Requirements for ARS2 in RNA Processing and Retina Development

by

Connor O'Sullivan
BSc, McMaster University, 2007

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

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

Supervisory Committee

Requirements for ARS2 in RNA Processing and Retina Development

by

Connor O'Sullivan
BSc, McMaster University, 2007

Supervisory Committee

Perry L. Howard, Department of Biochemistry and Microbiology
Supervisor

Robert D. Burke, Department of Biochemistry and Microbiology
Departmental Member

Christopher J. Nelson, Department of Biochemistry and Microbiology
Departmental Member

Robert L. Chow, Department of Biology
Outside Member

Abstract

Supervisory Committee
Perry L. Howard, Department of Biochemistry and Microbiology
Supervisor
Robert D. Burke, Department of Biochemistry and Microbiology
Departmental Member
Christopher J. Nelson, Department of Biochemistry and Microbiology
Departmental Member
Robert L. Chow, Department of Biology
Outside Member

ARS2 is a stable component of the nuclear cap-binding complex (CBC) and is critical for RNA Polymerase II transcript processing. As such, ARS2 functions in numerous RNA Polymerase II transcript processing events, which happen co-transcriptionally from initiation to termination, and post-transcriptionally during maturation and export into the cytoplasm. Developmentally, ARS2 is essential for stem cell maintenance and differentiation during embryogenesis and in neural stem cells. Two major questions in the field were: 1) how does ARS2 function in stem cell maintenance and/or differentiation? and 2) how does ARS2 distinguish between disparate RNA classes and processing complexes? In chapter 2, I show that ARS2 is required for the proliferation and cell fate decisions of progenitors in the mouse retina. Specifically, ARS2 knockdown delays cell cycle progression and leads to premature cell cycle exit. Additionally, ARS2 knockdown increases the proportion of cells expressing rod photoreceptor marker Nrl, and decreases Müller glial marker expression. Similarly, knockdown of FLASH, an essential component for replication-dependent histone transcript processing and cell cycle progression, increases the proportion of cells expressing the Nrl reporter, suggesting ARS2's role in histone processing is contributing to cell cycle progression and fate specification in the developing retina. In chapter 3, I used bioinformatics analysis and homology modeling to classify four structural domains of mammalian ARS2, including a newly identified RNA recognition motif (RRM), and performed mutagenesis to assess their functions. The unstructured C-terminus is required for interaction with the CBC, the Mid domain is implicated in binding DROSHA, which is required for microRNA biogenesis, while the zinc finger and RRM are involved in binding FLASH. Moreover,

the zinc finger is required for interacting with RNA. Collectively, this work establishes a model where ARS2 acts as a scaffold, using multiple domains to interact with distinct processing complexes in a mutually exclusive manner. It is also the first study describing the requirements of ARS2 in the developing retina. Understanding the molecular mechanisms governing progenitor proliferation and cell fate specification is crucial in order to design therapies for retinal degenerative diseases.

Table of Contents

List of Tables

List of Figures

List of abbreviations

m7G	7-methylguanosine
AGO	Argonaute
ARS2	Arsenite resistance protein 2
BrdU	Bromodeoxyuridine
CABP5	Calcium binding protein 5
CBC	Cap-binding complex
CBP20	Cap-binding protein 20 kDa
CBP80	Cap-binding protein 80 kDa
CDK	Cyclin-dependent kinase
CFI_m	Cleavage factor I_m
$CFII_m$	Ceavage factor II_m
ChIP	Chromatin immunoprecipitation
CldU	Chlorodeoxyuridine
CLRC	Clr4-Rik1-Cul4
CPSF	Cleavage and polyadenylation specificity factor
CRALBP	Cellular retinal binding protein
Cryo-EM	Cryogenic electron microscopy
CstF	Cleavage stimulation factor
CTD	RNA polymerase II C-terminal domain
CUTs	Cryptic unstable transcripts
D-bodies	Dicing bodies
DCL1	DICER-LIKE1
DGCR8	DiGeorge syndrome chromosomal region 8
DMEM	Dulbecco's modified eagle's medium
dsRNA	Double-stranded RNA
DSR	Determinant of selective removal
DUF	Domain of unknown function
eGFP	Enhanced green fluorescent protein
Erh1	Enhancer of rudimentary
EMC	Erh1-Mmi1 complex
EMSA	Electrophoretic mobility shift assay
EV	Empty vector
FACS	Fluorescence activated cell sorting
FARB	FLASH-ARS2 binding peptide
FLASH	FLICE-associated huge protein
FBS	Fetal bovine serum
GCL	Ganglion cell layer
H3K9me	Histone 3 lysine 9 methylation

HCC	Histone pre-mRNA cleavage complex
HDE	Histone downstream element
HOODs	Heterochromatin domains
HLBs	Histone locus bodies
hnRNP	Heterogeneous ribonucleoprotein
HYL1	HYPONASTIC LEAVES 1
IdU	Iododeoxyuridine
INL	Inner nuclear layer
INM	Interkinetic nuclear migration
IP	Immunoprecipitate
IPL	Inner plexiform layer
IVT	In vitro translation
kDa	kiloDalton
KD	Knockdown
LIF	Leukemia inhibitory factor
lncRNA	Long noncoding RNA
mRNA	Messenger RNA
miRNA	MicroRNA
MTREC	Mtl1-Red1 core
MEF2	Myogenic enhancer factor 2
NBL	Neuroblastic layer
NELF	Negative elongation factor
NEXT	Nuclear exosome targeting complex
NICD	Notch1 intracellular domain
NMD	Nonsense-mediated decay
NRL	Neural retina leucine zipper
NSCs	Neural stem cells
NURS	Nuclear RNA silencing complex
ONL	Outer nuclear layer
OPL	Outer plexiform layer
pA	Polyadenylation
PB	Sodium phosphate buffer
PBS	Phosphate buffered saline
PBTB	0.5% BSA, 0.1% Triton-X 100 in PBS
PFA	Paraformaldehyde
PHAX	Phosphorylated adaptor for RNA export
PI	Propidium iodide
Pre-miRNA	Precursor microRNA
Pri-miRNA	Primary microRNA
PROMPTs	Promoter upstream transcripts

qRT-PCR	Quantitative real-time polymerase chain reaction
RDH	Replication-dependent histone
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing complex
RNP	Ribonucleoprotein
RNAi	RNA interference
RNAP II	RNA polymerase II
RRM	RNA recognition motif
SD	Standard deviation
S. pombe	*Schizosaccharomyces pombe*
SE	SERRATE
SF3B4	Splicing factor 3b subunit 4
SLBP	Stem-loop binding protein
siRNA	Small interfering RNA
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein
snoRNA	Small nucleolar RNA
SOX2OT	SOX2 overlapping transcript
TBST	Tris-buffered saline-Tween 20 (0.5%)
TERRA	Telomeric repeat-containing RNA
tGFP	Turbo green fluorescent protein
TGH	TOUGH
TREX	Transcription export
ZnF	Zinc finger

Acknowledgments

I would like to give special thanks to my supervisor, Dr. Perry Howard, for his constant support, guidance, and feedback on all things related to my studies; without his mentorship this dissertation would not be possible. I would also like to thank my committee members, Dr. Bob Chow, Dr. Robert Burke, and Dr. Chris Nelson for their invaluable input and advice. I would like to thank all members of the Howard, Chow and Nelson labs, past and present, especially Phil Nickerson, Li-Li Chen, Spencer Alford, Jennifer Christie, Kevin Yongblah, Geoff Gudavicius and Dave Dilworth for all their help. I would also like to thank Dr. John Webb for his assistance with FACS sorting at the Deeley Research Centre.

I undoubtedly need to give an immense thank you to all my family members for their tremendous support and encouragement over the years. I also would like to thank the entire Adair/Acheson family for their overwhelming generosity, and for taking care of Cedar while Sarah and I are both working. I am especially indebted to my parents and grandparents, as they have nourished my sense of curiosity throughout my childhood and adult life, and their unwavering, astute guidance and advice made it possible for me to pursue science.

Dedication

For Sarah,

my perpetual inspiration, you make life a joy

Chapter 1: Introduction

* Excerpts from this introduction will be used in an invited review article:

O'Sullivan C., Howard P.L. The diverse requirements of ARS2 in nuclear cap-binding complex-dependent RNA processing. RNA&DISEASE. 2016;3:e1376.

1.1 Importance of ARS2 in stem cell maintenance and differentiation

Arsenite resistance protein 2 (ARS2) was first identified as a gene that conferred arsenic resistance to Chinese hamster ovary cells [1]. However, this was later shown to be due to expression of a truncated cDNA that results in a dominant-negative phenotype; instead, ARS2 expression correlates with arsenic sensitivity [2]. The first characterizations of ARS2 were genetic studies to determine requirements of ARS2, or its plant orthologue SERRATE (SE), during development. ARS2/SERRATE knockout results in lethality in plants, fission yeast, fruit flies, zebrafish, and mice [3–8]. *Ars2*$^{-/-}$ mouse embryos die peri-implantation, and fail to progress past the blastocyst stage. *Ars2* null embryos cultured *in vitro* with or without leukemia inhibitory factor (LIF), or following zona pellucida removal, failed to outgrow and became apoptotic [7]. These results suggest ARS2 is important for stem cell maintenance and/or differentiation [7]. To examine the developmental requirements in more detail, the Lai lab created a conditional *Ars2* knockout in mouse subventricular zone neural stem cells (NSCs) [9]. They found that ARS2 is necessary for NSCs to maintain their neurogenic and self-renewal capacity [9]. Conversely, high levels of ARS2 in NSCs correlates with increased neurogenesis and an elongated life span in mice [10]. Additionally, conditional *Ars2* knockout in hematopoietic tissues results in decreased cellularity in the bone marrow [2]. While less is known about the requirements for ARS2 in adult tissue, there are several recent reports of ARS2 dysregulation in human cancers [11–13], and ARS2 has been shown to regulate expression of several microRNAs (miRNAs) involved in transformation [2]. Together, these findings indicate ARS2 has a vital role in the development and maintenance of

diverse tissues and suggest that maintaining appropriate levels of ARS2 may be important for human health.

At the start of my program, this was essentially all that was known about ARS2 function. Most of the existing data came from studies using *Arabidopsis* SERRATE, implicating it in miRNA biogenesis [3,14]. The labs of Sara Cherry and Craig Thompson had shown that the function of ARS2/SERRATE in miRNA biogenesis was conserved in *Droshophila* and mammals [2,15]. Shortly thereafter, Kiriyama et al. identified ARS2 as a critical component in replication-dependent histone mRNA processing [16]. Thus, there were seemingly disparate results in the literature with no unifying hypothesis on ARS2 function. Tackling this problem requires an understanding of how ARS2 functions mechanistically, and how ARS2 loss impacts embryonic development in metazoans. Since that time, there has been a wealth of data showing that ARS2 plays a central role in the nuclear cap-binding complex and in RNA polymerase II (RNAP II) transcript processing.

After introducing the known functions of ARS2 in RNA polymerase II (RNAP II) transcript processing, I will discuss the early functional studies in diverse model organisms linking ARS2/SERRATE to the nuclear cap complex and its role in miRNA and replication-dependent histone mRNA (RDH) biogenesis, review what was known about ARS2 domain structure and function at the start of my thesis, as well as the recent "omic" data implicating ARS2 in linking transcription to the RNA surveillance, export, and silencing machineries. I will then expand on the essential requirements of ARS2 in progenitor cell proliferation and differentiation in the developing and adult central nervous system, provide background on the main model systems that I used during my thesis (the developing mouse retina and myoblast progenitor cells), and finally describe the central questions of my thesis and specific research objectives.

1.2 Overview of ARS2 functions in RNA polymerase II transcript processing

The early life of all RNA polymerase II (RNAP II) transcripts starts with the co-transcriptional addition of a 7-methylguanosine (m7G) cap at the 5'-end [17]. The 5'-m7G cap is co-transcriptionally bound by the nuclear cap-binding complex (CBC), which, in addition to protecting the transcript from degradation, interacts with several

multi-protein complexes to couple the 5'-cap with transcript processing, turnover, and export [2,18–26]. The interaction between the CBC and processing machinery is necessary for transcript maturation, which is thought to ensure only properly capped transcripts are further processed, and to limit the effects of promiscuous transcription. RNAP II generates several classes of RNA, including messenger RNA (mRNA), microRNA (miRNA), replication-dependent histone (RDH) mRNA, small nuclear RNA (snRNA), some small nucleolar RNA (snoRNA), and long noncoding RNA (lncRNA). Each of these RNA classes possess unique processing requirements. In addition, the CBC has been shown to play an integral role in transcription termination, splicing, 3'-end formation, exosomal degradation, intranuclear transport, and export to the cytoplasm [2,18–26]. The involvement of the cap complex in all of these steps is thought to provide critical quality control on RNAP II transcription, prevent errors in ribonucleoprotein (RNP) biogenesis, and ultimately in the case of mRNAs, prevent the translation of aberrant proteins. Precisely how the cap facilitates these disparate processes and discriminates between transcripts is not known. Over the past several years, it has become apparent that the cap-binding protein, ARS2, interacts constitutively with CBP20/80 and the 5'-end cap to form a complex called CBCA [2,25,26], which mediates interactions between the cap complex and RNA 3'-end processing, shuttling, exosome, and export machinery [2,16,24–27]. Thus, ARS2 is emerging as a critical factor, physically coupling multiple steps in the life of RNAP II transcripts (Figure 1).

ARS2 is an integral part of the nuclear cap complex and its function is intimately intertwined with cap-binding proteins CBP20/80. As mentioned, ARS2 interacts directly with CBP20/80 *in vitro* and *in vivo* and forms a complex with CBP20/80 and capped RNA termed CBCA [2,25,26]. Immunoprecipitation of either ARS2 or CBP20/80 from cell lysates pulls down a substantially overlapping set of proteins, implying a strong functional overlap [25,26]. This is supported by loss of function experiments demonstrating ARS2 and CBP20/80 regulate similar transcripts, and deficiencies in any of these proteins phenotypically resemble one another [2,15,25,26].

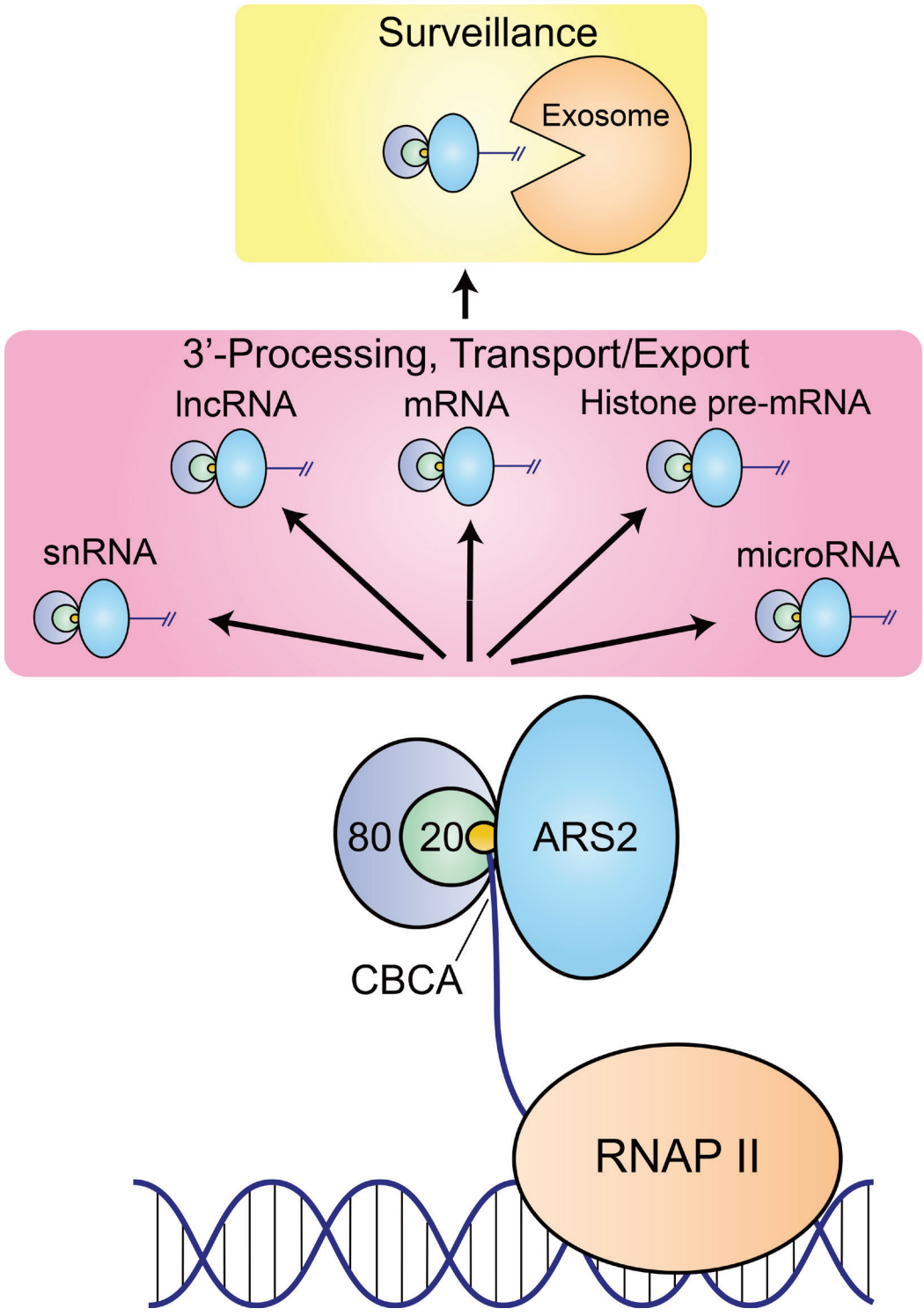

Figure 1 ARS2 functions in RNA polymerase II transcript processing

ARS2 is part of the nuclear cap-binding complex (CBC) along with cap-binding proteins CBP80 (80) and CBP20 (20). Together they form the CBCA complex, which co-transcriptionally binds the 5'-m7G cap (yellow circle) of RNA polymerase II (RNAP II) transcripts. CBCA then physically couples the 5'-cap to 3'-maturation events for distinct

RNA classes, such as snRNA, mRNA, replication-dependent histone pre-mRNA, and miRNA, and is required for intra-nuclear transport and export of RNA into the cytoplasm. CBCA has also been implicated in regulating lncRNA (pink box). CBCA also mediates RNA surveillance by targeting aberrant transcripts to the RNA exosome (yellow box). How ARS2 physically couples the 5'-cap to 3'-processing events for these disparate RNA classes is not understood.

1.3 SERRATE and microRNA biogenesis

The first studies to characterize ARS2 function were in *Arabidopsis* with its orthologue SERRATE (SE). Hypomorphic mutations of SE in *Arabidopsis* result in pleiotropic developmental defects, including a serrated leaf morphology, defects in leaf patterning, flower development and developmental phase transition, slower root growth, and increased shoot apical meristem size [28–30]. The pleiotropic defects observed with *se* mutants had overlapping phenotypes with mutants of genes involved in miRNA biogenesis, as well as with CBP80/ABH1 and CBP20 mutants [3,14,29–32].

Mature miRNAs are small noncoding RNAs (~21-24 nucleotides) that predominantly act to regulate gene expression post-transcriptionally by binding to the 3'-UTR of their target transcripts [33]. Canonical miRNAs are transcribed by RNAP II and begin as long stem-loop-containing primary miRNA (pri-miRNA) transcripts, which are capped and polyadenylated at the 5'- and 3'-ends, respectively (Figure 2) [34,35]. The pri-miRNA transcript is then cleaved by an RNase III-family enzyme to generate a precursor-miRNA (pre-miRNA) transcript (~65 nucleotides), which is exported by Exportin 5 to the cytoplasm where it is cleaved again by an RNase III enzyme to generate a mature miRNA/miRNA* duplex [36]. The double-stranded mature miRNA/miRNA* duplex is loaded onto an Argonaute (AGO)-containing RNA-induced silencing complex (RISC), and the passenger strand (miRNA*) is released, allowing the guide strand to bind to its antisense target and either inhibit translation or induce degradation of the target transcript [37–39]. The overall miRNA biogenesis pathway is evolutionary conserved in most eukaryotes, although the identity and location of a number of pathway components are species-specific, as discussed below.

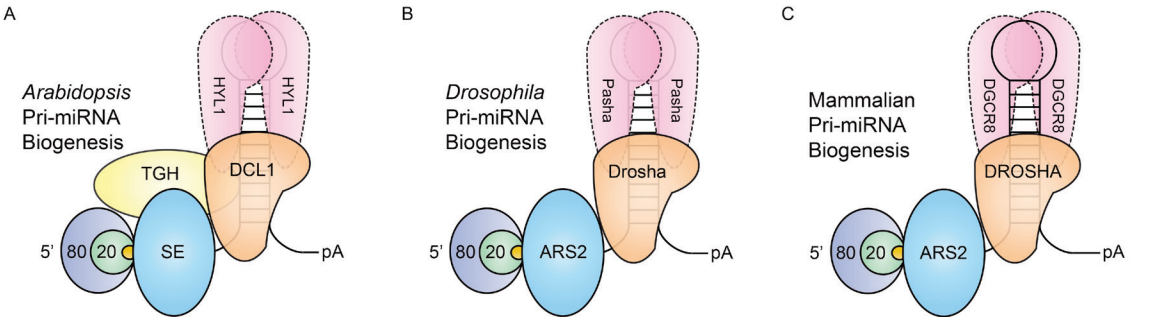

Figure 2 ARS2/SERRATE and microRNA biogenesis

A) In *Arabidopsis*, SERRATE (SE) interacts with CBP80/ABH1 (80) and CBP20 (20) at the 5' m7G cap (small yellow circle) and is required for efficient and accurate pri-miRNA cleavage by DCL1. SE also interacts with HYL1 and TGH. pA corresponds to the polyadenylation tail. B,C) In *Drosophila* and mammals, ARS2 is required for efficient pri-miRNA cleavage by the Microprocessor, which consists of Drosha and Pasha. C) DGCR8 is the mammalian orthologue of Pasha. In mammals, ARS2 interacts with DROSHA and is required for efficient processing and miRNA stability.

In *Arabidopsis*, pri-miRNAs are cleaved to miRNA in the nucleus by a complex composed of the RNase III enzyme DICER-LIKE 1 (DCL1), the double-stranded (ds) RNA binding protein HYPONASTIC LEAVES 1 (HYL1), and the scaffold TOUGH (TGH) (Figure 2A) [40]. Following DCL1 cleavage, the miRNA/miRNA* duplex is methylated by HEN1, which protects the miRNA from degradation, and is exported into the cytoplasm by HASTY, an Exportin 5 orthologue [41–43] . Once exported, miRNA are bound by ARGONAUTE 1 (AGO1) and induce degradation or inhibit translation of its target transcripts [44,45].

Early studies showed that SE, within the context of the nuclear cap complex, is required for proper miRNA biogenesis. SE is able to physically interact with pri-miRNA, as well as with the CBC [46–48]. Both *cbp* and *se* mutants accumulate levels of pri-miRNA [23,49], indicating their involvement at the early stages of processing. Indeed, SE and HYL1 directly interact with DCL1 and are required for the efficient and accurate cleavage of pri-miRNA and pre-miRNA in nuclear dicing bodies (D-bodies) [47,50–53]. SE also interacts with the scaffold TGH [54]. HEN1 interacts with the same regions of DCL1 and HYL1 as SE, and does not interact with SE itself, supporting the notion that SE acts upstream of miRNA methylation, and is possibly released in order for HEN1 to function [55]. This work established SE as a protein able to bridge interactions between

the cap complex, miRNA transcript, and the microRNA processing machinery (Figure 2A).

1.4 SERRATE structure/function

An important key to understanding how ARS2/SERRATE mediates interactions between the 5'-cap and 3'-processing machineries is its structure. Machida et al. published the crystal structure of *Arabidopsis* SE that provided the first glimpse of the domain architecture of an ARS2 orthologue [53]. The SE core structure consists of 3 domains in a walking man-like conformation, with the N-terminal, Mid, and zinc finger (ZnF) domains forming the leading leg, body, and lagging leg, respectively; the N and C-termini are unstructured (Figure 3) [53]. Deletion and point mutants have been used to assess the functions of these domains in *Arabidopsis*.

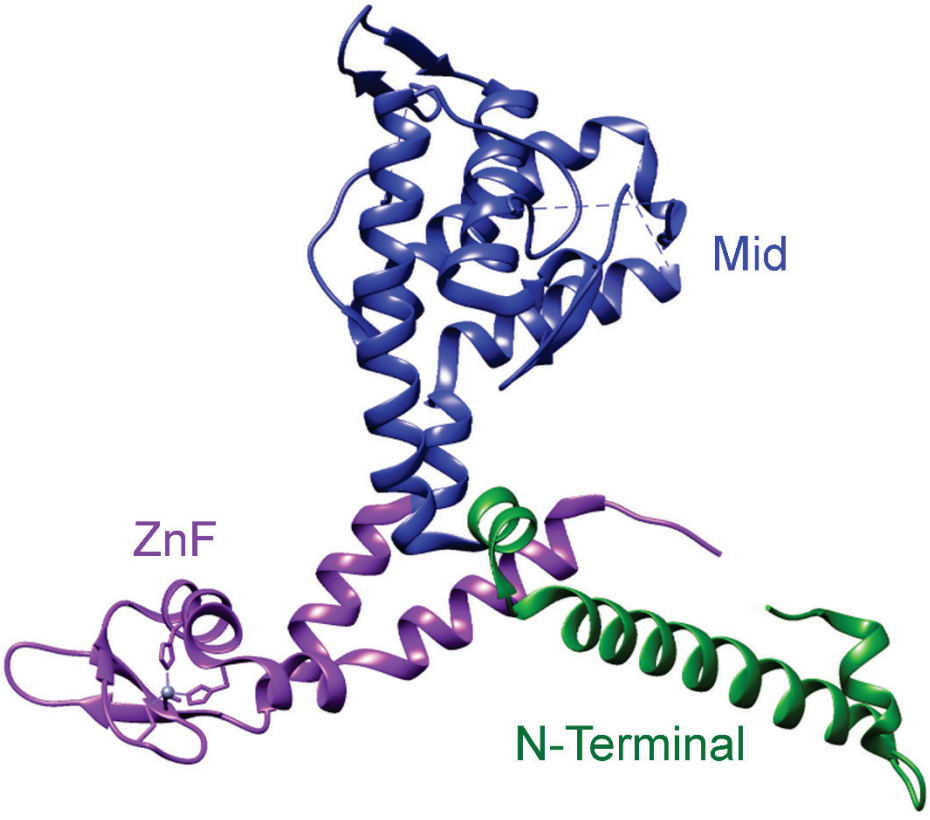

Figure 3 Structure of *Arabidopsis* SERRATE

Ribbon model of SERRATE (SE) structure, from PDB 3AX1 [53]. SE adopts a walking man-like topology, with the N-Terminal (green), Mid (blue) and zinc finger (ZnF) (purple) domains forming the leading leg, body, and lagging leg, respectively.

The interactions between SE, DCL1, and RNA are primarily mediated through the unstructured arginine-rich N-terminus, ZnF domain and unstructured C-terminus [47,53]. *In vitro* binding assays have mapped the RNA binding sites to a strong affinity site located within the unstructured N-terminus and a weaker site located within the ZnF domain and unstructured C-terminus [47,53]. Similarly, the plant microprocessor RNase III, DCL1, interacts both with the unstructured N-terminus and the ZnF domain [47]. Interestingly, the unstructured N-terminus is dispensable for pri-miRNA processing *in vitro* [47], suggesting the interactions through the ZnF and unstructured C-terminus are sufficient. In support of this hypothesis, the SE ZnF is essential for stimulating pri-miRNA cleavage *in vitro* [47]. In fact, the ZnF along with the C-terminal tail are sufficient to rescue the *se-1* mutant morphology and miRNA levels *in vivo* [53]. Thus, the ZnF and C-terminus form a critical core for miRNA processing, with additional non-essential interactions formed through the N-terminus.

These advances highlighted the structure/function relationship in *Arabidopsis* SE, yet there was a complete lack of corresponding structure/function information for metazoan ARS2. Additionally, the SE structural requirements had only been characterized for protein and RNA in the miRNA biogenesis pathway, but mammalian ARS2 had been implicated in processing both miRNA and replication-dependent histone (RDH) RNA (section 1.6 ARS2 and replication-dependent histone processing) by the time my project commenced, suggesting there was much more to be learned of ARS2 function.

1.5 ARS2 and microRNA biogenesis

The role of ARS2/SERRATE and the CBC in miRNA biogenesis is conserved in metazoans [2,15]. In mammals, miRNA are transcribed as precursors by RNAP II [34,35]. They are capped, and subsequently trimmed in the nucleus by the Microprocessor complex consisting of the RNase III DROSHA, and the RNA binding DiGeorge syndrome chromosomal region 8 (DGCR8) protein [56]. ARS2 interacts with DROSHA, and is required for the stability and efficient processing of primary miRNA (pri-miRNA) to precursor miRNA (pre-miRNA) by the Microprocessor [2]. Interestingly,

ARS2 is thought to act as a cofactor for DROSHA, as it influences the efficiency and specificity of DROSHA activity (Figure 2C) [2,24].

In *Drosophila*, the miRNA biogenesis process is very similar to mammals; pri-miRNAs are cleaved to pre-miRNA in the nucleus by the Microprocessor complex [57]. Following export into the cytoplasm, pre-miRNA are cleaved by Dicer-1 to generate mature miRNA, which are then loaded onto an Ago1-dependent RNA-induced silencing complex (RISC) [58–60]. As in mammals, *Drosophila* ARS2 (dArs2) interacts with the CBC and Microprocessor, and is required for pri-miRNA processing and stability (Figure 2B) [15]. *Drosophila* also use RNA interference (RNAi) for innate immunity to protect against viral infection [15]. Viral dsRNAs are processed into small interfering RNA (siRNA) by Dicer-2 in the cytoplasm [58,61]. Dicer-2, along with the dsRNA-binding protein R2D2, are required for loading Ago2-RISC, which mediates siRNA silencing [58,61,62]. Interestingly, dArs2 also interacts with cytoplasmic Dicer-2 and is required for the processing of presumably uncapped long dsRNA into siRNA [15]. Depletion of dArs2, CBP20, or CBP80 in flies results in an increased susceptibility to infections by RNA viruses, due to defective siRNA biogenesis [15]. Thus, ARS2, CBP20, and CBP80, may have cofactor roles in RNA processing that extend beyond their role in the nuclear cap complex. Collectively, this work established that ARS2/SE has a conserved role in miRNA biogenesis, and physically couples the CBC to the Microprocessor in insects and animals, or to DCL1 in plants (Figure 2).

