the DsRed open reading frame (ORF) in pIRES2-DsRed-Express (Clontech) using EcoRI and BstXI restriction sites. The internal ribosome entry site (IRES) was removed during cloning and H2A-Reporter and H2A-Control are

in-frame with the downstream DsRed sequence, therefore DsRed protein is expressed if processing does not occur.

The Firefly luciferase box b reporter (FLuc-5Xboxb) was generated by subcloning FLuc-5Xboxb from the *Drosophila* expression plasmid pAc5.1C-Fluc-STOP-5boxb (Addgene) into pcDNA3.1(+) (Life Technologies) using restriction sites KpnI and XhoI. The firefly luciferase control without the 5boxb (FLuc) was produced by subcloning FLuc from pAc5.1C-Fluc-STOP-5boxb into pcDNA3.1(+) using restriction sites KpnI and EcoRI. The *Renilla* luciferase control plasmid (RLuc) was provided courtesy of Dr. Rob Ingham. To produce ARS2-Myc-His-λN (designated ARS2- λN), λN was cloned onto the C-terminus of ARS2-Myc-His using BamHI and HindIII restriction sites. ARS2-Myc-His was cloned by Wilson *et al.* [1] as previously reported and the λN coding sequence was synthesized by Bio Basic Canada Inc. To clone 3XFLAG- λN-RNPS1 (designated λN-RNPS1), λN was PCR amplified using primers incorporating sites for the restriction enzymes NotI and NheI and RNPS1 was PCR amplified from 3XFLAG-RNPS1 with primers incorporating sites for the restriction enzymes NheI and EcoRI. The resulting PCR products were digested with the appropriate restriction enzymes and cloned by three-way ligation into a retroviral vector modified from the pMSCVpuro vector (BD Biosciences) to contain an N-terminal 3XFLAG sequence. Creation of the 3XFLAG vector was described by Wilson *et al.* [1]. Amanda Carette, an honours student in the Howard lab, generated the expression plasmid 3XFLAG-RNPS1. The plasmid expressing λN alone was created by cloning λN into pcDNA3.1(-) using the restriction sites NotI and EcoRI.

All PCR products used for cloning were amplified by Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in 50 µl volumes according to the manufacturer's instructions using primers listed in Table 2-3.

Table 2-3 Primer sequences used to generate λN-RNPS1.

Primer	Sequence (5' to 3')
λN_F_NotI	CATTGCGGCCGCACTTATGGACGCACAAACACGACG
λN_R_NheI	CCTAGCTAGCTGGAGGAGCGTAATCTGGAACATC
RNPS1_F_NheI	AACCAGCTAGCATGGATTTATCAGGAGTGAAAAAGAAGAGCTTGC
RNPS1_R_EcoRI	TTAAGAATTCGGTTATCGGGAGGAGTTGGAGCTGG

2.4 Recombinant DNA techniques

All recombinant cloning procedures including isolation of plasmid DNA, restriction enzyme digestion, ligations and agarose gel electrophoresis were performed according to methods described in Molecular Cloning [156]. The QIAquick PCR Purification Kit, Qiaex II Gel Extraction Kit, and Qiagen Plasmid Mini/Maxi Kit were used according to the manufacturer's instructions (Qiagen). Restriction endonucleases (New England Biolabs), Antarctic Phosphatase (New England Biolabs), T4 DNA ligase (Life Technologies) and the Zero Blunt® TOPO® PCR Cloning Kit (Life Technologies) were used according to the manufacturer's protocols.

2.5 DNA sequencing

All sequencing was done by Sequetech Corporation.

2.6 Transient transfection

Cells were transfected at 60-80% confluence using jetPRIME® (Polyplus Transfection™) according to the manufacturer's protocol with slight modifications. DNA-jetPRIME® complexes were mixed with GM prior to addition to the cells and were replaced with fresh GM 5 hours after application.

2.7 Immunocytochemistry of C2C12 myoblasts and myotubes

C2C12 myoblasts were cultured in GM to ~60% confluence. They were then fixed in 4% PFA/PBS for 10 minutes followed by ice-cold methanol for 10 minutes at room temperature. Cells were permeabilized in 0.2% Triton-X-100 for 10 minutes at room temperature and blocked in 3% BSA/PBS for 1 hour at room temperature. All primary antibodies were diluted 1:250 in 1% BSA/PBS and applied overnight at 4°C. The Ludwig Institute for Cancer Research generated all anti-ARS2 antibodies as previously reported [1]. The following primary antibodies were used: anti-ARS2 N-terminal polyclonal (XL12.2), anti-ARS2 C-terminal polyclonal (XL14.1), anti-ARS2 N-terminal monoclonal (LX186.3), anti-FLASH (Santa Cruz), anti-FLAG (Agilent Technologies) and anti-Myc (Abcam). After incubation with primary antibody, the cells were washed

3X for 5 minutes each in PBS and then incubated in secondary antibody diluted 1:500 in 1% BSA/PBS for 1 hour at room temperature. Secondary antibody was either Alexa Fluor® 488 goat-anti-mouse (Molecular Probes®), Alexa Fluor® 488 goat-anti-rabbit (Molecular Probes®), or Rhodamine Red-X™ goat-anti-rabbit (Molecular Probes®). ProLong Gold Antifade Reagent with DAPI (Life Technologies™) was used according to the manufacturer's protocol to preserve fluorophores and counterstain nuclei. All slides were sealed with nail polish after curing in the dark at room temperature for 24 hours. Images were captured using a Leica DMIREZ Microscope (Leica Microsystems Inc.) and OpenLab 5.0.2 (Improvision® Ltd.) except where indicated. Images were optimized for contrast and brightness in Adobe® Photoshop® CS5 (Adobe Systems Inc.).

For immunocytochemistry of myotubes C2C12 cells were cultured on 96-well plates in GM until they reached 90% confluence and then were induced to differentiate in low-serum DM for 5 days. The resulting myotubes were then washed 1X very gently with PBS and fixed in ice-cold methanol for 10 minutes at room temperature. This was followed by fixation in 3.7% PFA/4% sucrose/PBS for 10 minutes at room temperature and permeabilization in 0.2% Triton-X 100 for 10 minutes at room temperature. Cells were blocked in 3% BSA/PBS for 1 hour at 4°C and then incubated overnight in primary antibodies diluted in 1% BSA/PBS. Primary antibodies were used at the following dilutions: 1:200 anti-ARS2 N-terminal polyclonal (XL12.2), 1:200 anti-ARS2 C-terminal polyclonal (XL14.1), and 1:250 anti-myosin heavy chain (MHC) (Developmental Studies Hybridoma Bank MF20). After incubation in primary antibodies, cells were washed 3X for 5 minutes each in PBS and then incubated for 1 hour in secondary antibodies diluted 1:500 in 1% BSA/PBS at room temperature. Secondary antibodies used were Alexa Fluor® 488 goat-anti-mouse (Molecular Probes®) and Rhodamine Red-X™ goat-anti-rabbit (Molecular Probes). The cells were then washed 3X for 5 minutes each in PBS and 1X in sterile ddH_2O and stained with 1:16 000 Hoechst 33342 (Life Technologies). All images were taken using a Leica DMIREZ Microscope (Leica Microsystems Inc.) and OpenLab 5.0.2 (Improvision® Ltd.). Images were optimized for contrast and brightness using Adobe Photoshop CS5.

2.8 Differentiation assay

C2C12 myoblasts were seeded on a 96-well plate and transfected. GM was replaced with DM 24 hours post-transfection to induce fusion. After 5 days of fusion, cells transfected with shRNA plasmids were fixed, permeabilized and stained as described above. Cells transfected with expression plasmids were imaged live. The knockdown differentiation assay was performed with 6 replicates and images of 6 predetermined fields were taken per replicate resulting in a total of 36 images per condition. The number of myotubes, nuclei in myotubes and the total number of nuclei were quantified. Myotubes were defined as MHC-positive cells with 3 or more nuclei. Differentiation index (the number of nuclei in myotubes divided by the total number of nuclei) and fusion index (number of nuclei per myotube) were calculated for each image. The overexpression differentiation assay was performed 4 times with 6 replicates each. A single image at the centre of each well was taken and the number of transfected nuclei per myotube was counted. Adobe® Photoshop® CS5 (Adobe Systems Inc.) and ImageJ (National Institutes of Health) software were used to quantify the data.

2.9 SDS-PAGE and western blotting

Whole cell lysates were prepared with 1X SDS sample buffer (62.5 mM Tris pH 6.8, 2% SDS, 2.5% 2-mercaptoethanol, 10% glycerol, 0.05% bromophenol blue) and heated at 95°C for ten minutes prior to electrophoresis. SDS-PAGE was performed by the Laemmli method[157]. All cell lysates were resolved through 10% SDS-PAGE gels except where indicated and transferred to PVDF membranes using Towbin transfer buffer (25mM Tris pH 8.3, 192 mM glycine, 20% (v/v) methanol). Membranes were blocked in 5% skim milk in Tris-buffered saline-Tween 20 (TBS-T) (25 mM Tris pH 7.4, 151 mM NaCl, 2.7 mM KCl, 0.05% Tween-20). Primary antibodies anti-ARS2 N-terminal polyclonal 1:2000 (XL12.2), anti-β-Actin 1:4000 (Sigma), anti-FLAG 1:1000 (Agilent Technologies) and anti-RNPS1 (Santa Cruz) were diluted in 1% skim milk in TBS-T. Membranes were incubated in primary antibody for 18-24 h at 4°C. Blots were then washed 3X for 5 minutes each in TBS-T and probed with either anti-mouse IgG horseradish peroxidase (HRP) conjugate (R&D systems) or anti-rabbit IgG HRP conjugate (R&D systems) at a working dilution of 1:10 000. Proteins were detected using

an enhanced chemiluminescence western blotting substrate (Pierce) and blots were exposed to film that was developed using an automated X-ray film developer.

2.10 Immunoprecipitations

Cells were transfected with 3XFLAG-ARS2 alone or together with EGFP-FARB, EGFP-SCR or EGFP. Cells were lysed 24 hours after transfection in lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton-X 100) for 30 minutes on ice and centrifuged for 20 minutes at 20 000 x *g* to remove debris. The lysates were incubated with anti-FLAG beads (Sigma) for 2 hours at 4°C. Beads were washed 3X in lysis buffer diluted 1:4 in PBS and bound proteins were eluted with 3XFLAG peptide (Sigma). Immunoprecipitates were re-suspended in 2X SDS sample bluffer (125 mM Tris pH 6.8, 4% SDS, 5% 2-mercaptoethanol, 20% glycerol, 0.1% bromophenol blue) and heated at 95°C for 10 minutes. Proteins were resolved and detected by SDS-PAGE and western blot as described above except where indicated.

2.11 Replication-dependent histone (RDH) reporter assay

Reporter and control constructs were synthesized by BioBasic Canada Inc. The reporter consists of a partial H2A coding sequence followed by a histone processing signal which contains the histone 3' stem-loop and histone downstream element. The control consists only of the partial H2A sequence and lacks the processing signal. Both constructs were cloned into pIRES2-DsRed-Express in-frame with the downstream DsRed open reading frame. N2a cells were co-transfected with either the H2A-Reporter or H2A-Control and a knockdown plasmid encoding shRNA or expression plasmid encoding EGFP-ARS2, EGFP-FARB, EGFP-SCR or EGFP. Transfections were performed as outlined above. The cells were harvested 24 hours post-transfection by trypsin digestion. They were then washed once and re-suspended in PBS. All resuspended cells were passed through a 50 μm cell strainer to remove any aggregated cells or cellular debris. The cells were kept on ice throughout the harvesting process and run live on a BD FACSCalibur™ (BD Biosciences) flow cytometer. Analysis was performed using FlowJo™ 8.7 (Treestar Inc.).

2.12 Detection and cloning of *Ars2 -X5*

C2C12 cells were homogenized in TRIzol® reagent (Life Technolgoies) and total RNA was purified using Direct-zol™ RNA Miniprep (Zymo Research) according to the manufacturer's protocol. A NanoDrop ND-1000 (Thermo Scientific) was used to measure the concentration, 260/280 and 260/230 of the RNA samples. Complement DNA (cDNA) was synthesized from 1 µg RNA using oligo $d(T)_{23}VN$ primers (New England Biolabs) and the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions. Standard PCR amplifications were carried out using One*Taq*® DNA Polymerase (New England Biolabs) according to the manufacturer's protocol. Transcripts containing sequence in intron 5 were detected by standard PCR using primers ARS2_F5, Ars2_R6, and Ars2_R7. Positions of the primers in *Ars2* mRNA are indicated in Fig. 3-13B. The predicted isoform was amplified by PCR using Phusion® High-Fidelity DNA Polymerase (New England Biolabs) according to the manufacturer's instructions with the primers Ars2_F5 and Ars2_R3'UTR. The resulting products were then used as template in a subsequent nested PCR reaction using primers Nested_F5 and Nested_R3'UTR. All primers used are listed in Table 2-4. These reactions were resolved on a 1% agarose gel and two bands were excised. DNA was extracted from agarose using the QIAquick PCR Purification Kit (Qiagen). The purified DNA was cloned into pCR™Blunt-TOPO® using the Zero Blunt® TOPO® PCR Cloning Kit (Life Technologies) according to the manufacturer's instructions. Positive clones were sent for sequencing and multiple primers were used to achieve overlapping sequence reads. Sequence reads were assembled using the CAP3 sequence assembly program available on the Pôle Rhône-Alpes de Bioinformatique (PRABI) website[158]. Assembled sequences were translated to protein sequences using the Translate tool available at the ExPasy SIB Bioinformatics Resource Portal (http://web.expasy.org/translate/). Sequencing results and CAP3 assembly details are listed in Appendix D.

For the purposes of this study *Ars2* Refseq mRNA: NM_031405.2, known in the database as serrate RNA effector molecule homolog transcript variant 1, and ARS2 Refseq protein: NP_113582.1, known as serrate RNA effector molecule homolog isoform 1, will be the 'canonical' sequences and all positional data refers to them.

Table 2-4 Primer sequences used to detect and clone *Ars2-X5*

Primer	Sequence (5' to 3')
Ars2_F5	TGTGAGAGGCTATGCCCTGC
Ars2_R6	TCTAGCATCTTGACAATGGCATCAGC
Ars2_R7	ATCATTGGCTTTGCGCTCTCCTTC
Ars2_R3'UTR	AGTGGTCACGGGTACTGATGTAGC
Nested_F5	CCCACATTCTGGTCTTGACTGCT
Nested_R3'UTR	GAGGGGACATGGATGGCTCAA

2.13 Nonsense-mediated decay luciferase assay

HeLa cells were transfected with FLuc-5xboxb or Fluc; either ARS2, ARS2-λN, RNPS1, λN-RNPS1, λN, or empty pcDNA(+); and RLuc. Firefly and *Renilla* luminescence was analyzed 24 hours post-transfection using the Dual-Glo Luciferase Assay System (Promega) and a Perkin Elmer Victor^{3}V 1420 multi-label plate reader.

2.14 Statistics

All statistics were performed in GraphPad Prism 6 for Mac OS X (GraphPad Software, Inc.).

CHAPTER 3: RESULTS

3.1 Localization of ARS2 in proliferating myoblasts and post-mitotic myotubes

To study the localization and potential role of ARS2 during cellular differentiation I used the mouse skeletal muscle cell line C2C12. Muscle cell differentiation, known as myogenesis, is an advantageous system for studying differentiation because it can be easily monitored both visually and biochemically. When adult skeletal muscle sustains damage, mononucleated muscle precursor cells, known as myoblasts, align and fuse into multinucleated myotubes that express muscle-specific proteins. The C2C12 cell line is an immortalized myoblast cell line that faithfully recapitulates myogenesis *in vivo* in response to low-serum conditions. The resulting myotubes are easy to visualize using a light microscope and are readily identified by immunofluorescence staining of well-defined myotube markers such as myosin heavy chain (MHC). In addition, this approach is particularly relevant to the study of ARS2 because myogenesis is a miRNA-dependent process [44].

To gain insight into the role of ARS2 in myogenic cells, my first objective was to determine where ARS2 is expressed in C2C12 myoblasts. To achieve this goal I performed immunofluorescence staining using anti-ARS2 antibodies generated by the Ludwig Institute for Cancer Research: LX186.3, monoclonal antibody raised in mice to amino acids 65-370 of full-length ARS2 and XL12.2 and XL14.1, polyclonal antibodies raised in rabbits to amino acids 65-370 and 676-871 respectively. Our lab independently verified LX186.3 and XL12.2 by a combination of western blot and immunofluorescence. For example, LX186.3 and XL12.2 produced strong nuclear and diffuse cytoplasmic staining in cultured *Ars2*$^{+/-}$ and *Ars2*$^{+/+}$ E3.5 mouse embryos. In *Ars2*$^{-/-}$ E3.5 embryos there was no nuclear staining and the diffuse cytoplasmic signal was lost over time as maternal protein stores were depleted (Fig. A-1; data not shown for XL12.2). In addition, western blots using XL12.2 showed a band of expected molecular weight that decreased in intensity following knockdown of ARS2 expression by siRNA (Fig. A-2). Together these results indicate that XL12.2 specifically recognize ARS2 on western blots and in immunofluorescence. Although XL14.1 was not independently verified it produced an

identical localization pattern as XL12.2 in C2C12 myoblasts and myotubes (Fig. 3-1 and 3-2).

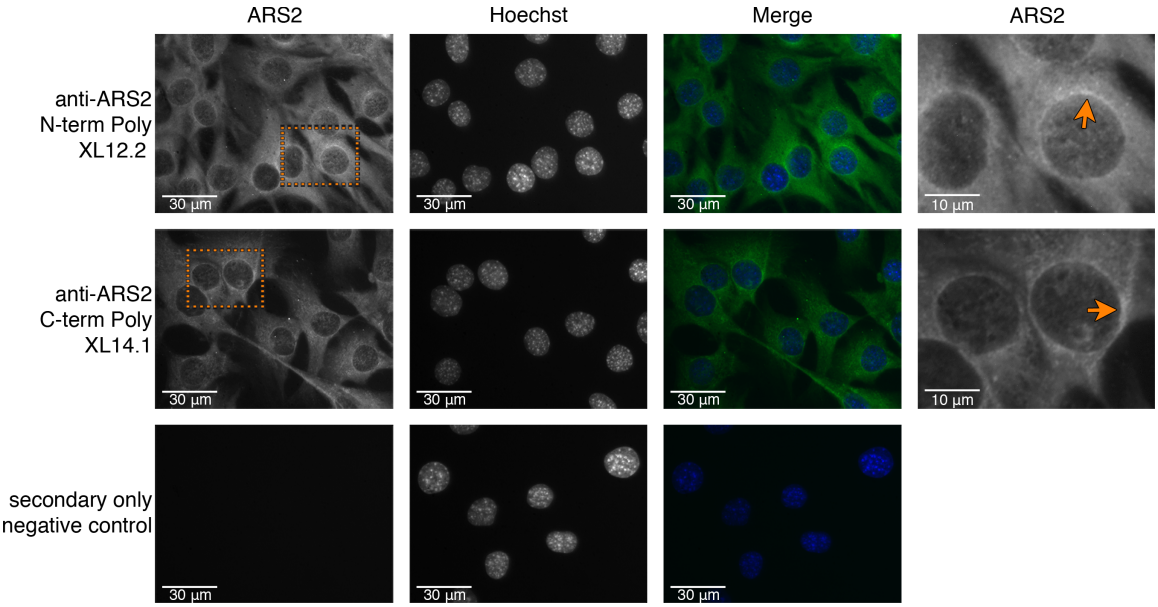

Figure 3-1 ARS2 is expressed in the cytoplasm and nucleus of C2C12 myoblasts. C2C12 myoblasts stained with polyclonal anti-ARS2 antibodies (green) and a negative control in which primary antibodies were omitted. Nuclei were stained with Hoechst 33342 (blue). ARS2 is expressed in both the cytoplasm and the nucleus of proliferating myoblasts. Dotted boxes indicate the regions that are magnified an additional 4X in the far-right panel. Arrows indicate areas of particularly strong signal in the perinuclear region.

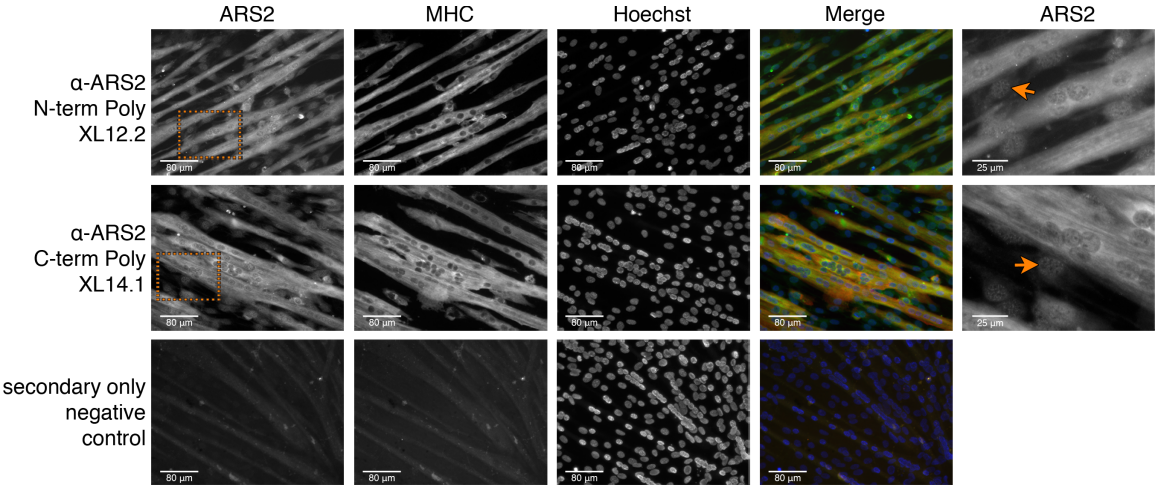

Figure 3-2 ARS2 is expressed in differentiating, post-mitotic myotubes. C2C12 myoblasts induced to differentiate in low serum conditions for 5 days and then fixed and stained with polyclonal anti-ARS2 antibodies (green), anti-myosin heavy chain (MHC) antibody (red) and a negative control in which primary antibodies were omitted. Nuclei were stained with Hoechst 33342 (blue). ARS2 is abundant in myotubes and localized in both the cytoplasm and the nucleus.

Dotted boxes indicate regions that are magnified an additional 4X in the far-right panel. Arrows indicate unfused, mononuclear cells that have a marked decrease in cytoplasmic expression compared to proliferating myoblasts.

Immunofluorescence staining with anti-ARS2 polyclonal antibodies (XL12.2 and XL14.1) showed ARS2 signal in the cytoplasm and the nucleus of proliferating myoblasts (Fig. 3-1). The cytoplasmic staining was widespread throughout the cell with areas of particularly strong signal in the perinuclear region. Nuclear staining showed weak ARS2 signal throughout the nucleoplasm with multiple small regions of exclusion that may represent nucleoli. This differs from the faint cytoplasmic and strong nuclear signal that was reported for exogenously expressed ARS2 in 3T3 MEFs and human kidney-derived BOSC-23 cells [1,4]. In contrast, immunofluorescence staining with anti-ARS2 monoclonal antibodies (LICR LX186.3 and LX187.4) resulted in a very different localization pattern compared to anti-ARS2 polyclonal antibodies (Fig. 3-6; data not shown for LX187.4). The monoclonal antibodies showed cytoplasmic, punctate localization and little to no nuclear localization. These results are confusing and call into question the specificity of the antibodies for ARS2. The simplest explanation is that all three antibodies cross-react with another antigen by immunofluorescence. An alternative explanation that accounts for differing localization is that the antibodies recognize different isoforms of ARS2 that have distinct localizations. Although, the antigens for monoclonal and polyclonal antibodies were the same, it is possible that they recognize different epitopes within the antigens.