1.6 ARS2 and replication-dependent histone processing

The demonstration that ARS2/SERRATE is part of the nuclear cap complex that physically bridges the cap to the Microprocessor raised the possibility that ARS2 would be required for other cap complex-dependent processes. The first indication of this came from work by Kiriyama et al. and Gruber et al. who showed that ARS2 is required for processing replication-dependent histone transcripts [16,24]. As cells enter S phase of the cell cycle, histone proteins are rapidly synthesized to package newly replicated DNA [63]. Replication-dependent histone (RDH) transcripts are transcribed by RNAP II, are m7G-capped, and are intronless [64]. At their 3’-ends, RDH transcripts contain a conserved stem-loop, and are the only known metazoan mRNA not polyadenylated [64].

Instead, RDH transcripts undergo endonucleolytic cleavage, via cleavage and polyadenylation specificity factor (CPSF) subunit CPSF73, between the conserved 3'-end stem-loop and a histone downstream element (HDE) (Figure 4) [65–67]. RDH processing is cap-dependent and requires the coordination of several multimeric complexes. For example, the stem-loop is bound by stem-loop binding protein (SLBP), while the HDE forms base pairs with U7 snRNA, which is part of a multi-subunit U7 small nuclear ribonucleoprotein (snRNP) complex that acts as a molecular ruler to guide endonucleolytic cleavage by the histone pre-mRNA cleavage complex (HCC), which consists of CPSF73, CPSF100, and symplekin [68–73]. In addition, negative elongation factor (NELF) not only associates with histone loci at the promoter, but is also required for processing RDH pre-mRNA at the 3'-end (Figure 4) [22]. Misprocessing of RDH transcripts results in aberrant read-through and polyadenylation [24,74].

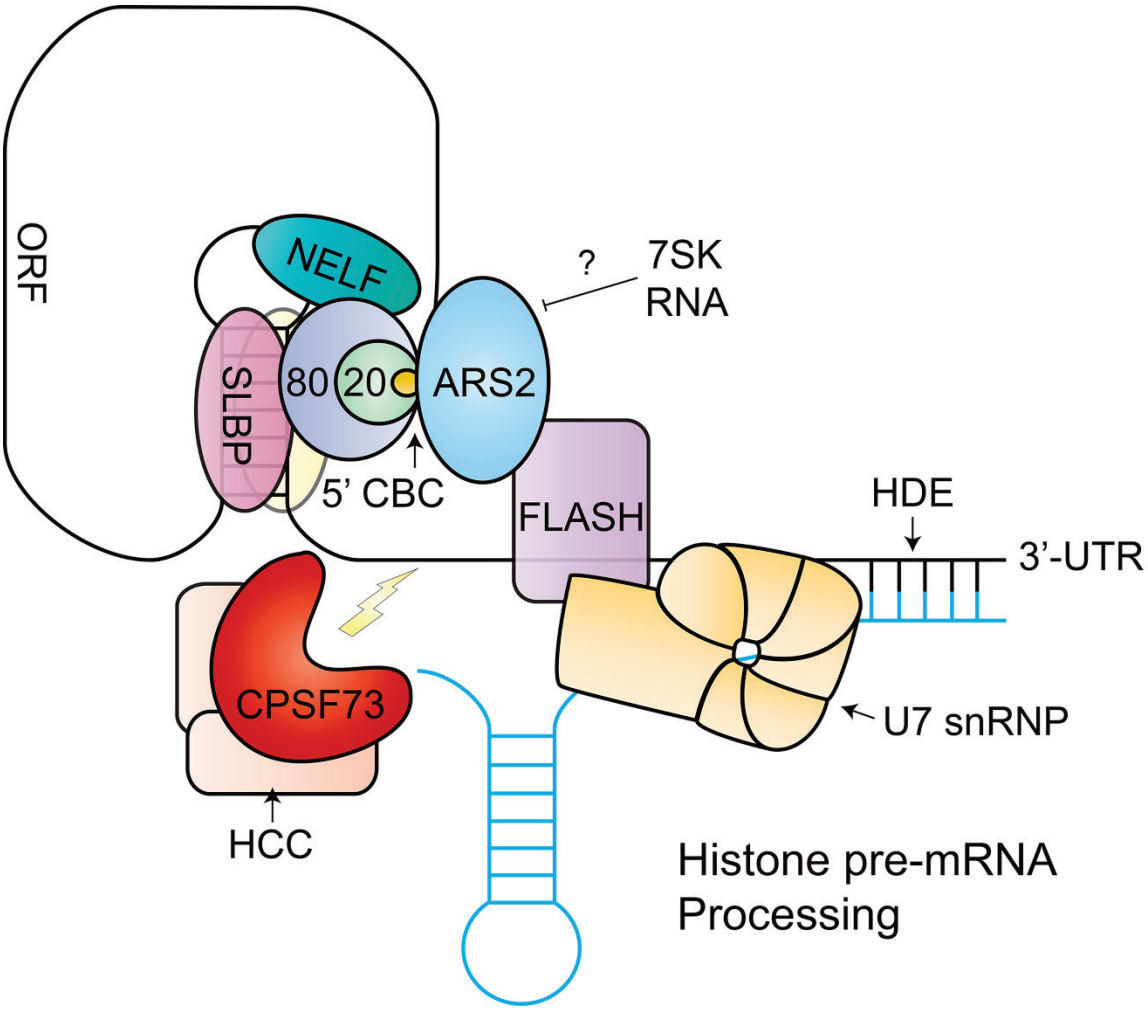

Figure 4 ARS2 and replication-dependent histone processing

Binding of the U7 snRNP to the HDE and FLASH are required for endonucleolytic cleavage by CPSF73, which is part of a histone pre-mRNA cleavage complex (HCC). CBCA makes multiple contacts with the 3' processing machinery, which are required for accurate cleavage. CBP80 interacts with SLBP and NELF, and ARS2 interacts with FLASH. Additionally, histone 3'-end formation is negatively regulated by the noncoding 7SK RNA through an interaction with ARS2. U7 snRNA is shown in blue.

ARS2 mediates interactions with RDH pre-mRNA and multiple components of the RDH 3'-processing machinery, and is required for RDH processing [16,24]. The CBC (CBP20/80) interacts with SLBP and NELF and is sandwiched between the two [22,26]. Meanwhile, mammalian ARS2 interacts with RDH RNA and FLICE-associated huge protein (FLASH) [16,24]. In turn, FLASH directly interacts with the U7 snRNP component LSM11, and is required for RDH pre-mRNA endonucleolytic cleavage [75–78]. Knockdown (KD) of either ARS2 or FLASH disrupts the formation of histone locus

bodies (HLBs), increases aberrant polyadenylated histone transcripts, and decreases histone protein levels [16,24,79]. Most likely as a result, cells deficient in either ARS2 or FLASH are delayed in S-phase progression [16,79]. As an added layer of complexity, the noncoding 7SK RNA, an important factor in regulating transcriptional elongation, interacts with ARS2 and negatively regulates RDH 3'-end processing, potentially by sequestering ARS2 [24]. The 7SK snRNP binds and inhibits cyclin-dependent kinase 9 (CDK9), which phosphorylates the RNAP II C-terminal domain (CTD) and NELF to promote transcriptional elongation [24,80]. Further work is needed to understand how 7SK RNA regulates ARS2, as well as the precise role of ARS2 in RDH 3'-end formation (Figure 4). Nevertheless, the finding that ARS2, as part of the nuclear cap complex, interacts with both miRNA and RDH processing machinery suggested ARS2 has a broader role in physically coupling RNAP II transcript processing.

1.7 ARS2 and transcription termination

Confirmation of a broader role in RNAP II transcription came with the demonstration that ARS2 and the CBC are required for inducing cap-proximal transcription termination for snRNA, RDH RNA, promoter upstream transcripts (PROMPTs) and mRNA. This was demonstrated through proteomic analysis using immunoprecipitations of the machinery involved in these processes, and knockdown experiments of ARS2 or CBP20/80, which increases 3' read-through transcripts for each of these RNA classes [25,26]. Interestingly, transcripts longer than ~1 kb were largely unaffected by ARS2 depletion with regards to 3' read-through [25], suggesting CBCA or a CBCA-interacting factor have a mechanism to limit their activity to promoter proximal areas. However, very little is currently known mechanistically about how ARS2 may interact with these processes.

1.8 ARS2 and the exosome

A major role of the cap complex is in RNA quality control, and in limiting the effects of promiscuous RNAP II transcription. Recently, ARS2 was shown to participate in targeting transcripts to the nuclear RNA exosome through interaction with the human nuclear exosome targeting (NEXT) complex [26,81]. This complex recruits the nuclear RNA exosome to degrade PROMPTs and 3'-extended snRNA [26,81–84]. NEXT is

composed of the RNA helicase hMTR4, the zinc-knuckle ZCCHC8, and the RNA recognition motif (RRM)-containing RBM7 [26,81]. Affinity capture mass spectrometry revealed a stoichiometric interaction between CBCA, NEXT, and another ZnF protein ZC3H18 [26]. Depletion of CBCA, NEXT components, or ZC3H18, results in accumulation of PROMPTs [25,26]. RBM7 associates with newly synthesized RNA, is enriched at regions close to the 5' cap, and RBM7-RNA interaction is disrupted following CBC KD [82]. These data indicate that ARS2 and the CBC recruit the NEXT complex to newly synthesized RNAP II transcripts that are destined for exosomal destruction (Figure 5A).

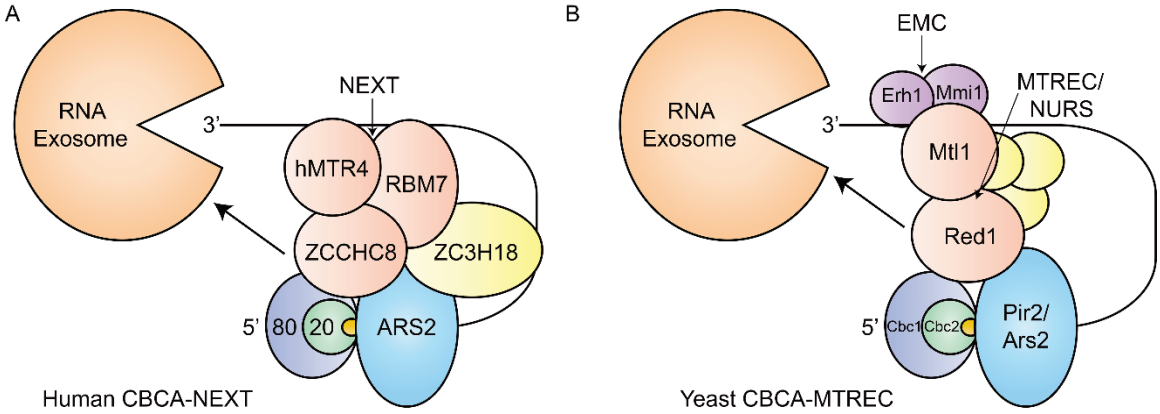

Figure 5 ARS2 and the exosome

A) CBC-ARS2 targets aberrant transcripts to the RNA exosome through the NEXT complex, composed of hMTR4, ZCCHC8 and RBM7, and the NEXT-associated protein ZC3H18. B) *S. pombe* Pir2/Ars2 and CBC proteins Cbc1/Cbc2 target aberrant transcripts to the exosome through the MTREC/NURS complex, which is minimally composed of Mtl1 and Red1, and is associated with several subcomplexes, including the Erh1-Mmi1 complex (EMC) (see 1.13 ARS2 and heterochromatin formation).

The role of ARS2 and the cap complex in RNA quality control is conserved in fission yeast. In *Schizosaccharomyces pombe* (*S. pombe*), Pir2/Ars2 (yeast orthologue of ARS2) and Cbc1-Cbc2 (orthologues of CBP80 and CBP20, respectively) interact with the Mtl1-Red1 core (MTREC) complex [85,86] (alternatively named nuclear RNA silencing (NURS) complex) [87]. Mtl1 is a Mtr4-like helicase and Red1 is a ZnF protein [85,88]. Despite not having sequence similarities, CBCA-MTREC is likely the fission yeast functional equivalent of the human CBCA-NEXT complex. MTREC, along with Pir2/Ars2 and the CBC, are essential for targeting RNA to the exosome in *S. pombe*

(Figure 5B) [85–88]. Similar to the fate of PROMPTs in mammals, MTREC in *S. pombe* delivers polyadenylated cryptic unstable transcripts (CUTs) to the exosome for destruction [86].

1.9 ARS2 and mRNA 3'-end formation

3'-end processing of mRNA consists of endonucleolytic cleavage followed by polyadenylation. The cleavage and polyadenylation machinery are recruited co-transcriptionally through the CTD of RNAP II [89]. A large number of factors are required for proper mRNA processing. These include the multi-subunit cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I_m (CFI_m) and $CFII_m$ complexes, as well as poly(A) polymerase [90,91]. $CFII_m$ bridges RNAP II and the nascent transcript to CPSF and CFI_m [92], and is required for pre-mRNA cleavage by CPSF73 prior to polyadenylation [90]. CBCA interacts with and stabilizes the pre-mRNA 3' processing machinery, and is required for efficient cleavage, but not polyadenylation, of short pre-mRNA [21,25]. Interestingly, ARS2 was shown to interact with $CFII_m$ component CLP1, and depletion of either ARS2, CLP1, or another $CFII_m$ component PCF11, induces 3' read-through reminiscent of CBCA KD, suggesting the interaction is functional (Figure 6A) [25,93].

1.10 ARS2 and snRNA 3'-end processing

ARS2 is also required for snRNA 3'-end formation [25], a process that shares NELF with RDH 3'-end processing machinery [22,94]. snRNA 3'-end formation is mediated by the twelve subunit Integrator complex that contains homologues of CPSF73 and CPSF100 (Int11 and Int9, respectively) [95]. NELF interacts with Integrator and is required for accurate snRNA 3'-processing [94]. Thus, similar to its role in RDH 3'-end formation, CBCA may be affecting snRNA processing through an interaction with NELF (Figure 6B).

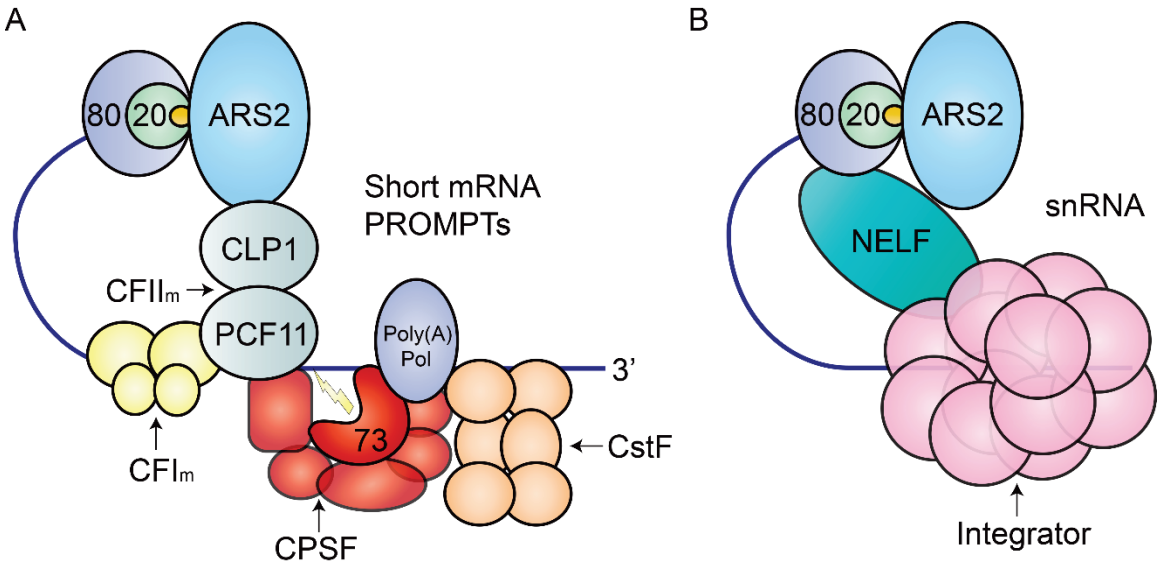

Figure 6 ARS2 and 3'-end formation

A) ARS2 is required for 3'-end cleavage by CPSF73 at short mRNA transcripts and PROMPTs, and interacts with CLP1/PCF11 of the $CFII_m$ complex, which bridges the CPSF and CFI_m complexes. The CstF complex and Poly(A) polymerase are also shown. B) ARS2 is also required for 3'-end formation by the Integrator complex at snRNA transcripts. CBCA may mediate an interaction with Integrator through NELF.

1.11 ARS2 and export

In metazoans, export of snRNA requires CBCA, phosphorylated adapter for RNA export (PHAX), the export receptor CRM1/XPO1, and RanGTP [96,97]. PHAX directly interacts with the CBC and snRNA [25,26,97,98]. Notably, ARS2 stimulates PHAX binding to the CBC, promoting the formation of a stable complex called CBCAP, composed of CBC, ARS2, and PHAX [25]. The mechanism underlying the allosteric regulation of PHAX binding by ARS2 is currently unclear and requires further investigation. However, PHAX is phosphorylated by CK2 kinase, and PHAX must be in its phosphorylated state in order for CRM1 to be recruited along with RanGTP, and for snRNA export to occur (Figure 7A) . In addition to snRNAs, the CBCAP complex can bind m7G capped snoRNAs, and PHAX binding is required for their intranuclear transport to Cajal bodies, where snoRNAs are further processed [25,100]. This suggests that ARS2 may also be involved in snoRNA transport, although this has not been tested.

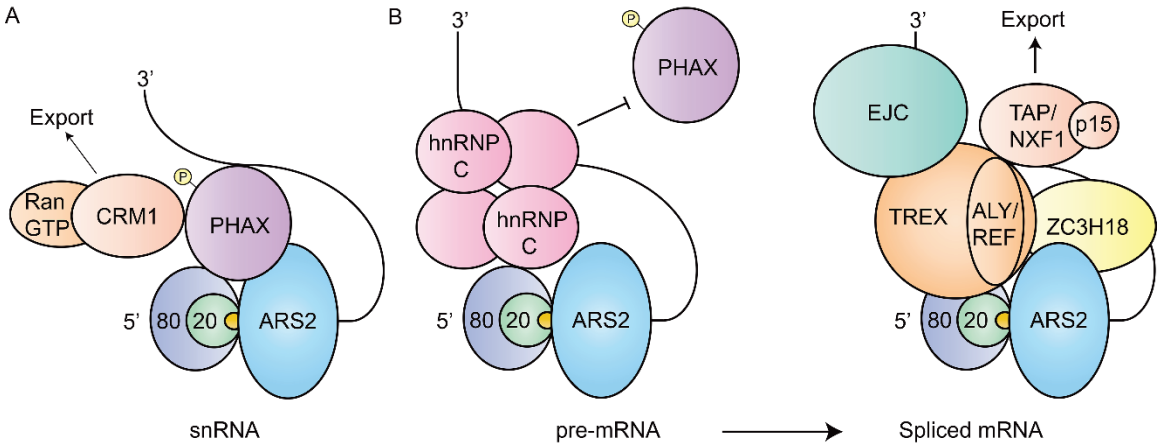

Figure 7 ARS2 and export

A) ARS2 mediates snRNA export by stimulating phosphorylated adaptor for RNA export (PHAX) binding to CBC, forming the CBCAP complex. snRNA bound to CBCAP is then exported by CRM1/RanGTP. B) Pre-mRNA transcripts longer than ~200-300 nt are bound by the hnRNP C tetramer, which interacts with the CBC and inhibits PHAX binding, thereby committing them to the mRNA export pathway (left). Splicing and the presence of CBCA, ZC3H18, and the exon junction complex (EJC) stimulates TREX recruitment to the 5'-end of mRNA. The export adaptor ALY/REF, a component of TREX, mediates handover to TAP/NXF1 and p15 for export into the cytoplasm (right).

The mechanism of the specific interaction between PHAX and snRNA provides insight into understanding how different types of transcripts are distinguished [25]. Differential processing of snRNAs is achieved through the preferential binding of the heterogeneous nuclear ribonucleoprotein (hnRNP) C tetramer to transcripts longer than 200-300nt [101]. hnRNP C directly interacts with CBP80 and RNA, and competitively inhibits PHAX binding [101]. Therefore, U snRNAs, which are typically <200 nt, bind PHAX by default and are exported via CRM1-RanGTP, while longer mRNA transcripts are inhibited from using this pathway through the action of hnRNP C (Figure 7B) [101,102].

CBCA also plays an important role in mRNA export, largely mediated by the multi-subunit transcription export (TREX) complex, which acts as an adaptor for the export receptor TAP/NXF1 [103–108]. Human TREX is composed of the THO complex (THOC1, THOC2, THOC5, THOC6, THOC7, Tex1), CIP29, UAP56 and ALY/REF [109–111]. TREX is involved in release of mRNA from nuclear speckle domains and coupling to the export receptor [112]. The presence of the exon junction complex (EJC), which is deposited following splicing, stimulates TREX recruitment

[103,104,107,109,113], ensuring correctly processed transcripts are exported. Following binding of TREX subunits to mRNA, there is an interaction between ALY/REF, THOC5, and the export receptor TAP/NXF1, which induces a conformational change in the receptor [114], and allows for a handover of the mRNA to the TAP/NXF1-p15 heterodimer [115], which then transports the mRNA through the nuclear pore [108,116].

ALY/REF directly interacts with CBP20/80 [101,117,118], while THOC2, UAP56, and CIP29 independently interact with ARS2 [110], and these interactions between TREX and CBCA are required for efficient mRNA export [118,119]. ARS2 KD results in accumulation of mRNA in nuclear speckle domains [119]. Interestingly, Zinc-knuckle protein ZC3H18, which interacts with ARS2 as part of the CBC-NEXT complex [26,81,82], also interacts with the TREX complex [119]. ZC3H18 KD prevents efficient TREX recruitment to RNA and also results in an accumulation of mRNA in nuclear speckle domains [119]. Although the details are less understood, intronless mRNA also relies on components of this pathway, including CBCA, TREX, ZC3H18, and TAP/NXF1 [105,106,118,119]. Thus, CBCA, through ZC3H18, may control recruitment of TREX and export of transcripts that are capped and correctly processed (Figure 7B).

Recently, an alternative mammalian CBC was discovered where NCBP3 could bind CBP80 and the m7G cap in place of CBP20 [120]. Interestingly, both NCBP3 and CBP20 bound many common factors, including ARS2, and only double KD of NCBP3 and CBP20 significantly disrupted mRNA export, suggesting some redundancy between the two complexes; however, NCBP3 preferentially interacted with TREX, while CBP20 exclusively bound snRNA and PHAX, indicating these two complexes may have developed specialized functions in mRNA and snRNA export [120]. Deciphering the roles of this alternative CBC, as well as how it interacts with ARS2, will be an exciting area of research.

1.12 ARS2/SERRATE and splicing

Pre-mRNA splicing, whereby introns are removed and exons are ligated together in a two-step transesterification reaction, is carried out by the spliceosome complex. For excellent, detailed reviews on splicing, please see [121–123]. In *Arabidopsis*, CBP80/ABH1, CBP20 and SE are required for cap-proximal splicing and alternative

splicing, as mutants primarily affect retention of the first intron and alternative 5'-splice site selection [23,48,124]. The mechanistic details of how SE mediates splicing have not been resolved. In mammals, the CBC facilitates cap-proximal splicing, interacts with the U4/U6·U5 tri-snRNP in a RNA-independent manner, is required for co-transcriptional spliceosome assembly, and is involved in alternative splicing [125–128]. Although mammalian ARS2 co-purifies with multiple splicing factors [26,129,130], whether ARS2 directly participates in splicing in metazoans has not been established.

1.13 ARS2 and heterochromatin formation

Recently, a requirement for ARS2 in heterochromatin formation has been shown in fission yeast. In *S. pombe*, Pir2/Ars2 is required for heterochromatin formation at centromeres, telomeres, heterochromatin domains (HOODs), and a subset of meiotic loci [87,131]. Pir2/Ars2 KD results in decreased histone H3 lysine 9 methylation (H3K9me), a hallmark of heterochromatin formation, at these sites [87,131]. The requirement of Pir2/Ars2 for heterochromatin formation at these diverse regions involves several partially overlapping pathways. The best-characterized pathway is centromeric silencing, which is mediated by RNA interference (RNAi) machinery and heterochromatin formation [132], and requires Pir2/Ars2 [87,131]. Transcription from repeat elements in these regions forms double-stranded RNAs (dsRNA), which are cleaved by Dicer to generate siRNAs that are loaded onto an Ago1-containing RNA-induced transcriptional silencing (RITS) complex [133]. RITS is then recruited to peri-centromeric regions for heterochromatin formation and silencing [133]. RITS is composed of Ago1, the chromodomain protein Chp1, and the GW protein Tas3 [133,134]. Heterochromatin formation is mediated through an association between the RITS complex and the Clr4-Rik1-Cul4 (CLRC) complex, comprised of Clr4, a histone methyltransferase, Rik1, a heterochromatin targeting protein, and Cul4, an E3 ubiquitin ligase [135]. Together, RITS and the CLRC complex are responsible for the initiation of H3K9 methylation [136]. Clr4 then binds to H3K9me and promotes the spread of heterochromatin [136]. Given the ability of dArs2 to associate with Dicer2 in *Drosophila*, and dArs2's role in siRNA biogenesis in this organism [15], it is possible that Pir2/Ars2 is regulating heterochromatin at the centromeres through interactions with Dicer. However, Pir2/Ars2

also associates with RITS-associated CLRC complex in pull-down assays, and therefore likely has additional roles in silencing beyond enhancing Dicer activity (Figure 8A) [87].

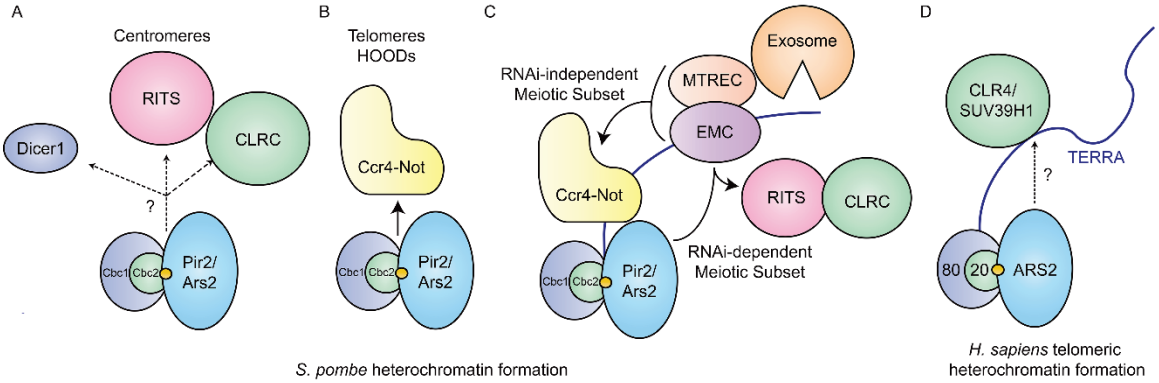

Figure 8 ARS2 and heterochromatin formation

A) *S. pombe* Pir2/Ars2 is required for heterochromatin formation at centromeres. Centromeric heterochromatin formation requires RNAi machinery (Dicer1 and RITS) as well as CLRC. Whether Pir2/Ars2 mediates centromeric heterochromatin through Dicer1, RITS and/or CLRC requires further study. B) Pir2/Ars2 also regulates heterochromatin formation at telomeres and heterochromatin domains (HOODs) in a Ccr4-Not dependent manner. C) Pir2/Ars2 regulates heterochromatin at a subset of meiotic loci through RNAi-independent and RNAi-dependent mechanisms, although the mechanisms are not fully understood. The RNAi-independent subset is thought to be mediated by the EMC through its ability to recruit MTREC and Ccr4-Not. The RNAi-dependent subset requires RITS and CRLC, which are thought to be recruited through the EMC. D) Human ARS2 regulates telomeric repeat-containing RNA (TERRA), which interact with the CLR4/SUV39H1 methyltransferase and are correlated with H3K9me levels. The mechanism of TERRA regulation by ARS2 and its involvement in telomeric heterochromatin formation is not known (indicated by question mark).

The second Pir2/Ars2-dependent pathway for heterochromatin formation is through the Ccr4-Not complex. This complex, along with Pir2/Ars2, mediates heterochromatin formation at telomeres, HOODs, and a subset of meiotic loci [131]. The Ccr4-Not complex in *S. pombe* is comprised of Ccr4, Caf1, Caf40, and five Not proteins (Not1-5) [137]. Ccr4-Not has deadenylation (Ccr4 and Caf1) and ubiquitination (Not4) activity and is responsible for mRNA turnover in the cytoplasm [137]. Interestingly, both enzymatic activities of the complex are required for heterochromatin formation [138]. The mechanism behind the requirement for Ccr4 deadenylation activity in establishing heterochromatin is unclear. Similarly, further work is required to understand the role of the ubiquitination activity of Not4. However, Not4 has been shown to trigger degradation of Jhd2, a histone demethylase, and thus may indirectly contribute to heterochromatin

formation through this mechanism [139]. Pir2/Ars2 co-purifies with all seven subunits of the Ccr4-Not complex, and heterochromatin formation at telomeres, HOODS, and some meiotic islands is Pir2/Ars2/Ccr4-Not-dependent (Figure 8B,C) [131]. A similar role for dArs2 and CCR4-NOT was found in retrotransponson silencing at telomeres in *Drosophila*, suggesting this silencing function is conserved [140,141].