According to Gruber *et al.*, ARS2 is expressed exclusively in proliferating cells and is absent in cells that have exited the cell cycle [4]. This report is inconsistent with a study of ARS2 in quiescent neural stem cells and post-mitotic progenitor cells that found no correlation between proliferative state and ARS2 expression [127]. When myogenic precursor cells are induced to differentiate they irreversibly exit the cell cycle early in the differentiation program and this process is recapitulated in the C2C12 cell line. Connor O'Sullivan, a PhD student in our lab, previously confirmed that there are no actively cycling cells in C2C12 cultures after 5 days of differentiation by observing a loss of BrdU incorporation into cellular DNA (O'Sullivan, in submission; Fig. A-3). O'Sullivan also found that ARS2 expression persists in differentiating C2C12 cells by western

blotting whole cell lysates 7, 8 and 9 days after inducing differentiation (O'Sullivan, in submission; Fig. A-3). Together this data indicates that ARS2 is expressed in non-cycling, differentiating C2C12 cells.

To determine the localization of endogenous ARS2 during differentiation and confirm that ARS2 is expressed in post-mitotic cells, I performed immunocytochemistry on C2C12 cells after 5 days of differentiation. The cells were fixed and stained using anti-ARS2 (XL12.2 and CL14.1) and anti-MHC, a commonly used marker for differentiated myotubes. I found that the presumed ARS2 signal is abundant in both maturing myotubes and unfused, mononuclear myocytes (Fig. 3-2). Like the undifferentiated myoblasts described above, staining was observed in both the cytoplasm and the nucleus of multi-nucleated myotubes (Fig. 3-2). Surprisingly, ARS2 signal in the cytoplasm of remaining unfused, mononuclear myocytes was markedly decreased compared to that in fused myotubes and cycling myoblasts. These cells also showed increased ARS2 signal in the nucleus. This result, combined with the fact that ARS2 contains putative nuclear localization and export signals, is consistent with the reported shuttling of ARS2 between the cytoplasm and the nucleus [4]. Although we cannot rule out cross reactivity with the antibodies, these data collectively show that ARS2 expression is maintained in post-mitotic cells and imply that ARS2 functions extend beyond the cell cycle.

As an alternative approach to addressing the question of full-length ARS2 localization in C2C12 cells, I performed immunofluorescence staining on cells transfected with plasmids expressing epitope tagged full-length *Ars2* cDNA. Full-length mouse *Ars2* cDNA produces a single band of the expected ~100 kDa on western blots when it is expressed as an N-terminal 3XFLAG fusion or a C-terminal Myc-His fusion (Fig. 3-3). 3XFLAG-ARS2 and ARS2-Myc-His both show strong nuclear localization with weak cytoplasmic staining in some cells (Fig. 3-4A). I found that N-terminal and C-terminal EGFP fusion proteins also localize predominantly to the nucleus (Fig. 3-4B). This is consistent with previous reports of exogenously expressed, epitope-tagged ARS2 in additional cell lines indicating that the localization of ARS2 is not cell type dependent [1,4]. The fact that N- and C-terminally tagged ARS2 fusion proteins show the same localization indicates that the position of the epitope tag does not interfere with ARS2

localization. My results suggest that full-length ARS2 is primarily a nuclear protein in C2C12 cells.

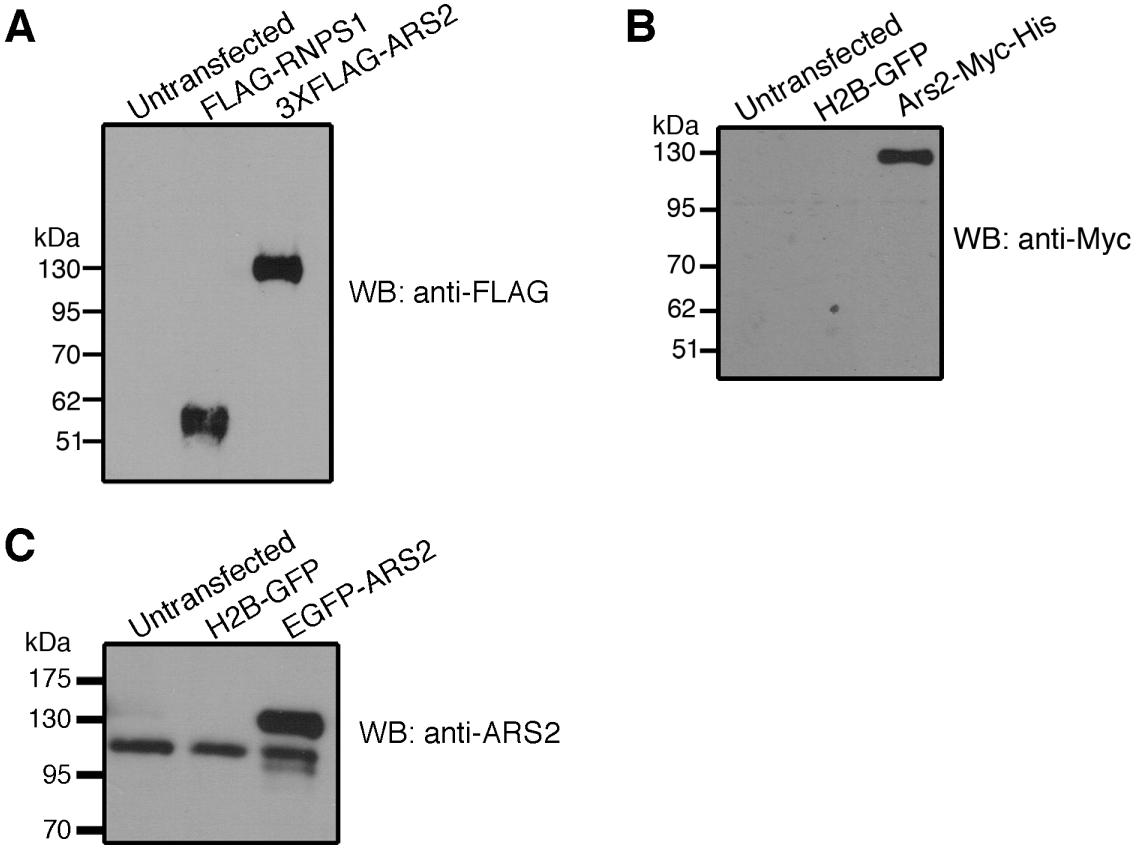

Figure 3-3 Exogenously expressed full-length ARS2 is expressed as a 100 kDa protein. Western blot analysis of **(A)** N-terminus FLAG tagged full-length ARS2 using anti-FLAG **(B)** C-terminus Myc tagged full-length ARS2 using anti-Myc and **(C)** N-terminus EGFP tagged full-length ARS2 using anti-ARS2 (XL12.2).

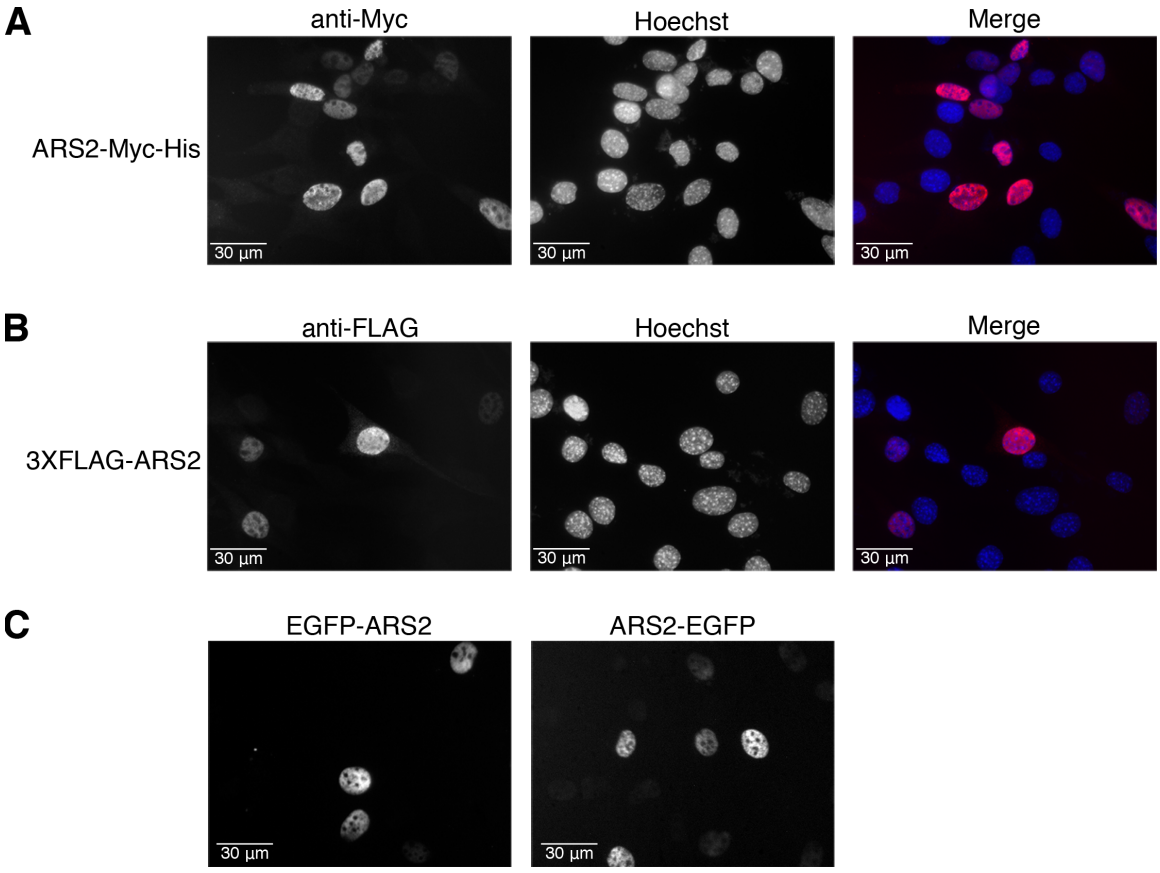

Figure 3-4 Exogenous full-length ARS2 is expressed predominantly in the nucleus. (A) C2C12 cells transfected with ARS2-Myc-His (red) and stained with anti-Myc (red). **(B)** C2C12 cells transfected with 3XFLAG-ARS2 and stained with anti-FLAG. Nuclei were stained with Hoechst 33342 (blue). **(C)** C2C12 cells transfected with N-terminally tagged EGFP-ARS2 or C-terminally tagged ARS2-EGFP and imaged live.

3.2 Localization of FLASH in proliferating C2C12 myoblasts

FLASH is a large nuclear protein required for replication-dependent histone (RDH) pre-mRNA processing that has been shown to interact with ARS2 in a number of cell lines [143,159]. Both FLASH and ARS2 are required for cell cycle progression in multiple cell types including C2C12 and this requirement has been linked to the maturation of RDH mRNA [5,143,159]. To determine if ARS2 and FLASH interact in C2C12 cells, I first examined the subcellular localization of FLASH. Immunofluorescence of C2C12 cells transiently transfected with a plasmid encoding HA-FLASH and stained with an anti-FLASH antibody showed punctate nuclear signals (Fig. 3-5). This pattern of localization is characteristic for FLASH and has been observed in

several different cell lines [143,155]. In human and *Drosophila* cell lines, FLASH-positive puncta correspond to dedicated subnuclear foci called histone locus bodies (HLBs) that localize to RDH gene clusters and mark sites of RDH transcription and pre-mRNA processing [72,154,155,160,161]. Immunofluorescence of untransfected C2C12 myoblasts using an anti-FLASH antibody revealed that endogenous FLASH has the same punctate, nuclear localization in C2C12 cells as exogenously expressed HA-FLASH (Fig. 3-6). Co-expression of HA-FLASH and EGFP-ARS2 in C2C12 cells showed that the two proteins are both localized to the nucleus and have overlapping, but distinct distributions (Fig. 3-5). EGFP-ARS2 is distributed throughout the nucleus whereas HA-FLASH is restricted to small punctate foci. Although ARS2 co-localizes with FLASH at these foci it is not limited to them, which is consistent with a broader role of ARS2 in RNAPII transcription. Consistent with published data on the ARS2/FLASH interaction, I found that FLASH co-immunoprecipitates with 3XFLAG-ARS2 in C2C12 cell lysates [143] (Fig. 3-7). Although the localization and immunoprecipitation data suggests that ARS2 and FLASH interact in C2C12 myoblasts, I was unable to detect colocalization of endogenous FLASH and ARS2 using a monoclonal anti-ARS2 antibody (LX186.3) (Fig. 3-6). This does not rule out the possibility of ARS2/FLASH colocalization since LX186.3 is the ARS2 antibody that displayed the least amount of nuclear localization. Unfortunately monoclonal anti-FLASH antibodies were not available to test for colocalization with polyclonal ARS2 antibodies XL12.2 and XL14.1.

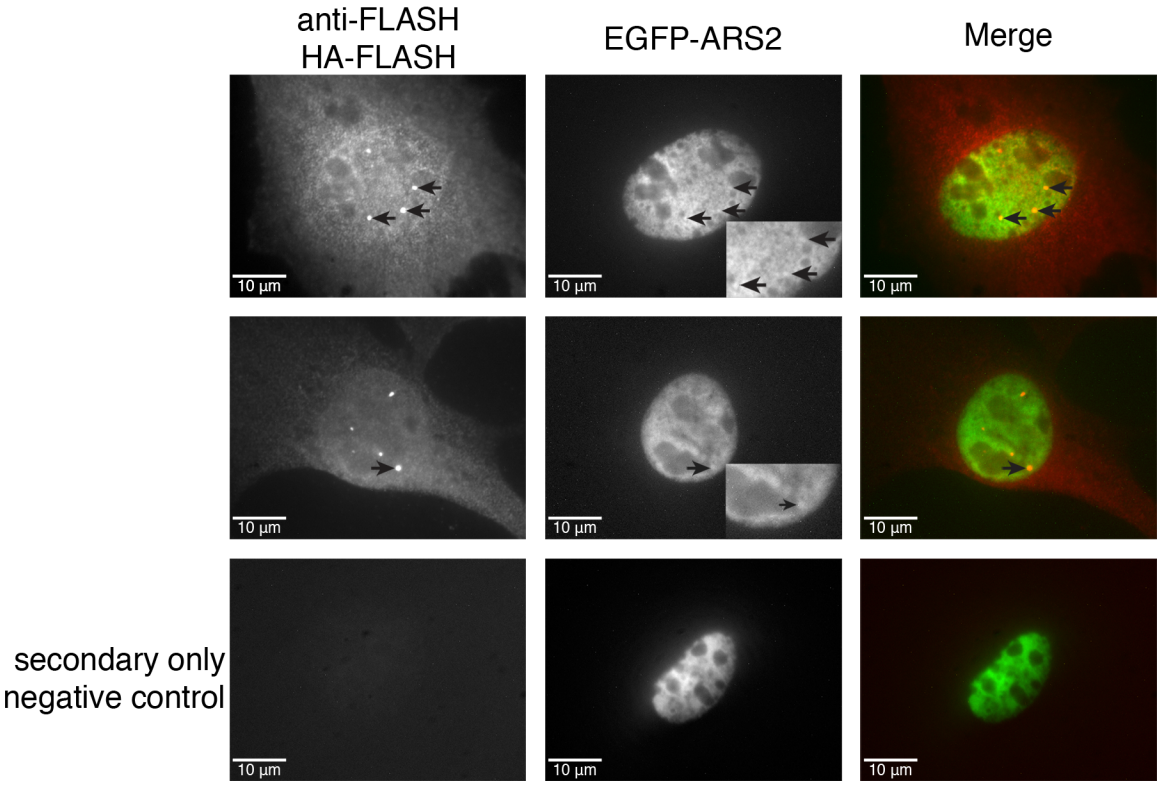

Figure 3-5 Exogenous ARS2 and FLASH co-localize in C2C12 myoblasts. C2C12 myoblasts transfected with EGFP-ARS2 and HA-FLASH. Cells were fixed 24h post-transfection and stained with an anti-FLASH antibody (red). EGFP-ARS2 signal is the native fluorescence retained following fixation. HA-FLASH shows punctate nuclear localization consistent with the localization of endogenous FLASH reported in multiple cell lines in the literature. EGFP-ARS2 is expressed throughout the nucleus but forms nuclear foci that co-localize with FLASH foci. Boxes in EGFP-ARS2 panels are magnified an additional 1.5X.

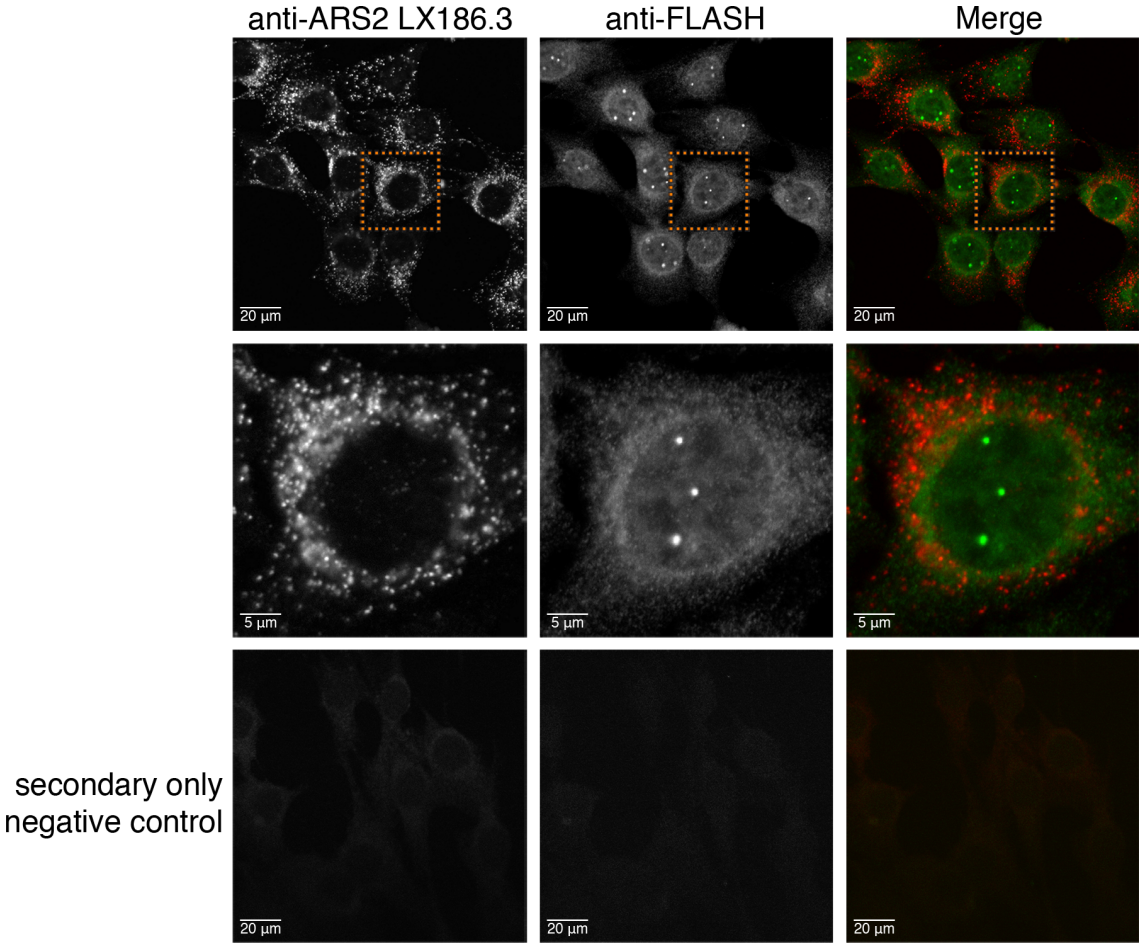

Figure 3-6 Endogenous ARS2 and endogenous FLASH do not appear to colocalize in C2C12 myoblasts. C2C12 myoblasts fixed and stained with an anti-ARS2 N-terminal monoclonal antibody (LX186.3) (red) and an anti-FLASH antibody (green). This ARS2 antibody gives a signal that is predominantly cytoplasmic and localized to distinct punctate bodies. The FLASH signal is found in characteristic, nuclear bodies. There does not appear to be any co-localization of the signals. Cells were observed by confocal microscopy, and a representative single optical section is shown. Images were taken by Dr. Bob Chow using a Nikon Digital Eclipse C1 plus confocal microscope and Nikon EZ-C1 3.6 confocal imaging software.

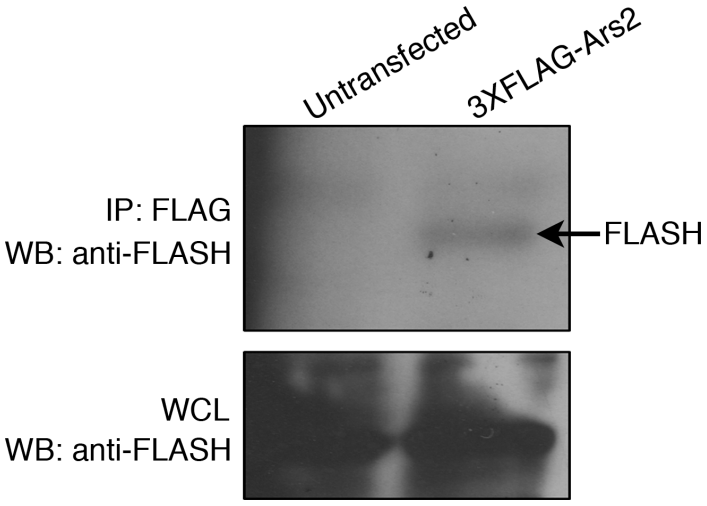

Figure 3-7 FLASH co-immunoprecipitates with 3XFLAG-ARS2 in C2C12 myoblasts. Cells were transfected with a plasmid overexpressing 3XFLAG-ARS2 and immunoprecipitation was performed with anti-FLAG beads. After elution with FLAG peptide, proteins were fractionated on a 4-12% gradient gel by SDS-PAGE and probed by western blot with anti-FLASH. Untransfected cells were used as a negative control.

3.3 ARS2 is required at specific levels for C2C12 myogenic differentiation.

The evidence that ARS2 is expressed in proliferating and differentiating myogenic cells (Fig. 3-1 and Fig. 3-2) suggests that it may be important for myogenic differentiation. To investigate the potential role of ARS2 in myoblast differentiation, I used shRNA to knock down ARS2 expression in C2C12 myoblasts immediately prior to differentiation. Knockdown of ARS2 was verified by western blot (Fig. A-4). C2C12 myoblasts were transfected with *Ars2*-targeting shRNA or non-targeting control shRNA and were induced to differentiate 24 hours post-transfection. Differentiation status was assessed 5 days later by calculating differentiation index and fusion index. Differentiation index represents the number of nuclei in myotubes divided by the total number of nuclei and fusion index represents the average number of nuclei per myotube. I found that fewer C2C12 myoblasts expressing *Ars2*-targeting shRNA formed myotubes than those transfected with control shRNA as illustrated by a decrease in mean differentiation index (0.109 vs. 0.331, $p<0.0001$; Fig. 3-8). Also, *Ars2* shRNA-transfected C2C12 myotubes had a lower mean fusion index than non-targeting shRNA-transfected C2C12 myotubes (5.686 vs. 9.067, $p<0.0001$; Fig. 3-8). We have previously shown that there is no difference in apoptosis between *Ars2*-shRNA- and control shRNA-expressing cells

(O'Sullivan, in submission). This data indicates that *Ars2* knockdown inhibits the differentiation of C2C12 myoblasts into myotubes. C2C12 cells overexpressing ARS2 as an EGFP fusion protein also show defects in differentiation, as measured by the contribution of transfected cells to multi-nucleated myotubes (Fig. 3-10). This is consistent with the observation that ARS2 overexpression in C2C12 cells produces a dominant negative phenotype (O'Sullivan, in submission). The fact that both ARS2 knockdown and overexpression show the same differentiation phenotype excludes the possibility that off-targeting of the shRNA produced nonspecific effects. These data have been independently confirmed by Amanda Carette, an honours student in the Howard lab, and they indicate that there is a specific requirement for ARS2 in myogenic differentiation.