Meiotic loci in fission yeast represent a third type of heterochromatin silencing. During vegetative growth, meiotic transcripts are recognized by Mmi1 through a determinant of selective removal (DSR) sequence found within the 3'UTR of meiotic transcripts [87,88,142]. Mmi1 forms a stable complex with Enhancer of rudimentary (Erh1) called Erh1-Mmi1 complex (EMC) [131]. This complex, through Mmi1, targets DSR-containing transcripts for degradation via the exosome (Figure 4) [85–88,131,142]. The EMC is also required for heterochromatin formation at meiotic loci during the vegetative state through both RNAi-independent and RNAi-dependent mechanisms [131]. The RNAi-independent process is facilitated through the ability of the EMC to recruit MTREC and the exosome [85–88,131,143]. This process requires transcription and is thought to rely on the ability of the MTREC complex to independently interact with the Ccr4-Not complex to mediate H3K9me [131,137,138]. An alternative mechanism of meiotic loci silencing is RNAi-dependent and is mediated through the EMC's ability to recruit the RITS complex and associated CLRC [131,144]. Both RNAi-dependent and -independent processes rely on Pir2/Ars2 (Figure 8C) [87,131]. However, it is currently unclear how Pir2/Ars2 is restricted to regulating heterochromatin formation at only a subset of meiotic loci.

In humans, ARS2 has been implicated in telomeric heterochromatin formation through its regulation of long noncoding telomeric repeat-containing RNA (TERRA) [145]. TERRA are transcribed by RNAP II, associate with telomeric DNA, and have been implicated in telomere maintenance [146]. ARS2 co-purifies with TERRA, and ARS2 KD in HeLa cells increases TERRA levels and the abundance of TERRA associated with telomeres [145]. Furthermore, TERRA itself directly interacts with the human homologue of the yeast Clr4 histone methyltransferase (SUV39H1) (Figure 8D) [147], and telomeric H3K9me levels are correlated with TERRA levels [147–149]. In *Arabidopsis*, SE and CBP80/CBP20 regulate the levels of hundreds of long noncoding RNAs (lncRNAs)

[150]. Thus, the role of ARS2 in lncRNA biogenesis is conserved, and it is likely ARS2 is important for the function of other lncRNAs in mammals. Taken together, ARS2 is tied to heterochromatin formation at diverse regions using multiple pathways in yeast and metazoans, although the mechanism remains unclear.

1.14 An anomalous role for ARS2 as a transcription factor

The vast majority of ARS2 functions involve its interactions with CBP80/20. However, there is one report of ARS2 as a transcription factor that does not fit this model. As mentioned, ARS2 has a key role in maintaining neural stem cells (NSCs) in the mouse brain [9]. Conditional *Ars2* knockout in subventricular zone NSCs decreased their self-renewal capacity and multipotency, with their fate skewed towards astroglial production [9]. This phenotype was rescued by SOX2, a transcription factor essential for NSC maintenance [9]. Curiously, ARS2 bound to a small region within the SOX2 enhancer in the presence of RNase [9]. The interaction between ARS2 and the enhancer was cell type-dependent [9]. Furthermore, the presence of this enhancer region was necessary for a SOX2 luciferase reporter to be expressed following ARS2 overexpression [9]. Expression of SOX2 was sufficient to rescue the defects in NSC self-renewal and multipotency of ARS2-deficient animals [9]. This work implicated ARS2 as a transcription factor. However, how ARS2 regulates SOX2 transcription is currently unclear. It is complicated by the fact that the *Sox2* gene is located within an intron of the lncRNA SOX2OT (overlapping transcript) [151], and SOX2OT positively regulates SOX2 [152–154]. As mentioned, ARS2 is implicated in lncRNA function [145,150]. Thus, it is plausible that some of the effects of ARS2 on SOX2 expression may be related to ARS2 regulation of SOX2OT biogenesis. Further work is needed to discern the precise role of ARS2 in SOX2 expression.

1.15 ARS2 and aging

The early embryonic lethality of *Ars2* null embryos and the NSC self-renewal and multipotency defect in the mouse brain demonstrated the requirement for ARS2 in stem cell maintenance and differentiation at early developmental time points [7,9]. Interestingly, the levels of ARS2 in the NSCs of aging mice also appear to be important. Mice with an extra copy of normally regulated Ink4/Arf/p53 have an elongated lifespan

and are protected from age-related decline [10,155]. Strikingly, old Ink4/Arf/p53 mice have increased ARS2 levels in their subventricular zone and dentate gyrus NSCs, which have increased self-renewal and neurogenic capacity [10]. This is consistent with ARS2's role in NSC maintenance, and demonstrates a correlation between ARS2 levels in NSCs and anti-aging. Collectively, this work established the importance of ARS2 in the NSC compartment.

1.16 Outstanding questions for ARS2 in progenitor cells

Early in my thesis work it became clear ARS2 was required in NSCs [9], but the ARS2-dependent mechanisms contributing to stem cell maintenance and differentiation remain unresolved. Moreover, there was controversy between whether ARS2 acts as a transcription factor involved in recruitment of RNAP II, or as a co-transcriptional regulator. Specifically, the Lai lab implicated ARS2 as a SOX2-specific transcription factor, and that ARS2 regulation of SOX2 was the main contributor to stem cell maintenance and multipotency [9]. While they presented compelling evidence, it conflicted with all prior and subsequent work showing ARS2 acts with the CBC to couple the 5'-cap with numerous maturation events on RNAP II transcripts [2,15,16,24–26,119,145]. A second controversial finding was by Gruber et al., who showed that ARS2 expression was exclusive to proliferating cells and was down-regulated in quiescent cells [2,24]. This did not seem consistent with the emerging role of ARS2/SE as part of the nuclear cap complex. They also suggested that ARS2 was required for all phases of cell cycle and did not see accumulation in any one phase [2]. This was in contrast to the results of Kiriyama et al., who saw an accumulation of cells in S phase following ARS2 KD in human cells, which they suggested was a result of RDH mRNA misprocessing [16]. However, neither group examined cell cycle kinetics, but instead relied on single time points [2,16]. For these reasons, there was a need to re-examine the role of ARS2 in cell cycle progression, and particularly how this related to the maintenance and differentiation of neural progenitor cells.

1.17 ARS2 and the retina

The developing mouse retina provides an excellent model system for examining ARS2's role in neural stem cells (NSCs). It is a portion of the central nervous system that

is highly accessible for *in vivo* and *in vitro* analysis, as a population of multipotent progenitors continue to divide and differentiate after birth. The one glial and six neuronal cell types of the mouse retina are derived from multipotent progenitors that are generated by exiting the cell cycle in an overlapping sequential order starting embryonically and continuing postnatally [156] (Figure 9 and Figure 10). Thus, retinal progenitor differentiation is coupled to their proliferation kinetics and timing of cell cycle exit. There are numerous intrinsic and extrinsic factors contributing to cell fate specification and terminal differentiation that are dynamically regulated over space and time [157]. For instance, the cell cycle duration increases throughout retinal development [158], and cell cycle length and timing of cell cycle exit can influence cell fate decisions. For example, deletion of cyclin D1 (*Ccnd*$^{-/-}$) in mice delays retinal cell cycle progression, hastens cell cycle exit, and increases the proportion of ganglion cells and photoreceptors [159]. Additionally, inducing premature cell cycle exit by misexpression of cyclin-dependent kinase inhibitors alters cell fate determination [160,161]. The postnatal period of cell proliferation and differentiation allows for the examination of the interplay between cell cycle progression and cell fate decisions. In addition, our collaborators in the Chow lab are experts in the field of retinal development, which provided us with a unique opportunity to examine the role of ARS2 in retinal progenitor cells.

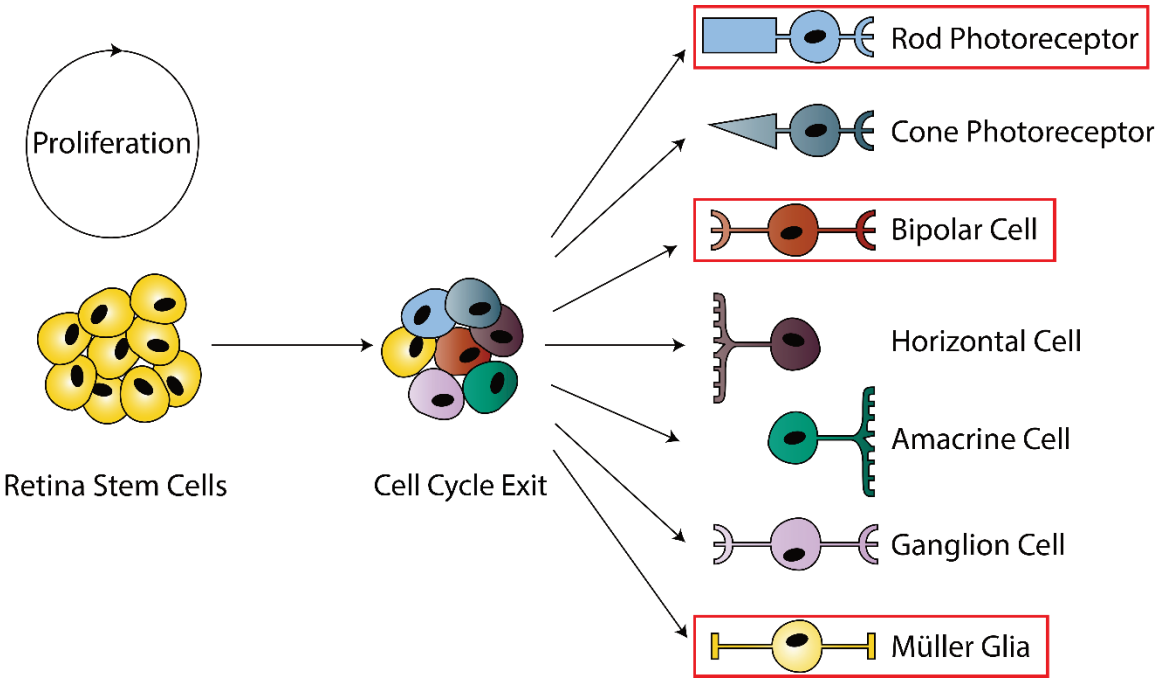

Figure 9 Retinal progenitor cells differentiate to produce 7 major cell types

Multipotent retinal stem cells (yellow, left) proliferate during embryonic development and some continue proliferating postnatally. Progenitors will eventually exit the cell cycle and differentiate to form the six major neuronal cell types (rod and cone photoreceptors, bipolar, horizontal, amacrine and ganglion cells) as well as Müller glial cells. Late-born progenitors (rod photoreceptors, bipolar cells and Müller glia), which are the predominant cell types generated postnatally, are highlighted with boxes.

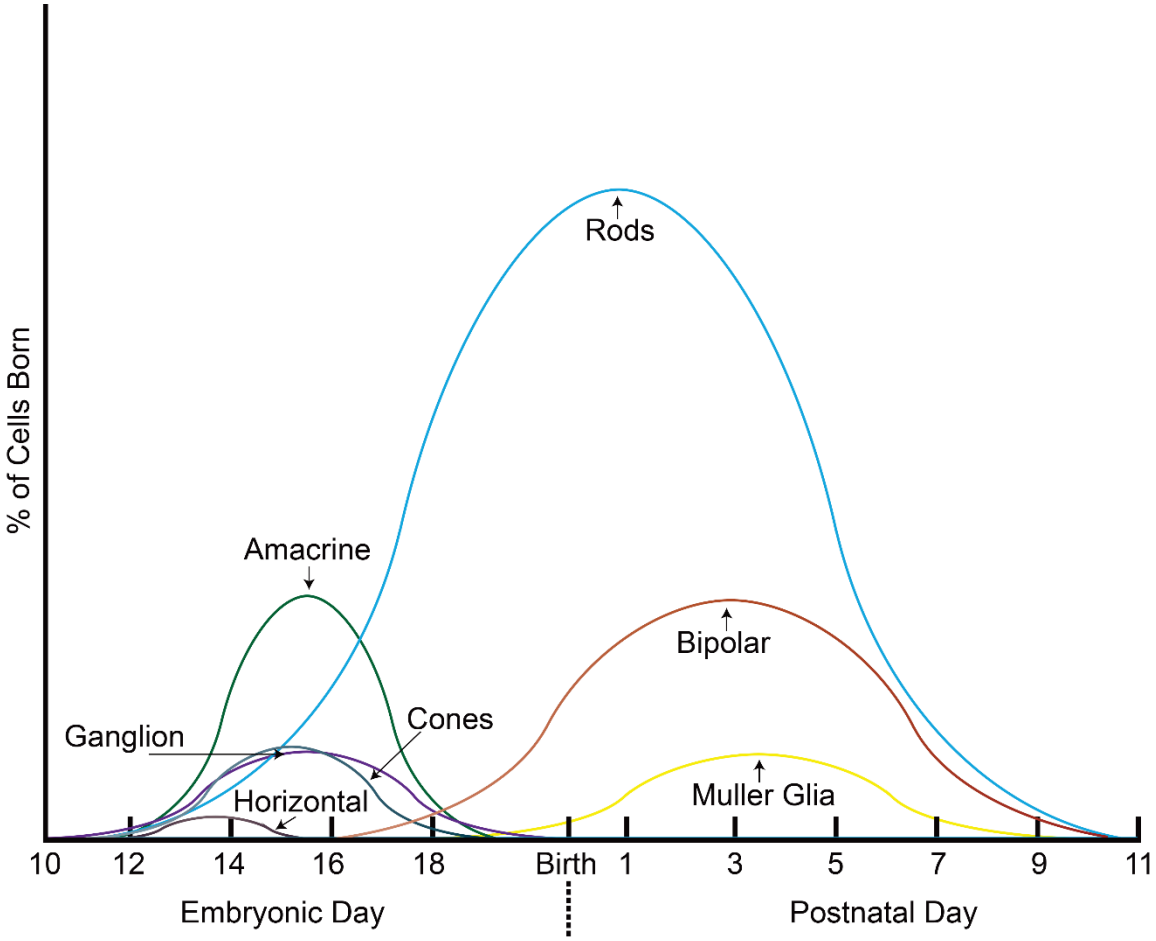

Figure 10 Approximate birth order in the developing mouse retina

The seven major cell types become postmitotic in an overlapping and sequential wave starting during embryogenesis and continuing postnatally. The approximate timing and proportion of cell types born are shown. The figure is adapted from [156,162].

Inextricably linked to the cell cycle in the mouse retina is interkinetic nuclear migration (INM), a process whereby nuclei of proliferating cells oscillate between the apical and basal positions of the neuroblastic layer, and the position correlates with the phases of the cell cycle [163]. Furthermore, cell cycle progression is a prerequisite for INM [164], and cells with nuclei that travel greater distances basally are significantly more likely to produce neurogenic daughter cells in Zebrafish retina [165]. Additionally,

NOTCH signaling is differentially regulated throughout the cell cycle [166] and exerts its effects in an apical-basal gradient [167], providing a link between cell cycle progression, INM, and differentiation in the retina (Figure 11). Importantly, our collaborators in the Chow lab discovered that disrupting ARS2 perturbs interkinetic nuclear migration (INM) in the mouse retina, resulting in a prolonged presence at a basal position, where S phase occurs [168]. This further supported a role for ARS2 in promoting progenitor proliferation, and highlighted the need to further investigate whether ARS2 is required for cell cycle progression and fate determination in retinal neural progenitor cells.

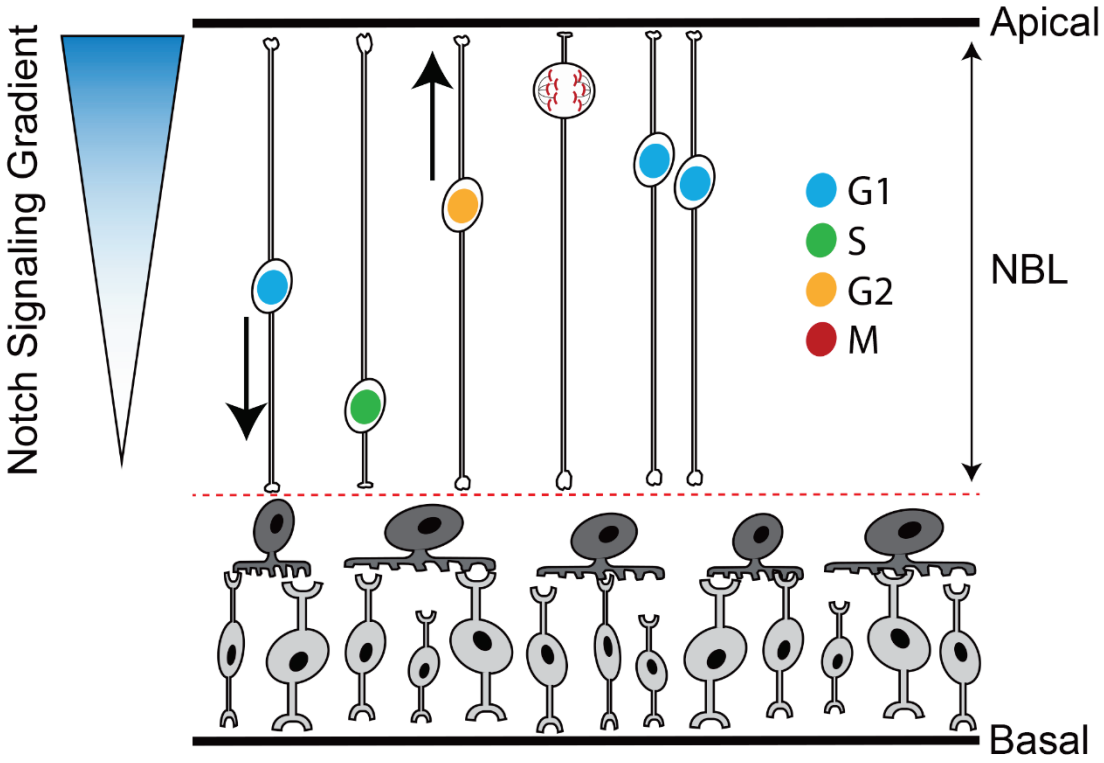

Figure 11 Interkinetic nuclear migration and NOTCH signaling

Nuclei of proliferating cells in the neuroblastic layer (NBL) undergo interkinetic nuclear migration (INM) as they progress through the cell cycle. During G1, nuclei (blue) migrate towards the basal side of the NBL and undergo DNA synthesis in S phase at a basal position (green). Upon completion of S phase, nuclei rapidly migrate towards the apical margin during G2 (yellow), and undergo mitosis (red) at the apical margin. Many intrinsic factors regulate this dynamic migration. Additionally, nuclei are subject to extrinsic signaling factors as they oscillate between the apical and basal position, such as NOTCH signaling, which is highest at the apical margin and lowest at the basal position.

1.18 ARS2 in myoblast progenitor cells

Myoblast progenitor cells represent another useful model system for studying ARS2 functions in progenitor proliferation and differentiation *in vitro*. Proliferative myogenic progenitors must exit the cell cycle prior to cell fusion to generate differentiated multinucleated myotubes [169]. Thus, like the developing retina, myogenic differentiation is coupled to cell cycle exit (Figure 12). Primary or C2C12 mouse myoblast cells can be cultured as proliferating progenitors by maintaining a low cell density and providing high serum levels in growth media [170–172], allowing us to study ARS2's role in replication-dependent histone (RDH) mRNA processing during progenitor proliferation. Confluent myoblast cells can be induced to differentiate in a well-characterized process by incubating cells in low serum conditions [171]. Importantly, Jennifer Christie, a student in our lab, found that ARS2 knockdown or overexpression during differentiation prevents myotube formation (unpublished observations), indicating a requirement for ARS2 in the differentiation process of this progenitor cell line.

Another advantage of studying ARS2 in myoblast progenitor cells is the established regulatory role miRNAs play in guiding myogenic differentiation [173]. For example, miRNA-155 inhibits myogenic differentiation by down-regulating myogenic enhancer factor 2 (MEF2), which is a transcription factor that activates the myogenic program (Figure 12) [174]. Conversely, miRNA-24 stimulates myogenesis by down-regulating numerous cell cycle genes, thereby promoting cell cycle exit (Figure 12) [175,176]. Since ARS2 had only been implicated in miRNA and RDH mRNA biogenesis at the time this work began, we felt C2C12 myoblast progenitor cells were an ideal model system to tease out the role of ARS2 in these two pathways.

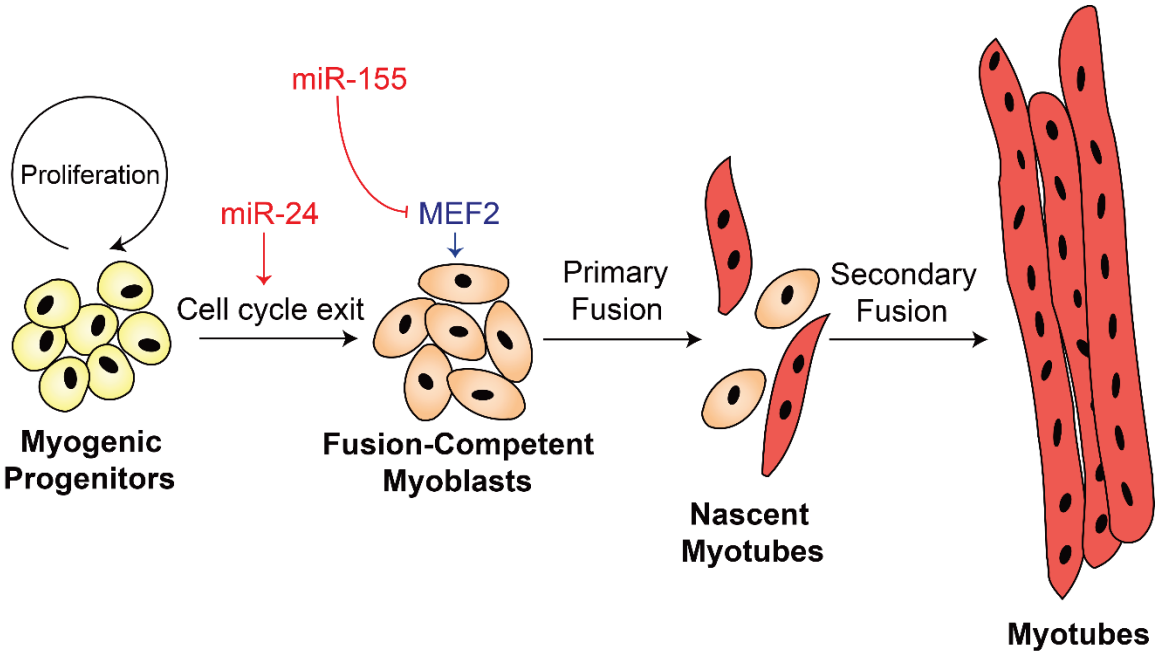

Figure 12 Myogenesis

Myogenesis is a process whereby proliferating myogenic progenitors exit the cell cycle and fuse to become multinucleated differentiated myotubes (left to right). Myogenesis is in part regulated by miRNAs, which can either inhibit or promote differentiation. For example, miRNA-155 (miR-155) inhibits differentiation by down-regulating its target MEF2, a pro-myogenic transcription factor, while miR-24 stimulates differentiation by promoting cell cycle exit.

1.19 Research objectives

ARS2 is essential for NSC maintenance and differentiation, and is required for INM in proliferating retinal progenitor cells [9,168]. Given the multiple functions ascribed to ARS2 in RNAP II transcript processing, it was difficult to precisely determine the mechanistic basis for ARS2's requirement in progenitor maintenance and differentiation. Moreover, the reported role for ARS2 as a SOX2-specific transcription factor in NSC contrasted with its reported functions in RNAP II transcript maturation [9]. Despite evidence to support ARS2's role in promoting progenitor cell proliferation [2,168], there was a lack of cell cycle kinetic analysis following ARS2 disruption, leading to a lack of consensus regarding ARS2's role in cell cycle progression [2,16]. Also unclear was the function of ARS2 within the cap complex and its role in multiple 3'-processing events for disparate RNA classes. At the time this dissertation began, ARS2 had only been shown to be involved in miRNA biogenesis and RDH pre-mRNA

processing, through interactions with DROSHA and FLASH, respectively [2,16]. Therefore, based on the combined evidence in the literature, I hypothesized that 1) ARS2 is required for the cell cycle progression of retinal progenitor cells and for retinal cell fate decisions, and 2) ARS2 acts as a scaffold that couples the 5'-RNA cap to multiple 3'-processing machineries. To test these hypotheses, my specific research objectives were as follows:

1) Determine whether ARS2 is required for cell cycle progression in retinal progenitor cells and determine the fate of late-born (postnatal) retinal progenitor cells deficient in ARS2 (chapter 2).
2) Assess the cell cycle kinetics following ARS2 knockdown or overexpression in a progenitor cell culture model (chapter 3).
3) Define the regions of ARS2 required for interacting with the 5'-CBC, components of the replication-dependent histone pre-mRNA, and pri-miRNA machineries (chapter 3).
4) Determine whether ARS2 binds RNA, and if so, which regions are required (chapter 3).

Chapter 2 – ARS2 is required for cell cycle progression and cell fate specification in the developing mouse retina

The following is an adapted manuscript in preparation for submission to PLOS ONE:

O'Sullivan C.[*], Nickerson, P.E.B. [*], Krupke, O., Chen, L.L., Zhu, M., Chow, R.L., Howard, P.L. PLOS ONE. *In preparation.*

[*] These authors contributed equally to this work.

Contributions: MZ cloned the pEF-N1ICD construct. LC performed *in vivo* electroporation in retinas used for Figure 15E, **CO** sectioned and imaged them. OK, performed *in vivo* electroporation in retinas used for Figure 18A and B, **CO** dissociated and analysed them using flow cytometry. PN electroporated and imaged retinas used in Figure 15A, PLH quantified fluorescence. PN performed ARS2 localization experiments in Figure 14B,C. For experiments shown in Figure 16A, and Figure 18C, PN performed electroporations, **CO** performed flow cytometry analysis. PN performed experiments in Figure 15C and Figure 17A-C, **CO** counted images to obtain quantification in Figure 17B and analyzed and graphed data for Figure 17A and C. **CO** also performed experiments in Figure 14A, Figure 15D,E, and Figure 16B,C. **CO**, RLC, and PLH wrote the manuscript.

2.1 Abstract

Proliferating progenitor cells in the mouse retina are specified to one of the seven major cell types as they exit the cell cycle in consecutive overlapping waves. Thus, cell cycle progression and exit are intricately coupled with cell fate specification to ensure the major cell types are generated in the correct proportions. ARS2 is required for efficiently processing a diverse set of RNA Polymerase II transcripts, and for maintaining self-renewal capacity and multipotency of neural stem cells in the mouse brain. Here, we show that ARS2 is expressed in all major cell types of the mouse retina. Postnatal retinal progenitors require ARS2 for proper cell cycle progression and ARS2 deficiency leads to

an early exit from the cell cycle. Cell identities are disrupted following ARS2 depletion, with an increase in the proportion of cells expressing a rod photoreceptor marker, and a loss of Müller glia marker expression. Knocking down FLASH, which interacts with ARS2 and is required for cell cycle progression and 3'-end processing of replication-dependent histone processing, recapitulates the proportional increase in rod photoreceptor marker expression following ARS2 knockdown. Therefore, we suggest a deficiency of properly processed histones contributes to the cell cycle and cell specification phenotypes following ARS2 depletion.

2.2 Introduction

As described in chapter 1, ARS2 has been implicated in cell cycle progression and in the maintenance and differentiation of neural progenitor cells. Furthermore, the Chow lab has shown that disrupting ARS2 perturbs interkinetic nuclear migration (INM) in the mouse retina, which results in accumulation at basal positions [168], suggesting ARS2 may be required for cell cycle progression. INM is intertwined with cell cycle progression in the retina, as nuclei of proliferating cells oscillate between the apical and basal positions of the neuroblastic layer, according to the phase of the cell cycle [163]. Additionally, cells with nuclei that travel greater distances basally during the preceding mitotic cycle are more likely to produce neurogenic daughter cells in Zebrafish retina [165]. NOTCH has been implicated in retinal progenitor cell fate and is expressed in an apical to basal gradient in the neural retina with highest levels found at apical positions [167]. Consistent with this, conditional *Notch1* knockout in the postnatal retina leads to an increase in rod photoreceptor cells and decrease in Müller glial cells [177,178]. Given the controversial findings of Andreu-Agullo et al., showing that ARS2 is required for neural stem cell maintenance and differentiation, and that ARS2 acts as a transcription factor controlling the expression of SOX2 [9], we decided to characterize the role of ARS2 in retinal progenitor cells. Here, we report ARS2 KD disrupts cell cycle progression of retinal progenitor cells and leads to premature cell cycle exit. We show that ARS2 is expressed in the neuroblastic layer during development and in all cell types of the adult retina. Furthermore, ARS2 KD increases the proportion of cells expressing a reporter specific for rod photoreceptors. At the same time, there is a decrease in Müller

glial cell reporters. The effect of ARS2 KD on cell fate reporters could not be rescued by NOTCH intracellular domain (NICD) expression, suggesting the effect is NOTCH signaling-independent. Instead, we show that the cell fate phenotype of ARS2-disrupted cells phenocopies deficiency in the RDH mRNA biogenesis pathway. This suggests that the delayed cell cycle progression and early cell cycle exit following ARS2 KD is contributing to the increase in rod photoreceptors.

2.3 Materials and Methods

2.3.1 Animals and ethics statement

This study was carried out in accordance with the guidelines of the Canadian Council on Animal Care. The protocol was approved by the University of Victoria Animal Care Committee (Permit Numbers: 2013-013(1), 2014-023). All surgery was performed under isoflurane anesthesia, and all efforts were made to minimize suffering. CD-1 mice were obtained from Charles River.

2.3.2 Plasmids

All shRNA plasmids were purchased from Origene, had a GFP cassette to assess electroporation efficiency, and their targeting sequences are shown in Table 1 (Appendix A). For Ars2 shRNA experiments, Ars2 sh79 was used unless otherwise indicated. pNrl-DsRed (Addgene plasmid # 13764) [179], pHes1-DsRed (Addgene plasmid #13767) [179], pCabp5-DsRed (Addgene plasmid # 11157) [180], and pCralbp-DsRed (Addgene plasmid # 11158) [180] were gifts from Dr. Connie Cepko. pEGFP-C1 was used to express enhanced GFP (eGFP) as a control in Figure 16B. The pEF-N1ICD construct was generated by subcloning a PmlI/XbaI fragment containing N1ICD from pBOS:N1ICD (a gift from Dr. Tasuku Honjo [181]) into pBluescript SK carrying the human EF1α promoter from pEF/myc/nuc (Thermo Fisher Scientific). pcDNA3.1(+) (Thermo Fisher Scientific) was used as an empty vector control.

2.3.3 *In vivo* electroporation

Newborn mice were electroporated as described in [180]. Briefly, once anesthetized using isoflurane, a small incision was made in the eyelid and sclera using a 30-gauge needle under a dissecting microscope. DNA solutions in PBS (~3-5 μg/μL)

containing 0.1% fast green were injected (~0.3 µL) into the subretinal space using a Hamilton syringe with a 32-gauge blunt-end needle. Electrodes (BTX Harvard Apparatus) briefly soaked in PBS were gently placed on the heads of the pups, and five square pulses of 50-ms duration with 950-ms intervals at 80 V were applied using a pulse generator (BTX Harvard Appartus) (Figure 13). Pups were then placed on a warm bed and monitored for recovery.

2.3.4 *In vitro* electroporation

Retinas were dissected and transferred to a micro-electroporation chamber containing DNA solution (~1 µg/µL) in Hank's Balanced Salt Solution (Thermo Fisher Scientific). Five square pulses of 50-ms duration with 950-ms intervals at 30 V were applied using a pulse generator (BTX Harvard Apparatus). Electroporated retinas were placed on Nucleopore track-etched membranes (Whatman, 1.0 µm pore) with Neurobasal medium (Thermo Fisher Scientific) supplemented with 1× Glutamax (Thermo Fisher Scientific), GS21 (Sigma-Aldrich), and Pen Strep (Thermo Fisher Scientific), and were incubated for the indicated time points at 37˚C (Figure 13).

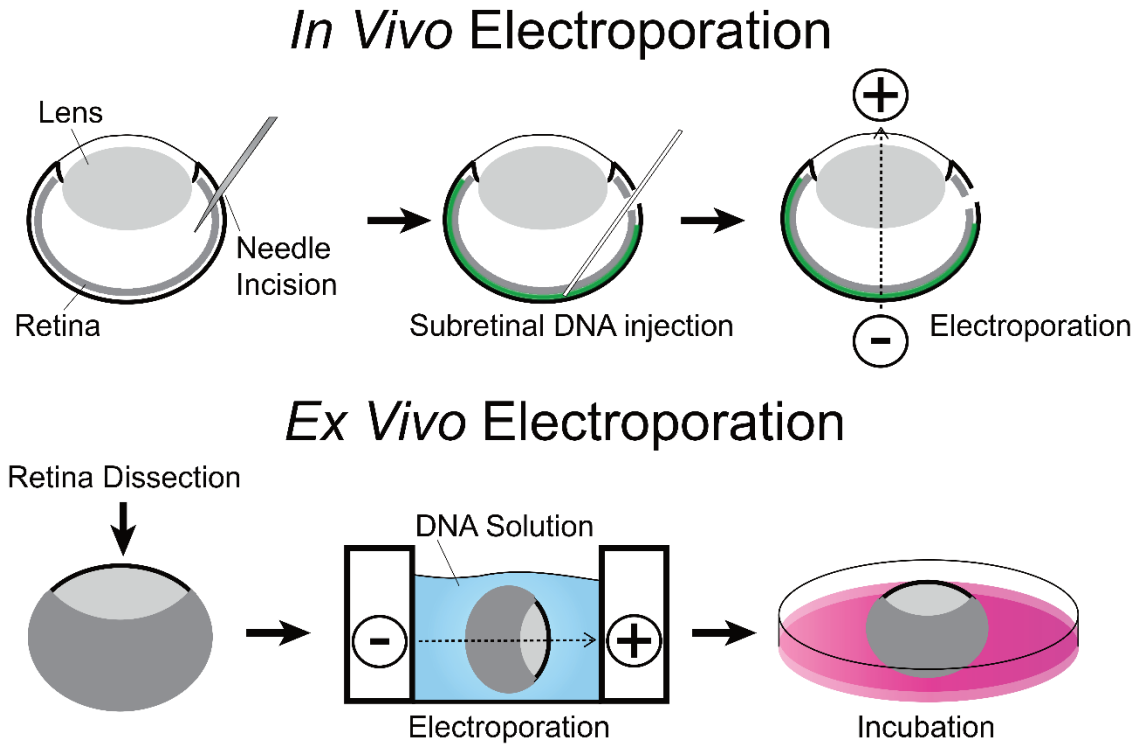

Figure 13 Retina electroporation

Schematic for *in vivo* electroporation (top) and *in vitro* electroporation (bottom) procedures in the newborn mouse retina.

2.3.5 Tissue fixation and sectioning

Retinas were dissected, fixed in 4% paraformaldehyde (PFA) in PB (0.2 M sodium phosphate buffer, pH 7.2) for ~ 4 h at 4˚C, washed in PBS, and incubated in 30% sucrose in PB overnight at 4˚C. Retinas were then washed twice in O.C.T compound (Sakura), placed in an O.C.T block, and flash frozen on dry ice with 70% ethanol. Retinas were then sectioned between 14-18 μm using a cryostat (Leica), placed onto Superfrost slides (Thermo Fisher Scientific), and let dry in the dark overnight at room temperature.

2.3.6 Immunohistochemistry and microscopy

Slides with sectioned retinas were washed three times for 10 mins in PBS. Retinas were permeabilized in 1% Triton-X 100 for 30 mins, washed in PBS, and incubated in either 1:500 rabbit anti-ARS2 (XL12.2) [7], 1:1,000 mouse anti-ARS2 (LX186.3) [7], 1:500 sheep anti-Vsx2 (Exalpha Biologicals), 1:10,000 rabbit anti-PKCα (Sigma-Aldrich), 1:1,000 rabbit anti-Calbindin-D-28K (EG-20, Sigma-Aldrich), 1:2,500 goat anti-Calretinin (EMD Millipore), 1:200 rabbit anti-Brn-3 (H-80, Santa Cruz Biotechnology), 1:1,000 mouse anti-Glutamine Synthetase (GS-6, EMD Millipore), or 1:1,000 rabbit anti-Recoverin (EMD Millipore) overnight at 4˚C. Next, slides were washed in PBS, and incubated in either 1:500 donkey anti-rabbit, anti-sheep, anti-goat or anti-mouse antibodies conjugated to either AlexaFluor 488 or 555 (Thermo Fisher Scientific) as appropriate for 1 hour at room temperature. Slides were then washed in PBS, nuclei were stained using 5 μM DRAQ5 (Thermo Fisher Scientific) in the dark for 5 mins at room temperature, and washed once more in PBS. Slides were then mounted with coverslips using Immu-Mount (Thermo Fisher Scientific), pH adjusted to ~8.0. Images were taken on a confocal microscope (Nikon).

2.3.7 Knockdown quantification

Retinas electroporated with control or Ars2 sh79 plasmids at P1 and dissected at P5 were fixed, sectioned, mounted to slides, stained for ARS2, and imaged using confocal microscopy. Corrected total cell ARS2 fluorescence, measured as integrated density – (Area of selected cell × Mean fluorescence of background) using ImageJ

software, was performed on GFP transfected cells individually, which were randomly sampled from sections taken from three biological replicates.

2.3.8 Western blotting

Retinas were dissected at the indicated time points, flash frozen in liquid nitrogen, and homogenized in 2× laemmli sample buffer with 1× protease inhibitor cocktail (Sigma-Aldrich) and 0.1 mM PMSF (Sigma-Aldrich). Samples were incubated in boiling water for 20 minutes and centrifuged at 12,000 ×g for 10 mins at 4˚C. Lysates were resolved using SDS-PAGE (10%) and transferred to a PVDF membrane (EMD Millipore). The membrane was blocked in 5% dehydrated milk in Tris-buffered saline–Tween 20 (0.5%) (TBST) for 1 h. Primary antibodies were diluted in TBST with 1% dehydrated milk at the following concentrations: 1:4,000 for mouse anti-Actin (Sigma-Aldrich), and 1:2,000 for rabbit anti-ARS2 (XL12.2) [7], and were incubated on the membrane with shaking overnight at 4˚C. Blots were washed 3× with TBST and incubated with 1:10,000 goat anti-mouse or goat anti-rabbit conjugated to HRP (Bio-Rad) for 1 h, washed with TBST, incubated in enhanced chemilumiscence (Thermo Fisher Scientific) and exposed on X-ray film (Thermo Fisher Scientific).

2.3.9 IdU/CldU pulse labeling experiments

Retinas were electroporated *in vitro* as described with the indicated plasmids, and incubated with either 37 μM IdU (Sigma-Aldrich) or 37 μM CldU (Sigma-Aldrich) in Neurobasal supplemented with 1 × Glutamax, GS21 and Pen Strep for the indicated time points. Retinas were then fixed and sectioned as described above. Slides were then washed three times for 10 mins in PBS, permeabilized in 1% Triton-X 100 for 30 mins, and incubated in 1:2000 rabbit anti-turboGFP (Origene) overnight at 4˚C. Next, slides were washed in PBS and incubated in 1:500 donkey anti-rabbit Alexa Fluor 488 (Thermo Fisher Scientific) for 1 h at room temperature. Slides were then washed in PBS, treated with 2N HCl for 45 mins, washed in PBS again, and incubated in 1:250 mouse anti-BrdU/IdU (BD Biosciences) and 1:250 rat anti-BrdU/CldU (Abcam) overnight at 4˚C. Slides were then washed in PBS, and incubated in 1:500 donkey anti-mouse Alexa Fluor 647 (Thermo Fisher Scientific) and 1:500 donkey anti-rat Cy3 (EMD Millipore) for 1 h at

room temperature. Slides were then washed, mounted and imaged using confocal microscopy.

2.3.10 Flow cytometry

Retinas were dissected at the indicated time points and dissociated to a single-cell suspension using the neural tissue dissociation kit for postnatal neurons (MACS Miltenyi Biotec) according to the manufacturer's instructions. For cell fate analysis using DsRed reporters, samples were immediately run on a BD FACSCalibur flow cytometer following dissociation, and ~1.0×10^6 events were acquired per sample.

2.4 Results

2.4.1 ARS2 is expressed in the developing and adult mouse retina

ARS2 expression is restricted to a subset of neural progenitor cells in the subventricular zone of the brain [9]. Based on this, we first sought to assess ARS2 expression in the developing and adult mouse retina. Western blot analysis using previously characterized ARS2 antibodies [7,9] showed ARS2 was expressed in the retina during embryogenesis (E14.5), postnatal differentiation (up to P14), and in the mature retina (8 weeks) (Figure 14A). Using immunohistochemistry, ARS2 could be detected in the nuclei and cytoplasm of cells in the developing ganglion and inner neuroblastic cell layers at postnatal day 0 (P0). We also detected strong cytoplasmic ARS2 signal at the apical margin, where the nuclei of progenitors undergo mitosis [182] (Figure 14B). By P3, strong nuclear localization was observed in presumptive post-mitotic neurogenic progenitors in the ganglion and inner neuroblastic layers, and in select cells in the neuroblastic layer (Figure 14B). By P10, weak nuclear localization was also observed in the outer nuclear layer (ONL) (Figure 14B). Consistent with ARS2's role in RNAP II transcript processing, these results show that ARS2 is widely expressed during retinal development.

We also examined ARS2 expression in adult retinas (≥ 3 months old). ARS2 was expressed in all layers of the adult retina, with nuclear localization observed in the ganglion, inner nuclear, and outer nuclear layers (Figure 14B). ARS2 was expressed in all major cell types of the adult retina, including ganglion cells (Brn3b, Calretinin), amacrine

cells (Calretinin), bipolar cells (Vsx2, PKCα), horizontal cells (Calbindin), and Müller glia (Glutamine Synthetase) (Figure 14C). Closer examination of the photoreceptor-containing outer nuclear layer (ONL) revealed a ring-shaped pattern, with ARS2 overlapping with the DRAQ5 nuclear stain around the inner perimeter of the nucleus (See Adult ONL in Figure 14B). This unusual nuclear staining pattern may reflect the inverted heterochromatin pattern in rod photoreceptors, where heterochromatin is centrally localized and euchromatin lines the nuclear periphery [183]. Taken together, these data demonstrate that ARS2 is expressed in retinal progenitor cells and in adult retinal neurons.

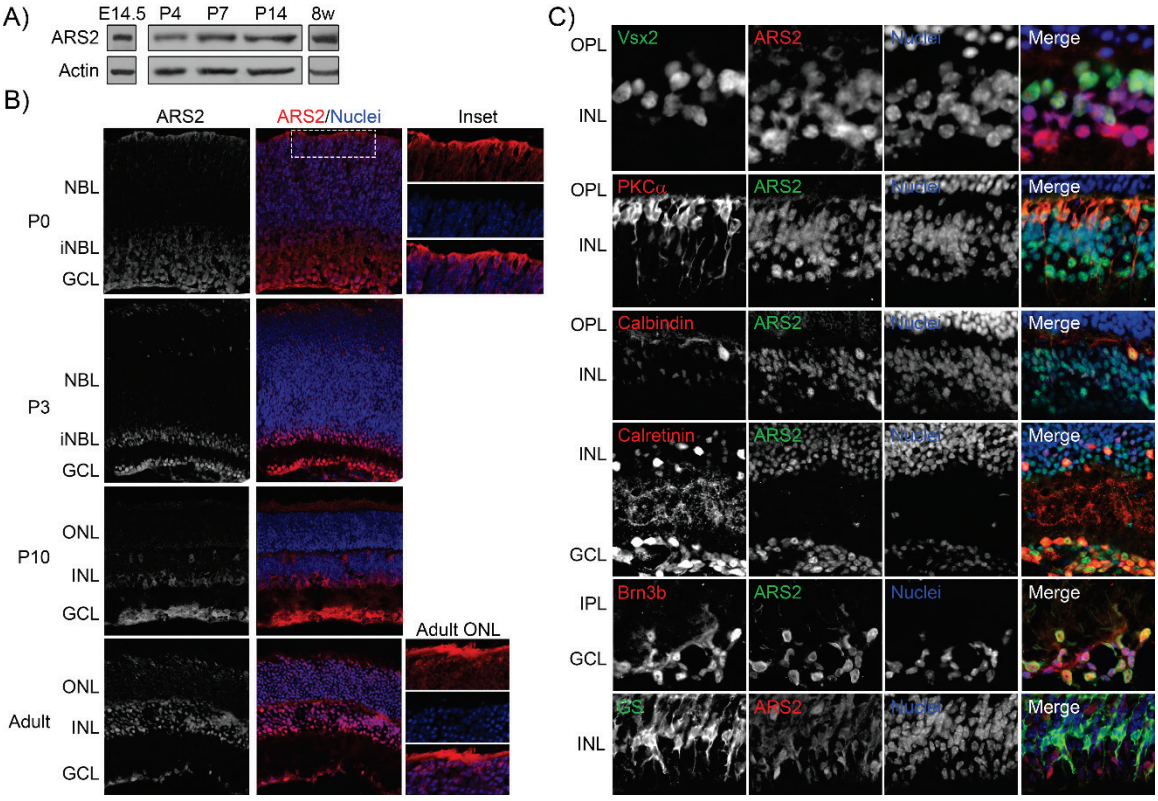

Figure 14 ARS2 is expressed in the developing and adult mouse retina

A) Western blots show endogenous ARS2 expression at embryonic day 14.5 (E14.5), postnatal day 0 (P0), P4, P7 and 8 weeks (8w). B) Immunohistochemistry shows ARS2 expression (red) at P0, P3, P10 and in adult retina (≥3 months). Nuclei stained with DRAQ5 are shown in blue. Inset at P0 highlights ARS2 expression at the apical margin of the NBL. Inset at Adult ONL exhibits ring-shaped expression pattern in photoreceptor cells. C) ARS2 expression (red or green, as indicated) in adult retina (≥3 months) bipolar cells (Vsx2 – green, PKCα - red), horizontal cells (Calbindin – red), amacrine cells (Calretinin – red), and ganglion cells (Calretinin, Brn3b – red). ARS2 expression in Müller glial cells using Glutamine Synthetase (GS – green) was analyzed at P10. Nuclei were stained with DRAQ5 (blue). NBL – neuroblastic layer, iNBL – inner neuroblastic

layer, GCL – ganglion cell layer, ONL – outer nuclear layer, INL – inner nuclear layer, OPL – outer plexiform layer, IPL – inner plexiform layer.

2.4.2 ARS2 knockdown increases expression of a rod photoreceptor marker

The P0-P8 postnatal period is dominated by the birth of rod photoreceptors, bipolar cells, and Müller glial cells [156,184], providing an excellent model system for studying differentiation of these cell types. Andreu-Agullo et al. demonstrated that conditional knockout of *Ars2* in subventricular zone neural stem cells leads to an overproduction of astroglial cells at the expense of oligodendrocytes and neurons [9], suggesting ARS2 can influence cell fate decisions in the brain. Therefore, we hypothesized that disrupting ARS2 in postnatal retinas would affect late-born progenitor cell fates.

To first confirm that we could deplete ARS2 in the retina, we electroporated retinas with a plasmid containing a GFP cassette and either control shRNA or a shRNA sequence targeting Ars2. Analysis of GFP positive transfected cells at 4 days post-electroporation showed that ARS2 levels were reduced ~3-fold relative to control shRNA (Figure 15A). To determine the requirement of ARS2 for late-born rod photoreceptor generation, we co-electroporated newborn pup retinas *in vivo* with either control or Ars2 shRNA along with a reporter containing the neural retina leucine zipper (NRL) promoter coupled to DsRed (Figure 15B) [179]. NRL is both necessary and sufficient for rod photoreceptor generation [185,186], and the Nrl-DsRed reporter is restricted to rod photoreceptors [179]. Analysis of retinas electroporated with pNrl-DsRed showed DsRed positive cells were localized to the ONL and were positive for recoverin, a marker for photoreceptors (Figure 15C) [187], corroborating the specificity of the reporter [179]. Significantly, ARS2 KD increased the proportion of Nrl-DsRed/GFP double positive cells approximately 3-fold when analyzed by flow cytometry or immunofluorescence of retinal sections at P8 *in vivo* (Figure 15D,E), suggesting ARS2 KD promotes rod photoreceptor fate.

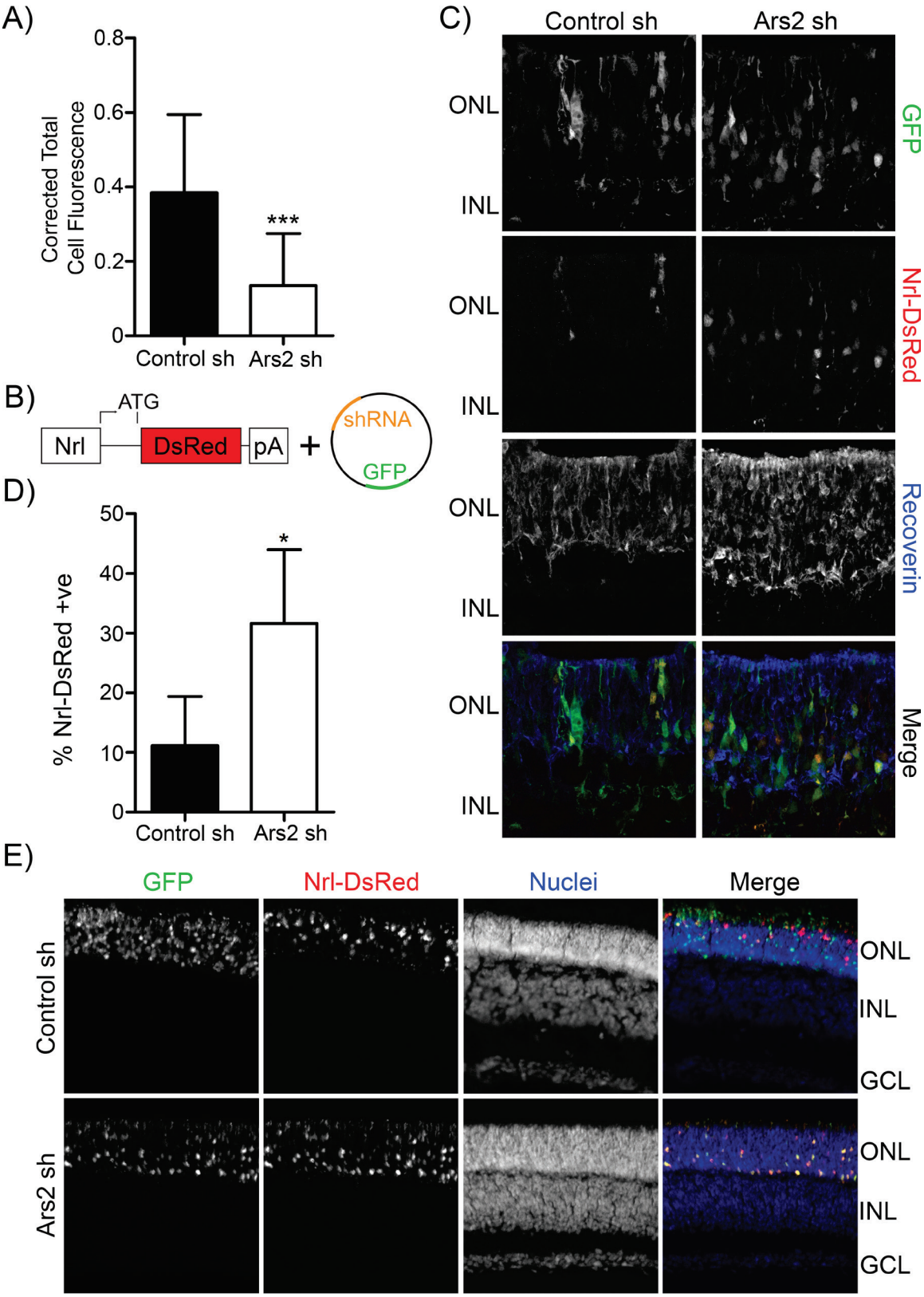

Figure 15 ARS2 knockdown increases expression of rod photoreceptor marker

A) Retinas were analyzed 4 days post-electroporation with either control shRNA (sh) or Ars2 sh79 using immunofluorescence. The corrected total cell ARS2 fluorescence of GFP transfected cells was calculated as described in the Methods, and a two-tailed unpaired *t* test was performed ($t(65) = 5.83$; $P<0.001$) . B) Schematic of the pNrl-DsRed and shRNA/GFP plasmids used to assess rod photoreceptor production. Nrl – rod photoreceptor-specific promoter, ATG – DsRed start codon, pA – polyA tail. C) Retinas co-electroporated with control sh (left panel) or Ars2 sh79 (right panel) and pNrl-DsRed (red) were sectioned at P8 and stained for Recoverin (blue). D) Retinas co-electroporated *in vivo* with control sh or Ars2 sh79 and pNrl-DsRed were analyzed at P8 using flow cytometry, and % Nrl-DsRed +ve is shown relative to total transfected cells. A two-tailed unpaired *t* test was used to determine whether there was a difference between control sh and Ars2 sh ($t(9) = 3.16$; $P<0.05$). E) Retinas electroporated as in D) were sectioned at P8 and analyzed using confocal microscopy. ONL – outer nuclear layer, INL – inner nuclear layer, GCL - ganglion cell layer. Error bars represent standard deviation (SD).

2.4.3 ARS2 knockdown decreases Müller glial cells

Since disrupting ARS2 impacts Nrl expression, a rod photoreceptor marker, we next tested whether ARS2 KD affects Müller glia and bipolar cell production, as they are also late-born cell types [184]. To assess Müller glial generation, a DsRed reporter containing the promoter for *Hes1* was initially used, which is expressed in progenitor and Müller glial cells [179]. Moreover, HES1 is required for Müller glial cell production in the mouse retina [188]. Mice were co-electroporated with control or Ars2 shRNA and pHes1-DsRed, incubated until P8 *in vitro*, and analyzed using flow cytometry. ARS2 KD resulted in a ~2.5-fold reduction in Hes1-DsRed positive cells relative to control (Figure 16A). Since HES1 is potentially expressed in a small pool of remaining progenitor cells at this time point and to corroborate this result *in vivo* with an additional Müller glial marker, a DsRed reporter containing the promoter for cellular retinal-binding protein (CRALBP) was used; the Cralbp promoter is restricted to Müller glial cells [180,189]. Mice were co-electroporated *in vivo* at P0 with control or Ars2 shRNA and pCralbp-DsRed *in vivo* and analyzed at P14 (Figure 16B,C). We noticed that the CMV-GFP expression from the shRNA plasmid was largely limited to the ONL in all cases, potentially indicating a limited ability to electroporate non-photoreceptor destined neurogenic progenitors (Figure 15E and Figure 16B). In contrast, the Cralbp-DsRed reporter was expressed in the inner nuclear layer and in cells displaying Müller glial morphology (Figure 16B), as previously reported [180], indicating that Müller glial destined progenitors were successfully targeted by electroporation. Therefore, it is likely

the restricted GFP expression is due to CMV promoter silencing in the other layers of the retina [190,191]. Consistent with this interpretation, ARS2 KD abolished Cralbp-DsRed activity as observed using flow cytometry or micrograph analysis (Figure 16B,C). No differences in the number of GFP fluorescent cells were observed between ARS2 knockdown and control conditions (data not shown), consistent with previous work showing that ARS2 knockdown in neural progenitors does not affect cell survival [9]. These results suggest there is a reduction in the production of Müller glial cells in the ARS2 knockdowns. Lastly, we did not observe a difference *in vivo* for bipolar cell production using a DsRed reporter under the control of the calcium binding protein 5 (CABP5) promoter, which is restricted to bipolar cells (data not shown) [180,192]. Overall, the data indicate that ARS2 KD results in an increased proportion of Nrl-DsRed positive rod photoreceptor cells at the expense of Müller glia.

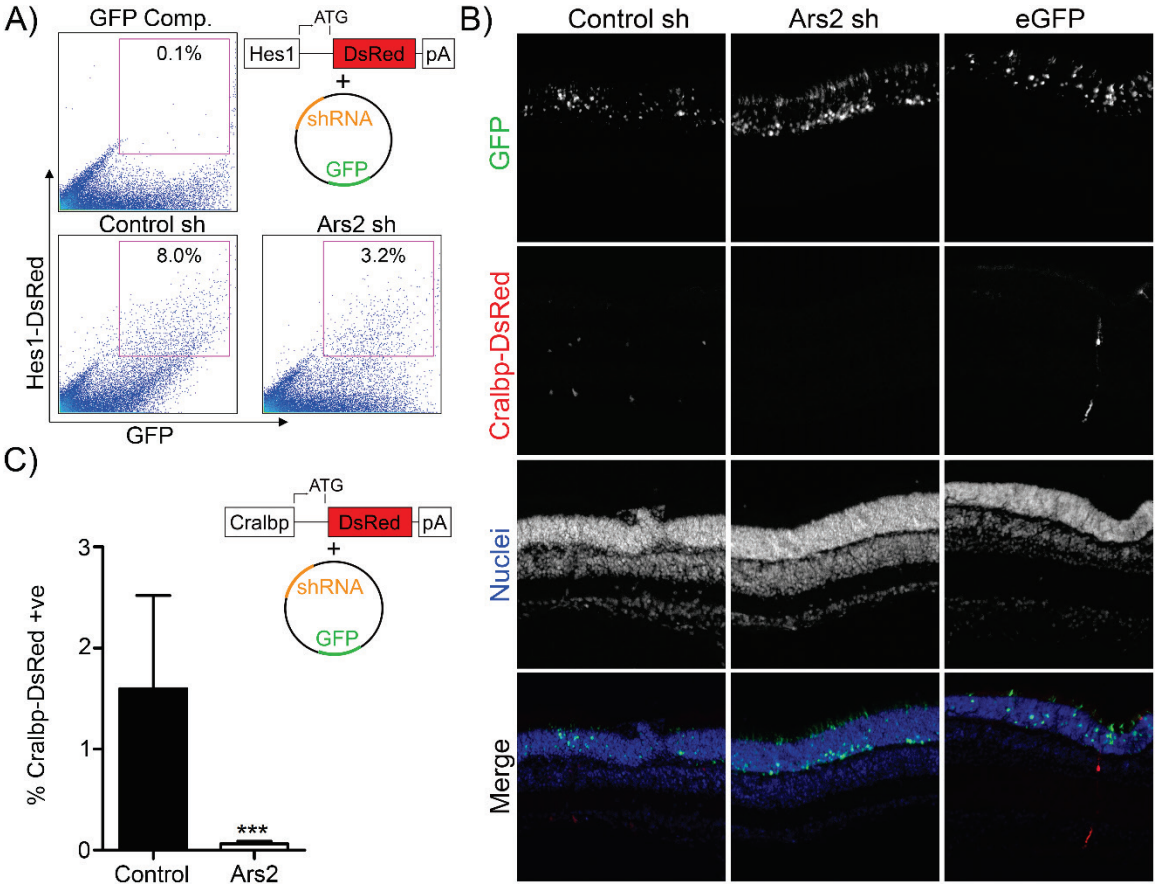

Figure 16 ARS2 knockdown decreases Müller glial cells

A) Retinas co-electroporated *in vitro* with either control sh or Ars2 sh and pHes1-DsRed were analyzed at P8 using flow cytometry. The GFP fluorescence compensation control (GFP Comp.) is shown in the upper left plot, and was used to draw the gate (pink box)

representing Hes1-DsRed +ve events. 4 retinas were pooled per experimental condition, and the proportion of Hes1-DsRed +ve events relative to total transfected cells is shown within the gates. B) Retinas co-electroporated *in vivo* with either control sh, Ars2 sh or eGFP and pCralbp-DsRed were sectioned on P14 and analyzed using confocal microscropy. Nuclei were stained using DRAQ5 and are shown in blue. C) Retinas co-electroporated *in vivo* with control sh or Ars2 sh79 and pCralbp-DsRed were analyzed at P14 using flow cytometry, and % Cralbp-DsRed +ve is shown relative to total transfected cells. A two-tailed unpaired *t* test was used to determine whether there was a difference between control sh and Ars2 sh ($t(8) = 4.21$; $P<0.001$). Error bars represent SD.