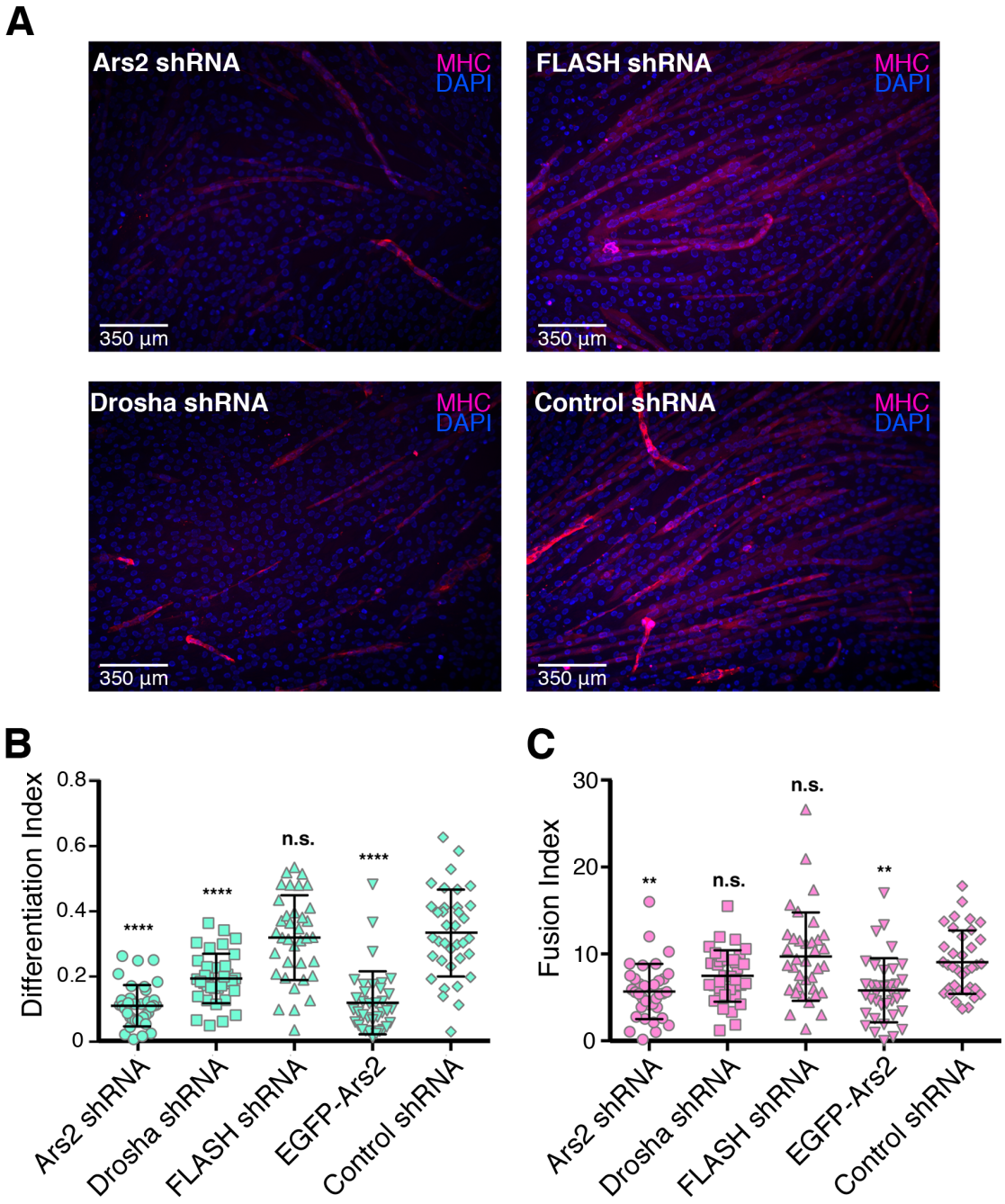

Figure 3-8 Knockdown of ARS2 expression results in decreased myoblast differentiation. **(A)** Representative images of C2C12 myoblasts that were transiently transfected with *Ars2*, *Drosha*, *FLASH* or Control shRNA and induced to differentiate for 5 days before fixation. Cells were stained with anti-myosin heavy chain (MHC) (red) to identify myotubes and Hoechst 33342 (blue) as a nuclear counterstain. This data was quantified by calculating **(B)** Differentiation index = # of MHC+ nuclei/Total # of nuclei and **(C)** Fusion index = # of nuclei in MHC+ cell with ≥ 3 nuclei/# of MHC+ cells. Data is from a single experiment with 6 replicates per condition. Images of 6 pre-determined fields of view per replicate were quantified. Statistical differences were determined by a one-way ANOVA followed by Tukey's multiple comparision post-test. Notations above columns indicate significance compared to Control shRNA. **** = $p \leq 0.0001$, ** = $p \leq 0.01$, n.s. = not significant

At the time of my thesis work ARS2 had been implicated in the biogenesis of two types of RNA: miRNA and RDH mRNA. miRNA was the first biogenesis pathway in which ARS2 was shown to operate. This was followed by reports that ARS2 functioned in RDH pre-mRNA processing. At this time it was not known that ARS2 is an integral component of the nuclear cap-binding complex (CBC) and there was some controversy over its function based on these differing reports. Therefore my thesis work focused, in part, on determining the relative contribution of these two pathways to myogenic differentiation. To assess whether the differentiation phenotype that results from ARS2 dysregulation is associated with ARS2's role in miRNA biogenesis, RDH pre-mRNA processing or both, I measured the differentiation status of cells transfected with shRNA targeting essential proteins in the two pathways and looked for phenocopies of ARS2 knockdown. DROSHA is an RNAse III-type enzyme required for miRNA biogenesis and FLASH is a large nuclear protein necessary for RDH pre-mRNA processing. The *Drosha* and *FLASH*-targeting shRNAs reduced the expression of their target proteins relative to a non-targeting control shRNA (Fig. A-4). C2C12 myoblasts transfected with *FLASH*-targeting shRNA showed no difference in differentiation or fusion index as compared to those transfected with non-targeting shRNA (Fig. 3-8). This suggests that FLASH is not required for differentiation of C2C12 myogenic cells. In contrast, myoblasts transfected with *Drosha*-targeting shRNA had a significantly lower differentiation index than those transfected with non-targeting shRNA (0.192 vs. 0.331, $p \leq 0.0001$) but showed no significant difference in fusion index (7.48 vs. 9.07, $p > 0.05$) (Fig. 3-8). Thus the effect of ARS2 knockdown on C2C12 myoblast differentiation more closely resembles the effect of DROSHA knockdown than FLASH knockdown, suggesting that deficiencies in miRNA biogenesis contribute to this phenotype, while FLASH-dependent functions of ARS2 do not. However, I cannot rule out the involvement of additional RNA biogenesis pathways, nor could I rule out that the FLASH knockdown was insufficient to compromise this pathway.

As an alternative approach to addressing the role of the FLASH/ARS2 interaction in the differentiation phenotype associated with ARS2 deficiency, I utilized a 13-amino acid peptide known as FARB [143]. FARB corresponds to the region of FLASH that interacts with ARS2 and it is both necessary and sufficient to achieve the interaction

[143]. Immunoprecipitation of 3XFLAG-ARS2 from C2C12 myoblast lysates specifically pulls down EGFP-FARB and does not pull down a scrambled control peptide (EGPF-SCR) or EGFP alone (Fig. 3-9). Moreover, data from our lab has shown that expression of FARB in developing mouse retinal explants phenocopies both ARS2 and FLASH knockdown, (Nickerson, in preparation), and implies that FARB can disrupt ARS2/FLASH interaction (Nickerson, in preparation). C2C12 myoblasts were transfected with plasmids expressing EGFP-FARB, EGFP-RevChrg, EGFP-ARS2 or EGFP and were induced to differentiate 24 hours later. The control peptide RevChrg consists of the FARB sequence in which all charged amino acids have been replaced with oppositely charged residues. Differentiation status was measured after 5 days of differentiation by counting the mean number of transfected nuclei per myotube. Unlike EGFP-ARS2, EGFP-FARB expressing cells showed no difference in differentiation when compared to EGFP-RevChrg, or EGFP expressing cells (Fig. 3-10). These data coupled with the FLASH knockdown data strongly suggest that the interaction between ARS2 and FLASH is not necessary for myoblast differentiation. However, more work is required to show that overexpression of FARB disrupts Ars2/FLASH interaction.

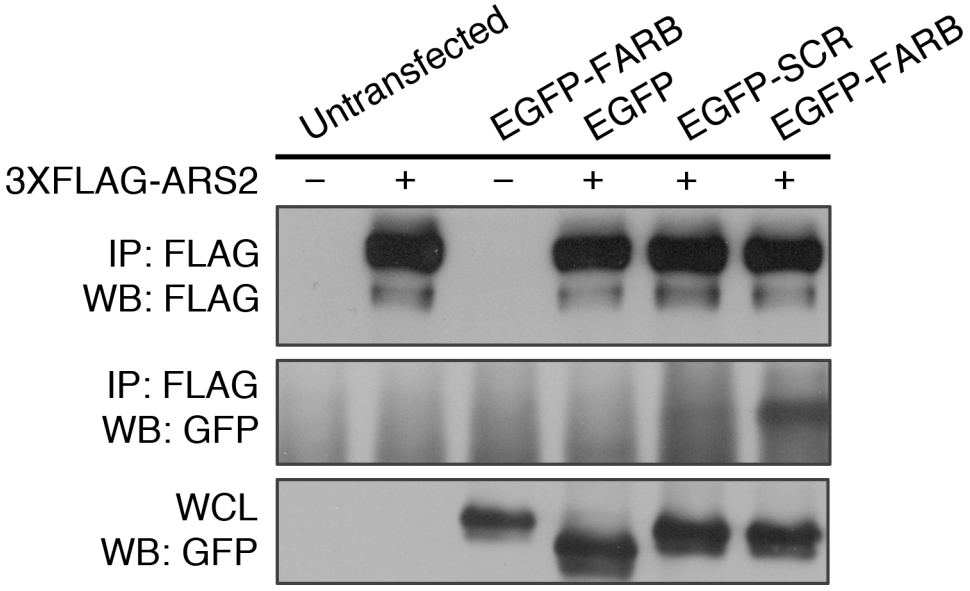

Figure 3-9 FARB specifically co-immunoprecipitates with 3XFLAG-ARS2. C2C12 myoblasts were transfected with 3XFLAG-ARS2 and/or EGFP-FARB, EGFP-SCR or EGFP. This was followed by immunoprecipitation with anti-FLAG beads and elution with FLAG peptide. The proteins were then resolved by SDS-PAGE and detected by western blot using the antibodies indicated. EGFP-FARB is only detected when 3XFLAG-ARS2 is present. EGFP and EGFP-SCR controls do not come down with 3XFLAG-ARS2.

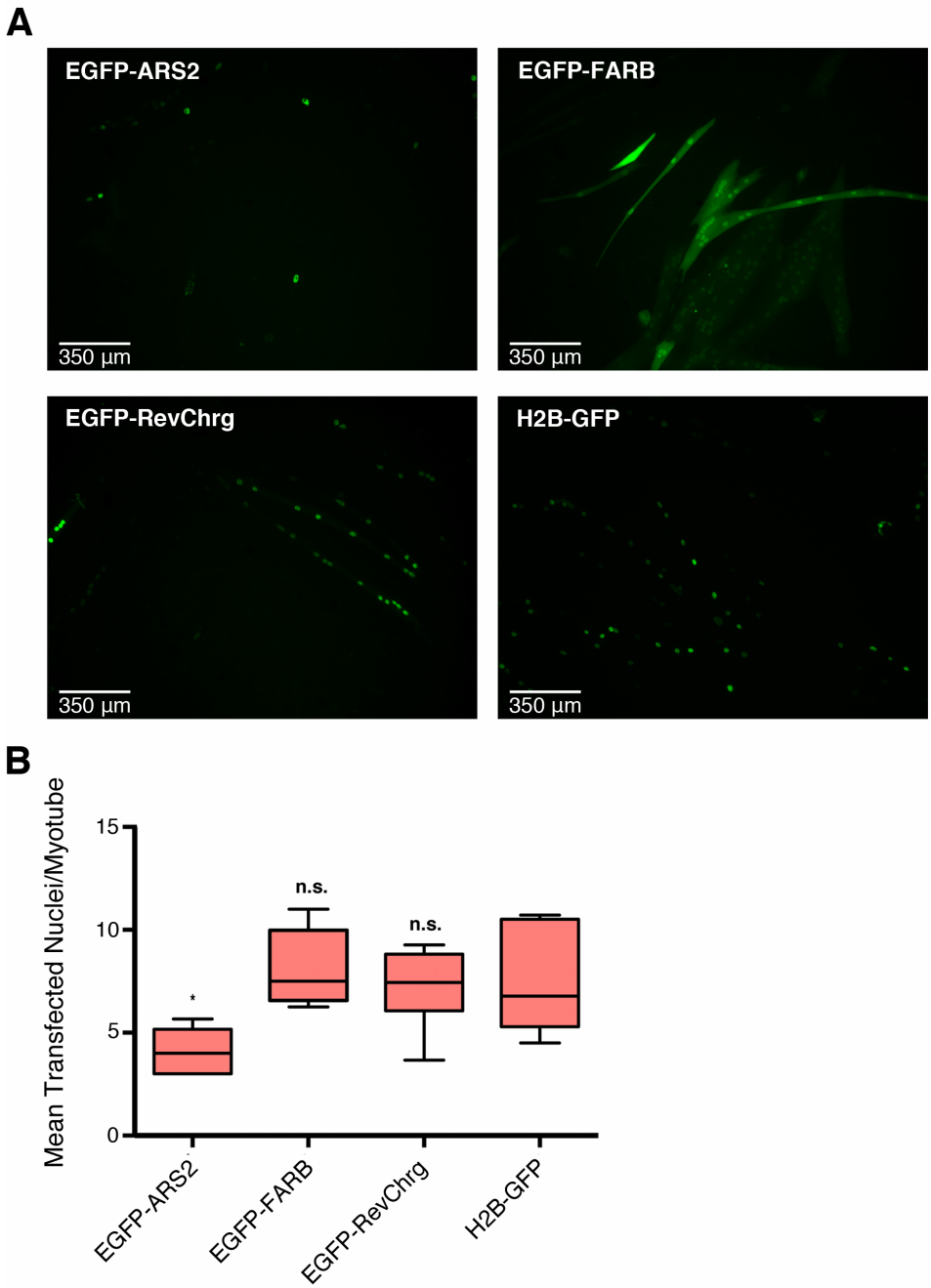

Figure 3-10 Overexpression of ARS2 results in decreased myoblast differentiation but expression of FARB has no effect. (A) Representative images of C2C12 myoblasts that were transiently transfected with EGFP-ARS2, EGFP-FARB, EGFP-RevChrg or H2B-GFP and induced to differentiate for 5 days before imaging. **(B)** Mean transfected nuclei/myotube of cells shown in (A). Data is representative of 3 separate experiments performed with 6 technical replicates each. A single image taken at the centre of each well was used for quantification. Statistical differences were determined by a one-way ANOVA followed by Tukey's multiple comparison post-test. Box = 25th to 75th percentile, whiskers = minimum to maximum data points, centre line = median. Notations above columns indicate significance compared to H2B-GFP control. * = p<0.05, n.s. = not significant.

3.4 FLASH knockdown by RNAi is sufficient to disrupt FLASH function in N2a cells

Although my data strongly suggests that FLASH/ARS2 interaction is not required for myogenic differentiation, to support the RNAi differentiation data I wanted to confirm that the level of FLASH knockdown is sufficient to disrupt FLASH function. To this end, I developed a fluorescent reporter assay to measure RDH pre-mRNA 3' cleavage by flow cytometry. This assay was adapted from a microscopy-based experiment previously performed exclusively in *Drosophila*[162]. The reporter (H2A-Reporter) consists of a partial histone H2A open reading frame (ORF) corresponding to the first 67 amino acids of the full-length protein followed by an RDH 3' processing signal – which includes the stem-loop, cleavage site, and histone downstream element (HDE) – located upstream of a DsRed ORF (Fig. 3-11). Theoretically, DsRed expression from the reporter should only occur under conditions in which RDH pre-mRNA 3' cleavage is impaired (Fig. 3-11).

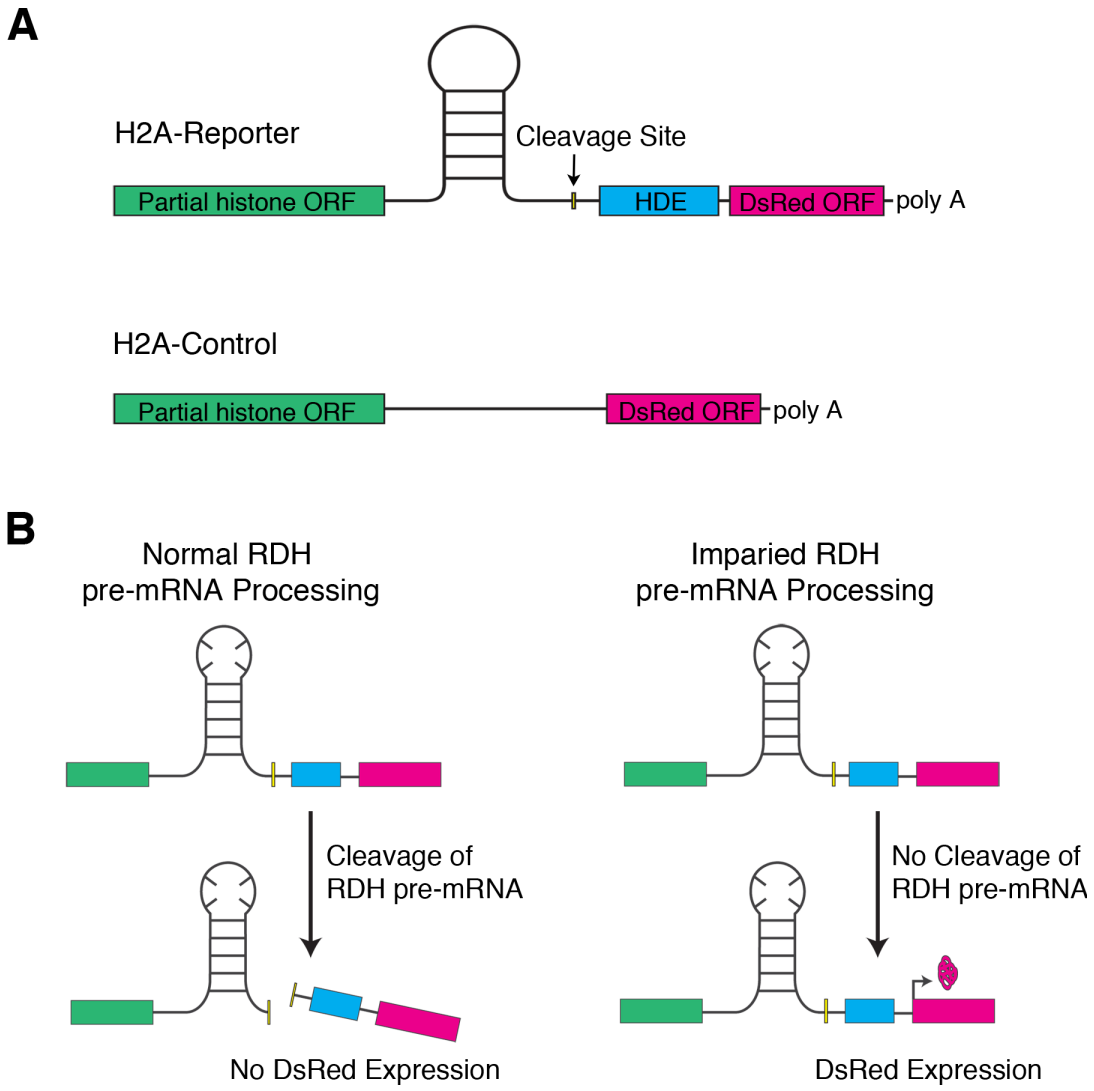

Figure 3-11 Replication-dependent histone (RDH) mRNA processing assay. (A) The reporter construct contains a partial histone H2a ORF corresponding to the first 67 amino acids of the full-length protein. This is followed by a replication-dependent histone pre-mRNA 3' processing signal which includes the stem loop, cleavage site and histone downstream element (HDE) and a DsRed ORF. The control construct contains the partial histone ORF and DsRed ORF but lacks the processing signal. **(B)** If there is a defect in RDH mRNA processing the reporter will not be cleaved and DsRed expression will occur. The control should produce consistent DsRed expression regardless of histone mRNA processing efficiency.

FLASH is a low abundance protein that is expressed exclusively during S-phase of the cell cycle and therefore it can be difficult to detect[160]. I have been able to detect FLASH protein in C2C12 cells by immunofluorescence staining and immunoprecipitation of 3XFLAG-ARS2 but not by western blot of C2C12 myoblast whole cell lysates. To avoid obstacles presented by low FLASH expression in C2C12

cells, I evaluated RDH pre-mRNA 3' cleavage in Neuro-2a (N2a) cells, which express FLASH at levels sufficient to detect by western blot (Fig. A-4).

N2a cells were co-transfected with H2A-Reporter and an shRNA plasmid targeting *Ars2, Drosha, FLASH, Lsm11* or a non-targeting control shRNA. LSM11 is an essential component of the RDH pre-mRNA processing machinery and was included in the analysis as a positive control. All shRNA plasmids contain a turboGFP cassette to allow for the identification of transfected cells. The cells were harvested 24 hours after transfection and analyzed by flow cytometry. RDH pre-mRNA processing efficiency was assessed by determining the number of DsRed and GFP double positive (double+) cells normalized to the total number of GFP+ cells. ARS2, FLASH and LSM11 knockdown significantly increased the population of double+ cells compared to control shRNA suggesting a loss of RDH pre-mRNA 3' processing (Fig. 3-12A). The effect on the reporter was specific to these RDH processing components and was not seen with DROSHA knockdown. These results suggest that FLASH knockdown by RNAi via shRNA is sufficient to disrupt FLASH function.

I also included in my analysis a new control (H2A-Control) that contains the partial histone H2A ORF and DsRed ORF but lacks the RDH 3' processing signal (Fig. 3-11). In an environment where RDH pre-mRNA 3' cleavage is compromised, H2A-Reporter should result in increased DsRed expression while H2A-Control should produce steady DsRed expression regardless of changes in RDH pre-mRNA 3' processing. Surprisingly, when N2a cells were co-transfected with H2A-Control and shRNA targeting *Ars2, Drosha, FLASH, Lsm11* or a non-targeting control shRNA, ARS2 FLASH, and LSM11 knockdown all increased DsRed expression while DROSHA knockdown did not (Fig. 3-12A). When the results were expressed as a ratio of H2A-Reporter to H2A-Control all significant differences compared to control shRNA were lost (Fig. 3-12B). These results suggest that ARS2, FLASH and LSM11 knockdown also affect transcription in addition to RDH 3' cleavage. The effect of LSM11 knockdown on plasmid expression was also seen on a related CMV-H2A-luciferase control indicating that this effect is independent of DsRed (data not shown).

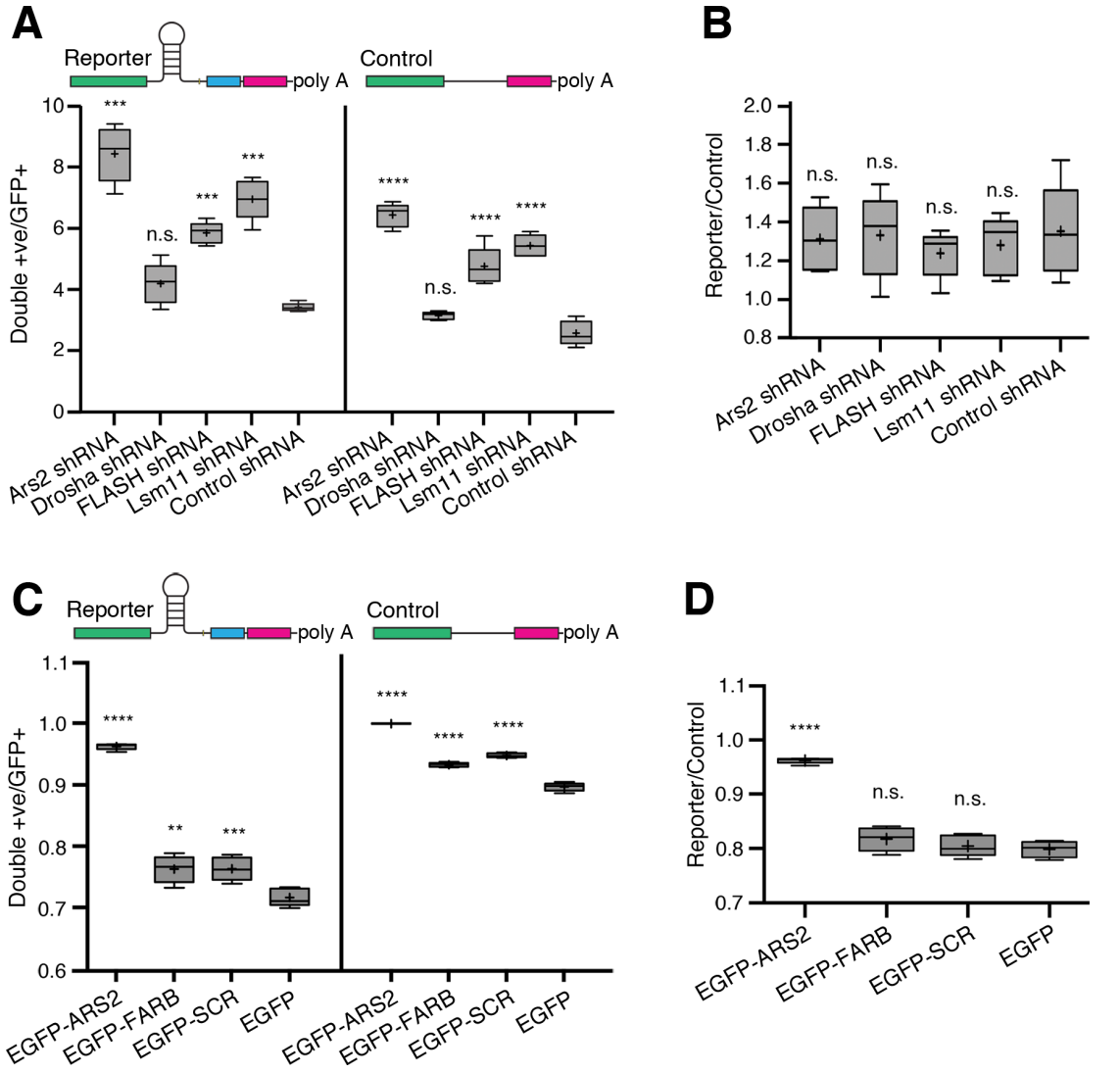

Figure 3-12 ARS2 is required at specific levels for RDH pre-mRNA processing. (A) N2a cells transiently co-transfected with either H2A-Reporter or H2A-Control along with *Ars2, Drosha, FLASH, LSM11* or control shRNA. All shRNA plasmids contain a TurboGFP cassette. The cells were analyzed by flow cytometry 48 h after transfection. TurboGFP+DsRed+ double positive cells are shown as a percentage of total TurboGFP+ cells. Data represents a single experiment performed with 5 replicates and 100 000 events captured per replicate. **(B)** Graph shows data from (A) as ratio of H2A-Reporter relative to H2A-Control. **(C)** N2a cells were transiently co-transfected with either H2A-Reporter or H2A-Control along with EGFP-ARS2, EGFP-FARB, EGFP-SCR or EGFP. The cells were analyzed by flow cytometry 24 h after transfection. EGFP+DsRed+ double positive cells are shown as a percentage of total EGFP+ cells. Data shown is from a single experiment that represents results from 3 separate experiments performed with 5 replicates and 100 000 events captured per replicate. **(D)** Graph shows data from (C) as ratio of histone reporter relative to histone control. Statistical differences were determined by a one-way ANOVA followed by Tukey's multiple comparison post-test. Box = 25th to 75th percentile, whiskers = minimum to maximum data points, centre line = median, + = mean. Notations above boxes indicate significance compared to (A-B) control shRNA or (C-D) EGFP. ** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p \leq 0.0001$, n.s. = not significant.

As previously indicated, overexpression of ARS2 results in dominant-negative phenotypes in the cell cycle and myoblast differentiation (O'Sullivan, in submission and Fig. 3-10). The dominant-negative has a more pronounced cell cycle arrest phenotype than RNAi knockdown (O'Sullivan, in submission). This led me to reason that knockdown via shRNA may not be complete enough for effects on RDH pre-mRNA 3' cleavage to avoid being masked by more general influences on DsRed transcription and that overexpression of ARS2 may be a more appropriate approach. I co-transfected N2a cells with H2A-Reporter and EGFP-ARS2, EGFP-FARB, EGFP-SCR or EGFP. As expected EGFP-ARS2 expressing N2a cells had an increased population of double+ cells compared to EGFP expressing cells, indicating a decrease in H2A-Reporter cleavage ($p \leq 0.0001$) (Fig. 3-12C). Although EGFP-FARB also significantly increased the double+ population, the effect of this peptide showed no difference compared to a scrambled peptide control (EGFP-SCR) suggesting the interaction between FLASH and ARS2 is not required (Fig. 3-12C). Unlike in the knockdown experiment, when H2A-Reporter was normalized to H2A-Control there is still a significant difference in double+ cells between EGFP-ARS2 and EGFP ($p \leq 0.0001$) (Fig. 3-12D). This indicates that ARS2 is an important factor for RDH pre-mRNA 3' cleavage in N2a cells. The reporter/control ratio for EGFP-FARB expressing cells is not significantly different when compared to EGFP expressing cells. Collectively, these data suggest that although ARS2 and FLASH are both required for RDH pre-mRNA 3'end processing, the interaction between ARS2 and FLASH is not necessary in N2a cells. However, an important caveat to this interpretation is the possibility that FARB does not disrupt the interaction sufficiently, even though it can be immunoprecipitated with ARS2 and phenocopies ARS2 and FLASH knockdown in the retina. However, the clear effect of FLASH knockdown on H2A-Reporter indicates that the level of knockdown is sufficient to disrupt FLASH function in RDH processing and supports the conclusion that FLASH is not involved in myogenic differentiation.

3.5 C2C12 cells express a truncated C-terminal *Ars2* transcript

When full-length *Ars2* cDNA is expressed the resulting protein is almost exclusively nuclear, yet staining of endogenous ARS2 produced both nuclear and cytoplasmic signals (Fig. 3-1). One potential explanation for this discrepancy is the presence of an alternative ARS2 isoform with cytoplasmic subcellular localization.

Western blotting of C2C12 whole cell lysates with the monoclonal anti-ARS2 antibody LX186.3, which was the antibody with the most cytoplasmic localization signal, produced two distinct bands: one at the predicted 100 kDa and one at ~60 kDa. Importantly, the additional 60 kDa band was specifically knocked down by *Ars2*-targeting siRNA which suggested that it represents either a degradation product of full-length ARS2 or a truncated ARS2 isoform that is expressed in C2C12 cells (Fig. 3-13A). The presence of a truncated ARS2 isoform could explain why immunofluorescence staining by different ARS2 antibodies produces different patterns of localization if the epitope for the monoclonal antibody is found on the 60 kDa isoform and the predominant epitopes for the polyclonal antibody are found on the 100 kDa but not the 60 kDa isoform (Fig. 3-1 and 3-6). Multiple alignments of *Ars2* (formally *SRRT*) in 40 placental mammals and measurements of evolutionary conservation revealed a highly conserved region in intron 5 (Fig. 1-7B). The conservation of this region in the middle of an intron suggested the possible presence of an internal promoter. Based on this observation I proceeded to search the NCBI-AceView database, a comprehensive database of alternative mRNA variants and alternative promoters compiled from experimental cDNA sequencing data, for alternative *Ars2* mRNAs. This analysis identified a partial mRNA that has a predicted transcriptional start site within intron 5 and is incomplete at the 3' end.

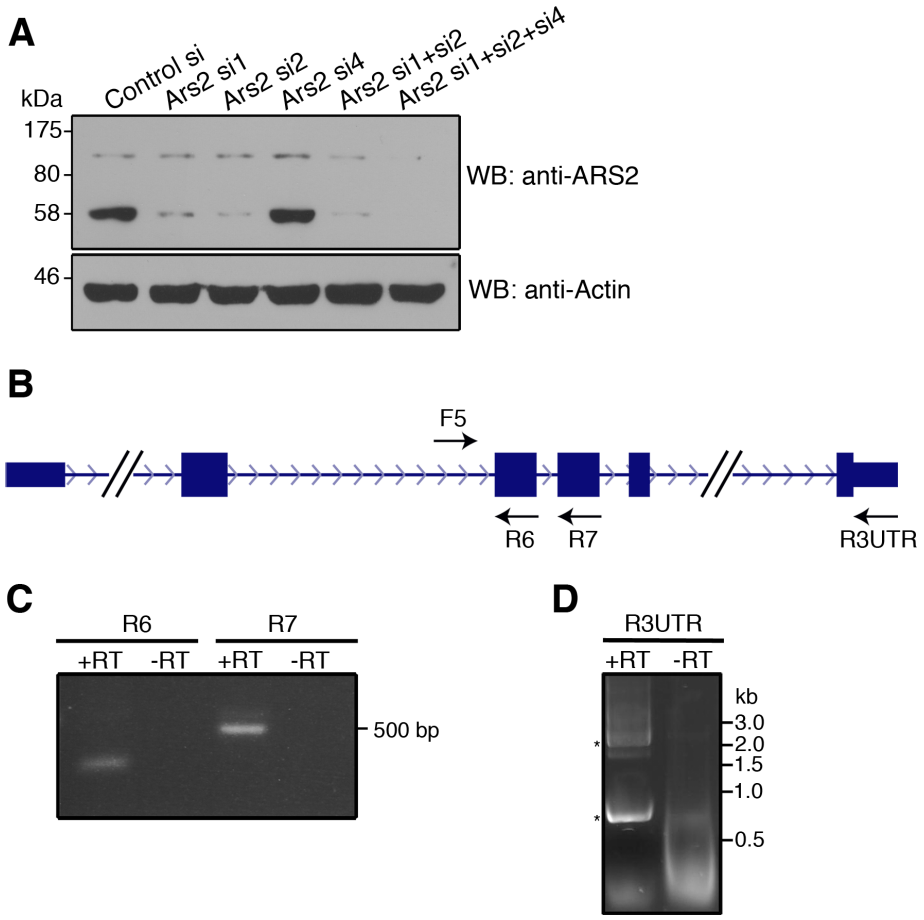

Figure 3-13 ARS2 isoform detection. (A) Western blot of C2C12 whole cell lysates with anti-ARS2 showing a band of approximately 60 kDa that is knocked down with siRNAs targeting ARS2. **(B)** Position of primers used for RT-PCR relative to *Ars2* mRNA sequence **(C)** RT-PCR of RNA isolated from C2C12 myoblasts identified evidence of a transcripts containing sequence from intron 5 of *Ars2*. **(D)** RT-PCT using primers complimentary to the canonical intron 5 and *Ars2* 3' UTR revealed two transcripts. Asterisks indicate bands that were excised and sent for sequencing.

To determine if this transcript is expressed in C2C12 cells, I performed RT-PCR on RNA that was reverse transcribed using oligo(dT) primers (Fig. 3-13C). I first amplified using a forward primer (Ars2_F5) located within intron 5 and a reverse primer (Ars2_R6) located in exon 6 (Fig. 3-13C). This reaction produced a band of the expected ~250 bp (Fig. 3-12C). Similarly, amplification using the same forward primer (Ars2_F5) and a reverse primer (Ars2_R7) located in exon 7 produced an expected band of ~500 bp (Fig. 3-13C). These results indicate that there is an alternative *Ars2* transcript expressed from a promoter within intron 5 of the *Ars2* gene in C2C12 cells. To characterize the full-length transcript, I next used the same Ars2_F5 forward primer with an additional reverse

primer (Ars2_R3'UTR) that is complementary to the 3'UTR of canonical *Ars2* mRNA based on the assumption that the two mRNAs would share a common 3'UTR. This amplification resulted in multiple weak bands of varying molecular weights (data not shown). All of the amplicons produced in this reaction were less than 4 kb ruling out the possibility that they were generated from unspliced RNA since amplification from unspliced RNA with these primers would be expected to produce a product >4 kb. To improve yield and increase specificity I next performed nested PCR, which resulted in 2 distinct bands with molecular weights of approximately 700 bp, and 2.1 kb (Fig. 3-13D). Sequencing revealed that the 700 bp transcript contains multiple stop codons preceding the longest ORF, suggesting that it is likely degraded by nonsense-mediated decay (NMD) and does not result in production of a protein. The 2.1 kb transcript contains a single ORF, with no preceding stop codons, that corresponds to amino acids 228-875 Δ809-815 of canonical ARS2. Following the naming convention utilized by Refseq, we have named this isoform ARS2-X5. We will need to perform 5' RACE in the future to complete the 5' UTR of this transcript. ARS2-X5 has an in-frame deletion of amino acids 809-815 (AGAVRPA) that is also missing in validated Refseq ARS2 isoform 2 (NP_113582.1) and is likely produced as a result of alternative splicing at an alternative splicing acceptor in exon 19. ARS2- X5 also has a large N-terminal deletion that includes the first 227 amino acids of full-length ARS2. In full-length ARS2 this region encodes an arginine rich region, a putative nuclear localization signal (NLS) and the DUF3546 domain (Fig 3-14A).

Loss of the putative NLS suggests that this isoform will be localized to the cytoplasm. To test this hypothesis Dr. Howard generated Δ227-ARS2 as an EGFP-fusion protein and I expressed it in C2C12 cells. One difference between this protein and the predicted ARS2-X5 is that Δ227-ARS2 includes amino acids 809-815 since we did not have the full ARS2-X5 sequence at the time of cloning. As shown in Fig. 3-14B, when expressed in C2C12 myoblasts, EGFP-Δ227-ARS2 is found almost exclusively in the cytoplasm. My data show that C2C12 cells express a truncated C-terminal *Ars2* transcript that codes for a predicted cytoplasmic ARS2 isoform. This result indicates that the cytoplasmic and nuclear localization detected with different ARS2 antibodies reflects the

localization of two different isoforms – one predominantly cytoplasmic and one predominantly nuclear.

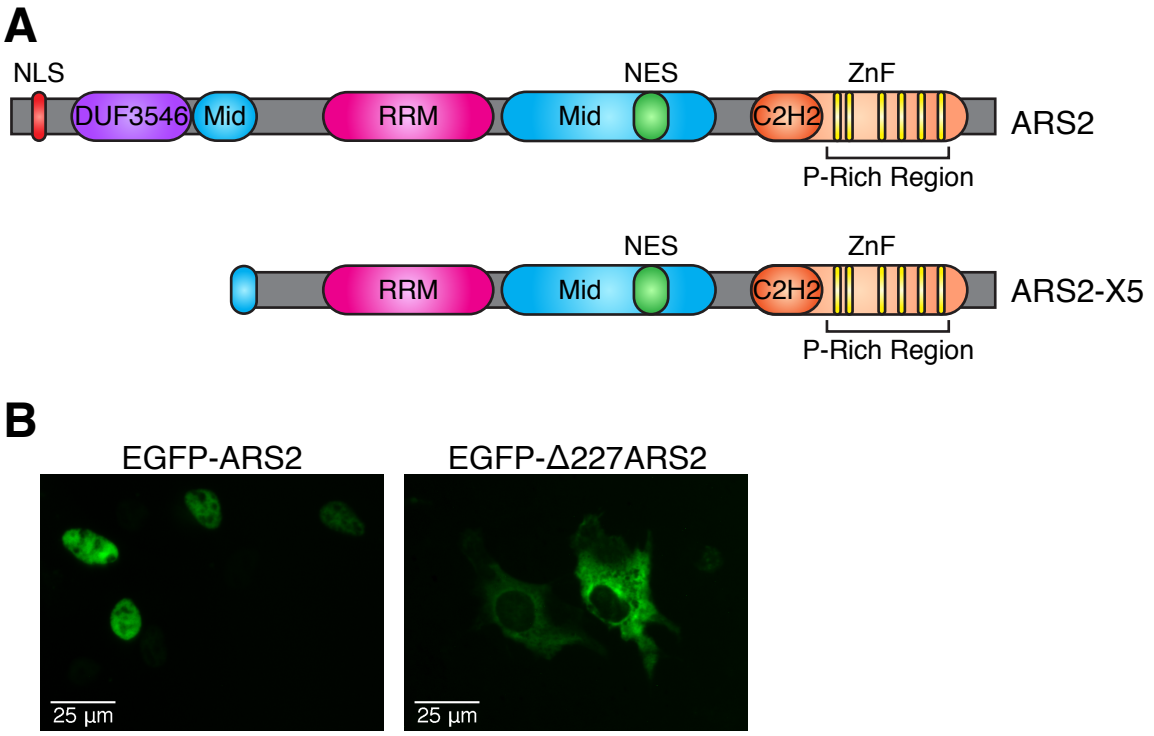

Figure 3-14 ARS2-X5. (A) The protein motif structure of ARS2 and the predictd isoform ARS2-X5. **(B)** C2C12 myoblasts transfected with EGFP-ARS2 or EGFP-Δ227-ARS2 and imaged live.

3.6 ARS2 stimulates a reduction in luciferase activity when tethered to the 3' UTR of a reporter mRNA

The discovery of a cytoplasmic ARS2 isoform raises the question – what does ARS2 do in the cytoplasm? All ARS2 functions described to date take place in the nucleus. The fact that ARS2 shows substantial cytoplasmic localization in C2C12 myoblasts and differentiated myotubes suggested that it also plays a role in the cytoplasm in these cells. A literature search for potential interaction partners of ARS2 that have a cytoplasmic function revealed that two separate high-throughput studies of protein-protein interactions identified an interaction between ARS2 and the RNA-binding protein RNPS1 [145,146]. The fact that RNPS1 is a nonsense-mediated decay protein (NMD), a predominantly cytoplasmic degradation pathway, suggested that ARS2 is involved in this

mRNA surveillance pathway. Our lab was the first to independently verify the interaction between ARS2 and RNPS1 in C2C12 lysates by immunoprecipitation (Fig. 3-15). To investigate ARS2's potential involvement in NMD, honours student Rachel Wong under my supervision, tethered full-length ARS2, using the λN/boxb tethering system, to the 3'UTR of an intronless luciferase reporter. In this assay the 22 amino acid RNA-binding domain of the bacteriophage λ antiterminator protein N (λN) is used to tag the protein of interest. The RNA binding site recognized by λN (boxb) is inserted into the 3' UTR of a reporter mRNA allowing the protein of interest to be tethered to the reporter. When the protein of interest is involved in NMD, tethering stimulates reporter mRNA degradation (Fig. 3-16). Tethering assays such as this are the conventional method for determining if a protein is involved in NMD. The human UPF proteins, major effectors of NMD, as well as RNPS1 have been shown to trigger NMD when tethered to the 3' UTR of a reporter mRNA [103,150]. An ARS2-λN fusion and a Firefly luciferase reporter containing 5 boxb sites in the 3' UTR were co-expressed in HeLa cells. A *Renilla* luciferase reporter without boxb sites was used to control for transfection efficiency. Luminescence was measured 24 hours after transfection and Firefly luciferase activity was normalized to *Renilla* luciferase activity. This analysis showed that tethering of ARS2 to the 3' UTR of a Firefly luciferase reporter significantly decreases the relative luminescence when compared to an empty vector control ($p \leq 0.0001$) (Fig. 3-17). As expected, the λN-RNPS1 positive control also showed a significant decrease in relative luminescence ($p \leq 0.05$). Expression of ARS2 without the λN peptide showed no significant difference compared to the empty vector control indicating that the loss of relative luminescence is dependent on tethering of ARS2 to the reporter mRNA (Fig. 3-17). Expression of ARS2-λN or λN-RNPS1 with a Firefly luciferase reporter without boxb sites produced no change in relative luminescence further indicating that the effect is tethering-dependent (Fig. 3-17). To ensure that the λN peptide has no independent effect on relative luminescence, HeLa cells were transfected with a plasmid encoding λN alone. This produced no significant changes compared to an empty vector control (Fig. 3-17). These results show that tethering ARS2 to the 3' UTR of a reporter transcript results in a loss of reporter protein activity. Further investigation is required to confirm that this is due to destabilization of

reporter mRNA through triggering NMD and whether ARS2-X5 is also involved in NMD.

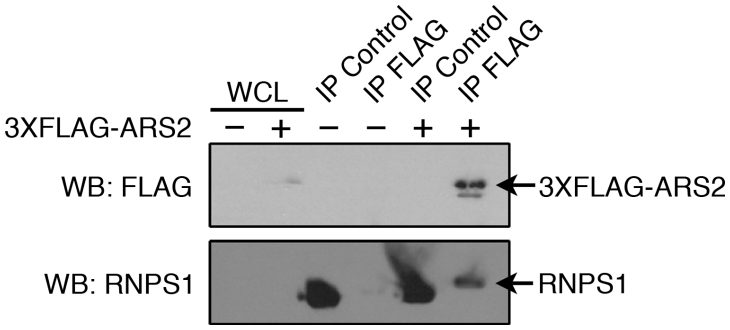

Figure 3-15 RNPS1 co-immunoprecipitates with 3XFLAG-ARS2. C2C12 myoblasts were transfected with 3XFLAG-ARS2 or GFP followed by immunoprecipitation with anti-FLAG beads and elution with FLAG peptide. Proteins were fractionated by SDS-PAGE and detected by western blot using the indicated antibodies. IP control was an isotype matched IgG.