2.4.4 ARS2 knockdown affects cell cycle progression and exit

Given the cell fate changes we observed in ARS2 deficient neural progenitors, we next investigated possible contributing mechanisms. ARS2 knockdown disrupts cell cycle progression in human cells *in vitro* [16]. Moreover, disrupting ARS2 in retinal explants results in an interkinetic nuclear migration (INM) defect, with an accumulation of nuclei at basal positions in the neuroblastic layer [168]. The position of the nuclei coincides with the cell cycle phases during INM, with S-phase occurring basally in the retina, and M-phase occurring at the apical margins [163]. Moreover, cell cycle progression is required for INM to proceed normally [164]. Therefore, we sought to assess the contribution of ARS2 towards the cell cycle progression of retina progenitors. Previous work has shown ARS2 deficient cells are delayed in S phase progression and hence accumulate in S phase [16]. To assess the proliferation status of retinal progenitors, we performed a thymidine analogue labeling experiment to determine the proportion of cells in S phase. Retinal explants were electroporated with control or Ars2 shRNA at P0 and were given a 30 minute IdU pulse at 21 hours post-electroporation. ARS2 KD increased the proportion of IdU +ve cells ~2-fold relative to control shRNA (Figure 17A), indicating an increased proportion of cells in S phase at this time point. Consistent with previous work showing an increase in the accumulation of cells at a basal position and reduced apical migrations [168], the increase in IdU incorporation (Figure 17A) supports the conclusion that ARS2 deficient cells are delayed in cell cycle progression through S-phase.

We next asked whether retinal progenitor cells deficient in ARS2 re-enter the cell cycle or if they prematurely exit the cell cycle. To this end, we performed an experiment in which we labeled the electroporated explanted retinas with IdU from 0 to 48 hours, and

with CldU from 48 to 192 hours post-electroporation, to determine whether actively cycling cells at the time of electroporation (IdU +ve), continue to cycle (IdU +ve/CldU +ve) or exit the cell cycle (IdU +ve/CldU -ve). The proportion of Idu+ve/CldU+ve cells was not significantly different following ARS2 KD relative to control (Figure 17B). However, the proportion of cells IdU+ve/CldU-ve was increased in the ARS2 knockdowns (Figure 17B), indicating that actively cycling progenitor cells deficient in ARS2 preferentially exit the cell cycle. Premature cell cycle exit would be expected to decrease the pool of late-born progenitors and result in a reduction in cellularity. Consistent with this, we observed that ARS2 deficient retinas were reduced in size relative to control knockdown retinas 96 hours post-electroporation (Figure 17C). Collectively, our results show that ARS2 KD cells are compromised in cell cycle progression and that dividing ARS2 deficient progenitors prematurely exit the cell cycle.

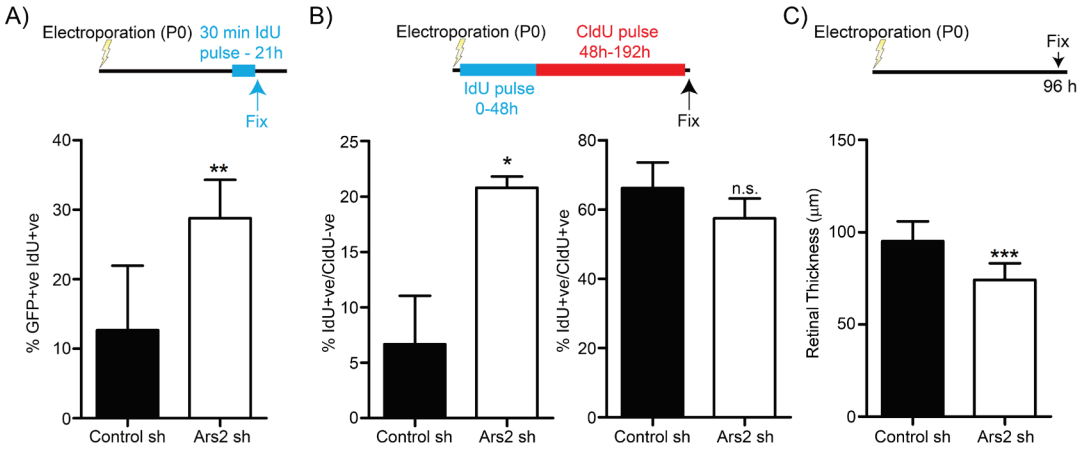

Figure 17 ARS2 knockdown affects cell cycle progression and exit

A) Retinas electroporated *in vitro* with either control or Ars2 shRNA at P0, and 21 hours post-electroporation were given a 30 minute IdU pulse. The % of electroporated (GFP +ve) IdU +ve cells is shown. A two-tailed unpaired *t* test was used to determine whether there was a difference between control sh and Ars2 sh ($t(10) = 3.67$, $P<0.01$) B) Retinas electroporated as in A) were given an IdU pulse from 0-48 hours post-electroporation, and a CldU pulse from 48-192 hours. The average proportion of cells that were IdU+ve/CldU–ve (left) or IdU+ve/CldU+ve (right) are shown for each condition. A two-tailed unpaired *t* test was used to determine whether there was a difference between control and Ars2 sh ($t(3) = 5.81$, $P<0.05$ for IdU+ve/CldU-ve), n.s. indicates not significant. C) Retinas electroporated as in A) were fixed 96 hours post-electroporation and retinal thickness was measured using confocal microscopy. A two-tailed unpaired *t* test was used to determine whether there was a difference between control and Ars2 sh conditions ($t(18) = 4.74$, $P<0.001$). Error bars represent SD.

2.4.5 ARS2 knockdown cell fate defect is not rescued by NOTCH signaling

An increase in rod photoreceptors and early cell cycle exit is reminiscent of the conditional *Notch1* knockout phenotype [177,178]. As mentioned, NOTCH signaling is strongest at the apical margin where cells undergo M phase [167]. Since disrupting ARS2 results in an impairment in S phase [16], accumulation of nuclei at a basal position in the developing retina [168], and decreased Hes1 reporter expression (Figure 16A), which is a downstream target of NOTCH [193], we hypothesized that the cell fate phenotype following ARS2 KD was due to an extrinsic deficiency in exposure to NOTCH signaling. To assess this possibility, we attempted to rescue the cell fate phenotype in ARS2 depleted cells by expressing the signal-transducing NOTCH1 intracellular domain (NICD) under the control of the ubiquitously expressed EF1α promoter. Retinas were co-electroporated *in vivo* with either control shRNA or Ars2 shRNA, pEF1α-NICD or pcDNA empty vector (EV), and pNrl-DsRed or pCralbp-DsRed and analyzed using flow cytometry. Overexpression of NICD did not appear to rescue the increase in Nrl-DsRed or the decrease in Cralbp-DsRed following ARS2 KD (Figure 18A,B), suggesting the increase in Nrl-DsRed positive rod photoreceptors and decrease in Müller glial cells seen following ARS2 KD is not due to a deficiency in NOTCH signaling.

2.4.6 ARS2 knockdown cell fate defect phenocopies FLASH

Intrinsic delaying of cell cycle progression or exiting the cell cycle prematurely can itself increase the proportion of rod photoreceptors, especially shortly after birth, which corresponds to the peak of rod photoreceptor production [159,161]. Therefore, we hypothesized that the cell cycle defect following ARS2 KD was directly contributing to the altered cell fate decisions. Previous work has shown that ARS2 cell cycle defects are in part caused by RDH deficiency during S phase due to defective RDH mRNA processing [16,24]. FLASH interacts with ARS2, and like ARS2, is required for RDH mRNA processing and cell cycle progression [16]. Therefore, we compared ARS2 and FLASH KD using the Nrl-DsRed reporters. Strikingly, ARS or FLASH KD increased the proportion of NRL-DsRed/GFP double positive cells approximately 2-fold relative to control shRNA (Figure 18C). Since ARS2 is also required for microRNA (miRNA) production [2,24], we also analyzed Nrl-DsRed following DROSHA KD, the enzyme

required for processing pri-miRNA [56]. Notably, DROSHA KD decreased the proportion of transfected cells expressing Nrl-DsRed relative to control shRNA, implying a miRNA biogenesis defect is not contributing to the cell fate phenotype following ARS2 KD (Figure 18C). Collectively, these data suggest the increased proportion of rod photoreceptors following ARS2 KD is coincident with a delayed cell cycle progression and early cell cycle exit.

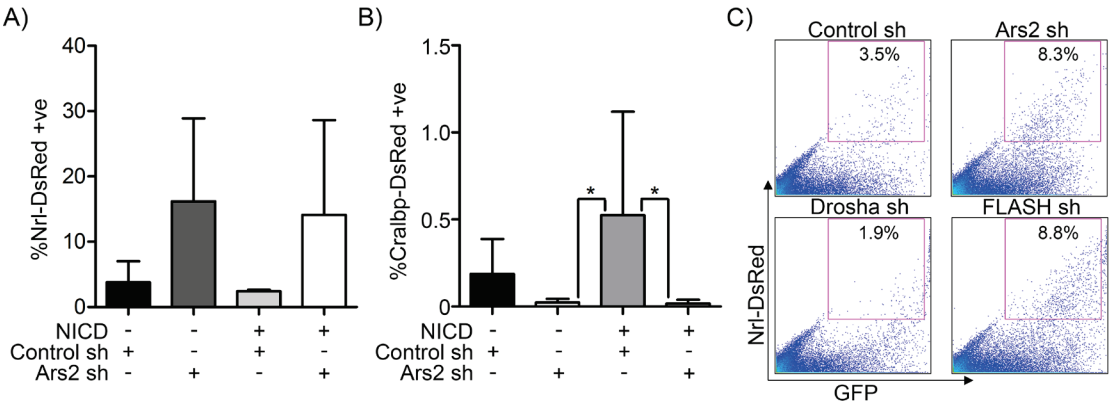

Figure 18 ARS2 knockdown cell fate defect phenocopies FLASH

A) Retinas were co-electroporated *in vivo* with either control shRNA or Ars2shRNA, pcDNA3.1 or pEF-NICD (NICD), and pNrl-DsRed, and were analyzed at P8 using flow cytometry. The proportion of Nrl-DsRed +ve cells is shown relative to total transfected cells. B) Retinas were co-electroporated *in vivo* as in A) except instead of pNrl-DsRed, pCralbp-DsRed was used, and samples were analyzed at P14 using flow cytometry. The proportion of Cralbp-DsRed +ve cells relative to total transfected cells is shown. A one way ANOVA was performed ($F(3,23)=3.74$, $P<0.05$) followed by Tukey's multiple comparision *post hoc* test to determine whether there were significant differences between groups. * = $P<0.05$. Error bars represent SD. C) Retinas were co-electroporated with either control, Ars2, FLASH or Drosha shRNA and pNrl-DsRed. Electroporated retinal explants were cultured *in vitro* to P8, at which point 4 retinas per condition were pooled and dissociated into a single cell suspension and analyzed using flow cytometry. The proportion of Nrl-DsRed +ve cells (pink gate) is shown relative to total transfected cells.

2.5 Discussion

We demonstrate that ARS2 is expressed in the developing and adult retina, with expression at the apical margin of the developing neuroblastic layer, and nuclear expression in all major cell types in the adult. The persistence of ARS2 expression in post-mitotic adult retinal cells suggests it has roles in non-dividing tissue, in contrast to previous reports of ARS2 downregulation in quiescent cells [2,24]. Gruber et al. characterized ARS2 expression in a specialized cell line ($Bax^{-/-}Bak^{-/-}$) that is strictly

dependent on IL3 for its growth [2,24]. Removal of IL3 results in a withdrawal from the cell cycle and loss of ARS2 expression in this model cell line [2,24]. However, restricted expression of ARS2 in actively cycling cells may be limited to $Bax^{-/-}Bak^{-/-}$ cells. Alternatively, ARS2 expression may be dependent on growth factors such as IL3, which may function outside of the cell cycle in differentiating cells *in vivo*. Further supporting ARS2 functions in non-dividing cells, Andreu-Agullo *et al.* showed that ARS2 expression in the subventricular zone of the mouse brain did not correlate with Ki67 expression, a marker of proliferating cells [9]. Instead, ARS2 was expressed in ependymal cells, a subset of astrocytes, and GFAP/CD133+ve quiescent stem cells [9]. Furthermore, I have observed ARS2 expression is maintained in post-mitotic, terminally differentiated C2C12 myotubes *in vitro* (Appendix A - Figure 32), providing additional support for the potential of ARS2 to perform important regulatory functions in post-mitotic cells. Our results showing ARS2 expression in post-mitotic retinal cells are consistent with the importance of ARS2 in the nuclear cap-binding complex (CBC), which co-transcriptionally binds the 7-methylguanosine cap at the 5'-end of RNAP II transcripts [17]. ARS2, through its interaction with the CBC, has been shown to play a role in efficient processing, turnover, and trafficking for disparate RNAP II transcripts, including snRNA, replication-dependent histone (RDH) pre-mRNA, miRNA, and mRNA [2,16,24–26,119]. Thus, the expression of ARS2 in non-dividing cells that we and others have found is consistent with the established role of ARS2 as an integral component of the nuclear cap complex.

We determined that depleting ARS2 in proliferating retinal progenitor cells leads to an increase in Nrl positive rod photoreceptors at the expense of Cralbp/Hes1 positive Müller glial cells. The pleiotropic nature of ARS2 makes it difficult to precisely establish the mechanistic details contributing to this phenotype. However, since DROSHA KD did not increase rod photoreceptor production, we ruled out ARS2's role in miRNA biogenesis as a contributing factor. Conversely, we observed an increase in the proportion of cells expressing Nrl-DsRed following KD of either ARS2 or FLASH. Both ARS2 and FLASH are required for replication-dependent histone processing, and FLASH KD also delays progression through S phase [16,24,75,77,79,194]. Notably, mutation of SLBP, which is also required for RDH mRNA processing [195], delays cell cycle progression

and neurogenesis in Zebrafish retina [196]. Of course, we cannot preclude that functions of ARS2 other than RDH mRNA biogenesis are contributing to the observed phenotype in the retina.

Here, we report ARS2 KD results in increased IdU incorporation, which is consistent with a delay in S phase in the postnatal mouse retina. This requirement of ARS2 for S phase agrees with a previous report showing S phase accumulation following ARS2 KD [16]. Retinal progenitor cells depleted of ARS2 do not re-enter the cell cycle and instead prematurely exit cycling, which we predict influences their neurogenic capacity. This interpretation is consistent with loss of cyclin D1 in the retina, which regulates the transition from G1 into S phase, and its deletion results in a delayed cell cycle progression, early cell cycle exit, and increase in rod photoreceptor cells [159]. Furthermore, overexpression of the cell cycle inhibitor $p27^{Kip1}$, which interacts with cyclin D1, also leads to early cell cycle exit and an increase in rod photoreceptors and decrease in Müller glia [161]. Therefore, disrupting cell cycle progression and the timing of cell cycle exit has consequences on cell fate specification. The phenocopy we observe between ARS2 and FLASH depletion in terms of Nrl reporter expression, leads us to speculate that delayed cell cycle progression, histone deficiency, followed by early cell cycle exit, is contributing to the increase in Nrl reporter activity.

At first glance, our results are in contrast to those of Lai's lab, which showed that multipotency and self-renewal of *Ars2* null NSCs could be rescued by ectopic expression of SOX2, a transcription factor required for NSC maintenance [9]. However, further work is required before we can rule out this mechanism in the retina, and it is plausible that our results are not mutually exclusive to those of the Lai lab. In the developing retina, SOX2 is expressed in proliferating progenitors in the neuroblastic layer, and in the adult retina SOX2 expression is restricted to Müller glia and cholinergic amacrine cells [197,198]. SOX2 is required for the maintenance and maturation of Müller glia, and prevents quiescent Müller glia from aberrantly entering the cell cycle and eventually undergoing apoptosis [198–201]. Thus, in progenitor and glia cells, there is overlapping expression patterns between ARS2 and SOX2. The effect of ARS2 KD on the Cralbp reporter is consistent with SOX2's role in these cells. Interestingly, SOX2 directly regulates NOTCH1 expression by interacting with the *Notch1* enhancer [197], and

concomitant expression of NICD with *Sox2* conditional knockout at P0 in the mouse retina rescues the loss of Müller glia seen in the *Sox2* null cells [200].

In our study, NICD expression did not appear to rescue the Müller glial deficit following ARS2 KD, suggesting ARS2 is contributing to this phenotype independent of SOX2-Notch. However, it is important to note limitations of our data concerning the NICD cell fate rescue experiments. Despite having a reasonable sample size per experimental condition (n~6), there were no statistical differences between control and Ars2 shRNA using the Nrl-DsRed reporter along with empty vector DNA plasmid (See Figure 18A,B, compare lanes 1 and 2). This is in contrast to the differences seen when retinas were co-electroporated with 2 plasmids: the shRNA and DsRed reporters (See Figure 15D,E, Figure 16A,C,D, and Figure 18C). I hypothesize this discrepancy is due to limitations in the amount of plasmid capable of being delivered in the retina. In the 2 plasmid experiments (only shRNA and DsRed), I used a ratio of 30:70 shRNA:DsRed when electroporating, so as to maximize the amount of DsRed reporter activity observed, as these reporters typically have weak expression. In the 3 plasmid experiments, I used a 1:1:1 ratio of shRNA:DsRed:NICD/pcDNA. Therefore, there was ~2-fold less DsRed reporter plasmid used in these experiments, resulting in lower amounts of Nrl-DsRed or Cralbp-DsRed reporter activities and higher variabilities between samples. It is also possible that the EF1α used to drive NICD is not working as expected, as Valerie Wallace's lab observed ~50% of cells expressing NICD were Cralbp-DsRed positive [202]. Therefore, at this point we cannot formally rule out that SOX2/NOTCH signaling is contributing to the cell fate phenotype seen following ARS2 KD. In addition, activation of NOTCH signaling through NICD is insufficient to prevent the aberrant cell cycle re-entry in Müller glia following conditional *Sox2* deletion [200], indicating SOX2 has NOTCH-independent roles in regulating proliferation in the retina. Further highlighting the complex interplay between proliferation and differentiation, $p27^{Kip1}$ directly represses SOX2 by binding to an enhancer following the induction of embryonic stem cell differentiation [203]. Thus, cells exiting the cell cycle downregulate SOX2 expression. Therefore, we favour a model in which early cell cycle exit may lead to defects in SOX2 signaling, which in turn alters cell fate in favour of photoreceptor development.

To approximate electroporation efficiency, we utilized the GFP cassette on the shRNA plasmids under the control of a CMV promoter, which is restricted in expression to the ONL postnatally up to P14 [190], corresponding to the time points used to assess cell fate DsRed proportions. Since rod photoreceptors account for ~85% of all cells born postnatally in the mouse retina [156,204], %GFP only approximates the electroporation efficiency. The shRNA targeting sequences were under the control of the U6 promoter, which is ubiquitously expressed [180]. Additionally, the pCralbp-DsRed and pCabp5-DsRed reporters displayed the expected morphology and localization patterns, indicating our electroporation technique could target all late-born retinal cell types. The fact that Ars2 shRNA abolished Cralbp-DsRed expression further supports the notion that cells outside of the ONL could be targeted with our constructs. Unfortunately, the Nrl-DsRed reporter used in our assay is not expressed as widely as expected in the ONL. In principle, all GFP positive cells in the ONL should be positive for NRL and express the DsRed reporter. In practice, only a fraction of the GFP positive cells expressed the reporter. This may be a reflection of transfection conditions, in which not all GFP positive cells received the DsRed reporter, or alternatively, the reporter construct may be weakly expressed and below the limit of detection in some cells.

The early cell cycle exit would be expected to result in a reduction in the overall number of retinal progenitors. Consistent with this, we observed a reduction in retinal thickness as a consequence of ARS2 deficiency. However, unexpectedly, we did not see a reduction in the overall number of GFP positive cells between the ARS2 deficient condition and the control. The reason for this is likely a selective loss of neurogenic progenitor contribution to the Müller glial population in the inner nuclear layer following ARS2 knockdown. Additionally, we predict that ARS2 deficient cells are proportionally more likely to end up in the ONL as rod photoreceptors. Thus, we hypothesize that although there is an overall reduction of cells following ARS2 knockdown, since GFP expression is restricted to the ONL, the overall population of GFP positive cells appears the same between conditions.

This study establishes the requirement of ARS2 in mouse retinal progenitor cells, which is consistent with its known role in the subventricular zone NSC compartment in

the brain [9]. Depleting ARS2 postnatally results in a delayed cell cycle progression, premature cell cycle exit, loss of Müller glia and increase in rod photoreceptor marker expression. We hypothesize the proliferation defect is coupled to the observed cell fate phenotype, as FLASH KD recapitulates the increase in Nrl-DsRed positive rod photoreceptors. ARS2 and FLASH interact and are required for RDH mRNA processing [16,24], which suggests a deficit of properly processed histones during S phase is contributing to the increase in rod photoreceptors. A deficiency in properly processed histone transcripts could be verified using qRT-PCR on electroporated cells sorted using fluorescence activated cell sorting (FACS). In our working model, we posit that delayed cell cycle progression and early cell cycle exit is at least partially due to histone misprocessing, and that the timing of cell cycle exit biases the progenitors to adopt a rod fate (Figure 19). It is important to note that we cannot rule out other ARS2-dependent processes, such as SOX2 regulation, are contributing to this phenotype, and that our results do not preclude a downstream involvement of SOX2 in this phenotype.

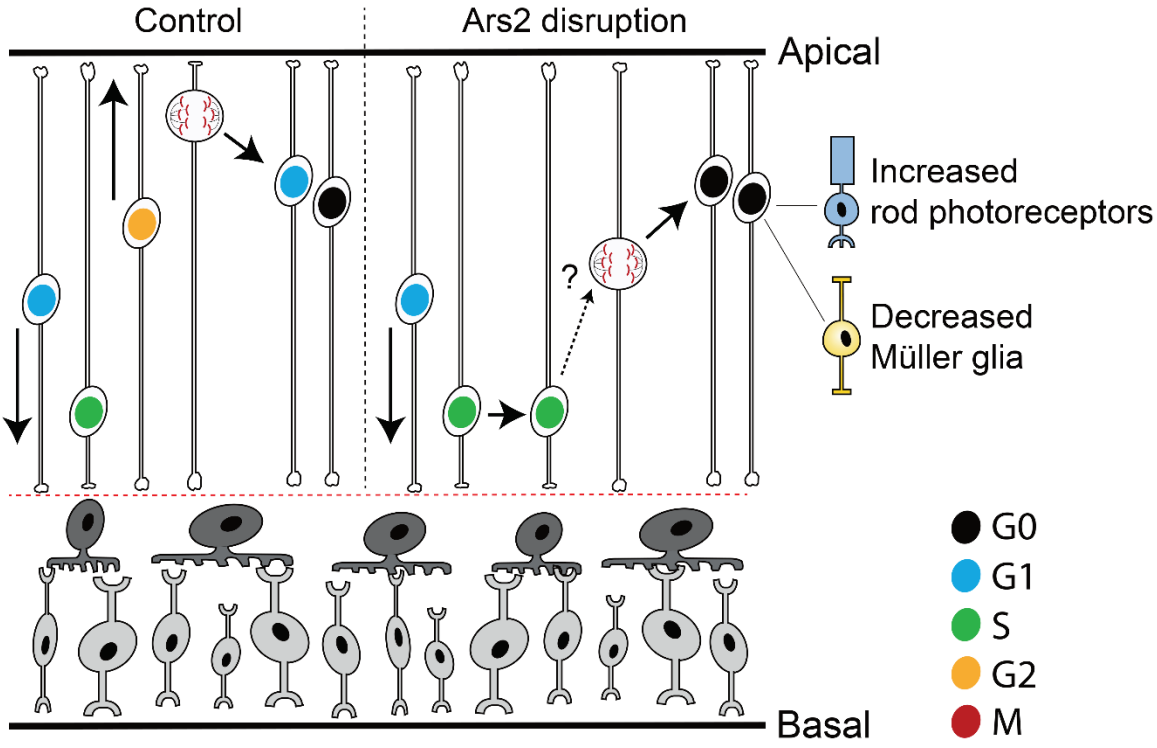

Figure 19 Working model of ARS2 depletion phenotype in the mouse retina

In control retinas (left panel) nuclei of proliferating progenitors undergo interkinetic nuclear migration and exit the cell cycle to produce the correct proportions of the major cell types. Following ARS2 disruption (right panel) cell cycle progression is delayed,

nuclei spend more time at a basal position, and exit the cell cycle prematurely (extra green and black nuclei, respectively). Whether mitosis occurs at the apical margin or at an ectopic position is unknown and is indicated with a question mark. Additionally, ARS2 depletion increases rod photoreceptor reporter expression, and decreases Müller glia reporter expression. We hypothesize that the delayed cell cycle progression and early cell cycle exit is at least partially due to a defect in histone processing, and the increase in rod photoreceptors is a consequence of the timing of cell cycle exit.

Chapter 3 – ARS2 domain function in RNA polymerase II transcript processing

Adapted from the following publication, with permission from the American Society for Microbiology:

O'Sullivan, C., Christie, J., Pienaar, M., Gambling, J., Nickerson, P.E.B., Alford, S.C., Chow, R.L., Howard, P.L. Mutagenesis of ARS2 domains to assess possible roles in cell cycle progression and microRNA and replication-dependent histone mRNA biogenesis. Molecular and Cellular Biology. 2015;35;3753-3767.

Contributions:

JC cloned Histone-DsRed, eGFP-FARB, and eGFP-ScrambledFARB, and performed the Histone-DsRed experiment shown in Figure 22D. Under the supervision of **CO**, MP performed the IP experiments shown in Figure 25C,E, and JG performed the IP experiments shown in Figure 25A,D. PEBN imaged cells labeled with BrdU in Figure 20E, **CO** counted these images. SCA cloned 3×FLAG-ARS2 RRM Mutant. **CO** performed all other experiments. **CO**, PLH and RLC wrote the manuscript.

3.1 Abstract

ARS2 is a regulator of RNA polymerase II transcript processing through its role in the maturation of distinct nuclear cap-binding complex (CBC)-controlled RNA families. Here, we examined ARS2 domain function in transcript processing. Structural modelling based on the plant ARS2 orthologue, SERRATE, revealed 2 previously uncharacterized domains in mammalian ARS2: an N-terminal domain of unknown function (DUF3546), which is also present in SERRATE, and an RNA recognition motif (RRM) that is present in metazoan ARS2, but not in plants. The DUF3546 or the Zinc Finger domain (ZnF) were required for association with microRNA and replication-

dependent histone mRNA. Mutations in the ZnF disrupted interaction with FLASH, a key component in histone pre-mRNA processing. Mutations targeting the Mid domain implicated it in DROSHA interaction and miRNA biogenesis. The unstructured C-terminus was required for interaction with the CBC protein CBP20, while the RRM was required for cell cycle progression and for binding to FLASH. Together, our results support a scaffold model, in which ARS2 plays a central role in RNA recognition and processing through multiple protein and RNA interactions.

3.2 Introduction

The generation of mature RNA in the nucleus is a highly coordinated process that requires distinct complexes for the biogenesis of different RNA families. For RNA Polymerase II (RNAP II) transcripts, a critical step is the co-transcriptional addition of a 7-methylguanosine (m7G) cap to the nascent transcript [17], which is subsequently bound by CBP20 and CBP80 of the nuclear cap-binding complex (CBC) [19]. CBC-controlled transcripts include messenger RNA (mRNA), microRNA (miRNA), replication-dependent histone (RDH) mRNA, small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA), each with its own unique processing requirements. Binding of the CBC to the m7G cap of these RNAs protects the transcripts from degradation, and plays a central role in recruiting the appropriate machinery for processing different RNA families [2,20–23,25]. However, exactly how distinct RNA families are differentially recognized to allow for correct processing complex formation is not understood. As discussed, ARS2 is involved in most CBC functions. In addition to the above roles, ARS2 has been shown to promote snRNA 3'-end processing and transport through PHAX [25], cap-proximal processing of the 3'-end of mRNA [25], RNA export through TREX [205], and RNA degradation through the exosome [26]. However, little is known about the mechanistic details of ARS2's role in these processes.

I hypothesized that ARS2 acts as a scaffold to bridge the CBCA to the appropriate processing machinery by interacting with protein and RNA elements. This hypothesis is based on studies of the ARS2 plant orthologue SERRATE in miRNA biogenesis. As described, SE consists of 3 domains in a walking man-like conformation, with the N-terminal, Mid, and zinc finger (ZnF) domains forming the leading leg, body, and lagging

leg, respectively (Figure 3); the unstructured N- and C- termini, as well as the ordered domains, interact with the Microprocessor, nuclear cap-binding complex, and pri-miRNA, and promote miRNA processing [47,53]. In metazoans, ARS2 function is more complex. In addition to binding to the nuclear cap-binding complex and DROSHA of the Microprocessor [2,24], ARS2 interacts with the 3'-end processing machinery through the protein FLASH of the RDH mRNA processing pathway [16]. FLASH has been shown to interact with ARS2 through a 13 amino acid sequence called FARB (FLASH ARS2 Binding), which is necessary and sufficient for interaction [16]. It has also been hypothesized that ARS2 recognizes the stem loop found in the 3'UTR of RDH transcripts [206]. In the absence of ARS2, there is a reduction of properly processed histone mRNA and protein [24], which impairs cell cycle progression through S phase [16].

I have shown that ARS2 knockdown (KD) results in an accumulation of IdU positive cells in the developing mouse retina (Figure 17). We and others have interpreted this result as a delay in S-phase progression. However, to date there has been no kinetic analysis of the cell cycle to support this conclusion. Therefore, we sought to analyze cell cycle dynamics in ARS2 deficient progenitor cells to precisely determine the nature of the cell cycle defect.

Here, we determined the importance of ARS2 and its domains towards miRNA and RDH biogenesis in the myogenic C2C12 cell line. We show depletion or overexpression of ARS2 causes cells to stall in S phase, indicating ARS2 levels must be tightly regulated. To facilitate structure-function analysis, we modeled the structure of ARS2 based on the plant SERRATE structure and identified 2 new domains in mammalian ARS2: an N-terminal DUF3546 domain, which is conserved in SERRATE, and an RRM domain that is present in metazoan ARS2, but not in plants. This structural model information provided the basis for deletion and point mutant analysis to assess the functions of these domains. We performed systematic mutagenesis to decipher which of ARS2's domains mediate interactions with the miRNA and RDH mRNA processing pathways. Our results support a scaffold model, in which ARS2 plays a central role in RNA recognition and processing through protein and RNA interactions.

3.3 Materials and methods

3.3.1 Cell culture and transfection

Primary mouse myoblasts were isolated from P0 C57BL/6 mice as described [172]. Primary and C2C12 myoblasts were cultured in growth media (DMEM supplemented with 10% FBS and 1% penicillin/streptomycin). Transfections were performed using JetPrime Reagent (VWR) as per the manufacturer's instructions. Ars2 siRNAs were purchased from Qiagen and had their targeting sequences are shown in Table 1 (Appendix A). AllStars Negative control siRNA (Qiagen) was used as a control siRNA. shRNAs were purchased from Origene and their targeting sequences are shown in Table 1 (Appendix A).

3.3.2 Plasmids

Mir-24 and miR-155 firefly luciferase reporters were obtained from Signosis. Renilla luciferase was obtained from Promega. For the histone DsRed reporter, the first 67 amino acids encoded by the H2A ORF were fused to the 3'UTR of histone H2A, including the histone downstream element (HDE), but not the downstream polyadenylation signals (adapted from [207]). This was cloned into pIRES2-DsRed-Express using EcoRI and BstX1 restriction sites to generate histone DsRed. EnhancedGFP-C1 (eGFP-C1) was used as a GFP expression control, and ARS2 (wildtype and mutants) was synthesized and cloned into pEGFP-C1 by Bio Basic to generate eGFP-ARS2. For eGFP-FARB, the FARB peptide sequence [16] was cloned into pEGFP-C1. For scrambled-FARB, a random generator scrambled the FARB sequence to EGEIESELDEDDR, which was then cloned into pEGFP-C1. 3×FLAG-ARS2 and ARS2-mychis were generated as described previously [7].

3.3.3 Cell cycle analysis

For BrdU pulse-chase experiments, C2C12 cells were seeded on 6 cm plates at 6.0×10^5 cells/plate, and 18 h later were transfected with the indicated plasmids. 24 h post-transfection, cells were passaged onto a 10 cm plate to allow growth. The next day, cells were processed according to the protocol in [208]. Briefly, cells were given a 30 min BrdU pulse and fixed at intervals between 0 and 10 h. Cells were labeled with propidium

iodide (PI) to measure DNA content and 1:50 mouse anti-BrdU-AF488 conjugate (Life technologies). For cells transfected with eGFP vectors, cells were also labeled with 1:400 chicken anti-GFP (Abcam) and 1:100 donkey anti-chicken-AF647 (Jackson ImmunoResearch) since HCl treatment abolished GFP fluorophore activity. 1.0×10^4 GFP +ve events per sample were collected on a BD FACSCalibur flow cytometer.

3.3.4 Luciferase reporter assays

C2C12 cells were seeded on 96-well plates at 1.0×10^4 cells/well, 18 h later were co-transfected with the indicated eGFP vector, firefly luciferase reporter and Renilla luciferase vector. 24 h post-transfection, samples were analyzed for luminescence using the Dual-Glo Luciferase Assay System (Promega) on a Perkin Elmer Victor^{3}V 1420 multilabel plate counter.

3.3.5 DsRed reporter assays

For the histone DsRed reporter assay, cells were seeded onto 12-well plates at 2.5×10^5 cells /well. 18 h later, cells were co-transfected with shRNA (containing a turbo GFP (tGFP) cassette) or eGFP/eGFP-ARS2 along with histone DsRed. 48 h post-transfection, 1.0×10^5 events were acquired using a BD FACSCalibur. Reporter activity was quantified as the % double +ve (DsRed/GFP)/total transfected cells. For the let-7c DsRed reporter, C2C12 cells were seeded at 5.0×10^4 cells/well, 18 h later were co-transfected with shRNA and let-7c DsRed, and 48 h post-transfection were analyzed as described for histone DsRed.

3.3.6 Apoptosis assay

C2C12 cells were seeded on 24-well plates at 4.0×10^4 cells/well, and transfected with either siRNA or shRNA 18 h later. 24 h post-transfection, cells were analyzed for apoptosis using BD Biosciences Annexin V-FITC Apoptosis Detection Kit according to the manufacturer's instructions using a BD FACSCalibur.

3.3.7 RNA immunoprecipitation

Immunoprecipitation (IP) was performed using EZview M2 anti-FLAG beads (Sigma-Aldrich) as described in [25]. Briefly, cells were extracted in HNTG buffer for 30 min at 4°C, and cellular debris was removed by centrifugation (20 min at 20,000g).

Extracts were incubated with beads for 2 h at 4°C. Beads were washed twice with HNTG buffer and three times in PBS. For samples treated with a high salt wash, beads were washed once in low salt buffer (50mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5% Triton X-100, 0.5% NP-40), once in high salt buffer (50mM Tris-HCl, pH 7.5, 450 mM NaCl, 0.5% Triton X-100, 0.5% NP-40), and again in low salt buffer. Beads were resuspended in TRIzol (Life Technologies), RNA was isolated using Direct-zol RNA MiniPrep (Zymo research), and cDNA was synthesized using random primers with the high capacity cDNA reverse transcription kit (Life technologies).

3.3.8 qRT-PCR

Histone transcripts were amplified using Ssofast EvaGreen Supermix (Bio-rad) on a Stratagene MX3000P qPCR system. Primers are listed in Table 1 (Appendix A – Supplementary Information). Pri-miRNAs were amplified using TaqMan Universal PCR Master Mix, no AmpErase UNG (Life technologies) and TaqMan primers. Pri-miR-24 was amplified using mmu-mir-24-2 and pri-miR-155 was amplified using mmu-mir-155 (Life technologies). For samples sorted for high GFP, cells were sorted on a BD Influx at the Deeley Research Centre, and relative quantification of RNA was analyzed using the $\Delta\Delta C_t$ method. For RNA IPs, fold enrichment was quantified as $2^{(Ct\ untransfected\ -\ Ct\ ARS2/ARS2Mutant)}$. Equal pulldown efficiency was verified by Western blotting.

3.3.9 Immunoprecipitation

Cells co-transfected with Drosha-FLAG or 3×FLAG-ARS2 and eGFP-ARS2 or eGFP-FARB were lysed in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, pH 7.5) for 30 min at 4°C, and debris was removed by centrifugation (20 min at 20, 000g). For samples treated with RNase, RNase A was added to the lysis buffer at a concentration of 100 μg/mL. Extracts were incubated with anti-FLAG beads for 2 hours at 4°C. Beads were washed three times with 1:4 lysis buffer:PBS, and eluted with 3×FLAG peptide (Sigma-Aldrich). For samples treated with a high salt wash, beads were washed once in low salt buffer, once in high salt buffer, and again in low salt buffer before elution.

3.3.10 Western blotting

Cell lysates were resuspended in 2× Laemmli sample buffer, resolved by SDS-PAGE (10%) and transferred to PVDF membranes. The membrane was blocked in 5% dehydrated milk in Tris-buffered saline-Tween 20 (0.5%) (TBST) for 1 h. Antibodies were diluted in TBST 1% dehydrated milk, at the following concentrations: 1:4000 mouse anti-Actin (Sigma), 1:2000 rabbit anti-ARS2 (XL12.2) 1:1000 mouse anti-FLAG M2 (Agilent Technologies), 1:1000 mouse anti-GFP (Roche), 1:500 rabbit anti-CBP20 (Novus Biologicals).

3.3.11 Immunofluorescence

To measure histone H3 levels, cells were fixed for 15 min with 1% paraformaldehyde (PFA)-PBS at 4°C, washed, and incubated in 70% ethanol for ≥1 h. Samples were washed and incubated in 1:2000 rabbit anti-histone H3 (Abcam) overnight at 4°C. Samples were then washed and incubated in 1:100 goat anti-rabbit AF647 (Life technologies) for 2 h at 4°C, resuspended in PBS/BSA, and analyzed on a BD FACSCalibur. For fluorescence microscopy, cells were transfected on 6-well dishes with the indicated eGFP-ARS2 constructs. 24 h post-transfection, cells were live imaged at 40× magnification using a Leica DMIRE2 inverted fluorescent microscope.

3.3.12 Silver stain

Cell lysates were resuspended in 2× Laemmli sample buffer and resolved by SDS-PAGE (8%). Gel was washed in H_20, fixed for 1 hr (40% ethanol, 10% acetic acid, 50% H_20), and washed overnight in H_20 on a shaker. Gel was sensitized for 1 min (0.02% $Na_2S_2O_3$), washed 3× in H_20, and incubated in 0.1% $AgNO_3$ for 20 mins at 4°C. Gel then washed 3× in H_20, developed in 3% Na_2CO_3 with 0.05% formaldehyde, and once developed, washed once in H_20, incubated with 5% acetic acid for 5 mins, and washed and incubated in gel drying solution (30% methanol, 5% glycerol). Gel was dried overnight in cellophane.

3.4 Results

3.4.1 ARS2 is required for cell cycle progression

To investigate the role of ARS2 in progenitor cell proliferation, we used the C2C12 myoblast progenitor cell line [171]. This cell line was chosen as it is diploid and has progenitor-like properties, including the ability to differentiate *in vitro*. Furthermore, C2C12 progenitor cells are amenable to culture and transfection, were available in our lab, and our lab had expertise in working with muscle cells *in vitro*. Given the important role of ARS2 in stem and progenitor cells [3–5,7,9], we felt it was important to examine its function in this context. Knockdown of ARS2 with multiple siRNA or shRNA (Figure 20A-C) reduced the growth of primary and C2C12 myoblasts (Figure 20D). ARS2 KD increased the accumulation of cells in S phase 48 hours post-transfection, as measured by an increased proportion of BrdU positive cells (Figure 20E) and by propidium iodide (PI) incorporation using flow cytometry (Figure 20F). Additionally, C2C12 stable clones expressing Ars2 shRNA accumulated in S phase relative to non-targeting control shRNA-expressing stable clones (Figure 20G). To demonstrate the decrease in growth was the consequence of improper cell cycling and not due to an increase in cell death, we also examined apoptosis following ARS2 knockdown using both siRNA and shRNA. As shown in Figure 20H, decreased expression of ARS2 in C2C12 cells did not increase apoptosis. ARS2 deficient cells do eventually progress through the cell cycle (Figure 21), which is consistent with a previous report [2], our results in the developing retina (see Chapter 2), and the fact that we were able to maintain stable knockdown cell lines and did not observe an increase in apoptosis. Taken together, these results indicate ARS2 deficient cells display slower cell cycle kinetics as a result of an S phase delay.

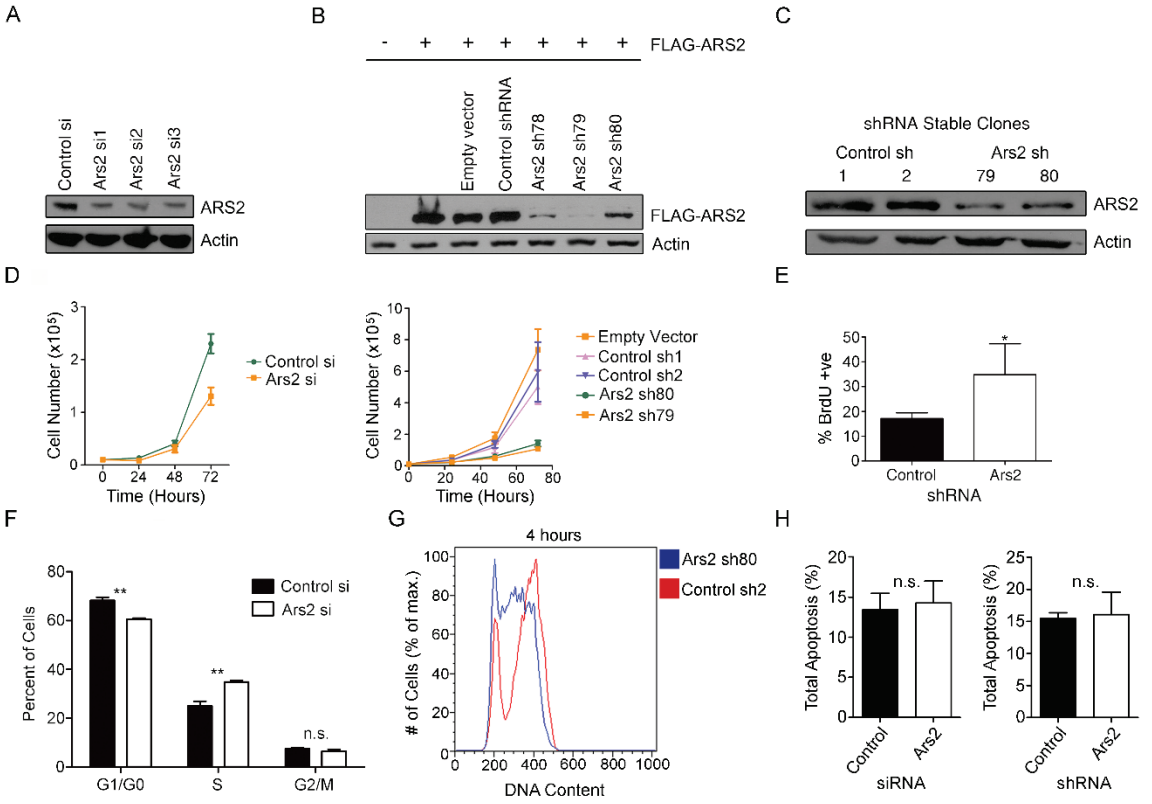

Figure 20 ARS2 is required for cell cycle progression

Western blots showing A) knockdown of endogenous ARS2 in C2C12 myoblasts using three different siRNA B) knockdown of FLAG-ARS2 using three different shRNA (lanes 5-7) and C) Knockdown of endogenous ARS2 in C2C12 clones stably expressing *Ars2* shRNA79 or shRNA80. D) ARS2 knockdown decreased cell number in primary myoblasts using siRNA (left) and in C2C12 stable shRNA clones (right). E) ARS2 knockdown increased the proportion of BrdU positive cells as measured by immunocytochemistry and fluorescence microscopy 48 h post-transfection. F) Cell cycle analysis using propidium iodide (PI) and flow cytometry in primary mouse myoblasts 48 h after siRNA transfection shows ARS2 knockdown increases the proportion of cells in S phase. An unpaired samples t-test at each phase was used to compare means in control and knockdown groups. A Bonferroni correction method was used to modify the p value cut-off to 0.05/3 = 0.017 to reflect error due to multiple comparisons. For the S phase comparison, t=8.27, p<0.017. G) C2C12 stable shRNA clones were pulsed with BrdU for 30 min, 4 h later were fixed, labeled with PI and BrdU antibody, and analyzed using flow cytometry. Histogram shows BrdU positive events. H) Knockdown of ARS2 with either siRNA (left) or shRNA (right) did not increase levels of apoptosis 24 h post-transfection as measured by PI and annexin V-FITC using flow cytometry. A two-tailed unpaired t-test was used to determine whether there was a significant difference between control and Ars2 knockdown in E) (n=3, p<0.05) and H) (n=3, p>0.05). * = p<0.05 ** = p<0.01, n.s. = not significant, and error bars represent standard deviation (SD).

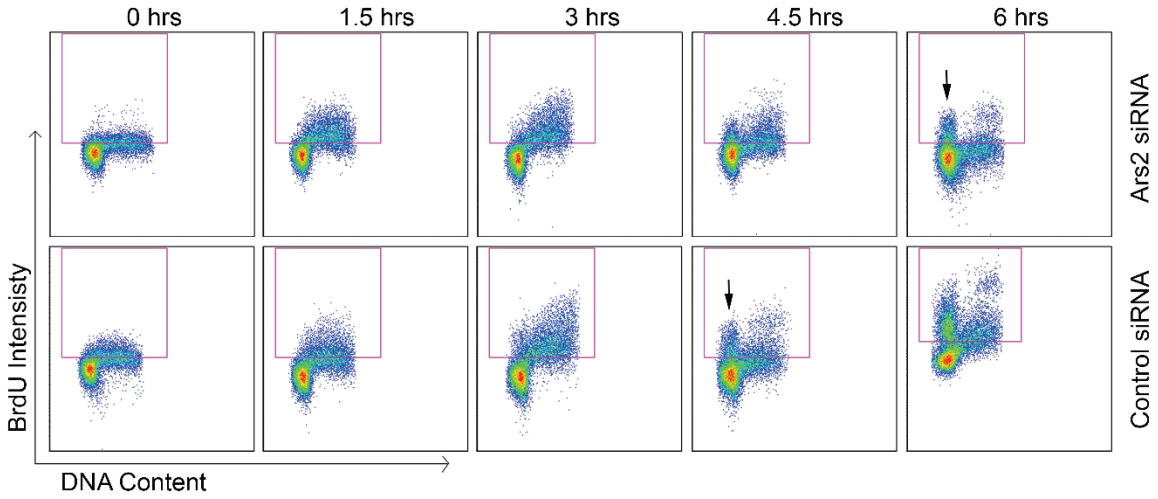

Figure 21 ARS2 knockdown delays cell cycle progression

C2C12 myoblasts transfected with Ars2 or control siRNA (top and bottom panels, respectively) were given a 30 min BrdU pulse at time 0 to mark cells in S phase, fixed every 1.5 h, labeled with PI to measure DNA content, and BrdU +ve cells were tracked over time. Cells were gated for GFP, and total BrdU positive events are in the large gate. Cells in S phase at the time of the pulse took longer to progress through to G1/G0 following ARS2 KD relative to control (indicated by black arrows).

3.4.2 ARS2 overexpression arrests cells in early S phase

Since ARS2 is a critical component of several multiprotein/RNA complexes, we reasoned that overexpression of ARS2 may generate a dominant-negative phenotype. Such dominant-negative effects through overexpression are common in genes encoding components of multi-protein complexes, where misexpression of one component can disrupt complex stoichiometry [209]. To determine if ARS2 overexpression conferred a dominant-negative phenotype, myoblasts were transfected with GFP-ARS2, pulsed with BrdU for 30 minutes to identify the cells in S phase, and the DNA content and BrdU incorporation were determined for the GFP population at various time points. As shown in Figure 22A, overexpression of GFP-ARS2, but not GFP alone, generated a high-intensity BrdU population. This high BrdU population accumulated in early S phase immediately post-pulse (0hr) and failed to progress through the cell cycle in the subsequent 10 hours. This effect was evident in the high GFP expressing cells, which presumably also express the highest levels of ARS2 (Figure 22B), whereas low GFP/ARS2 levels did not show this high BrdU early S phase accumulation (Figure 22C). Thus, while reducing the levels of ARS2 slows cell cycle progression, the high-level

overexpression of ARS2 results in an early S phase cell cycle arrest through a dominant-negative effect.

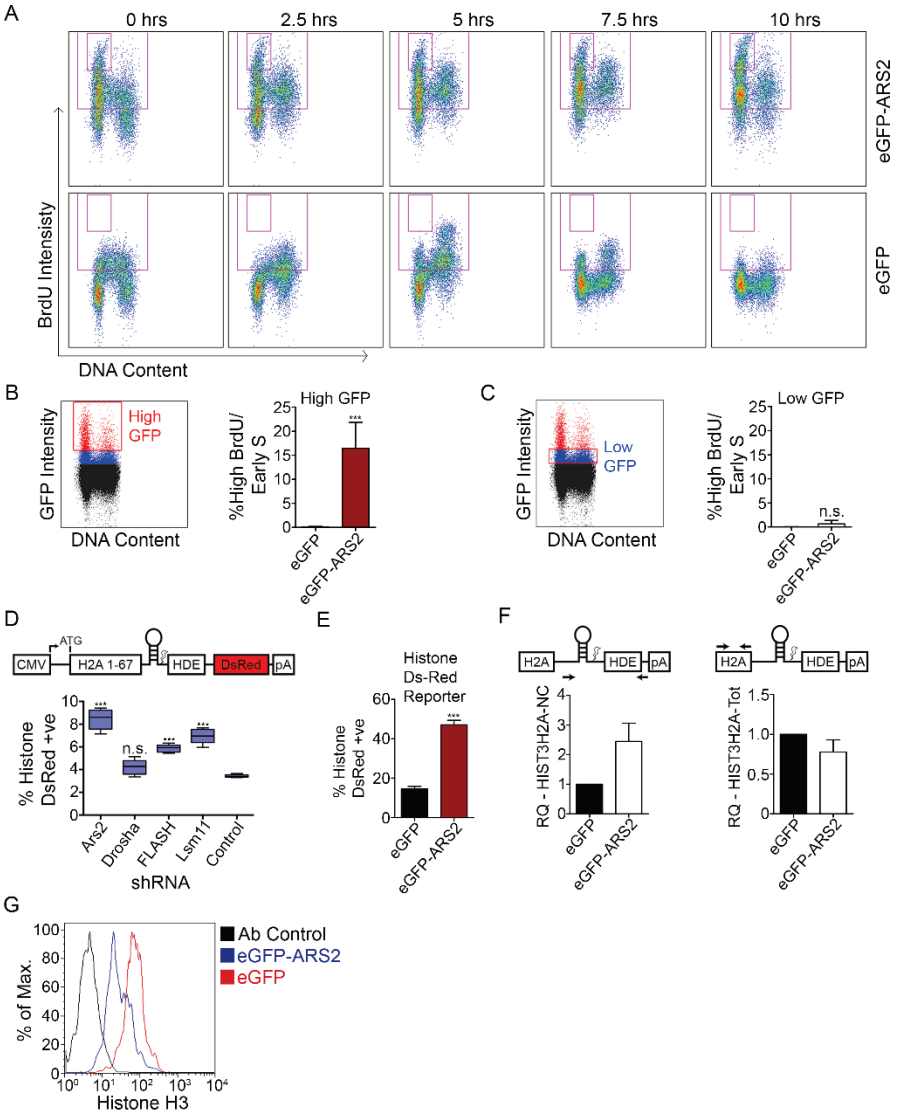

Figure 22 ARS2 dominant negative affects histone processing and expression

A) C2C12 myoblasts transfected with eGFP-ARS2 or eGFP (top and bottom panels, respectively) were given a 30 min BrdU pulse, fixed every 2.5 h, labeled with PI to measure DNA content, and BrdU +ve cells were tracked over time. Cells were gated for GFP, total BrdU positive events are in the large gate, and high BrdU/early S phase cells are in the small inset gate. B) Cells treated as in A) were gated for high GFP intensity (red gate). Graph shows the % of high GFP gated cells (average % of each time point) contributing to the high BrdU/early S phase arrested population. C) Cells treated as in A) were gated for low GFP intensity (red gate). D) Cells were co-transfected with shRNA containing a tGFP cassette and a histone DsRed reporter that expresses DsRed upon disrupted processing [25,207]. The % double positive (DsRed, GFP) is shown relative to total transfected cells. E) Cells were co-transfected with eGFP or eGFP-ARS2 and the

histone DsRed reporter, and the % double positive is quantified as in D). F) Non-cleaved HIST3H2A (left) and HIST3H2A-Total (Tot) mRNA (right) was quantified using qPCR with the $\Delta\Delta C_t$ method from cells sorted for high GFP intensity. GAPDH was used as a control RNA and samples were normalized to eGFP samples. G) Cells were co-transfected with eGFP or eGFP-ARS2, gated for high GFP intensity, and histone H3 levels were measured using flow cytometry. A two-tailed unpaired t-test was used to determine whether there was a significant difference between eGFP and eGFP-ARS2 in B) ($n=5$, $p<0.001$), C) ($n=5$, $p>0.05$) and E) ($n=5$, $p<0.001$). A one way ANOVA ($F(4,20)=56.79$, $p<0.001$) followed by Tukey's multiple comparison *post hoc* test was used to determine whether there were significant differences between groups in D). *** = $p<0.001$, n.s. = not significant, error bars represent SD.

3.4.3 ARS2 is required for histone processing and expression

The high BrdU population associated with ARS2 overexpression is intriguing, since arrest in early S phase should decrease the amount of BrdU incorporated per cell. However, the increase in BrdU labeling suggested a likely mechanism for the arrest. Since ARS2 is required for RDH expression [16,24], we reasoned that the increase in BrdU signal we observe may simply reflect an increase in the accessibility of antibody to BrdU-DNA due to a histone deficiency, and under-packaged chromatin. To examine this, we knocked down ARS2, or components of the RDH mRNA processing machinery, FLASH and LSM11, and assessed histone mRNA processing using a reporter in which DsRed expression is dependent on improper histone 3'-UTR processing [25] (Figure 22D). This assay revealed elevated levels of reporter expression following ARS2, FLASH and LSM11 knockdowns, but not in DROSHA knockdowns (Figure 22D), confirming DROSHA is not involved in RDH mRNA processing. Overexpression of ARS2 similarly increased expression of the reporter, suggesting improper processing of histone transcripts (Figure 22E). To further evaluate this, we isolated the high GFP populations and assessed the levels of endogenous non-cleaved histone H2A transcripts using qRT-PCR with primers that amplify across the 3'-UTR cleavage site. Overexpression of GFP-ARS2 increased the levels of misprocessed H2A transcripts (Figure 22F). Consistent with previous reports, the levels of total H2A transcripts declined slightly (Figure 22F) [16,24]. In addition, defective RDH transcript processing led to a decrease in histone H3 protein levels following ARS2 overexpression (Figure 22G). Although we cannot rule out contributions to the cell cycle phenotype from other types of RNA processed by ARS2 with known roles in the cell cycle [210–212], histone

deficiency during S phase and the inability to package chromatin caused by improper processing of histone RNAs are likely contributing to the S phase arrest associated with high levels of ARS2 overexpression.

3.4.4 Predicted structures of ARS2 and RRM domains

To help understand how SERRATE functions, Machida et al. determined its crystal structure [53], revealing a unique "walking man-like" topology of 3 domains: an N-terminal, Mid, and Zinc Finger (ZnF) domain, which form the leading leg, body, and lagging leg, respectively [53] (Figure 23A). The N and C-termini of the protein are unstructured. The recent crystal structure of *Arabidopsis* SERRATE created an opportunity to model the mouse ARS2 sequence and re-examine the conserved sequences of ARS2 for additional functional groups. Although the overall identity between mouse ARS2 and *Arabidopsis* SERRATE is relatively low (~28%), bioinformatic analysis showed the DUF3546, Mid, and Zinc finger (ZnF) domains are conserved (Figure 23D). However, ARS2 diverges considerably from SERRATE at the unstructured C-terminus, and interestingly, contains an insert within the Mid domain that is not present in SERRATE (Figure 23D). Alignment of this insert within metazoans shows it contains an unstructured glutamate-rich region that is followed by an RNA recognition motif (RRM). Homology modelling of the mouse ARS2 RRM predicted a fold similar to the RRM domain of Splicing Factor 3B subunit 4 (SF3B4) (Z score 14.9, r.m.s.d. 0.1 Å, 73 Cα) (Figure 23B).

We next modeled the regions that were conserved between mouse ARS2 (minus the unstructured N and C-termini, glutamate-rich and RRM domains) and SERRATE. The computational model conforms very well to the previously described SERRATE structure (Z-score 30.1, r.m.s.d 0.7 Å, 279 Cα) (Figure 23A,C). Importantly, the glutamate-rich and RRM domains are expected to insert into the region of ARS2 between the predicted α–helix 4 and 5 (Figure 23C,D). The orientation of the RRM domain relative to the core ARS2 structure is unknown. C-terminal to the ZnF in metazoans is a proline-rich region. Eight PXXP motifs are present within the unstructured C-terminus of mouse ARS2 (amino acids 763-875) (Figure 23D). These motifs often mediate protein interactions, suggesting the C-terminus may serve this role. In summary, the core of

metazoan ARS2 consists of 4 domains: a DUF3546 domain, a newly identified RRM domain, a Mid domain and a ZnF domain.

The strong dominant-negative phenotype associated with ARS2 overexpression created an opportunity for screening mutations in ARS2 functional domains. Specifically, we asked if conserved amino acids in these regions could abrogate or exacerbate the ARS2 overexpression phenotype. If so, these mutations would reveal the importance of the different ARS2 domains. Since deletions beyond the unstructured N and C-termini affect the overall fold of the protein [53], we decided to mutate several conserved motifs within the DUF3546, RRM, Mid and ZnF domains. We also generated deletions of the N and C-termini. Mutation sites, shown in Figure 23D and Figure 24B, were selected based on their conservation and predicted surface exposure. Because of the large size of the protein and paucity of interaction data, we mutated 3-4 residues at a time to ensure mutations would abolish function. Each mutant was expressed as a GFP-ARS2 fusion and its localization was observed (Figure 24A). Only mutants that exhibited wildtype localization, expression level, and retained some ability to associate with protein and/or RNA were included in our phenotypic analysis. Deletion of the N-terminal residues (amino acids 1-227) resulted in cytoplasmic localization (Figure 24A), confirming the presence of a nuclear localization signal within this region [7]. We focused our analysis on mutants with nuclear expression and their effects on the RDH and miRNA biogenesis pathways, as well as the CBC, since ARS2 has been shown to form direct protein interactions with these complexes.