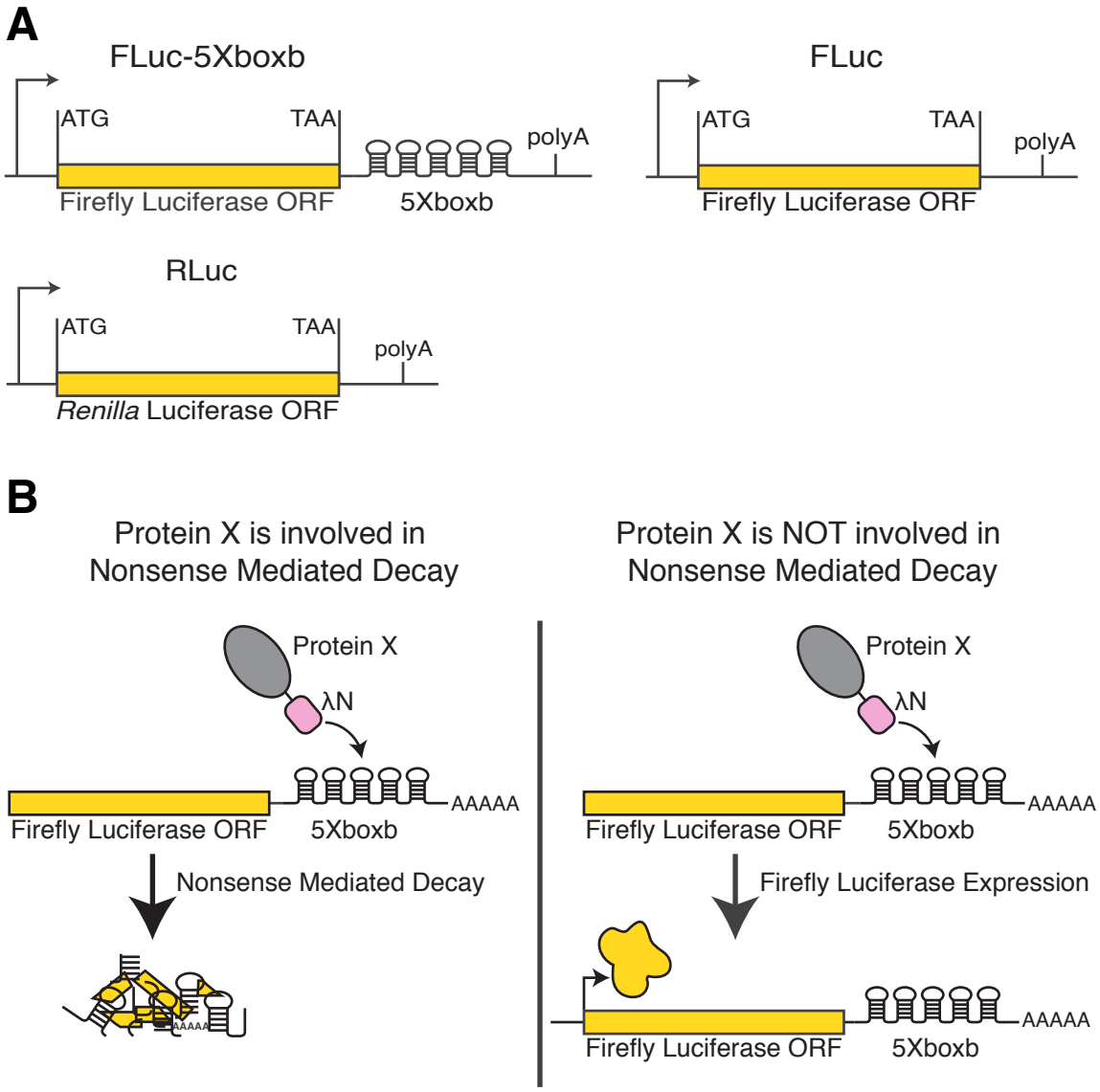

Figure 3-16 NMD tethering assay. (A) Schematic representation of reporter and control constructs used in the NMD luciferase assay. **(B)** λN/boxb-based tethering system. Protein X-λN fusions are tethered to boxb sites in the 3'UTR of a Firefly luciferase reporter. If Protein X is involved in NMD, tethering to the reporter results in a loss of Firefly luciferase activity.

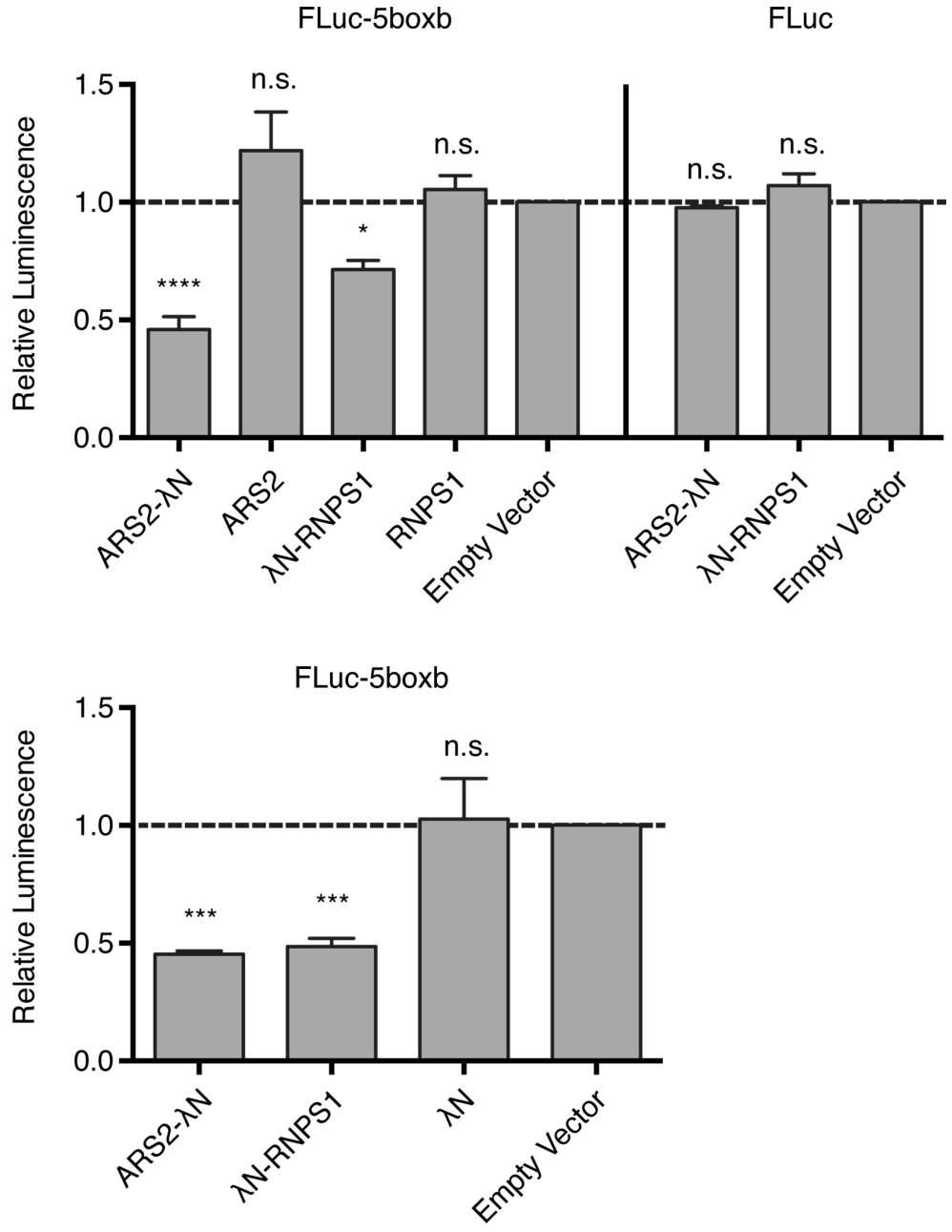

Figure 3-17 ARS2 triggers a decrease in luciferase activity when tethered to the 3' UTR of reporter mRNA. HeLa cells were transfected with the Firefly luciferase 5Xboxb reporter (FLuc-5Xboxb) or the Firefly luciferase control without boxb sites (FLuc), the *Renilla* luciferase control (RLuc) and the indicated construct. The luciferase activites were measured 24 h after transfection and Firefly luciferase activity was normalized to *Renilla* luciferase activity. Data presented is from a single experiement and shows the means of 3 technical replicates per condition. Error bars represent SD. Significant differences were identified by one-way ANOVA follwed by Tukey's multiple comparison post-test. Notations above bars indicate significance compared to empty vector control. * = $p \leq 0.05$, ** = $p \leq 0.01$, **** = $p \leq 0.0001$, n.s. = not significant.

CHAPTER 4: DISCUSSION AND FUTURE DIRECTIONS

4.1 Discussion

Recently several studies reporting on ARS2 function have emerged, adding to our knowledge of this previously uncharacterized protein. However, none of these studies included analysis of the subcellular localization of endogenous ARS2. The localization of proteins defines the environment in which they operate and as such can provide essential clues to their function. In previous reports, ARS2 localization was investigated using exogenously expressed, epitope-tagged ARS2 only. Immunofluorescence detection of V5-tagged ARS2 expressed in 3T3 MEFs and 3XFLAG-tagged ARS2 expressed in BOSC cells revealed that full-length ARS2 localizes predominantly to the nucleus [1,4]. To better understand the functions of ARS2 in myogenic cells I sought to determine the subcellular localization of endogenous ARS2 in proliferating myoblasts and differentiating myotubes. Using immunofluorescence detection, I observed abundant ARS2 expression in the nucleus and the cytoplasm of myogenic cells (Fig. 3-1 and 3-2). This differs from the previous reports in which only faint cytoplasmic staining was seen [1,4]. One possible explanation for the difference in localization is that the antibodies used do not specifically recognize ARS2. However, this is unlikely since we rigorously verified the specificity of one of the polyclonal ARS2 antibodies (XL12.2) used in this analysis through a combination of immunofluorescence detection and western blotting. Although the second polyclonal ARS2 antibody (XL14.1) was not independently verified, it produced an identical localization pattern to XL12.2 indicating that it specifically recognizes ARS2. Another possibility is that the localization of ARS2 is cell-type specific. Expression of full-length ARS2 in C2C12 cells showed strong nuclear signals and little to no cytoplasmic signals, which is consistent with the studies mentioned previously (Fig. 3-4) [1,4]. N-terminal and C-terminal EGPF-ARS2 fusion proteins also showed the same localization when visualized in living C2C12 cells, reducing the likelihood that cytoplasmic ARS2 was lost during fixation in the aforementioned experiments. These results indicate that the cytoplasmic staining I observed with anti-ARS2 antibodies is not due to cell-type specific, differential localization of full-length ARS2 in myogenic cells.

Another possibility is that the abundant signal in the cytoplasm is due to the presence of a cytoplasmic ARS2 isoform. Here I uncovered a C-terminal *Ars2* transcript expressed in C2C12 cells that has an open reading frame (ORF) coding for a predicted 72 kDa protein consisting of 641 amino acids, in contrast to the 100 kDa 875 amino acid protein coded for by the full-length *Ars2* cDNA. The molecular weight of this predicted isoform, designated ARS2-X5, is larger than the 60 kDa band that was observed on western blots of C2C12 lysates. Further study is required to determine if the 60 kDa band is the result of proteolytic degradation of full-length ARS2 or if it is the shortened isoform predicted here. ARS2-X5 retains most of the MID domain, the RNA recognition motif (RRM), the ZnF domain and the unstructured C-terminus (Fig. 3-13). The missing N-terminal region includes the domain of unknown function (DUF3546), the nuclear localization signal (NLS) and the unstructured N-terminus. Unsurprisingly, when this transcript was expressed as an EGFP fusion protein in C2C12 cells it was localized exclusively in the cytoplasm. Additional work is needed to confirm that *Ars2-X5* represents a true *Ars2* isoform in C2C12 cells. It is unlikely that we would be able to generate antibodies specific for ARS2-X5 since its entire amino acid sequence is included in the full-length protein. A better approach would be to use RNAi to specifically knockdown *Ars2-X5* expression while leaving full-length *Ars2* expression intact. Taken together my data shows that C2C12 cells express a truncated C-terminal *Ars2* transcript that contains an open reading frame coding for a predicted cytoplasmic ARS2 isoform.

Although ARS2 is capable of shuttling between the nucleus and the cytoplasm, to date all of its reported functions take place in the nucleus. Abundant ARS2 staining in the cytoplasm of myogenic cells combined with the discovery of a predicted cytoplasmic isoform raised the question: what does ARS2 do in the cytoplasm? I postulated that ARS2 could be involved in nonsense-mediated decay (NMD), an mRNA surveillance pathway that occurs in the cytoplasm. This was based on evidence that ARS2 interacts with RNPS1, a peripheral member of the exon-junction complex (EJC) that has been implicated in NMD and the known role of the cap-binding complex (CBC) in NMD [145,146,150]. My results suggest that ARS2 is capable of inducing NMD when tethered to the 3' UTR of a reporter transcript. This is consistent with previous studies that showed that tethering essential effectors of NMD, such as the UPF proteins, significantly

decreases transcript abundance by triggering NMD [103,163]. Our lab has since combined the tethering assay with knockdown of UPF1 and found that tethered ARS2 cannot induce a reduction in luciferase activity in UPF1-depleted cells (O'Sullivan, unpublished). This indicates that ARS2 requires UPF1 to reduce reporter levels and further suggests that ARS2 is involved in NMD. Additional work is needed to confirm that this result is due to UPF1-depletion and not off-target effects of the shRNA. This is the first time that ARS2 has been implicated in a cytoplasmic process. The existence of nuclear NMD has been suggested, however since NMD is translation-dependent and the current consensus is that there is no nuclear translation, such a pathway seems improbable[164-166].

My results show that ARS2 is expressed in post-mitotic, differentiating myotubes (Fig. 3-2 and A-3), which is inconsistent with previous work by Gruber *et al.* that concluded that ARS2 expression is limited to actively proliferating cells [4]. This conclusion was based on studies using an IL-3-dependent cell line derived from *Bax*$^{-/-}$ *Bak*$^{-/-}$ mice. Gruber *et al.* found that when these cells exited the cell cycle in response to IL-3 withdrawal there was depletion in ARS2 levels [4]. It is possible that the absence of ARS2 in post-mitotic cells is limited to this cell line or that downregulation of ARS2 expression is a response to loss of IL-3 signalling. The results reported here show that ARS2 is expressed in myogenic cells that have exited the cell cycle through terminal differentiation.

Our lab has previously shown that ARS2 is essential in the early stages of mouse embryonic development. *Ars2* knockout in mice is embryonic lethal and *Ars2*$^{-/-}$ embryos die around the time of implantation at ~ embryonic day 3.5-5.5[1]. This is consistent with *Ars2* deletion in *Drosophila* [126] and zebrafish [124,125], which also exhibit early embryonic-lethal phenotypes. *Ars2*$^{-/-}$ mouse embryos cultured *in vitro* do not hatch from the zona pellucida and fail to develop beyond the blastocyst stage. Analysis of lineage-specific markers showed that *Ars2*$^{-/-}$ embryos contain both trophectoderm-positive (Cdx2) and inner cell mass-positive (Oct4) cells, suggesting that loss of ARS2 does not impede early cell fate decisions[1]. Although, these results suggest that the early embryonic lethality seen in *Ars*$^{-/-}$ embryos is not caused by defects in early cellular differentiation

the early embryonic lethal phenotype left many questions unanswered about the role of ARS2 in later differentiation processes.

I have shown that specific levels of ARS2 are required for myogenic differentiation of C2C12 cells as ARS2 knockdown or overexpression resulted in differentiation defects (Fig. 3-7 and Fig. 3-8). To better understand what functions of ARS2 contribute to the differentiation phenotype I compared the phenotype of ARS2 knockdown to the phenotypes of DROSHA and FLASH knockdowns and looked for phenocopy. I found that the differentiation phenotype associated with ARS2 knockdown phenocopied DROSHA knockdown while FLASH depletion had no significant effect on myoblast differentiation (Fig. 3-7). In addition, uncoupling the ARS2/FLASH interaction through expression of FARB had no effect on differentiation suggesting that ARS2's role in myogenic differentiation is independent of its interaction with FLASH (Fig. 3-8). Altogether, these data suggest that the myogenic differentiation defect associated with disruption of ARS2 expression levels is unlikely to be a result of impaired replication-dependent histone (RDH) pre-mRNA processing and is probably caused in part by defective miRNA biogenesis. This is consistent with reports that miRNA play key roles in myogenesis [32-36]. Rescue experiments are necessary to definitively show that miRNA deficiency is responsible for this phenotype. However, this will be complicated by the fact that multiple miRNA have been implicated in myogenesis, and it is unlikely that a single miRNA will rescue the phenotype. Additionally, contributions from additional ARS2-dependent RNAPII transcript processing cannot be ruled out.

While several studies have reported on RDH pre-mRNA processing, I document here the first use of a flow cytometry-based fluorescent reporter assay for measuring RDH 3' cleavage efficiency. The results of this assay indicate that RDH pre-mRNA 3' cleavage in N2a cells is impaired following knockdown of ARS2, FLASH or LSM11, whereas the knockdown of DROSHA had no effect (Fig. 3-12). This was expected since LSM11 is an integral component of the U7 snRNP, FLASH's involvement in this processing pathway has been well documented and ARS2 has been implicated in RDH pre-mRNA processing in human cancer cell lines [2,5,80,82,155,159]. I also found that overexpression of ARS2 acts as a dominant negative in this assay, which is consistent with findings that ARS2 overexpression has a dominant negative effect on cell cycle

progression and myogenic differentiation of C2C12 mouse myoblasts (O'Sullivan; in submission and Fig. 3-8). Together these results indicate that specific levels of ARS2 are required for proper 3' cleavage of RDH pre-mRNA transcripts in N2a cells. The fact that FARB had a smaller effect on reporter cleavage than ARS2 overexpression suggests that ARS2's role in RDH processing is not mediated by its interaction with FLASH. This is further supported by the fact that there was no significant difference in reporter cleavage between FARB and a scrambled peptide control (SCR). This is surprising since Kiriyama *et al.* reported that disruption of the ARS2/FLASH interaction results in impaired S-phase progression and a decrease in total histone mRNA levels, implying that ARS2 and FLASH share a role in histone expression [143]. My results suggest that the ARS2/FLASH interaction is not required for RDH pre-mRNA 3' cleavage. A recent publication reported that ARS2 does not interact with FLASH and is not required for correct RDH pre-mRNA cleavage in *Drosophila* cells, suggesting that ARS2's role in this process is restricted to mammalian cells[167]. These data support the notion that ARS2 and FLASH play necessary but independent roles in RDH pre-mRNA 3' end processing. A more likely alternative theory is that ARS2 contributes to RDH pre-mRNA cleavage as an effector of the CBC. Consistent with this is a report from Hallais *et al.* in which depletion of ARS2 produced a similar reduction in RDH pre-mRNA reporter cleavage when compared to depletion of CBP20 or CBP80 [2]. Given that ARS2 has been shown to stimulate 3' end-processing of a number of different RNAs and has been implicated in additional RNA processing pathways, it appears that ARS2's primary role in the nucleus is as a major effector of the CBC. Therefore, it is probable that the loss of RDH mRNA processing and subsequent reduction in histone protein levels seen following ARS2 knockdown contributes to the accumulation of ARS2-depleted cells in S-phase and is not a consequence of impaired cell cycle progression.

In an effort to examine the nature of ARS2's influence on RDH expression more closely, I also developed a control that contained all the elements of the reporter with the exception of the RDH 3' cleavage signal (Fig. 3-11). The control therefore consisted of a partial H2A ORF followed immediately by a DsRed ORF all driven by a CMV promoter. Surprisingly, knockdown of ARS2, FLASH and LSM11 all resulted in increased expression of the control plasmid while knockdown of DROSHA had no significant

effect. It is interesting to note that the three proteins that had an effect on the control are all involved in RDH 3' processing. The effect is likely to be independent of the DsRed ORF as LSM11 knockdown had a similar effect on a CMV-H2A-luciferase control (data not shown). In addition, Hallais *et al.* saw little to no effect on a CMV-luciferase control following knockdown of ARS2, CBP20 or CBP80, suggesting that the effect of ARS2, FLASH and LSM11 knockdown is not dependent on the CMV promoter however this cannot be confirmed [2]. It seems that the increase in the H2A-control plasmid may be the result of some kind of regulation on the partial H2A coding sequence. Alternative explanations include the possibility that disruption of this pathway has a general effect on transcriptional activation or translation. Another possibility is that this is an artefact of decreased histone production. There is evidence that plasmid DNA undergoes chromatization *in vivo* suggesting that decreased histone expression could lead to a reduction in the amount of histone protein deposited on the control plasmid DNA leading to increased expression due to increased accessibility of the DNA[168]. When the reporter results were normalized to the control the differences between ARS2, FLASH or LSM11 knockdown and non-targeting control shRNA were no longer significant (Fig. 3-12). This could suggest that there was no effect on RDH 3' cleavage in response to the knockdowns or, the more likely possibility, that a general influence on expression swamped any specific disruption of RDH 3' cleavage. I decided to repeat the reporter assay using overexpression of ARS2 instead of knockdown as ARS2 overexpression produces a more severe cell cycle phenotype. ARS2 overexpression resulted in a significant increase in reporter expression and in control expression. However, when the reporter was normalized to the control the difference between EGFP-ARS2 expressing cells and EGFP expressing cells remained significant. This indicates that overexpression of ARS2 in N2a cells results in a decrease in RDH pre-mRNA processing.

4.2 Future Directions

4.2.1 ARS2 and myogenesis in vivo

My thesis work establishes that ARS2 is required for myogenic differentiation of the C2C12 cell line, however additional studies are required to determine if ARS2 is necessary for myogenic differentiation *in vivo*. To further characterize ARS2 expression

and function in skeletal muscle it will be advantageous to develop an inducible, skeletal muscle specific *Ars2* knockout mouse. This will allow us to bypass the early embryonic requirement of ARS2 and examine the effects of ARS2 loss on adult skeletal muscle *in vivo*. Embryonic stem (ES) cells with a floxed allele of *Ars2* can be obtained from the European Mouse Mutagenesis Project (EUCOMM) and used to generate mice that carry an $Ars2^{flox}$ allele. Subsequent breeding with $Ars2^{+/-}$ mice will allow us to establish mice with an $Ars2^{flox/-}$ genotype. Next, these mice can be crossbred with mice harboring a tamoxifen-inducible Cre recombinase transgene driven by a muscle-specific promoter. There are two such mouse strains available through The Jackson Laboratory. Both express Cre recombinase fused to a mutant estrogen receptor (Cre-ER) that cannot bind its natural ligand at physiological concentrations but retains high affinity for tamoxifen. The Cre-ER fusion protein can only enter the nucleus and catalyze recombination when it is bound to tamoxifen, allowing for inducible Cre activity. In one mouse strain Cre-ER is under the control of the endogenous *Myf5* promoter/enhancer elements and in the other mouse strain it is under the control of the *Pax7* promoter/enhancer elements. *Myf5* and *Pax7* are both expressed during muscle development and regeneration. PAX7 is a marker for satellite cells, the adult muscle stem cells that expand and differentiate in order to regenerate muscle after injury. When either of these mice is bred to $Ars2^{flox/-}$ mice, Cre-mediated recombination will result in deletion of the floxed sequence in cells of the myogenic lineage when tamoxifen is administered. These mice will allow us to examine the effect of muscle-specific ARS2 knockout during development and regeneration following muscle injury. In addition, primary myoblasts cultured from these mice can be used to individually analyze cell survival, proliferation, senescence and differentiation following loss of ARS2.

4.2.2 ARS2 and miRNA expression in myoblasts

The results of my differentiation assays suggest that ARS2 contributes to C2C12 differentiation through miRNA biogenesis. It would be interesting to know if depletion of ARS2 in myogenic cells is accompanied by a decrease in miRNA expression. I propose that we use commercially available miRNA microarrays to determine if ARS2-deficient myoblasts have an altered miRNA expression profile. Our lab previously performed microarray analysis using total RNA from C2C12 cells transfected with either *Ars2*-

targeting or control siRNA to probe a custom mouse cDNA microarray of genes important for self-renewal and maintenance. This analysis revealed that several genes including transcription factors and signaling molecules undergo significant changes in expression in ARS2-deficient cells (Howard, unpublished). It would be interesting to identify miRNA-mRNA target pairs that are inversely affected by a loss of ARS2. Such a result would indicate that ARS2 is required for miRNA biogenesis in myoblasts and further support the idea that the differentiation defect in ARS2-deficient myoblasts is due to a loss of miRNAs. To confirm this we could add back specific miRNAs in an attempt to rescue the differentiation phenotype.