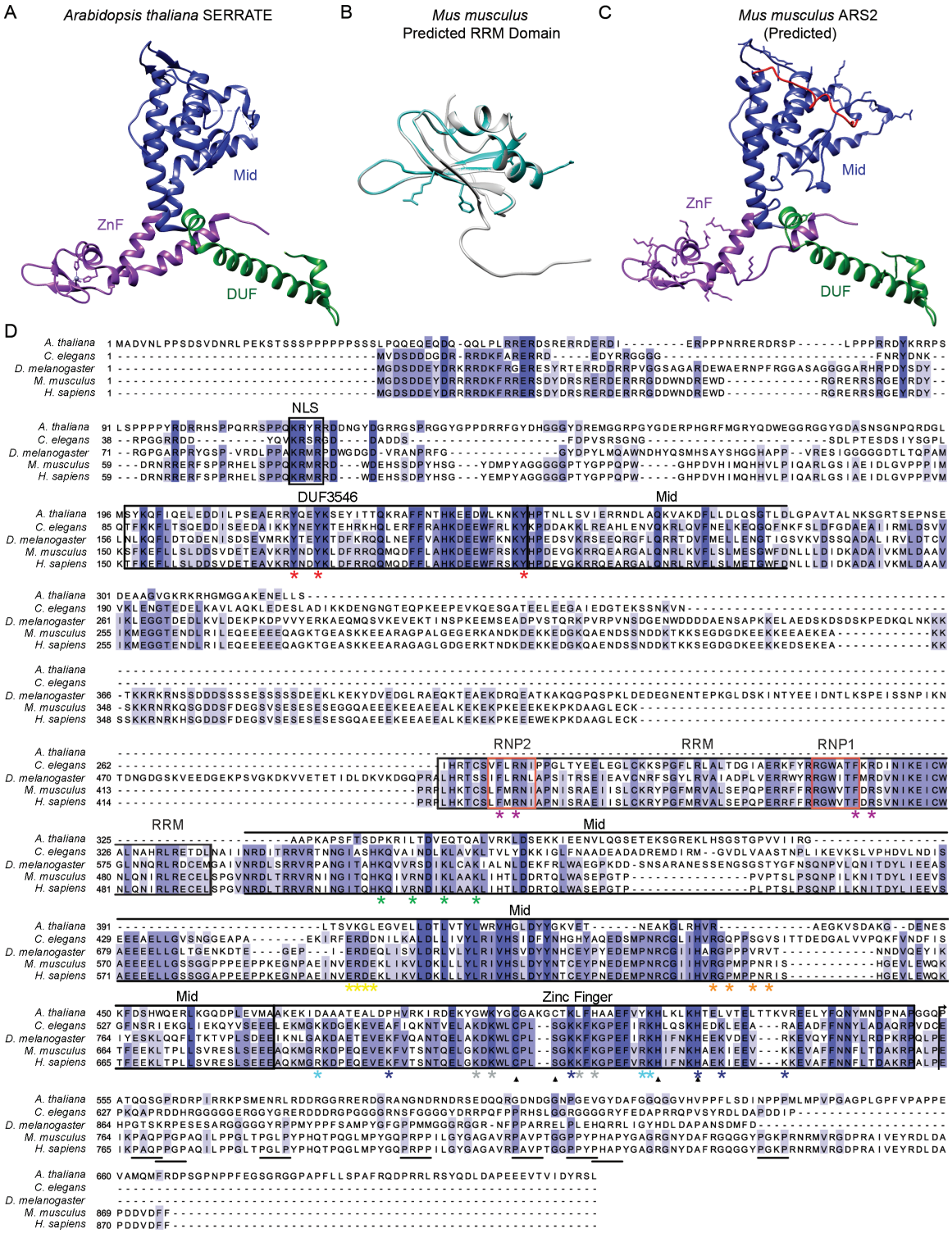

Figure 23 Predicted structures of ARS2 and RRM domains

A) Crystal structure of *Arabidopsis thaliana* SERRATE [53] (PDB accession 3AX1) highlighting the DUF3546 domain (green) Mid domain (blue) and ZnF domain (purple). B) *Mus musculus* ARS2 RRM domain (blue-green) was modeled using SF3B4 as a reference (PDB accession 1X5U, shown in white) [213]. The mutated residues of ARS2

RRM are highlighted. C) *Mus musculus* ARS2 amino acid sequence (without the glutamate-rich and RRM domain insert) was modeled using *Arabidopsis* SERRATE as a reference [213]. Domains are highlighted as in A) with the exception of the disordered region coloured in red, which is where the glutamate-rich region and RRM domain are predicted to be inserted. Also highlighted are the residues chosen to be mutated to alanines, in addition to the residues predicted to coordinate Zinc. D) Bioinformatic alignment of SERRATE/ARS2 using Clustal Omega [214] and manual alignment. Asterisks indicate amino acids that were mutated, arrowheads indicate residues predicted to coordinate Zinc, and the arrow indicates the C-terminal region that was deleted in ARS2-ΔC. The regions that correspond to structural motifs or domains are outlined with boxes and labeled above the alignment. NLS indicates a predicted nuclear localization sequence. PXXP motifs are underlined below the alignment.

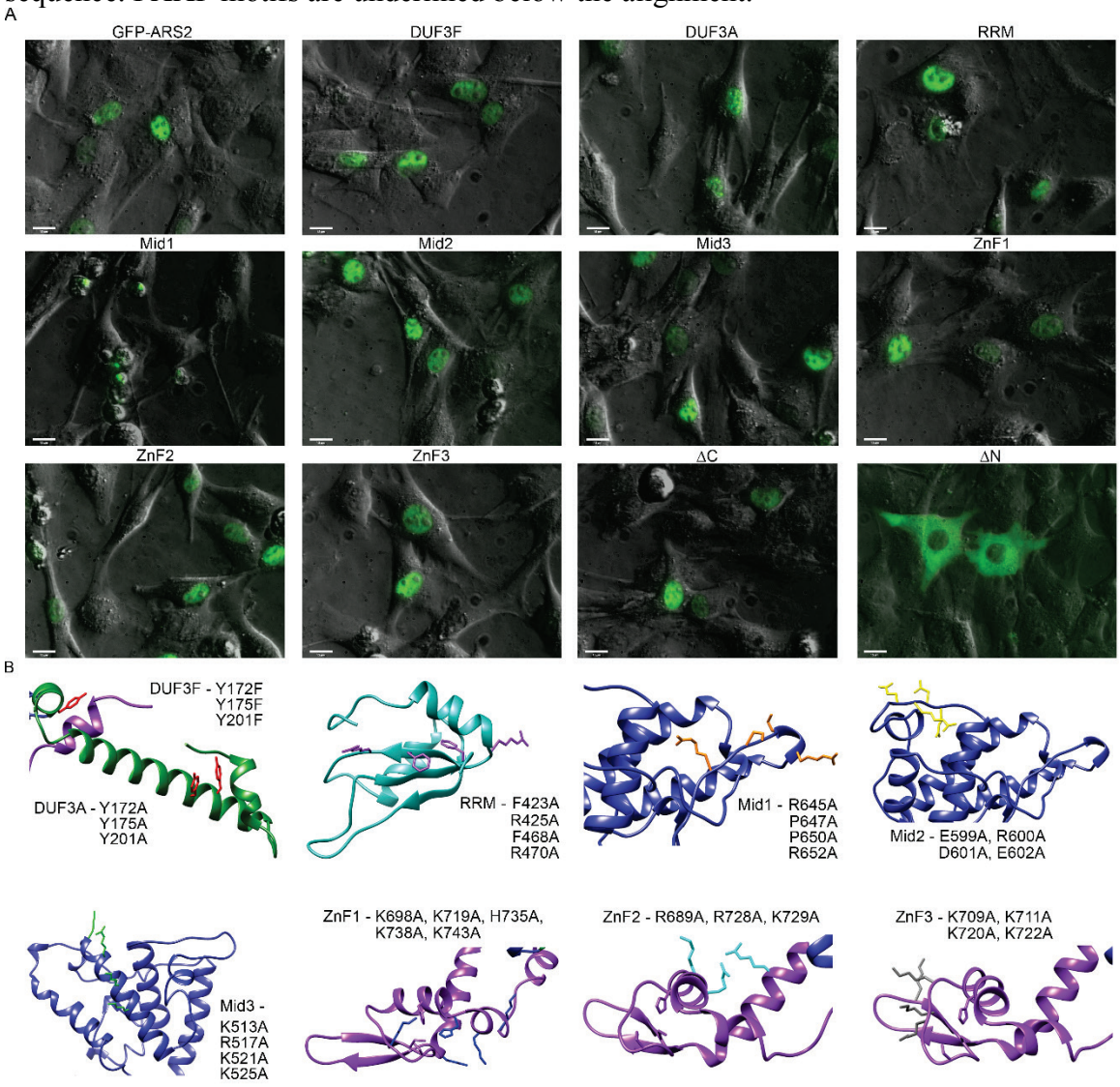

Figure 24 ARS2 mutant localization

A) C2C12 cells transfected with eGFP-ARS2 or the indicated mutants were live imaged at 40× magnification. Shown is an overlay of GFP and phase contrast. Scale bar = 15 µm.

B) Using the predicted model structures of ARS2 from Figure 23B and C, the positions of the mutated amino acids are highlighted using the same colour scheme as in Figure 23D, and the names of the specific mutations are shown adjacent to their respective locations within the structure.

3.4.5 The Zinc Finger domain mediates interactions with FLASH and RNA

Amino acids 550-876 of ARS2 (ZnF and proline-rich region) have previously been shown to be important for RNA and FLASH/FARB binding [16]. To refine the region of ARS2 required for FLASH interaction, we tested the ability of ARS2 ZnF mutants to interact with FLASH. Three groups of mutations were generated within the ZnF, either in the predicted dsRNA interface (ZnF1), or in basic regions (ZnF2 and ZnF3) that might associate with the highly acidic FARB peptide (DELEEGEIRSDDE) (Figure 23C,D and Figure 24B). As endogenous FLASH protein was difficult to detect in C2C12 cells, we used FARB peptide to map the ARS2 region required for FLASH interaction. As shown in Figure 25A and B, ZnF1 and ZnF3 mutants lost the ability of ARS2 to co-immunoprecipitate (co-IP) with FARB. RNase treatment demonstrates that this effect is not mediated by RNA (Figure 25A,B). Since the ZnF2 did not express well (Figure 25A, lane 10), it was not pursued further. By contrast, ZnF1 and ZnF3 maintained the ability to bind to DROSHA (Figure 25C,D) and CBP20 (Figure 25E), confirming these proteins remain folded. Thus, FLASH interaction with ARS2 requires these conserved features of the ZnF domain.

We also examined whether the ZnF mutants were able to induce ARS2 dominant-negative phenotypes. Consistent with the ZnF participating in RDH mRNA processing and FLASH interaction, ZnF1 and ZnF3 were less effective at inducing the ARS2 S phase arrest (Figure 26A). We next determined whether the ZnF was important for RDH mRNA recognition. Interestingly, mutation of the ZnF domain reduced the ability of ARS2 to IP endogenous histone mRNA (Figure 26B) and partially rescued RDH transcript processing, as measured by the DsRed reporter assay (Figure 26C). Thus, the ARS2 ZnF plays a critical role in the recognition, and possibly processing, of RDH mRNA.

The ZnF of SERRATE has been implicated in miRNA biogenesis in plants [47,53]. We examined whether the ZnF domain was required for miRNA biogenesis in mammalian ARS2. Mutations in the ZnF abrogated miRNA interaction (Figure 26D) and

overexpression of these mutants did not elicit a miRNA biogenesis defect (Figure 26E, F), indicating an additional role for this region in miRNA biogenesis. Since FLASH potentially binds to the same domain of ARS2 as miRNA, we next asked whether FLASH was required for miRNA biogenesis using a Let-7c reporter. While knockdown of DROSHA or ARS2 resulted in the expected decrease in mature miRNA, knockdown of FLASH did not affect these miRNA levels (Figure 26G), suggesting FLASH is part of the RDH machinery and is in a mutually exclusive complex from the pri-miRNA processing machinery. Together, these results indicate that the ARS2 ZnF mediates interactions with miRNA, RDH mRNA, and FLASH.

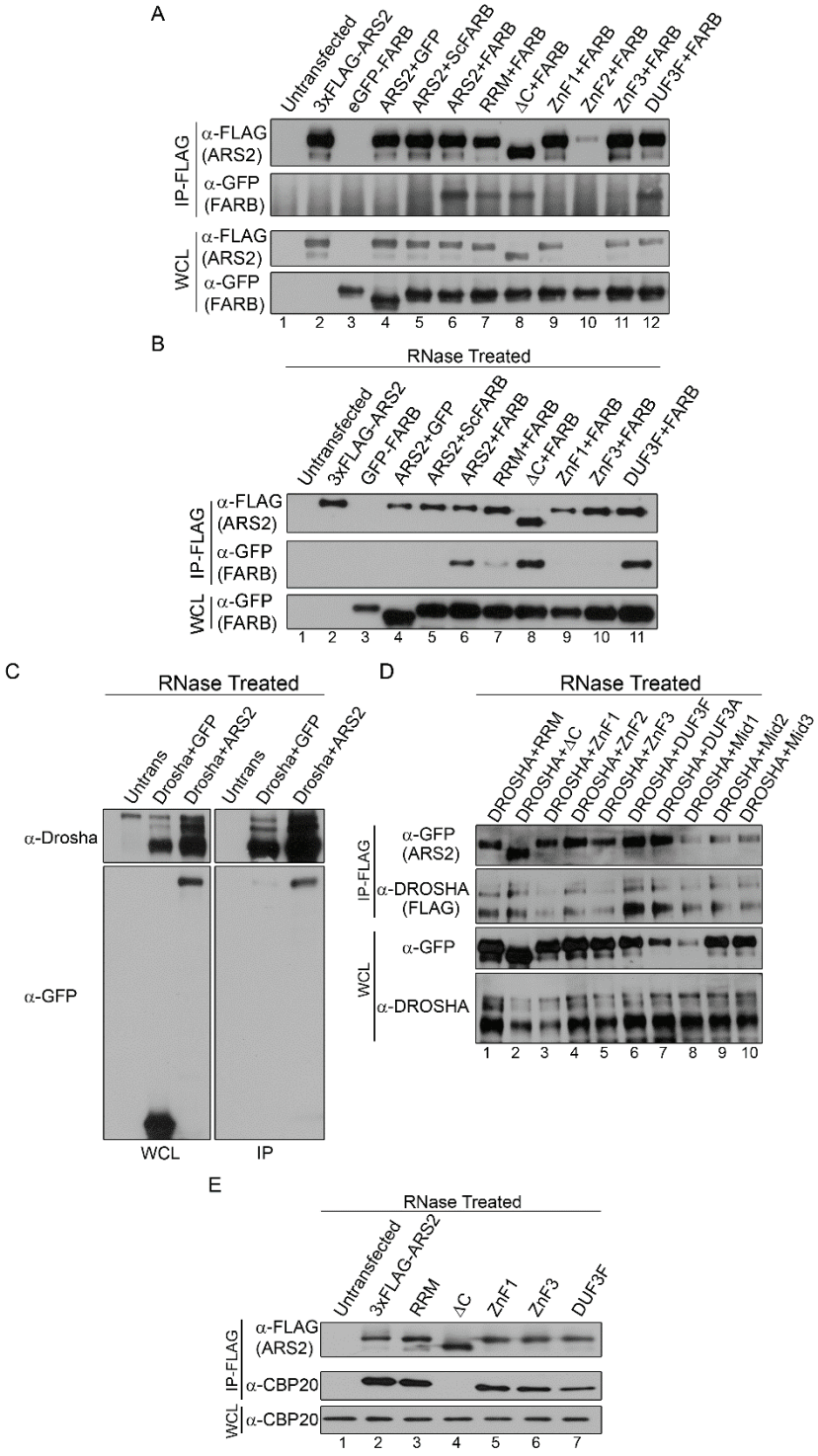

Figure 25 Mapping ARS2 protein interactions

A) Cells were co-transfected with 3×FLAG-ARS2 (or 3×FLAG-ARS2 mutants) and either eGFP control (GFP), eGFP-ScrambledFARB (ScFARB) or eGFP-FARB. IP was carried out using 3×FLAG beads. B) Cells were transfected and immunoprecipitated (IP'd) as in A) but were treated with 100 µg/mL RNase A. C-D) Cells co-transfected with DROSHA-FLAG and eGFP, eGFP-ARS2 or eGFP-ARS2 mutants were treated with 100

μg/mL RNase A and IP'd using 3×FLAG beads. E) C2C12 cells transfected with 3×FLAG-ARS2 or the indicated mutants were treated with 100 μg/mL RNase A and IP'd using FLAG beads.

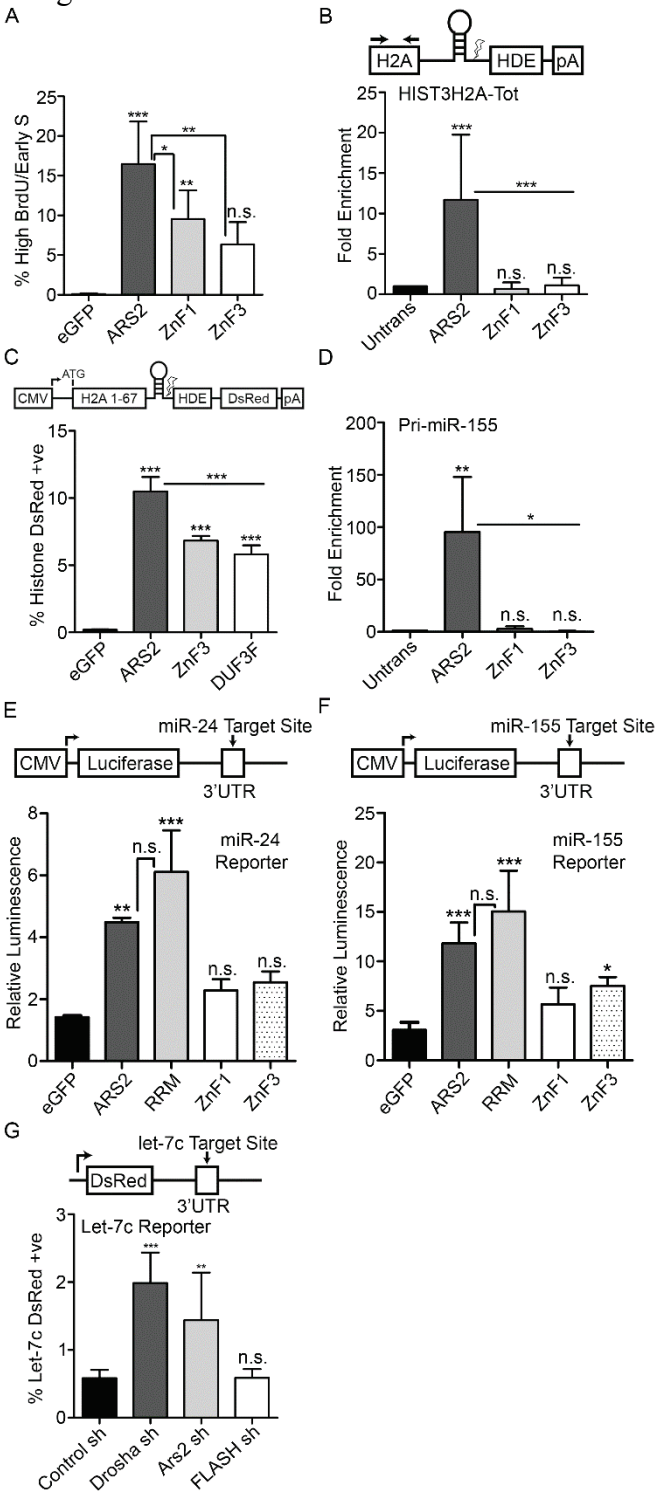

Figure 26 The zinc finger domain mediates interaction with RNA

A) Cells transfected with eGFP, eGFP-ARS2, -ZnF1 or –ZnF3 were pulsed with BrdU, fixed every 2.5 h, and gated for high GFP. Shown is the % of gated cells in the high BrdU/early S phase arrested population (average % of each time point). B) Cells transfected with 3×FLAG-ARS2 or the indicated mutants were IP'd, and the fold enrichment of RNA pulldown was quantified relative to untransfected samples using qPCR for HIST3H2A-Tot, primers upstream of the cleavage site. C) Cells co-transfected with eGFP, eGFP-ARS2, -DUF3F, or -ZnF3 and histone DsRed were quantified as in Figure 22E,D) Samples were treated as in B) and amplified using primers for Pri-miR-155. E-F) Cells were transfected with eGFP, eGFP-ARS2, -RRM, -ZnF1 or -ZnF3 along with the indicated firefly luciferase reporter and Renilla luciferase. Each firefly luciferase reporter contains a miRNA target site of perfect complementarity in its 3' UTR. Firefly luminescence was normalized to Renilla luminescence. G) Cells were co-transfected with shRNA containing a tGFP cassette and a let-7c DsRed reporter which contains a let-7c binding site in the 3'-UTR. The % double positive (DsRed, GFP) is shown relative to total transfected cells. A one way ANOVA was performed in A) ($F(3,16)=18.66$, $p<0.001$), B) ($F(3,18)=11.28$, $p<0.001$), C) ($F(3,16)=210.4$, $p<0.001$) , D) ($F(3,8)=9.65$, $p<0.01$), E) ($F(4,10)=26.18$, $p<0.001$), F) ($F(4,25)=26.90$, $p<0.001$), and G) ($F(3,20)=22.84$, $p<0.001$), followed by Tukey's multiple comparison *post hoc* test to determine whether there were significant differences between groups. *=$p<0.05$, ** = $p<0.01$, *** = $p<0.001$, n.s. = not significant, and error bars represent SD.

3.4.6 The DUF3546 domain is required for miRNA and histone mRNA pulldown

The DUF3546 domain forms the leading leg of the protein and consists of a helix-turn-helix-turn-helix motif. Notably, it contains 3 invariant tyrosine residues in helix α2 and α3 (Tyr 172, 175, 201), which we mutated to alanine or phenylalanine (Figure 23C, D and Figure 24B). Mutating these residues to alanine resulted in poor expression and protein mislocalization, with ARS2 forming nuclear aggregates indicative of protein misfolding (Figure 24A). However, mutating these tyrosines to phenylalanine yielded an ARS2(3F) mutant with distribution and expression levels similar to that of wildtype GFP-ARS2 (Figure 24A).

To assess the function of the DUF3546 domain's conserved tyrosines, we first examined the phenotypic consequences of the DUF3F mutant in cell cycle progression. Upon overexpression, the DUF3F mutation was not able to induce the cell cycle arrest associated with wildtype ARS2 overexpression (Figure 27A). This demonstrates that these amino acids in the DUF3546 domain are required for the dominant-negative effect of ARS2 overexpression. To investigate the mechanism behind this effect, we first examined the ability of this mutant to interact with RDH mRNA. Interestingly, the 3F mutations impaired ARS2's ability to IP both total and non-cleaved misprocessed RDH

mRNAs (Figure 27B,C). This mutation also partially rescued RDH mRNA processing, as measured by the DsRed reporter (Figure 26C). We next assessed whether the 3F mutant's inability to pull down RDH mRNA was specific for this type of RNA or was impaired in miRNA interaction as well. Similar to the RDH mRNA, this mutant was impaired in its ability to IP miRNA (Figure 27D), and as a consequence, the ability of this mutant to affect miRNA biogenesis was significantly reduced, as measured by a miR-155 luciferase reporter (Figure 27E). In contrast, the ability of this mutant to interact with ARS2 protein binding partners - the FLASH peptide, FARB (Figure 25A,B), DROSHA (Figure 25C,D), or CBP20 (Figure 25E) - was not affected, even in the presence of RNase, confirming these interactions were RNA-independent and the mutant remained folded. Together, these data indicate the DUF3546 domain is necessary, directly or indirectly, for ARS2 interactions with miRNA and histone mRNA.

To assess whether ARS2 interacts directly with histone mRNA and miRNA, IPs were subject to a high salt wash, which effectively removed all detectable ARS2 interacting proteins as determined by silver staining (Figure 27F). Wildtype ARS2 remained associated with miRNA and histone mRNA after a high salt wash (Figure 27G,H). Consistent with its inability to IP RDH and miRNAs under milder conditions (Figure 26B,D), the high-salt washed ZnF mutant did not associate with RDH mRNA or miRNA (Figure 27G,H). However, we find that when co-purifying factors are removed from ARS2, the DUF3F mutant retains both miRNA and RDH mRNA binding ability (Figure 27G, H). The contrasting results between the low and high salt washes with the DUF3F mutant suggests that this domain may facilitate ARS2-RNA interactions *in vivo*, but is not a necessary component of the protein-RNA interface. Regardless, since purified ARS2 captures RNA as efficiently as ARS2-multiprotein complexes, it is likely that ARS2 binds directly to miRNA and RDH mRNA, and this interaction requires the ZnF domain.

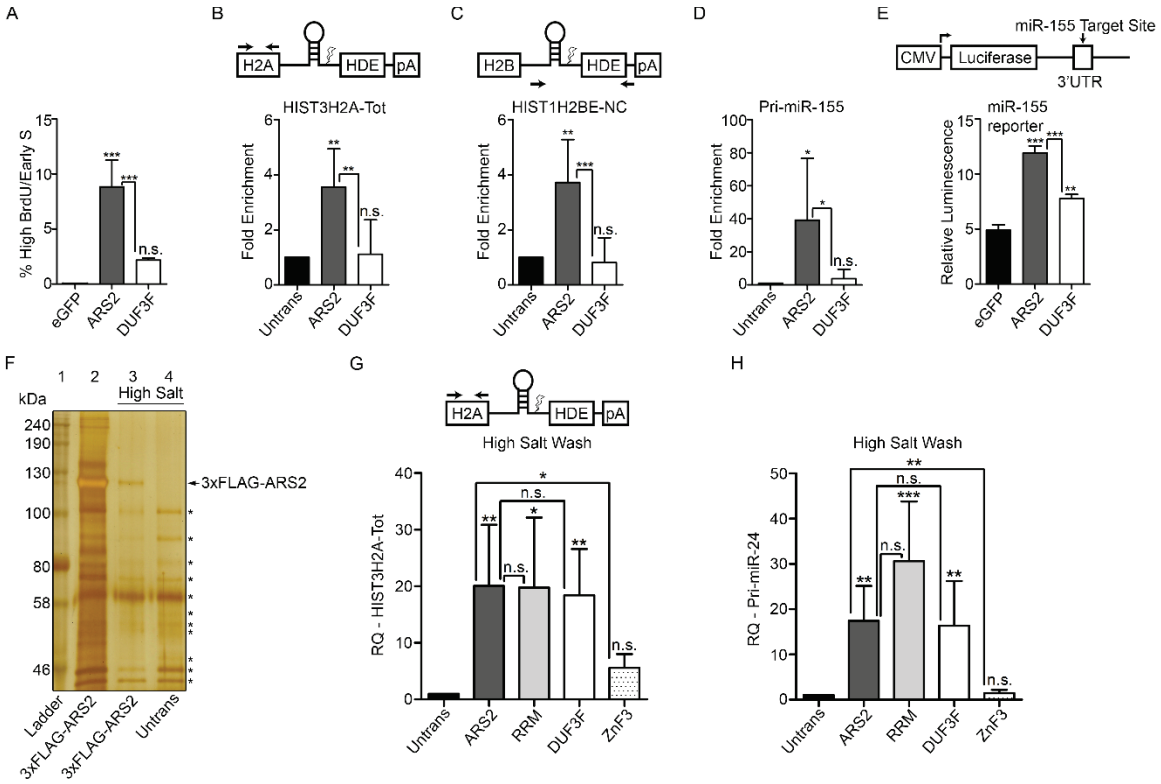

Figure 27 The DUF3546 domain is required for miRNA and histone mRNA pulldown

A) C2C12 Cells transfected with eGFP, eGFP-ARS2 or -DUF3F were treated as in Figure 22A and gated as in Figure 22B. B-D) Cells transfected with 3×FLAG-ARS2 or –DUF3F, protein IP'd using 3×FLAG beads, RNA was extracted from IP, and the fold enrichment of RNA pulldown was quantified relative to untransfected samples using qPCR for B) HIST3H2A-Tot C) HIST1H2BE-noncleaved (NC) and D) pri-miRNA-155. For HIST3H2A-Tot, primers amplified the histone mRNA upstream of the cleavage site, and for HISTH2BE-NC, primers span the cleavage site. E) Cells were co-transfected with eGFP, ARS2, or DUF3F, miR-155 firefly luciferase, and Renilla luciferase. Firefly luminescence was normalized to Renilla luminescence. F) Silver stain of elution following IP of 3×FLAG-ARS2 under normal conditions (lane 2) or following a high salt wash (lane 3). An untransfected sample treated with a high salt wash was used as a negative control (lane 4). Asterisks indicate non-specific bands present in both ARS2 and untransfected conditions (lane 3 and 4). G-H) 3×FLAG-tagged proteins were IP'd using 3×FLAG beads from cells transfected with the indicated plasmids. Following a high salt wash, RNA was extracted, and HIST3H2A-Tot (G) or pri-miR-24 (H) was quantified using qPCR with the $\Delta\Delta C_t$ method. GAPDH was used as a control RNA and samples were normalized to the untransfected control. A one way ANOVA was performed in A) ($F(2,12)=52.73$, $p<0.001$), B) ($F(2,15)=10.82$, $p<0.01$), C) ($F(2,15)=14.81$, $p<0.001$), D) ($F(2,15)=5.75$, $p<0.01$), E) ($F(2,6)=142.1$, $p<0.001$), G) ($F(4,22)=7.89$, $p<0.001$), and H) ($F(4,22)=13.33$, $p<0.001$) followed by Tukey's multiple comparision *post hoc* test to determine whether there were significant differences between groups. * = $p<0.05$), ** = $p<0.01$, *** = $p<0.001$, n.s. = not significant, and error bars represent SD.

3.4.7 The RRM domain is required for cell cycle progression and involved in FLASH interaction

During our bioinformatics analysis, we identified a previously undescribed RRM domain within metazoan ARS2. Canonical RRM domains have two conserved sequence motifs: RNP1 ([RK]-[G]-[**FY**]-[GA}-[**FY**]-[ILV]-[X]-[FY]) and RNP2 ([ILV]-[**FY**]-[ILV]-X-N-L) [215]. Critical aromatic residues (in bold) of RNP1 and RNP2 form stacked pi interactions with RNA bases and the arginine/lysine of RNP1 forms multiple hydrogen bonds with nitrogenous bases [216]. Non-canonical RRMs diverge from the consensus as they utilize other interfaces to interact with RNA or protein [217]. Analysis of ARS2 indicates that it contains a conserved RNP2-like motif (LFMRNI). The RNP1 (RGWVTFDR) deviates from the consensus (Figure 23D) and it is therefore unclear whether the ARS2 RRM mediates RNA binding.