4.2.3 Further validation of ARS2-X5

To further corroborate the existence of a shortened, cytoplasmic isoform of ARS2 I recommend that we next prepare nuclear and cytoplasmic fractions of C2C12 cells and analyze them for ARS2 expression by western blot. This experiment would allow us to confirm that a smaller ARS2 isoform is localized mainly in the cytoplasm of C2C12 cells and that, in contrast the full-length protein is found almost exclusively in the nuclear fractions. A clear separation between the two forms of ARS2 across the nuclear and cytoplasmic fractions would suggest that the smaller 60 kDa band is a true isoform of ARS2. To further validate that the *Ars2-X5* transcript identified here encodes the 60 kDa isoform observed on western blots, I suggest that we attempt to specifically knockdown expression of *Ars2-X5* in C2C12 cells by targeting its unique 5' UTR by RNAi. This would be followed by analysis by qPCR, western blotting and immunofluorescence. Loss of the 60 kDa band or cytoplasmic localization corresponding with a decrease in *Ars2-X5* transcript levels would indicate that *Ars2-X5* encodes a cytoplasmic, truncated isoform of ARS2.

Previous studies have identified several proteins that interact with ARS2 but there has been no high-throughput study focused on characterizing ARS2's entire network of interactions. In addition, it is unlikely that ARS2 and ARS2-X5 share all the same interacting partners due to the truncation in ARS2-X5 and their differential localization. It would be interesting to compare and contrast the interactomes of these two isoforms. To do this I propose that we use affinity purification followed by quantitative high-resolution mass spectrometry. Individual interactions could be independently verified by

subsequent co-immunoprecipitations coupled with western blotting and a yeast-2-hybrid (Y2H) approach. Next we would use targeted mutagenesis to determine which regions of ARS2 or ARS2-X5 are responsible for specific interactions. Connor O'Sullivan recently developed the first structural model of ARS2 based on the crystal structure of the ARS2 ortholog SERRATE. O'Sullivan has used this model to design mutations in ARS2 that target the function of each of ARS2's 4 domains. Using these mutants he has found that the unstructured C-terminus of ARS2 is required for binding to CBP20, the DUF3546 and ZnF domain are essential for interaction with miRNA and histone mRNA, the MID domain is necessary for binding to DROSHA and the ZnF domain is required for interaction with FLASH (O'Sullivan, in submission). I propose that we use the current mutants and generate similar mutants of ARS2-X5 to identify the regions responsible for any novel interactions that are identified.

4.2.4 Future NMD experiments

While my data suggests that ARS2 is involved in NMD, further experiments are required to confirm this. As mentioned above, recent work in our lab indicates that ARS2's ability to trigger a decrease in luciferase activity is dependent on UPF1. To ensure that the loss of ARS2-mediated reduction in luciferase activity is due to depletion of UPF1 and not off-target effects of the shRNA I suggest that we next attempt to rescue the effect of ARS2 on the reporter by expressing an RNAi resistant version of UPF1. We should also repeat the assay in the presence of a translation inhibitor such as cycloheximide. NMD is translation-dependent, thus blocking translation of the luciferase reporter should abolish the effect of tethered ARS2 if the effect is due to triggering of NMD. Furthermore, to verify that the decrease in luciferase activity following ARS2 tethering is due to degradation at the transcript level it would be prudent to repeat the assay using qPCR to quantify Firefly luciferase mRNA levels directly. Ideally, all of these experiments would be repeated with the ARS2-X5 isoform. Another interesting approach to studying ARS2's involvement in NMD would be to examine the effect of ARS2 knockdown on a number of *bona fide* physiological NMD targets. A recent publication has compiled a list of *bona fide* NMD targets in HeLa cells that were selected based on a minimum of two criteria: up-regulation following UPF1-depletion and longer

half-life following UPF1-depletion[169]. If ARS2 is essential for NMD loss of ARS2 should also result in up-regulation and increased stabilization of these transcripts.

REFERENCES

1. Wilson MD, Wang D, Wagner R, Breyssens H, Gertsenstein M, et al. (2008) ARS2 is a conserved eukaryotic gene essential for early mammalian development. Mol Cell Biol 28: 1503-1514.
2. Hallais M, Pontvianne F, Andersen PR, Clerici M, Lener D, 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.
3. Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, et al. (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nat Struct Mol Biol 20: 1367-1376.
4. Gruber JJ, Zatechka DS, Sabin LR, Yong J, Lum JJ, et al. (2009) Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. Cell 138: 328-339.
5. Gruber JJ, Olejniczak SH, Yong J, La Rocca G, Dreyfuss G, et al. (2012) Ars2 promotes proper replication-dependent histone mRNA 3' end formation. Mol Cell 45: 87-98.
6. Charge SB, Rudnicki MA. (2004) Cellular and molecular regulation of muscle regeneration. Physiol Rev 84: 209-238.
7. Mauro A. (1961) Satellite cell of skeletal muscle fibers. J Biophys Biochem Cytol 9: 493-495.
8. Gamble HJ, Fenton J, Allsopp G. (1978) Electron microscope observations on human fetal striated muscle. J Anat 126: 567-589.
9. Darr KC, Schultz E. (1987) Exercise-induced satellite cell activation in growing and mature skeletal muscle. J Appl Physiol (1985) 63: 1816-1821.
10. Schultz E, Gibson MC, Champion T. (1978) Satellite cells are mitotically quiescent in mature mouse muscle: An EM and radioautographic study. J Exp Zool 206: 451-456.
11. Snow MH. (1977) Myogenic cell formation in regenerating rat skeletal muscle injured by mincing. II. an autoradiographic study. Anat Rec 188: 201-217.
12. Snow MH. (1978) An autoradiographic study of satellite cell differentiation into regenerating myotubes following transplantation of muscles in young rats. Cell Tissue Res 186: 535-540.
13. Snow MH. (1983) A quantitative ultrastructural analysis of satellite cells in denervated fast and slow muscles of the mouse. Anat Rec 207: 593-604.
14. Davis RL, Weintraub H, Lassar AB. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51: 987-1000.
15. Braun T, Buschhausen-Denker G, Bober E, Tannich E, Arnold HH. (1989) A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. EMBO J 8: 701-709.
16. Edmondson DG, Olson EN. (1989) A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes Dev 3: 628-640.
17. Rhodes SJ, Konieczny SF. (1989) Identification of MRF4: A new member of the muscle regulatory factor gene family. Genes Dev 3: 2050-2061.

18. Braun T, Bober E, Winter B, Rosenthal N, Arnold HH. (1990) Myf-6, a new member of the human gene family of myogenic determination factors: Evidence for a gene cluster on chromosome 12. EMBO J 9: 821-831.
19. Miner JH, Wold B. (1990) Herculin, a fourth member of the MyoD family of myogenic regulatory genes. Proc Natl Acad Sci U S A 87: 1089-1093.
20. Buckingham M, Rigby PW. (2014) Gene regulatory networks and transcriptional mechanisms that control myogenesis. Dev Cell 28: 225-238.
21. Olguin HC, Olwin BB. (2004) Pax-7 up-regulation inhibits myogenesis and cell cycle progression in satellite cells: A potential mechanism for self-renewal. Dev Biol 275: 375-388.
22. Seale P, Sabourin LA, Girgis-Gabardo A, Mansouri A, Gruss P, et al. (2000) Pax7 is required for the specification of myogenic satellite cells. Cell 102: 777-786.
23. Zammit PS, Relaix F, Nagata Y, Ruiz AP, Collins CA, et al. (2006) Pax7 and myogenic progression in skeletal muscle satellite cells. J Cell Sci 119: 1824-1832.
24. Black BL, Olson EN. (1998) Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. Annu Rev Cell Dev Biol 14: 167-196.
25. Naya FJ, Olson E. (1999) MEF2: A transcriptional target for signaling pathways controlling skeletal muscle growth and differentiation. Curr Opin Cell Biol 11: 683-688.
26. Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, et al. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 38: 228-233.
27. Zhao Y, Samal E, Srivastava D. (2005) Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature 436: 214-220.
28. Ge Y, Chen J. (2011) MicroRNAs in skeletal myogenesis. Cell Cycle 10: 441-448.
29. Baskerville S, Bartel DP. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11: 241-247.
30. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, et al. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol 5: R13.
31. O'Rourke JR, Georges SA, Seay HR, Tapscott SJ, McManus MT, et al. (2007) Essential role for dicer during skeletal muscle development. Dev Biol 311: 359-368.
32. Goljanek-Whysall K, Sweetman D, Abu-Elmagd M, Chapnik E, Dalmay T, et al. (2011) MicroRNA regulation of the paired-box transcription factor Pax3 confers robustness to developmental timing of myogenesis. Proc Natl Acad Sci U S A 108: 11936-11941.
33. Dey BK, Gagan J, Dutta A. (2011) miR-206 and -486 induce myoblast differentiation by downregulating Pax7. Mol Cell Biol 31: 203-214.
34. Chen JF, Tao Y, Li J, Deng Z, Yan Z, et al. (2010) microRNA-1 and microRNA-206 regulate skeletal muscle satellite cell proliferation and differentiation by repressing Pax7. J Cell Biol 190: 867-879.
35. Hirai H, Verma M, Watanabe S, Tastad C, Asakura Y, et al. (2010) MyoD regulates apoptosis of myoblasts through microRNA-mediated down-regulation of Pax3. J Cell Biol 191: 347-365.

36. Crist CG, Montarras D, Pallafacchina G, Rocancourt D, Cumano A, et al. (2009) Muscle stem cell behavior is modified by microRNA-27 regulation of Pax3 expression. Proc Natl Acad Sci U S A 106: 13383-13387.
37. Lu J, McKinsey TA, Zhang CL, Olson EN. (2000) Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases. Mol Cell 6: 233-244.
38. Greco S, De Simone M, Colussi C, Zaccagnini G, Fasanaro P, et al. (2009) Common micro-RNA signature in skeletal muscle damage and regeneration induced by duchenne muscular dystrophy and acute ischemia. FASEB J 23: 3335-3346.
39. Sun Y, Ge Y, Drnevich J, Zhao Y, Band M, et al. (2010) Mammalian target of rapamycin regulates miRNA-1 and follistatin in skeletal myogenesis. J Cell Biol 189: 1157-1169.
40. Wang H, Garzon R, Sun H, Ladner KJ, Singh R, et al. (2008) NF-kappaB-YY1-miR-29 regulatory circuitry in skeletal myogenesis and rhabdomyosarcoma. Cancer Cell 14: 369-381.
41. Yuasa K, Hagiwara Y, Ando M, Nakamura A, Takeda S, et al. (2008) MicroRNA-206 is highly expressed in newly formed muscle fibers: Implications regarding potential for muscle regeneration and maturation in muscular dystrophy. Cell Struct Funct 33: 163-169.
42. Wong CF, Tellam RL. (2008) MicroRNA-26a targets the histone methyltransferase enhancer of zeste homolog 2 during myogenesis. J Biol Chem 283: 9836-9843.
43. Winbanks CE, Wang B, Beyer C, Koh P, White L, et al. (2011) TGF-beta regulates miR-206 and miR-29 to control myogenic differentiation through regulation of HDAC4. J Biol Chem 286: 13805-13814.
44. Shenoy A, Blelloch RH. (2014) Regulation of microRNA function in somatic stem cell proliferation and differentiation. Nat Rev Mol Cell Biol 15: 565-576.
45. Salditt-Georgieff M, Harpold M, Chen-Kiang S, Darnell JE,Jr. (1980) The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. Cell 19: 69-78.
46. Rasmussen EB, Lis JT. (1993) In vivo transcriptional pausing and cap formation on three drosophila heat shock genes. Proc Natl Acad Sci U S A 90: 7923-7927.
47. Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, et al. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell 78: 657-668.
48. Calero G, Wilson KF, Ly T, Rios-Steiner JL, Clardy JC, et al. (2002) Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. Nat Struct Biol 9: 912-917.
49. Gonatopoulos-Pournatzis T, Cowling VH. (2014) Cap-binding complex (CBC). Biochem J 457: 231-242.
50. Flaherty SM, Fortes P, Izaurralde E, Mattaj IW, Gilmartin GM. (1997) Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc Natl Acad Sci U S A 94: 11893-11898.
51. Izaurralde E, Lewis J, Gamberi C, Jarmolowski A, McGuigan C, et al. (1995) A cap-binding protein complex mediating U snRNA export. Nature 376: 709-712.

52. Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW. (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell 101: 187-198.
53. Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, et al. (2006) Human mRNA export machinery recruited to the 5' end of mRNA. Cell 127: 1389-1400.
54. Laubinger S, Sachsenberg T, Zeller G, Busch W, Lohmann JU, et al. (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in arabidopsis thaliana. Proc Natl Acad Sci U S A 105: 8795-8800. 10.1073/pnas.0802493105.
55. Sabin LR, Zhou R, Gruber JJ, Lukinova N, Bambina S, et al. (2009) Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in drosophila. Cell 138: 340-351.
56. Hosoda N, Kim YK, Lejeune F, Maquat LE. (2005) CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. Nat Struct Mol Biol 12: 893-901.
57. Hwang J, Sato H, Tang Y, Matsuda D, Maquat LE. (2010) UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. Mol Cell 39: 396-409.
58. Carthew RW, Sontheimer EJ. (2009) Origins and mechanisms of miRNAs and siRNAs. Cell 136: 642-655.
59. Lee Y, Ahn C, Han J, Choi H, Kim J, et al. (2003) The nuclear RNase III drosha initiates microRNA processing. Nature 425: 415-419.
60. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. (2004) Processing of primary microRNAs by the microprocessor complex. Nature 432: 231-235.
61. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in drosophila. Cell 130: 89-100.
62. Ruby JG, Jan CH, Bartel DP. (2007) Intronic microRNA precursors that bypass drosha processing. Nature 448: 83-86.
63. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. (2004) Nuclear export of microRNA precursors. Science 303: 95-98.
64. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, et al. (2001) Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell 106: 23-34.
65. Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, et al. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev 15: 2654-2659.
66. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, et al. (2001) A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. Science 293: 834-838.
67. Vermeulen A, Behlen L, Reynolds A, Wolfson A, Marshall WS, et al. (2005) The contributions of dsRNA structure to dicer specificity and efficiency. RNA 11: 674-682.
68. Kim VN. (2005) MicroRNA biogenesis: Coordinated cropping and dicing. Nat Rev Mol Cell Biol 6: 376-385.
69. Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. (2002) The human and mouse replication-dependent histone genes. Genomics 80: 487-498.

70. Marzluff WF, Wagner EJ, Duronio RJ. (2008) Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. Nat Rev Genet 9: 843-854.
71. Zhao J, Kennedy BK, Lawrence BD, Barbie DA, Matera AG, et al. (2000) NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. Genes Dev 14: 2283-2297.
72. Nizami Z, Deryusheva S, Gall JG. (2010) The cajal body and histone locus body. Cold Spring Harb Perspect Biol 2: a000653.
73. Wang ZF, Whitfield ML, Ingledue TC,3rd, Dominski Z, Marzluff WF. (1996) The protein that binds the 3' end of histone mRNA: A novel RNA-binding protein required for histone pre-mRNA processing. Genes Dev 10: 3028-3040.
74. Bond UM, Yario TA, Steitz JA. (1991) Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro. Genes Dev 5: 1709-1722.
75. Kolev NG, Yario TA, Benson E, Steitz JA. (2008) Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. EMBO Rep 9: 1013-1018.
76. Sullivan KD, Steiniger M, Marzluff WF. (2009) A core complex of CPSF73, CPSF100, and symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. Mol Cell 34: 322-332.
77. Davila Lopez M, Samuelsson T. (2008) Early evolution of histone mRNA 3' end processing. RNA 14: 1-10.
78. Tan D, Marzluff WF, Dominski Z, Tong L. (2013) Structure of histone mRNA stem-loop, human stem-loop binding protein, and 3'hExo ternary complex. Science 339: 318-321.
79. Pillai RS, Will CL, Luhrmann R, Schumperli D, Muller B. (2001) Purified U7 snRNPs lack the sm proteins D1 and D2 but contain Lsm10, a new 14 kDa sm D1-like protein. EMBO J 20: 5470-5479.
80. Pillai RS, Grimmler M, Meister G, Will CL, Luhrmann R, et al. (2003) Unique sm core structure of U7 snRNPs: Assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev 17: 2321-2333.
81. Dominski Z, Erkmann JA, Yang X, Sanchez R, Marzluff WF. (2002) A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. Genes Dev 16: 58-71.
82. Yang XC, Sabath I, Debski J, Kaus-Drobek M, Dadlez M, et al. (2013) A complex containing the CPSF73 endonuclease and other polyadenylation factors associates with U7 snRNP and is recruited to histone pre-mRNA for 3'-end processing. Mol Cell Biol 33: 28-37.
83. Dominski Z, Yang XC, Marzluff WF. (2005) The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell 123: 37-48.
84. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, et al. (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature 444: 953-956.
85. Schoenberg DR, Maquat LE. (2012) Regulation of cytoplasmic mRNA decay. Nat Rev Genet 13: 246-259.

86. Lelivelt MJ, Culbertson MR. (1999) Yeast upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. Mol Cell Biol 19: 6710-6719.
87. He F, Li X, Spatrick P, Casillo R, Dong S, et al. (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. Mol Cell 12: 1439-1452.
88. Rehwinkel J, Letunic I, Raes J, Bork P, Izaurralde E. (2005) Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. RNA 11: 1530-1544.
89. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 36: 1073-1078.
90. Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, et al. (2005) Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol Cell 20: 65-75.
91. Wittmann J, Hol EM, Jack HM. (2006) hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. Mol Cell Biol 26: 1272-1287.
92. Johansson MJ, He F, Spatrick P, Li C, Jacobson A. (2007) Association of yeast Upf1p with direct substrates of the NMD pathway. Proc Natl Acad Sci U S A 104: 20872-20877.
93. Gardner LB. (2008) Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. Mol Cell Biol 28: 3729-3741.
94. Gaba A, Jacobson A, Sachs MS. (2005) Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. Mol Cell 20: 449-460.
95. Colak D, Ji SJ, Porse BT, Jaffrey SR. (2013) Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. Cell 153: 1252-1265.
96. Dahlseid JN, Lew-Smith J, Lelivelt MJ, Enomoto S, Ford A, et al. (2003) mRNAs encoding telomerase components and regulators are controlled by UPF genes in saccharomyces cerevisiae. Eukaryot Cell 2: 134-142.
97. Kebaara B, Nazarenus T, Taylor R, Atkin AL. (2003) Genetic background affects relative nonsense mRNA accumulation in wild-type and upf mutant yeast strains. Curr Genet 43: 171-177.
98. Viegas MH, Gehring NH, Breit S, Hentze MW, Kulozik AE. (2007) The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the nonsense mediated decay pathway. Nucleic Acids Res 35: 4542-4551. 10.1093/nar/gkm461.
99. Kervestin S, Jacobson A. (2012) NMD: A multifaceted response to premature translational termination. Nat Rev Mol Cell Biol 13: 700-712.
100. He F, Brown AH, Jacobson A. (1997) Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. Mol Cell Biol 17: 1580-1594.
101. Cui Y, Hagan KW, Zhang S, Peltz SW. (1995) Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes Dev 9: 423-436.

102. He F, Jacobson A. (1995) Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. Genes Dev 9: 437-454.
103. Lykke-Andersen J, Shu MD, Steitz JA. (2000) Human upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell 103: 1121-1131.
104. Gonzalez CI, Ruiz-Echevarria MJ, Vasudevan S, Henry MF, Peltz SW. (2000) The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. Mol Cell 5: 489-499.
105. Gonzalez CI, Wang W, Peltz SW. (2001) Nonsense-mediated mRNA decay in saccharomyces cerevisiae: A quality control mechanism that degrades transcripts harboring premature termination codons. Cold Spring Harb Symp Quant Biol 66: 321-328.
106. Le Hir H, Izaurralde E, Maquat LE, Moore MJ. (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J 19: 6860-6869.
107. Kim VN, Kataoka N, Dreyfuss G. (2001) Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. Science 293: 1832-1836.
108. Dostie J, Dreyfuss G. (2002) Translation is required to remove Y14 from mRNAs in the cytoplasm. Curr Biol 12: 1060-1067.
109. Palacios IM, Gatfield D, St Johnston D, Izaurralde E. (2004) An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. Nature 427: 753-757.
110. Tange TO, Nott A, Moore MJ. (2004) The ever-increasing complexities of the exon junction complex. Curr Opin Cell Biol 16: 279-284.
111. Ivanov PV, Gehring NH, Kunz JB, Hentze MW, Kulozik AE. (2008) Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. EMBO J 27: 736-747.
112. Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, et al. (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev 20: 355-367.
113. Yamashita A, Izumi N, Kashima I, Ohnishi T, Saari B, et al. (2009) SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. Genes Dev 23: 1091-1105.
114. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J 20: 4987-4997.
115. Buchwald G, Ebert J, Basquin C, Sauliere J, Jayachandran U, et al. (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. Proc Natl Acad Sci U S A 107: 10050-10055.
116. Chamieh H, Ballut L, Bonneau F, Le Hir H. (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol 15: 85-93.

117. Okada-Katsuhata Y, Yamashita A, Kutsuzawa K, Izumi N, Hirahara F, et al. (2012) N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. Nucleic Acids Res 40: 1251-1266.
118. Huntzinger E, Kashima I, Fauser M, Sauliere J, Izaurralde E. (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA 14: 2609-2617.
119. Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH. (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol 16: 49-55.
120. Unterholzner L, Izaurralde E. (2004) SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. Mol Cell 16: 587-596.
121. Fukuhara N, Ebert J, Unterholzner L, Lindner D, Izaurralde E, et al. (2005) SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol Cell 17: 537-547.
122. Machida S, Chen HY, Adam Yuan Y. (2011) Molecular insights into miRNA processing by arabidopsis thaliana SERRATE. Nucleic Acids Res 39: 7828-7836.
123. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12: 996-1006.
124. Amsterdam A, Nissen RM, Sun Z, Swindell EC, Farrington S, et al. (2004) Identification of 315 genes essential for early zebrafish development. Proc Natl Acad Sci U S A 101: 12792-12797.
125. Golling G, Amsterdam A, Sun Z, Antonelli M, Maldonado E, et al. (2002) Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. Nat Genet 31: 135-140.
126. Oh SW, Kingsley T, Shin HH, Zheng Z, Chen HW, et al. (2003) A P-element insertion screen identified mutations in 455 novel essential genes in drosophila. Genetics 163: 195-201.
127. Andreu-Agullo C, Maurin T, Thompson CB, Lai EC. (2011) Ars2 maintains neural stem-cell identity through direct transcriptional activation of Sox2. Nature 481: 195-198.
128. Grigg SP, Canales C, Hay A, Tsiantis M. (2005) SERRATE coordinates shoot meristem function and leaf axial patterning in arabidopsis. Nature 437: 1022-1026. 10.1038/nature04052.
129. Prigge MJ, Wagner DR. (2001) The arabidopsis serrate gene encodes a zinc-finger protein required for normal shoot development. Plant Cell 13: 1263-1279.
130. Dong Z, Han MH, Fedoroff N. (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc Natl Acad Sci U S A 105: 9970-9975. 10.1073/pnas.0803356105.
131. Raczynska KD, Stepien A, Kierzkowski D, Kalak M, Bajczyk M, et al. (2014) The SERRATE protein is involved in alternative splicing in arabidopsis thaliana. Nucleic Acids Res 42: 1224-1244.
132. Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, et al. (2011) Interaction profiling identifies the human nuclear exosome targeting complex. Mol Cell 43: 624-637.

133. Chi B, Wang K, Du Y, Gui B, Chang X, et al. (2014) A sub-element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. Nucleic Acids Res 42: 7305-7318.
134. Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, et al. (2000) The protein aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature 407: 401-405.
135. Masuda S, Das R, Cheng H, Hurt E, Dorman N, et al. (2005) Recruitment of the human TREX complex to mRNA during splicing. Genes Dev 19: 1512-1517.
136. Dias AP, Dufu K, Lei H, Reed R. (2010) A role for TREX components in the release of spliced mRNA from nuclear speckle domains. Nat Commun 1: 97.
137. Mourao A, Varrot A, Mackereth CD, Cusack S, Sattler M. (2010) Structure and RNA recognition by the snRNA and snoRNA transport factor PHAX. RNA 16: 1205-1216.
138. Boulon S, Verheggen C, Jady BE, Girard C, Pescia C, et al. (2004) PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. Mol Cell 16: 777-787.
139. Rappsilber J, Ryder U, Lamond AI, Mann M. (2002) Large-scale proteomic analysis of the human spliceosome. Genome Res 12: 1231-1245. 10.1101/gr.473902.
140. Zhou Z, Licklider LJ, Gygi SP, Reed R. (2002) Comprehensive proteomic analysis of the human spliceosome. Nature 419: 182-185. 10.1038/nature01031.
141. Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev 10: 1683-1698.
142. Pabis M, Neufeld N, Steiner MC, Bojic T, Shav-Tal Y, et al. (2013) The nuclear cap-binding complex interacts with the U4/U6.U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. RNA 19: 1054-1063.
143. Kiriyama M, Kobayashi Y, Saito M, Ishikawa F, Yonehara S. (2009) Interaction of FLASH with arsenite resistance protein 2 is involved in cell cycle progression at S phase. Mol Cell Biol 29: 4729-4741.
144. Narita T, Yung TM, Yamamoto J, Tsuboi Y, Tanabe H, et al. (2007) NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. Mol Cell 26: 349-365.
145. Ewing RM, Chu P, Elisma F, Li H, Taylor P, et al. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol 3: 89.
146. Tange TO, Shibuya T, Jurica MS, Moore MJ. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. RNA 11: 1869-1883.
147. Badolato J, Gardiner E, Morrison N, Eisman J. (1995) Identification and characterisation of a novel human RNA-binding protein. Gene 166: 323-327.
148. Mayeda A, Badolato J, Kobayashi R, Zhang MQ, Gardiner EM, et al. (1999) Purification and characterization of human RNPS1: A general activator of pre-mRNA splicing. EMBO J 18: 4560-4570.
149. Sakashita E, Tatsumi S, Werner D, Endo H, Mayeda A. (2004) Human RNPS1 and its associated factors: A versatile alternative pre-mRNA splicing regulator in vivo. Mol Cell Biol 24: 1174-1187.

150. Lykke-Andersen J, Shu MD, Steitz JA. (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. Science 293: 1836-1839.
151. Imai Y, Kimura T, Murakami A, Yajima N, Sakamaki K, et al. (1999) The CED-4-homologous protein FLASH is involved in fas-mediated activation of caspase-8 during apoptosis. Nature 398: 777-785.
152. Chen S, Evans HG, Evans DR. (2012) FLASH knockdown sensitizes cells to fas-mediated apoptosis via down-regulation of the anti-apoptotic proteins, MCL-1 and cflip short. PLoS One 7: e32971.
153. Koonin EV, Aravind L, Hofmann K, Tschopp J, Dixit VM. (1999) Apoptosis. searching for FLASH domains. Nature 401: 662; discussion 662-3.
154. Barcaroli D, Bongiorno-Borbone L, Terrinoni A, Hofmann TG, Rossi M, et al. (2006) FLASH is required for histone transcription and S-phase progression. Proc Natl Acad Sci U S A 103: 14808-14812.
155. Yang XC, Burch BD, Yan Y, Marzluff WF, 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.
156. Maniatis T, Fritsch EF, Sambrook J. (1982) Molecular cloning. A laboratory manual. New York: Cold Spring Harbor Laboratory Press.
157. Laemmli UK. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.
158. Huang X, Madan A. (1999) CAP3: A DNA sequence assembly program. Genome Res 9: 868-877.
159. Yang XC, Xu B, Sabath I, Kunduru L, Burch BD, et al. (2011) FLASH is required for the endonucleolytic cleavage of histone pre-mRNAs but is dispensable for the 5' exonucleolytic degradation of the downstream cleavage product. Mol Cell Biol 31: 1492-1502. 10.1128/MCB.00979-10.
160. Barcaroli D, Dinsdale D, Neale MH, Bongiorno-Borbone L, Ranalli M, et al. (2006) FLASH is an essential component of cajal bodies. Proc Natl Acad Sci U S A 103: 14802-14807.
161. Bongiorno-Borbone L, De Cola A, Vernole P, Finos L, Barcaroli D, et al. (2008) FLASH and NPAT positive but not coilin positive cajal bodies correlate with cell ploidy. Cell Cycle 7: 2357-2367.
162. White AE, Burch BD, Yang XC, Gasdaska PY, Dominski Z, et al. (2011) Drosophila histone locus bodies form by hierarchical recruitment of components. J Cell Biol 193: 677-694.
163. Lykke-Andersen J. (2001) mRNA quality control: Marking the message for life or death. Curr Biol 11: R88-91.
164. Dahlberg J, Lund E. (2012) Nuclear translation or nuclear peptidyl transferase? Nucleus 3: 320-321.
165. Buhler M, Wilkinson MF, Muhlemann O. (2002) Intranuclear degradation of nonsense codon-containing mRNA. EMBO Rep 3: 646-651.
166. Wilkinson MF, Shyu AB. (2002) RNA surveillance by nuclear scanning? Nat Cell Biol 4: E144-7.

167. Sabath I, Skrajna A, Yang XC, Dadlez M, Marzluff WF, et al. (2013) 3'-end processing of histone pre-mRNAs in drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. RNA 19: 1726-1744.
168. Riu E, Chen ZY, Xu H, He CY, Kay MA. (2007) Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo. Mol Ther 15: 1348-1355.
169. Kurosaki T, Li W, Hoque M, Popp MW, Ermolenko DN, et al. (2014) A post-translational regulatory switch on UPF1 controls targeted mRNA degradation. Genes Dev 28: 1900-1916.

APPENDIX A: ADDITIONAL FIGURES

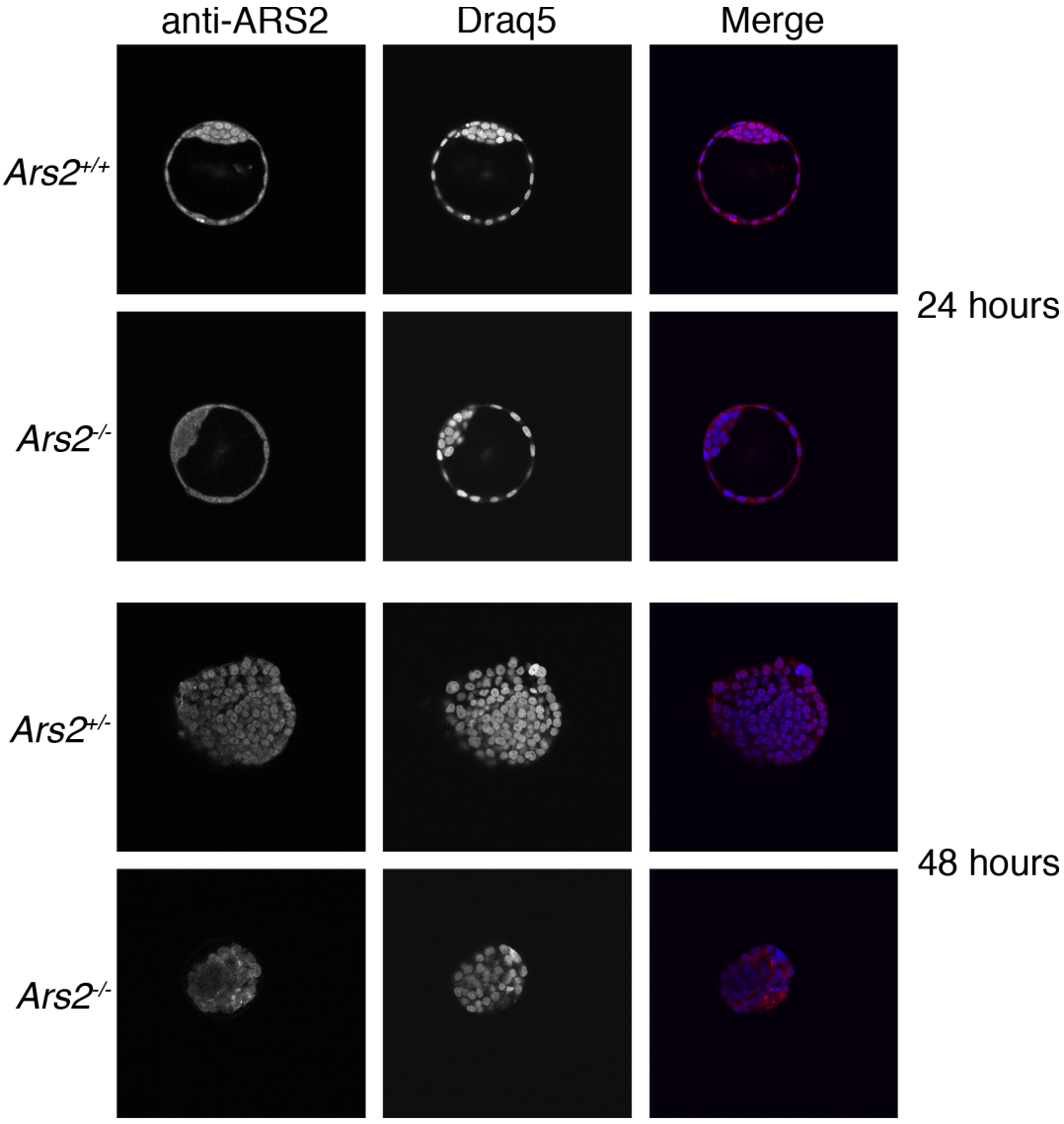

Figure A-1 ARS2 distribution in E3.5 embryos after 24 and 48 hours of culture. The nuclear localization of ARS2 was detected with anti-ARS2 monoclonal antibody (LX186.3) and Alexa Fluor 568 goat anti-rabbit IgG (red) in the wild type and was present in smaller amounts or even absent in null embryos. Nuclear DNA was stained using DRAQ5 (blue). Embryos were observed by confocal microscopy, and a representative single optical section is shown. After 24 hours of culture ARS2$^{-/-}$ embryos had completely lost their nuclear ARS2 localization. After 48 hours of culture, the null blastocyst had collapsed and several abnormal nuclei could be observed.

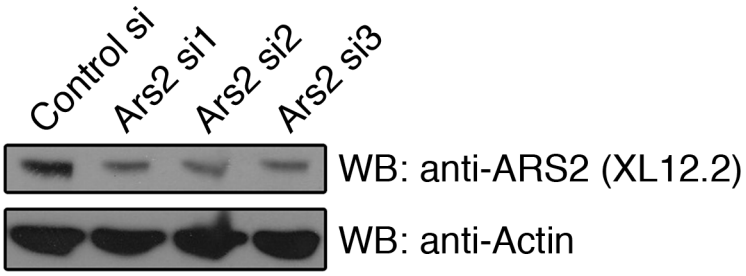

Figure A-2 ARS2 knockdown in C2C12 cells. Knockdown of endogenous ARS2 in C2C12 myoblasts detected with anti-ARS2 polyclonal antibody (XL12.2).

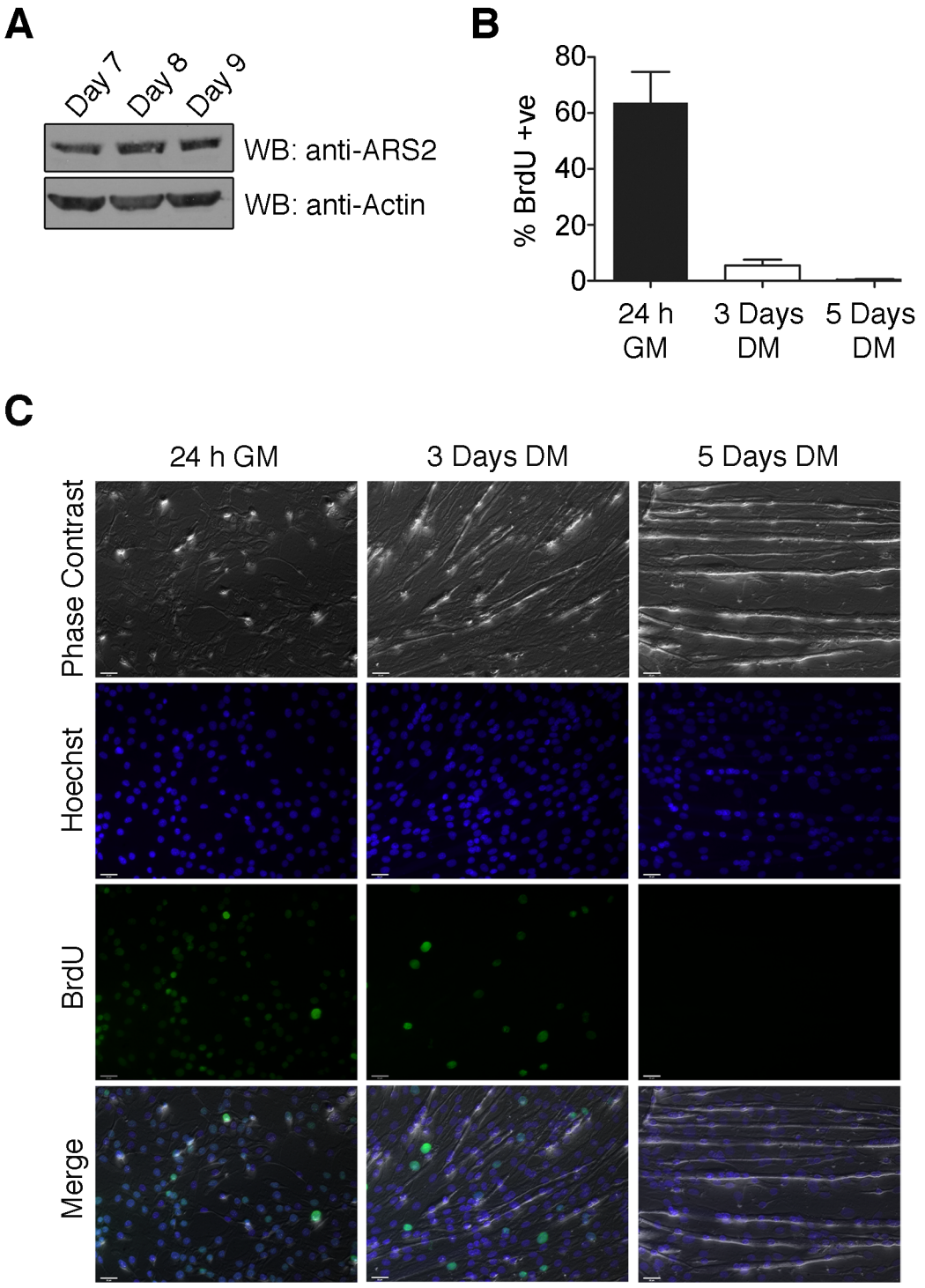

Figure A-3 ARS2 is expressed in post-mitotic C2C12 myotubes. (A) Western blot detection of ARS2 in C2C12 lysates after 7, 8 and 9 days of differentiation. **(B)** C2C12 cells exit the cell cycle after 5 days in differentiation medium (DM) as evidenced by a reduction in BrdU incorporation. **(C)** Representative images of data quantified in (B).

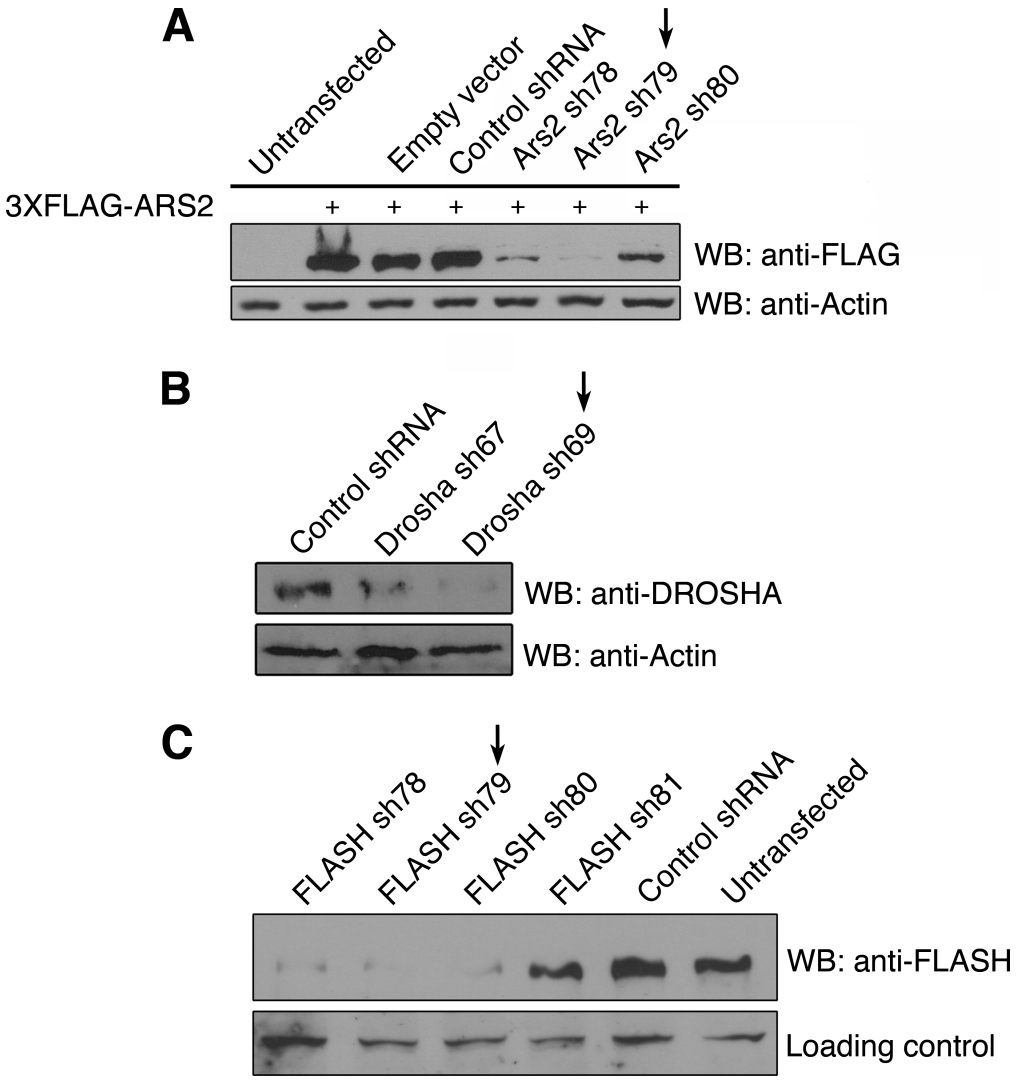

Figure 4-4 Validation of knockdowns. Western blots of C2C12 myoblast lysates showing shRNA knockdown of **(A)** transiently expressed 3XFLAG-ARS2 **(B)** DROSHA and **(C)** FLASH. Arrows indicate shRNA used in this thesis.

APPENDIX B: KNOCKDOWN DIFFERENTIATION RAW DATA

Ars2 shRNA	Total Nuclei in MHC+ Cells	Total Nuclei	Diff Index	Nuclei in Myotubes	Myotubes	Fusion Index
1-1	130	998	0.130	112	29	3.86
1-2	26	1335	0.019	16	16	1.00
1-3	6	1153	0.005	1	6	0.17
1-4	116	1051	0.110	100	14	7.14
1-5	124	1096	0.113	108	26	4.15
1-6	264	1063	0.248	257	47	5.47
2-1	27	1172	0.023	20	19	1.05
2-2	152	1266	0.120	101	26	3.88
2-3	96	1111	0.086	81	38	2.13
2-4	16	1120	0.014	16	9	1.78
2-5	134	1073	0.125	111	18	6.17
2-6	141	1278	0.110	102	24	4.25
3-1	239	1163	0.206	230	47	4.89
3-2	127	1187	0.107	111	17	6.53
3-3	55	1303	0.042	47	18	2.61
3-4	89	1223	0.073	71	28	2.54
3-5	94	1191	0.079	70	13	5.38
3-6	182	1455	0.125	169	34	4.97
4-1	127	1123	0.113	96	6	16.00
4-2	71	1059	0.067	55	12	4.58
4-3	129	1301	0.099	116	42	2.76
4-4	215	1277	0.168	180	26	6.92
4-5	200	1277	0.157	178	20	8.90
4-6	74	1124	0.066	62	7	8.86
5-1	112	961	0.117	100	14	7.14
5-2	300	1153	0.260	287	33	8.70
5-3	108	1269	0.085	90	16	5.63
5-4	142	1315	0.108	123	19	6.47
5-5	231	1391	0.166	218	29	7.52
5-6	246	1000	0.246	239	31	7.71
6-1	129	1237	0.104	113	11	10.27
6-2	80	1111	0.072	72	6	12.00
6-3	69	1100	0.063	65	11	5.91
6-4	76	1217	0.062	57	9	6.33
6-5	194	1073	0.181	163	29	5.62

6-6	65	1256	0.052	54	10	5.40
	Mean Diff Index		0.109	Mean Fusion Index		5.69
	Standard Deviation		0.064	Standard Deviation		3.19

Drosha shRNA	Total Nuclei in MHC+ Cells	Total Nuclei	Diff Index	Nuclei in Myotubes	Myotubes	Fusion Index
1-1	179	1434	0.125	152	17	8.94
1-2	155	1244	0.125	127	20	6.35
1-3	261	1281	0.204	246	48	5.13
1-4	71	1505	0.047	54	29	1.86
1-5	229	1020	0.225	204	32	6.38
1-6	214	1203	0.178	193	22	8.77
2-1	181	942	0.192	165	18	9.17
2-2	187	985	0.190	157	21	7.48
2-3	253	1322	0.191	230	45	5.11
2-4	250	1403	0.178	226	35	6.46
2-5	131	1164	0.113	107	32	3.34
2-6	371	1249	0.297	345	37	9.32
3-1	451	1329	0.339	431	36	11.97
3-2	84	1410	0.060	51	42	1.21
3-3	232	1315	0.176	207	34	6.09
3-4	252	1366	0.184	221	48	4.60
3-5	312	1360	0.229	292	35	8.34
3-6	182	1333	0.137	161	25	6.44
4-1	235	1380	0.170	211	31	6.81
4-2	167	1084	0.154	136	37	3.68
4-3	272	1103	0.247	245	21	11.67
4-4	403	1282	0.314	388	25	15.52
4-5	98	1236	0.079	86	13	6.62
4-6	258	1137	0.227	237	32	7.41
5-1	187	1151	0.162	157	16	9.81
5-2	378	1253	0.302	370	38	9.74
5-3	201	1242	0.162	184	31	5.94
5-4	285	1424	0.200	262	29	9.03
5-5	300	1052	0.285	282	26	10.85
5-6	86	1349	0.064	74	14	5.29
6-1	197	1163	0.169	169	40	4.23
6-2	212	1075	0.197	196	21	9.33
6-3	217	1192	0.182	201	19	10.58
6-4	182	1106	0.165	162	23	7.04
6-5	283	1066	0.265	263	26	10.12

6-6	391	1082	0.361	380	43	8.84
	Mean Diff Index		0.192	Mean Fusion Index		7.48
	Standard Deviation		0.076	Standard Deviation		2.96

FLASH shRNA	Total Nuclei in MHC+ Cells	Total Nuclei	Diff Index	Nuclei in Myotubes	Myotubes	Fusion Index
1-1	135	1402	0.096	116	26	4.46
1-2	419	1323	0.317	403	55	7.33
1-3	420	1263	0.333	391	46	8.50
1-4	110	1158	0.095	87	29	3.00
1-5	270	1105	0.244	238	38	6.26
1-6	174	898	0.194	157	18	8.72
2-1	498	963	0.517	482	23	20.96
2-2	475	1287	0.369	458	53	8.64
2-3	376	1182	0.318	341	66	5.17
2-4	235	1176	0.200	233	19	12.26
2-5	220	1174	0.187	204	34	6.00
2-6	611	1276	0.479	595	38	15.66
3-1	336	1149	0.292	309	37	8.35
3-2	418	1134	0.369	396	34	11.65
3-3	509	1246	0.409	496	44	11.27
3-4	272	1321	0.206	255	36	7.08
3-5	648	1263	0.513	643	37	17.38
3-6	607	1269	0.478	586	22	26.64
4-1	456	1157	0.394	436	41	10.63
4-2	516	1193	0.433	496	43	11.53
4-3	435	1131	0.385	417	28	14.89
4-4	490	1352	0.362	462	49	9.43
4-5	363	1161	0.313	347	33	10.52
4-6	464	1345	0.345	452	37	12.22
5-1	552	1035	0.533	535	58	9.22
5-2	526	1093	0.481	507	37	13.70
5-3	248	1034	0.240	228	41	5.56
5-4	306	1135	0.270	283	24	11.79
5-5	540	1128	0.479	511	45	11.36
5-6	349	1133	0.308	318	42	7.57
6-1	358	1131	0.317	335	57	5.88
6-2	46	1369	0.034	29	21	1.38
6-3	233	1441	0.162	207	39	5.31
6-4	132	1063	0.124	102	34	3.00
6-5	289	1192	0.242	261	47	5.55

6-6	413	1089	0.379	397	38	10.45
	Mean Diff Index		0.317	Mean Fusion Index		9.70
	Standard Deviation		0.130	Standard Deviation		5.09

EGFP-Ars2	Total Nuclei in MHC+ Cells	Total Nuclei	Diff Index	Nuclei In Myotubes	Myotubes	Fusion Index
1-1	119	985	0.121	97	21	4.62
1-2	43	1252	0.034	37	9	4.11
1-3	136	1391	0.098	95	30	3.17
1-4	13	1291	0.010	0	11	0.00
1-5	210	1183	0.178	195	35	5.57
1-6	125	1036	0.121	115	18	6.39
2-1	138	1047	0.132	126	20	6.30
2-2	57	994	0.057	36	28	1.29
2-3	496	1033	0.480	486	53	9.17
2-4	256	1342	0.191	235	27	8.70
2-5	101	1054	0.096	90	21	4.29
2-6	110	1262	0.087	91	48	1.90
3-1	323	1179	0.274	313	48	6.52
3-2	40	1346	0.030	23	5	4.60
3-3	39	1257	0.031	13	28	0.46
3-4	144	895	0.161	132	15	8.80
3-5	177	1023	0.173	160	12	13.33
3-6	76	1085	0.070	60	11	5.45
4-1	175	1013	0.173	164	13	12.62
4-2	49	1030	0.048	39	8	4.88
4-3	56	1007	0.056	51	8	6.38
4-4	375	1032	0.363	357	33	10.82
4-5	41	1425	0.029	29	30	0.97
4-6	132	1152	0.115	112	18	6.22
5-1	102	1101	0.093	85	5	17.00
5-2	40	934	0.043	33	4	8.25
5-3	24	1075	0.022	17	5	3.40
5-4	118	1017	0.116	98	14	7.00
5-5	116	1309	0.089	104	16	6.50
5-6	89	1204	0.074	76	30	2.53
6-1	93	1171	0.079	77	16	4.81
6-2	76	1184	0.064	74	25	2.96
6-3	173	923	0.187	162	20	8.10
6-4	184	1282	0.144	163	31	5.26
6-5	50	1466	0.034	34	20	1.70

6-6	167	1116	0.150	130	24	5.42
	Mean Diff Index		0.117	Mean Fusion Index		5.82
	Standard Deviation		0.097	Standard Deviation		3.69

Control shRNA	Total Nuclei in MHC+ Cells	Total Nuclei	Diff Index	Nuclei In Myotubes	Myotubes	Fusion Index
1-1	194	1208	0.161	165	37	4.46
1-2	358	1411	0.254	323	37	8.73
1-3	303	1328	0.228	262	43	6.09
1-4	147	1332	0.110	129	9	14.33
1-5	490	1194	0.410	463	47	9.85
1-6	466	1140	0.409	440	82	5.37
2-1	351	1149	0.305	320	52	6.15
2-2	346	1099	0.315	318	53	6.00
2-3	363	1227	0.296	345	38	9.08
2-4	266	942	0.282	242	42	5.76
2-5	36	1274	0.028	24	2	12.00
2-6	460	1160	0.397	444	41	10.83
3-1	358	1190	0.301	347	58	5.98
3-2	179	1073	0.167	159	29	5.48
3-3	168	1228	0.137	154	40	3.85
3-4	429	1140	0.376	410	64	6.41
3-5	236	1239	0.190	210	57	3.68
3-6	423	1072	0.395	416	26	16.00
4-1	533	1100	0.485	523	38	13.76
4-2	384	973	0.395	373	46	8.11
4-3	553	1162	0.476	550	43	12.79
4-4	623	1182	0.527	604	43	14.05
4-5	804	1287	0.625	792	58	13.66
4-6	291	1117	0.261	266	30	8.87
5-1	473	1147	0.412	456	35	13.03
5-2	269	1092	0.246	252	47	5.36
5-3	522	1122	0.465	501	43	11.65
5-4	596	1022	0.583	581	69	8.42
5-5	330	1099	0.300	306	35	8.74
5-6	462	1114	0.415	429	40	10.73
6-1	425	1201	0.354	401	49	8.18
6-2	552	1176	0.469	535	30	17.83
6-3	469	1321	0.355	460	54	8.52
6-4	437	1378	0.317	413	54	7.65
6-5	355	1329	0.267	332	33	10.06

6-6	253	1292	0.196	218	44	4.95
	Mean Diff Index		0.331	Mean Fusion Index		9.07
	Standard Deviation		0.133	Standard Deviation		3.65

APPENDIX C: OVEREXPRESSION DIFFERENTIATION RAW DATA

EGFP-ARS2	Total Transfected Nuclei in Myotubes	Total Myotubes	Mean Transfected Nuclei per Myotube	Total Transfected Nuclei
1-1	9	3	3.000	12
1-2	18	7	2.571	22
1-3	23	5	4.600	33
1-4	12	4	3.000	24
1-5	11	3	3.667	18
1-6	0	0	0.000	4
2-1	5	1	5.000	12
2-2	8	2	4.000	17
2-3	9	3	3.000	12
2-4	3	1	3.000	7
2-5	32	8	4.000	36
2-6	17	3	5.667	23
3-1	5	2	2.500	6
3-2	5	1	5.000	6
3-3	13	2	6.500	14
3-4	0	0	0.000	0
3-5	0	0	0.000	2
3-6	0	0	0.000	0
4-1	0	0	0.000	2
4-2	4	1	4.000	4
4-3	6	1	6.000	12
4-4	0	0	0.000	0
4-5	3	1	3.000	5
4-6	0	0	0.000	2
	Mean		2.854	
	Standard Deviation		2.140	

EGFP-FARB	Total Transfected Nuclei in Myotubes	Total Myotubes	Mean Transfected Nuclei per Myotube	Total Transfected Nuclei
1-1	34	8	4.250	45
1-2	35	6	5.833	48
1-3	17	3	5.667	28
1-4	32	5	6.400	38

1-5	18	6	3.000	29
1-6	31	5	6.200	42
2-1	76	10	7.600	91
2-2	74	10	7.400	92
2-3	25	4	6.250	38
2-4	40	6	6.667	46
2-5	121	11	11.000	133
2-6	135	14	9.643	170
3-1	9	1	0.111	13
3-2	6	2	0.333	11
3-3	17	3	0.176	19
3-4	60	3	0.050	62
3-5	56	5	0.089	62
3-6	3	1	0.333	4
4-1	53	5	0.094	57
4-2	5	1	0.200	5
4-3	0	0	0.000	4
4-4	15	2	0.133	19
4-5	8	2	0.250	9
4-6	14	2	0.143	14
	Mean		3.409	
	Standard Deviation		3.636	

EGFP-SCR	Total Transfected Nuclei in Myotubes	Total Myotubes	Mean Transfected Nuclei per Myotube	Total Transfected Nuclei
1-1	43	8	5.375	59
1-2	113	12	9.417	124
1-3	47	9	5.222	55
1-4	28	3	9.333	35
1-5	85	11	7.727	94
1-6	44	9	4.889	58
2-1	33	9	3.667	41
2-2	39	5	7.800	52
2-3	55	8	6.875	67
2-4	102	11	9.273	114
2-5	78	11	7.091	82
2-6	130	15	8.667	145
3-1	23	3	0.130	27
3-2	62	7	0.113	81
3-3	62	8	0.129	64
3-4	29	4	0.138	32

3-5	4	1	0.250	7
3-6	27	3	0.111	33
4-1	35	4	0.114	38
4-2	6	1	0.167	7
4-3	29	3	0.103	36
4-4	0	0	0.000	0
4-5	68	6	0.088	73
4-6	64	7	0.109	64
	Mean		3.616	
	Standard Deviation		3.816	

H2B-GFP	Total Transfected Nuclei in Myotubes	Total Myotubes	Mean Transfected Nuclei per Myotube	Total Transfected Nuclei
1-1	55	11	5.000	82
1-2	20	4	5.000	27
1-3	23	4	5.750	33
1-4	66	12	5.500	97
1-5	60	8	7.500	85
1-6	76	16	4.750	102
2-1	115	11	10.455	143
2-2	41	6	6.833	54
2-3	27	6	4.500	64
2-4	74	11	6.727	82
2-5	50	9	5.556	73
2-6	75	7	10.714	98
3-1	10	4	0.400	22
3-2	32	7	0.219	46
3-3	10	3	0.300	23
3-4	60	8	0.133	70
3-5	17	4	0.235	22
3-6	35	3	0.086	42
4-1	48	6	0.125	63
4-2	37	4	0.108	39
4-3	21	3	0.143	22
4-4	71	6	0.085	82
4-5	86	9	0.105	131
4-6	84	6	0.071	110
	Mean		3.346	
	Standard Deviation		3.558	

APPENDIX D: ARS2-X5 SEQUENCE DATA

>2.1kb;M13For
AGCGGCCGCGAATTTCGCCCTTTCCCACATTCTGGTCTTGACTGCTCGTAACTGTTTTGTCTCTC
CCATCCTGCTTGCTTTCTTGTCCAGGTTCCGATCTAAGTACCACCCTGATGAGGTGGGAAAGCGT
CGGCAGGAGGCCCGGGGGGCCCTGCAGAACCGGCTGAAGGTGTTCCTGTCCCTCATGGAGAGTGG
CTGGTTTGATAACCTTCTCTTGGACATAGACAAAGCTGATGCCATTGTCAAGATGCTAGATGCAG
CTGTCATTAAGATGGAAGGTGGCACAGAGAACGATCTCCGAATTTTGGAGCAGGAGGAGGAGGAG
GAACAGGCAGGCAAGACTGGGGAGGCCAGCAAGAAAGAGGAGGCCCGTGCTGGACCAGCCCTGGG
GGAAGGAGAGCGCAAAGCCAATGATAAGGATGAGAAGAAAGAAGATGGAAAACAGGCTGAGAATG
ACAGTTCCAACGATGACAAAACTAAAAAATCTGAGGGTGATGGGGACAAGGAGGAGAAGAAAGAA
GAGGCTGAGAAGGAAGCCAAAAAGAGCAAGAAGCGGAACAGGAAGCAGAGTGGCGATGACAGCTT
CGATGAGGGCAGTGTGTCCGAGTCTGAGTCCGAGTCTGAGGGTGGCCAGGCCGAGGAGGAGAAGG
AGGAGGCCGAAGAAGCACTTAAAGAAAAGGAGAAGCCCAAAGAGGAGGAGAAGGAGAAGCCTAAN
GATGCTGCAGGGTTGGAGTGTAAGCCCCGGCCCTTGCATAAGACTTGCTCTCTCTTCATGCGCAA
CATCGCACCCAACATTTCAAGGGCAGAGATCATTTCTCTTTGTAAACGATACCCAGGCTTTATGC
GAGTGGCACTGTCAGAGCCCC

>2.1kb;7For
ACGATGACAAAACTAAAAAATCTGAGGGTGATGGGGACAAGGAGGAGAAGAAAGAAGAGGCTGAN
AAGGAAGCCAAAAAGAGCAAGAAGCGGAACAGGAAGCAGAGTGGCGATGACAGCTTCGATGAGGG
CAGTGTGTCCGAGTCTGAGTCCGAGTCTGAGGGTGGCCAGGCCGAGGAGGAGAAGGAGGAGGCCG
AAGAAGCACTTAAAGAAAAGGAGAAGCCCAAAGAGGAGGAGAAGGAGAAGCCTAAGGATGCTGCA
GGGTTGGAGTGTAAGCCCCGGCCCTTGCATAAGACTTGCTCTCTCTTCATGCGCAACATCGCACC
CAACATTTCAAGGGCAGAGATCATTTCTCTTTGTAAACGATACCCAGGCTTTATGCGAGTGGCAC
TGTCAGAGCCCCAGCCAGAGAGGAGGTTTTTTCGCCGTGGCTGGGTGACTTTTGACCGCAGTGTT
AACATTAAGGAGATCTGTTGGAACCTGCAGAACATTCGGCTCCGGGAGTGTGAACTGAGTCCCGG
TGTGAACAGAGACCTGACCCGTCGTGTCCGCAACATAAATGGCATTACACAGCACAAGCAGATAG
TGCGCAATGACATCAAGTTGGCAGCCAAGCTAATCCACACACTGGATGACAGGACCCAGCTCTGG
GCCTCTGAGCCTGGGACGCCTCCTGTGCCCACAAGCCTCCCCTCGCAAAACCCCATCCTGAAGAA
CATCACTGACTACCTGATTGAGGAAGTGAGTGCGGAGGAGGAGGANCTTCTGGGGAGCAGTGGGG
GACCCCCTCCTGAGGAGCCTCCCAAGGAGGCAACCCAGCCGANATCAATGTGGANAGGGATGAGA
AGCTGATCAAGTCTGGATAAACTTCTNCTCTATTTGCGT

>2.1kb;Ars2i_11For
AGGAGGTTTTTTCGCCGTGGCTGGGTGACTTTTGACCGCAGTGTTAACATTAAGGAGATCTGTTG
GAACCTGCAGAACATTCGGCTCCGGGAGTGTGAACTGAGTCCCGGTGTGAACAGAGACCTGACCC
GTCGTGTCCGCAACATAAATGGCATTACACAGCACAAGCAGATAGTGCGCAATGACATCAAGTTG
GCAGCCAAGCTAATCCACACACTGGATGACAGGACCCAGCTCTGGGCCTCTGAGCCTGGGACGCC
TCCTGTGCCCACAAGCCTCCCCTCGCAAAACCCCATCCTGAAGAACATCACTGACTACCTGATTG
AGGAAGTGAGTGCGGAGGAGGAGGAGCTTCTGGGGAGCAGTGGGGGACCCCCTCCTGAGGAGCCT
CCCAAGGAGGGCAACCCAGCCGAGATCAATGTGGAGAGGGATGAGAAGCTGATCAAGGTCTTGGA
TAAACTTCTTCTCTATTTGCGTATTGTGCATTCTCTGGATTATTATAACACCTGTGAGTACCCTA
ATGAAGACGAGATGCCCAACCGCTGTGGCATAATCCACGTTCGGGGGCCCATGCCTCCCAACCGA
ATTAGTCACGGAGAAGTGCTGGAGTGGCAGAAGACATTTGAGGAGAAACTGACTCCACTGTTGAG
TGTGCGTGAATCCCTTTCTGAGGAAGAGGCCCAGAAGATGGGTCGANNGACCCAGAGCAGGAAGT
GGAGAAGTTTGTCACCTCCAACACGCAGGAACTGGGCAAGGATAAGTGGCTATGTCCTCTCAGTG
GCAAGAAATTCAAGGGCCCGGAGTTTGTGCGCAAGCATATCTTCAATAAGCATGCCGAGAAGATC
GAGGAGGTGAAGAAGGAGGTGGCGTTCTTCAATAACTTTCTCACAGACGCCAAGCGCCCAGCTTT
GCCTGAGATCAAGCCAGCTCA

>2.1kb;M13Rev
GAGAAGTGCTGGAGTGGCAGAAGACATTTGAGGAGAAACTGACTCCACTGTTGAGTGTGCGTGAA
TCCCTTTTTGAGGAAGAGGCCCAGAAGATGGGTCGAAAAGACCCAGAGCAGGAAGTGGAGAAGTT
TGTCACCTCCAACACGCAGGAACTGGGCAAGGATAAGTGGCTATGTCCTCTCAGTGGCAAGAAAT
TCAAGGGCCCGGAGTTTGTGCGCAAGCATATCTTCAATAAGCATGCCGAGAAGATCGAGGAGGTG
AAGAAGGAGGTGGCGTTCTTCAATAACTTTCTCACAGACGCCAAGCGCCCAGCTTTGCCTGAGAT
CAAGCCAGCTCAGCCACCTGGCCCTGCCCAGATACTCCCCCCAGGCCTGACCCCAGGACTTCCCT
ACCCACATCAGACGCCACAGGGCTTGATGCCGTATGGTCAGCCCCGGCCTCCCATCTTGGGCTAT
GGAGTCCCAACAGGAGGGCCTCCATACCCCCATGCTCCATATGGTGCCGGCCGTGGGAACTATGA
TGCTTTTCGAGGCCAAGGCGGTTATCCTGGGAAACCTCGGAACAGGATGGTTCGAGGAGACCCAA
GGGCCATAGTGGAGTATCGGGACCTGGATGCCCCGGATGATGTTGACTTCTTTTGAGCCATCCAT
GTCCCCTC

>*Ars2-X5* cDNA sequence
CCCACATTCTGGTCTTGACTGCTCGTAACTGTTTTGTCTCTCCCATCCTGCTTGCTTTCTTGTCC
AGGTTCCGATCTAAGTACCACCCTGATGAGGTGGGAAAGCGTCGGCAGGAGGCCCGGGGGGCCCT
GCAGAACCGGCTGAAGGTGTTCCTGTCCCTCATGGAGAGTGGCTGGTTTGATAACCTTCTCTTGG
ACATAGACAAAGCTGATGCCATTGTCAAGATGCTAGATGCAGCTGTCATTAAGATGGAAGGTGGC
ACAGAGAACGATCTCCGAATTTTGGAGCAGGAGGAGGAGGAGGAACAGGCAGGCAAGACTGGGGA
GGCCAGCAAGAAAGAGGAGGCCCGTGCTGGACCAGCCCTGGGGGAAGGAGAGCGCAAAGCCAATG
ATAAGGATGAGAAGAAAGAAGATGGAAAACAGGCTGAGAATGACAGTTCCAACGATGACAAAACT
AAAAAATCTGAGGGTGATGGGGACAAGGAGGAGAAGAAAGAAGAGGCTGAGAAGGAAGCCAAAAA
GAGCAAGAAGCGGAACAGGAAGCAGAGTGGCGATGACAGCTTCGATGAGGGCAGTGTGTCCGAGT
CTGAGTCCGAGTCTGAGGGTGGCCAGGCCGAGGAGGAGAAGGAGGAGGCCGAAGAAGCACTTAAA
GAAAAGGAGAAGCCCAAAGAGGAGGAGAAGGAGAAGCCTAAGGATGCTGCAGGGTTGGAGTGTAA
GCCCCGGCCCTTGCATAAGACTTGCTCTCTCTTCATGCGCAACATCGCACCCAACATTTCAAGGG
CAGAGATCATTTCTCTTTGTAAACGATACCCAGGCTTTATGCGAGTGGCACTGTCAGAGCCCCAG
CCAGAGAGGAGGTTTTTTCGCCGTGGCTGGGTGACTTTTGACCGCAGTGTTAACATTAAGGAGAT
CTGTTGGAACCTGCAGAACATTCGGCTCCGGGAGTGTGAACTGAGTCCCGGTGTGAACAGAGACC
TGACCCGTCGTGTCCGCAACATAAATGGCATTACACAGCACAAGCAGATAGTGCGCAATGACATC
AAGTTGGCAGCCAAGCTAATCCACACACTGGATGACAGGACCCAGCTCTGGGCCTCTGAGCCTGG
GACGCCTCCTGTGCCCACAAGCCTCCCCTCGCAAAACCCCATCCTGAAGAACATCACTGACTACC
TGATTGAGGAAGTGAGTGCGGAGGAGGAGGAGCTTCTGGGGAGCAGTGGGGGACCCCCTCCTGAG
GAGCCTCCCAAGGAGGGCAACCCAGCCGAGATCAATGTGGAGAGGGATGAGAAGCTGATCAAGGT
CTTGGATAAACTTCTTCTCTATTTGCGTATTGTGCATTCTCTGGATTATTATAACACCTGTGAGT
ACCCTAATGAAGACGAGATGCCCAACCGCTGTGGCATAATCCACGTTCGGGGGCCCATGCCTCCC
AACCGAATTAGTCACGGAGAAGTGCTGGAGTGGCAGAAGACATTTGAGGAGAAACTGACTCCACT
GTTGAGTGTGCGTGAATCCCTTTCTGAGGAAGAGGCCCAGAAGATGGGTCGAAAAGACCCAGAGC
AGGAAGTGGAGAAGTTTGTCACCTCCAACACGCAGGAACTGGGCAAGGATAAGTGGCTATGTCCT
CTCAGTGGCAAGAAATTCAAGGGCCCGGAGTTTGTGCGCAAGCATATCTTCAATAAGCATGCCGA
GAAGATCGAGGAGGTGAAGAAGGAGGTGGCGTTCTTCAATAACTTTCTCACAGACGCCAAGCGCC
CAGCTTTGCCTGAGATCAAGCCAGCTCAGCCACCTGGCCCTGCCCAGATACTCCCCCCAGGCCTG
ACCCCAGGACTTCCCTACCCACATCAGACGCCACAGGGCTTGATGCCGTATGGTCAGCCCCGGCC
TCCCATCTTGGGCTATGGAGTCCCAACAGGAGGGCCTCCATACCCCCATGCTCCATATGGTGCCG
GCCGTGGGAACTATGATGCTTTTCGAGGCCAAGGCGGTTATCCTGGGAAACCTCGGAACAGGATG
GTTCGAGGAGACCCAAGGGCCATAGTGGAGTATCGGGACCTGGATGCCCCGGATGATGTTGACTT
CTTTTGAGCCATCCATGTCCCCTC

>ARS2-X5 protein sequence
MESGWFDNLLLDIDKADAIVKMLDAAVIKMEGGTENDLRILEQEEEEEQAGKTGEASKKEEARAG
PALGEGERKANDKDEKKEDGKQAENDSSNDDKTKKSEGDGDKEEKKEEAEKEAKKSKKRNRKQSG
DDSFDEGSVSESESESEGGQAEEEKEEAEEALKEKEKPKEEEKEKPKDAAGLECKPRPLHKTCSL
FMRNIAPNISRAEIISLCKRYPGFMRVALSEPQPERRFFRRGWVTFDRSVNIKEICWNLQNIRLR

ECELSPGVNRDLTRRVRNINGITQHKQIVRNDIKLAAKLIHTLDDRTQLWASEPGTPPVPTSLPS
QNPILKNITDYLIEEVSAEEEELLGSSGGPPPEEPPKEGNPAEINVERDEKLIKVLDKLLLYLRI
VHSLDYYNTCEYPNEDEMPNRCGIIHVRGPMPPNRISHGEVLEWQKTFEEKLTPLLSVRESLSEE
EAQKMGRKDPEQEVEKFVTSNTQELGKDKWLCPLSGKKFKGPEFVRKHIFNKHAEKIEEVKKEVA
FFNNFLTDAKRPALPEIKPAQPPGPAQILPPGLTPGLPYPHQTPQGLMPYGQPRPPILGYGVPTG
GPPYPHAPYGAGRGNYDAFRGQGGYPGKPRNRMVRGDPRAIVEYRDLDAPDDVDFF