Based on the modeled RRM structure, we mutated conserved residues within RNP1 (F468A, R470A) and RNP2 (F423A, R425A) (Figure 23B,D and Figure 24B). This mutant expressed at wildtype levels and showed normal localization (Figure 24A). However, these mutations had no effect on histone mRNA or miRNA pulldown under low or high salt wash conditions (Figure 27G,H, Figure 28A-D).

We next asked if the RRM is required for the ARS2 cell cycle progression phenotype. Expression of the GFP-ARS2 RRM mutant resulted in an increase in the early S phase-arrested high BrdU population compared to wildtype GFP-ARS2 (Figure 28E). With this mutant, even low GFP-ARS2 expressing populations resulted in high BrdU incorporation and cell cycle arrest in early S phase (Figure 28F). Additionally, the RRM mutant resulted in decreased levels of histone H3 protein relative to wildtype ARS2 (Figure 28G,H), in agreement with histone deficiency being a major contributor to the cell cycle phenotype.

Consistent with the strong effect of this mutation on cell cycle progression and histone expression, the ability of the RRM mutation to interact with FLASH/FARB was reduced, but not lost, under RNase treated conditions (Figure 25B, lane 7). In contrast, the interactions with DROSHA (Figure 25D, lane 1), and CBP20 (Figure 25E, lane 3) were not affected, indicating that global destabilization of the protein as a result of the RRM mutations did not occur. In addition, two independent miRNA biogenesis reporters

did not show a significant change as a result of this mutation, suggesting this domain is not required for miRNA biogenesis (Figure 26E,F). Finally, the reduced binding of the RRM mutant to FARB suggests this domain, and the ZnF domain, are each important for FLASH interaction. Further mutagenesis is required to determine the site of interaction and how these two domains mediate this interaction.

3.4.8 The Mid domain is important for miRNA biogenesis

In *Arabidopsis* SERRATE, the Mid domain comprises a 3 helical bundle (α5-α7) orientated orthogonally against a long fourth alpha helix (α4) [53]. The long unstructured loop regions between each of these three helices were predicted to mediate protein-protein interactions [53]. Within the Mid domain, we generated three groups of mutations that are each based on well-conserved residues: Mid1 (R645A, P647A, P650A, R652A), Mid2 (E599A, R600A, D601A, E602A), and Mid3 (K513A, R517A, K521A, K525A) (Figure 23C,D and Figure 24B). DROSHA specifically co-immunoprecipitates (co-IPs) wild-type GFP-ARS2 in the presence of RNase, confirming this interaction is not RNA-dependent [2] (Figure 25C). Under the same conditions, all Mid domain mutations reduced, but did not abolish, DROSHA binding (Figure 25D, lanes 8-10), suggesting these conserved residues may contribute to the DROSHA interaction surface. The Mid1 mutant did not express well and displayed mislocalization (Figure 24A and Figure 25D, lane 8), likely explaining its reduced binding to DROSHA and rescue of the miR-24 and miR-155 reporters (Figure 28I,J). However, Mid2 and Mid3 mutants expressed and localized normally (Figure 24A, Figure 25D), suggesting their ability to interact with DROSHA is compromised. Consistent with this interpretation, Mid2 and Mid3 mutants had a markedly reduced miRNA-misprocessing phenotype (Figure 28I,J); i.e. an impaired ability to elicit a dominant-negative effect on this pathway. Taken together, these results suggest that the DROSHA binding site likely lies within the Mid domain. Further mutagenesis will be required to determine the precise binding site.

3.4.9 The ARS2 C-terminus is required for CBP20 interaction

Hallais, et al. recently showed that the C-terminal amino acids 502-871 of mammalian ARS2 are sufficient for binding to the CBC [25]. Using a deletion spanning amino acids 763-875, we refined this region and found that the unstructured C-terminal

proline-rich region is necessary for CBP20 co-IP (Figure 25E, lane 4). We next utilized the cell cycle defect associated with ARS2 overexpression to assay the involvement of the C-terminus in this phenotype. Perhaps surprisingly, deletion of the proline-rich C-terminus did not affect the cell cycle arrest, as overexpression of the deletion mutant blocked cell cycle progression to the same degree as wildtype ARS2 (Figure 28K). This was despite the C-terminal deletion abolishing the ability of ARS2 to IP histone mRNA (Figure 28A,B). Notably, these data indicate that sequestration of RDH mRNA by ARS2 is not the major mechanism contributing to the dominant-negative cell cycle phenotype. This suggests that while the C-terminus is required for RDH mRNA interaction, the intact domains in ARS2-ΔC (i.e. the RRM, ZnF, and/or DUF3546 domains) may be sequestering proteins necessary for RDH processing and/or expression. Our results with the RRM and ZnF mutants strongly implicate these domains. Furthermore, while the ZnF and DUF3546 domains are necessary for RDH mRNA pulldown, they are not sufficient, as ARS2 requires an interaction with the CBC to tightly associate with this RNA class (Figure 28A,B).

We next examined whether the C-terminal proline-rich region (amino acids 763-875) was required for miRNA biogenesis. As shown in Figure 28L and M, ARS2 requires this region for miRNA biogenesis, since deletion of the C-terminus results in a more severe dominant-negative. However, the C-terminal deletion mutant can still interact with pri-miRNA (Figure 28C,D). This suggests that although ARS2-ΔC retains the ability to interact with pri-miRNA, likely through the DUF3546 and ZnF domains, the C-terminus of ARS2, and presumably the interaction with the CBC, are required for pri-miRNA processing.

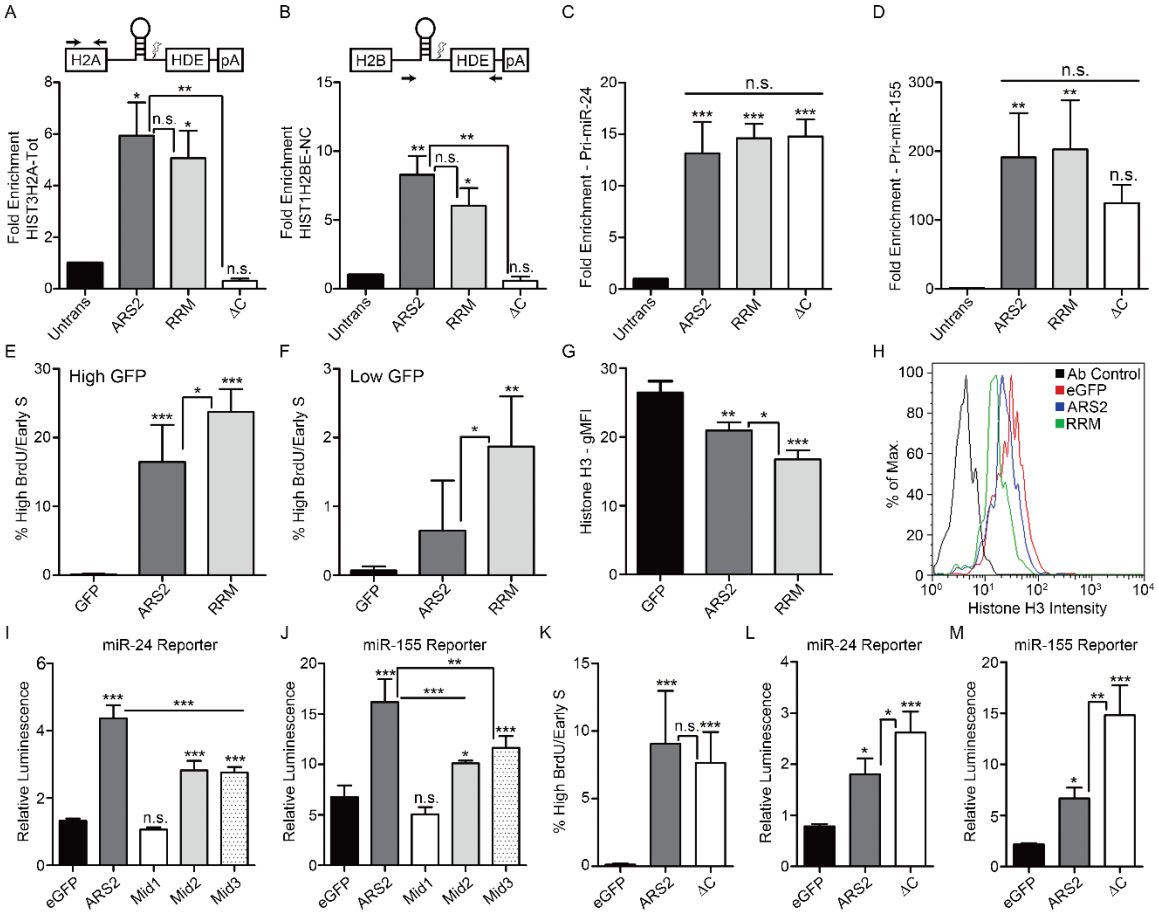

Figure 28 The RRM is required for cell cycle progression, and the Mid domain and C-terminus are required for miRNA biogenesis

A-D) Cells transfected with 3×FLAG-ARS2, -RRM Mut (RRM), and -ΔC were IP'd, and the fold enrichment of RNA pulldown was quantified as in Figure 26B for A) HIST3H2A-Tot, B) HIST1H2BE-NC, C) Pri-miR-24 and D) Pri-miR-155. E-F) Cells transfected with eGFP, eGFP-ARS2, or -RRM were treated as in Figure 22A. Cells were either gated for high GFP (E) or low GFP intensity (F), and the % of gated cells in the high BrdU/early S phase arrested population is shown (average % of each time point). G) Cells were transfected as in E) and histone H3 levels were analyzed using flow cytometry. The geometric mean fluorescence intensity (gMFI) is shown for the high GFP gated population. H) Representative samples from G) are shown as a histogram overlay for histone H3 fluorescence intensity. I-J) Cells were co-transfected with eGFP, eGFP-ARS2, or -Mid mutants along with the indicated firefly luciferase reporter and Renilla luciferase. Firefly luminescence was normalized to Renilla luminescence. K) Cells transfected with the indicated plasmids were treated as in Figure 22A, gated and quantified as in Figure 22B. L-M) Cells were transfected with eGFP, eGFP-ARS2, or -ΔC along with the indicated firefly luciferase reporter and Renilla luciferase. A one way ANOVA was performed in A) ($F(3,8)=16.00$, $p<0.01$), B) ($F(3,8)=11.48$, $p<0.01$), C) ($F(3,8)=37.96$, $p<0.001$), D) ($F(3,8)=10.33$, $p<0.01$), E) ($F(2,12)=54.86$, $p<0.001$), F) ($F(2,12)=11.97$, $p<0.01$), G) ($F(2,6)=36.23$, $p<0.001$) I) ($F(4,25)=192.9$, $p<0.001$), J)

($F(4,14)=44.05$, $p<0.001$), K) ($F(2,12)=16.78$, $p<0.001$), L) ($F(2,6)=28.62$, $p<0.001$), and M) ($F(2,6)=38.14$, $p<0.001$), followed by Tukey's multiple comparison *post hoc* test to determine whether there were significant differences between groups. * = $p<0.05$), ** = $p<0.01$, *** = $p<0.001$, n.s. = not significant, and error bars represent SD.

3.5 Discussion

ARS2 knockdown has been shown to disrupt RDH mRNA processing and cause a reduction in histone expression, leading to disruption of the cell cycle [16,24]. Consistent with this, we found that myogenic cells deficient in ARS2 cycle more slowly, and spend an increased time in S phase. The more severe S phase arrest phenotype associated with ARS2 overexpression reflects the ability of the dominant-negative to more severely compromise ARS2 function versus RNAi. Many types of RNAP II transcripts controlled by ARS2 likely contribute to the cell cycle phenotype of altering ARS2 levels. Indeed, microRNA [210], mRNA splicing [211,212], and replication-dependent histones are all important for cell cycle progression [16,79]. However, the distinctive phenotype of the high BrdU-containing cells in early S phase points to defective chromatin packaging as a consequence of histone deficiency as a major contributor for this unique population within the overexpressing cells. Alternatively, ARS2 overexpression may alter heterochromatin modifications, which could make it more accessible to the BrdU antibody. These two possibilities are not mutually exclusive.

We have generated the first models of mammalian ARS2 and its RRM domain and designed mutations that disrupt function of each of ARS2's domains. These models, coupled with the finding that ARS2 overexpression generates a dominant-negative, allowed us to screen these mutations for the ability to disrupt known protein interactions, or to change the magnitude of the dominant-negative effect of ARS2 overexpression on miRNA and RDH mRNA biogenesis, RNA binding, and cell cycle progression. Through this process, we identified two types of mutations: those that exacerbate the dominant-negative effect by forming non-functional complexes better than wildtype, and those that lose the ability to evoke a dominant-negative effect because they lose the ability to sequester key interacting partners. Both of these types of mutations are expected to represent regions in ARS2 that mediate protein or RNA interactions. Although our models proved useful for our experimental rational design, a crystal structure of

mammalian ARS2 is needed to gain insight into how the RRM is oriented relative to the core structure.

Our data show the DUF3546 domain of ARS2 indirectly contributes to RDH mRNA and miRNA binding. Although the DUF3546 was required for ARS2 to associate with these classes of RNA under low salt conditions where ARS2 complexes were intact, this domain was not required when ARS2 was purified under high salt wash conditions. This suggests that the DUF3546 domain likely facilitates protein interactions, or conformations of ARS2, that permit RNA interaction with the ZnF domain. For example, it is possible there is an indirect effect through an RNA binding cofactor at the DUF3546 domain. Interestingly, published mass spectrometry data show ARS2 can be phosphorylated at Y172 [218] and Y175 [219] within the DUF3546 domain, suggesting the function of this domain may be regulated post-translationally by phosphorylation. However, we did not detect a difference in Tyr phosphorylation in the DUF3546 mutant using a 4G10 antibody (data not shown). A similar requirement for both the N-terminus and ZnF of SERRATE for miRNA interaction was found in plants [47,53]. In fission yeast, mutation of F165 in the DUF3546 domain contributes to a temperature sensitive allele [131]. How the DUF3546 domain regulates interactions with miRNA and RDH mRNA requires further investigation.

The ZnF domain of ARS2 is critical for both miRNA and RDH mRNA interaction. The mutations in this domain obviated the ARS2 dominant-negative in miRNA and RDH mRNA biogenesis, as well as the high BrdU/cell cycle phenotypes. This domain was necessary for miRNA and RDH mRNA interaction under both low and high salt wash conditions. Thus, these RNA interactions are not dependant on other interacting proteins, which is evidence that the ZnF domain is the principle interface for these classes of RNA. Further work with recombinant ZnF domain is required to determine the mechanism of RNA binding and confirm a direct interaction.

The ARS2 complexes involved in processing the different RNA classes are thought to be mutually exclusive [206]. We speculate that differences in the manner in which the different classes of RNA interact with ARS2 may induce conformational changes. Interestingly, the ZnF domain was also found to be necessary for binding the FLASH peptide FARB. Since it is unlikely that FLASH binds to the same interface of the

ZnF at the same time as histone mRNA, one possibility that accounts for the loss of FLASH binding with the ZnF mutations is that histone mRNA binding is a prerequisite for FLASH binding at another site, possibly within the RRM domain, or another site within the ZnF.

The proline-rich C-terminus of ARS2 is required for CBC interaction [25]. We refined the region involved to amino acids 743-875 of ARS2. Although deletion of the C-terminus impaired RDH mRNA pulldown, the overexpression of ARS2-ΔC still generated the dominant-negative high BrdU cell cycle phenotype. This suggests that ARS2-ΔC remains capable of sequestering key proteins involved in RDH processing through its other intact regions. Conversely, the ARS2 C-terminus was not required to IP miRNA, but was involved in miRNA biogenesis, as ARS2-ΔC exacerbated this phenotype, suggesting that ARS2 interaction with the CBC is critical for proper miRNA biogenesis.

The RRM domain is found only in metazoan ARS2, implying that it confers unique functionality to ARS2 not found in the plant orthologue SERRATE. We showed that the RRM mutant exacerbated the high BrdU cell cycle phenotype and further reduced histone H3 levels relative to wildtype ARS2, suggesting this domain may be specifically required for RDH mRNA processing. Consistent with this interpretation, we found no observable effects of this mutation on the miRNA biogenesis phenotypes. Our data suggests the RRM domain is involved in interactions with histone mRNA 3'-end processing machinery, since the FARB peptide showed reduced binding to the RRM mutant in the presence of RNase. We found no effect of the RRM mutant on binding to RDH mRNA or miRNA, nor does it appear that the RRM/RNA interaction is masked by binding to an interacting partner, since the high salt wash also showed no effect on ARS2 interaction with RNA. Since we mutated the canonical RNA interface, we cannot rule out the RRM interacts with these RNA through another interface. However, the lack of conservation of RNP1 and the effect on FARB binding suggests that the RRM domain likely mediates protein interactions. In this manner, the ARS2 RRM bears similarity with the Y14 RRM, which specifically interacts with MAGOH protein, and the UPF3 RRM, which interacts with UPF2, but not RNA [220,221].

The Mid domain of ARS2 is the largest part of the protein and was predicted to be a platform for protein-protein interactions [53]. We found that this domain was important for miRNA biogenesis, as the mutations we made partially negated the dominant-negative effect of ARS2 on miR-24 and miR-155 levels. One possible mechanism to account for this finding is that this domain is involved in DROSHA interaction. None of the mutations completely disrupted the binding to DROSHA, but we found reduced binding in the presence of RNase with two mutations affecting different regions of the protein. One potential explanation for this result is that these mutations locally disrupt the structure of the Mid domain and thereby diminish the interaction with DROSHA.

We have shown that ARS2 can interact with at least two different types of RNA, likely through its ZnF domain. Additionally, ARS2 interacts with at least two RNA processing complexes, via the ZnF, RRM, and Mid domains (Figure 29). The ability to recognize different RNA and processing complexes suggests ARS2 has a key role in linking the nuclear cap to the appropriate RNA biogenesis machinery. A key function of the CBC is thought to be the coupling of RNA processing steps [206]. ARS2 appears to be uniquely situated to do this. For example, we found that RDH mRNA pulldown, and presumably processing, requires both CBC-ARS2 interaction through the C-terminus of ARS2, as well as additional interactions through the DUF3546 and ZnF domains. Similarly, miRNA biogenesis requires both CBC-ARS2 interaction through the C-terminus, and miRNA binding, likely through the ARS2 ZnF/DUF3546. Since cap assembly on nascent transcripts is thought to occur co-transcriptionally within the first 20 nucleotides, it is likely that ARS2 binds first to the CBP20/80-m7G cap complex through its proline-rich C-terminus. Next, the type of RNA may be recognized through the ZnF domain. We speculate the binding of either miRNA or histone mRNA alters the conformation of ARS2 to allow for either FLASH binding or DROSHA binding to the ZnF/RRM or Mid domains, respectively. The interaction of ARS2 with the 5'-cap and 3'-processing site thereby ensures only capped RNA are processed. Taken together, these results show that ARS2 is a multifunctional platform for protein and RNA interactions suited for directing primary RNA transcripts to the appropriate processing pathways (Figure 29).

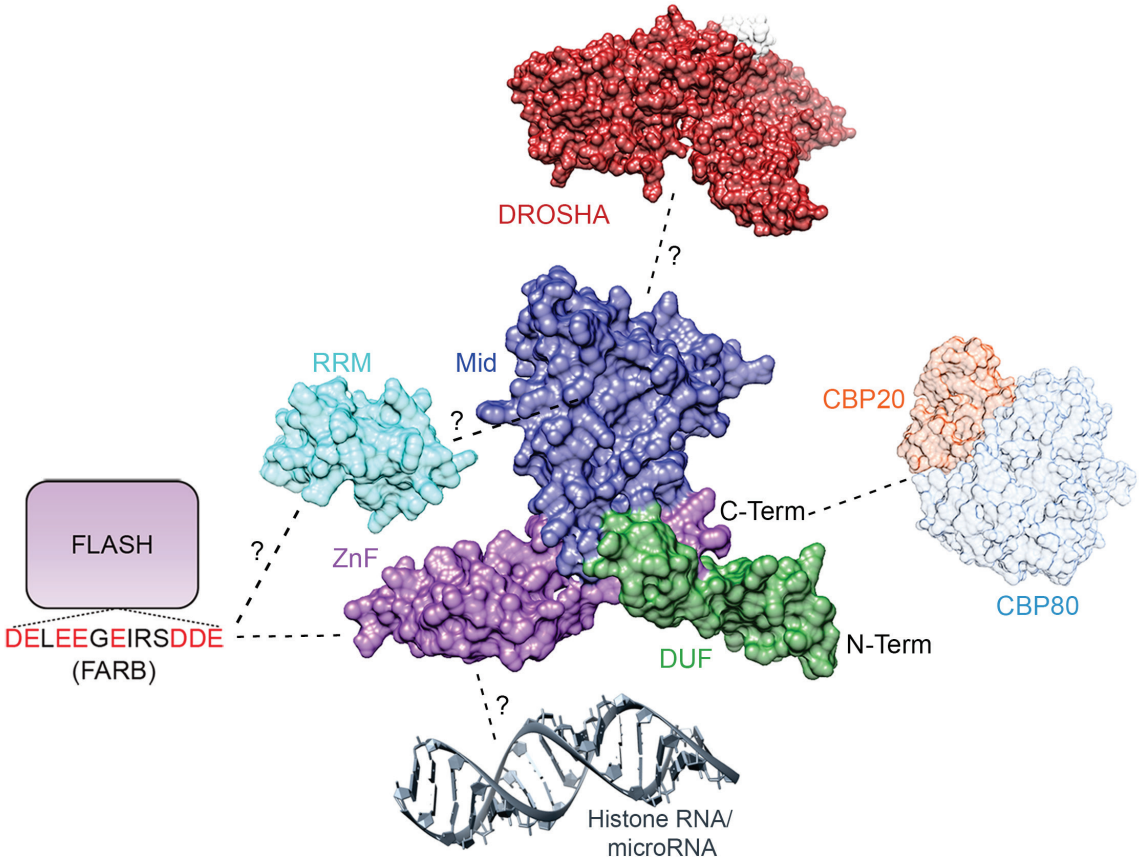

Figure 29 Model of mammalian ARS2 structure/function.

A representation of mouse ARS2 domains (center) along with their interacting partners. The core consists of the domain of unknown function (DUF3546) shown in green, Mid domain in dark blue, and ZnF domain in purple. The orientation of the RRM (shown in cyan) relative to the core is unknown. The unstructured C-terminus (C-term) is required for interaction with CBP80/20, the Mid domain is implicated as being important for interaction with DROSHA, the ZnF is required for interaction with replication-dependent histone RNA and microRNA, and FLASH through the FLASH ARS2 binding (FARB) peptide interacts with the ZnF and RRM. How ARS2 precisely contacts RNA/protein is not understood. CBP80/20 structure corresponds to PDB 1N52, DROSHA to PDB 5B16, and the RNA structure is from PDB 5DV7.

Chapter 4 – Discussion and Future Directions

4.1 Summary of research objectives

A major goal for regenerative medicine is to understand the mechanisms regulating stem cell maintenance and differentiation, which could enable tissue restoration with the correct proportions and connections of differentiated cells. This is an enormous undertaking across the field, as there are numerous intrinsic and extrinsic factors that are tightly controlled during development and into adulthood. The retina represents a particularly attractive model to study progenitor maintenance and differentiation, as it is experimentally amenable, is well characterized anatomically and physiologically for the central nervous system, and a population of progenitors continues to divide and differentiate postnatally in the mouse. In addition, retinal degenerative diseases such as retinitis pigmentosa can arise due to mutations of genes leading to loss of rod photoreceptors and ultimately blindness [222]. Notably, cell-based therapies are currently in clinical trials for treating patients with retinal degenerative diseases [222]. Thus, studies providing insights into the factors governing progenitor cell proliferation and fate specification are an important contribution to the field.

ARS2 had been implicated as an essential factor for stem cell maintenance and differentiation during early embryogenesis, and in the hematopoietic and neural stem cell compartments during adulthood [2,7,9]. The work I have described in this dissertation extends the importance of ARS2 in progenitor cells to the retina, where ARS2 knockdown results in altered proliferation kinetics, timing of cell cycle exit, and cell fate specification. Additionally, this work establishes ARS2 expression patterns in the mouse retina, with expression retained in all major cell types of the adult retina, suggesting it retains function in terminally differentiated cells. We saw a similar retention of ARS2 expression in terminally differentiated myotubes (Appendix A – Supplementary Figure 32). This is an important point, as other groups have reported ARS2 downregulation in quiescent $Bax^{-/-}Bak^{-/-}$ cells *in vitro* [2,24], and absence in multiple cell types in the subventricular zone *in vivo* [9]. The increase in rod photoreceptor and decrease in Müller glia marker expression following ARS2 depletion demonstrates it is required for cell fate specification decisions in the mouse retina. Moreover, the finding that FLASH KD

recapitulates the increased proportion of cells expressing Nrl-DsRed points to a defect in histone processing and cell cycle progression as a contributor to the phenotype that results.

Similarly, ARS2 KD disrupts the cell cycle progression in proliferating retinal progenitor cells. My working model, based on this work and the literature, is that ARS2 KD disrupts RDH processing, which is needed for the proliferating progenitors to package chromatin, and at least partially contributes to a delay in the cell cycle. The report demonstrating a prolonged duration of nuclei at a basal position in the NBL following ARS2 disruption [168], combined with the increase in thymidine analogue-labeling shown in this work, suggests there is a delay in S phase progression. Based on the double-labeling experiment, ARS2 deficient cells do not enter a subsequent cell cycle to the same extent as controls, suggesting premature cell cycle exit. The established link between cell cycle exit and cell fate in the retina suggests that exiting the cell cycle due to ARS2 KD at ~P1, the peak of rod photoreceptor production [162], skews cell fate specification towards rod photoreceptor generation at the expense of Müller glia. The multifunctional nature of ARS2 makes it particularly challenging to precisely determine the major ARS2-regulated pathways contributing to this phenotype, but the results shown here suggest that it is not due to a deficiency in miRNA or NOTCH1 signaling exposure. This is an important step towards understanding the roles of ARS2 in the dynamically regulated developing retina.

In recent years it has become clear that ARS2 is part of the nuclear cap-binding complex, and is implicated in regulating the maturation, turnover, and transport of most, if not all, capped RNAP II transcripts. However, very little is known about how ARS2, at the 5'-end, mediates interactions with the 3'-end processing machinery, nor how it distinguishes between distinct RNA classes or complexes. In addition, at the time my project started, there was a complete absence of structural information for ARS2, and it had only been shown to be involved in miRNA biogenesis and RDH processing. Kiriyama et al., who first showed ARS2 was required for RDH processing, reported ARS2 KD increased the accumulation of cells in S phase [16], while Gruber et al., who were the first to report the requirement of ARS2 in pri-miRNA processing by DROSHA, reported a general delay in cell cycle progression but no phase-specific accumulation [2].

I was able to address this discrepancy using cell cycle kinetics experiments, and demonstrate that while depleting ARS2 delays cell cycle progression, overexpressing ARS2 at high levels arrests cells in early S phase.

The ARS2 overexpression constructs provided a useful tool, as they were capable of inducing a strong dominant-negative phenotype, and could be visualized and IP'd using fluorescent and epitope tags. Careful bioinformatic analysis, combined with the SE structural information, allowed me to identify and model the four structured domains of mammalian ARS2. In addition, it enabled me to identify and mutate critical residues within each domain, and to begin to map the regions used by ARS2 to interact with the 5'-CBC, 3'-miRNA and RDH processing machinery, as well as RNA itself. I was able to refine the region required for interacting with the CBC to the proline-rich unstructured C-terminus. Furthermore, I demonstrated a requirement for the ZnF domain for interaction with miRNA or RDH RNA. Whereas a FLASH peptide requires an intact ZnF and RRM to bind ARS2, DROSHA interaction with ARS2 requires an intact Mid domain. This suggests that ARS2 is interacting with the 5'-CBC, and 3'-miRNA and RDH processing complexes using different interfaces. In addition, FLASH KD did not affect miRNA biogenesis, and mutating the newly identified RRM affected histone, but not miRNA levels. Conversely, Gruber et al. demonstrated that depletion of the Microprocessor does not affect histone biogenesis [24]. Together, this suggests the interactions between ARS2 and FLASH or DROSHA are mutually exclusive, a point that has been reinforced experimentally through purification of independent ARS2-containing complexes [25,26]. Collectively, the work described in this dissertation establishes ARS2 as a scaffold protein, capable of bridging the 5'-CBC to distinct 3'-processing machineries using multiple interfaces. Thus, it appears to be at the hub of RNAP II transcript maturation in the nucleus, and its roles in these processes are essential for proper stem cell maintenance and differentiation.

4.2 ARS2 and the retina

In the mouse retina, I used electroporation of shRNA targeting vectors to KD proteins of interest. This technique proved useful, as I was able to achieve KD, which resulted in observable phenotypic consequences, and highlighted the importance of ARS2

in the developing retina *in vitro* or *in vivo*. The limitations of this technique included variable and sometimes low electroporation efficiency, which could result in variable levels of KD, and my analysis was largely limited to late-born progenitor cells. Using an *Ars2*$^{fl/fl}$ mouse to generate various conditional knockouts would be an ideal tool for analysing ARS2 function in the retina. This mouse has already been generated by Craig Thompson's lab, but since it is not publically available, an ARS2$^{fl/fl}$ mouse would have to be generated either commercially or in-house. However, with the advances in CRISPR/Cas9 technology, it is now more efficient, relative to conventional gene-targeting methods using homologous repair in embryonic stem cells, to generate conditional alleles in mice [223,224]. Once generated, these mice could be crossed to various mice containing *Cre* recombinase under the control of retina-specific promoters. For example, the *mRx-Cre; ROSA26R-EYFP* mouse, could be used to assess ARS2 function in early eye development, as Cre-mediated recombination occurs in virtually all cells in the retina by ~E10.5 [225]. To assess cell-autonomous effects in the postnatal retina, a replication-incompetent virus containing *Cre* recombinase and a reporter to mark infected cells could be used. Following infection and *Ars2* deletion, visualization of the reporter would allow the assessment of late-born cell fate decisions using morphology, or proliferation capacity using clone sizes [177,178].

In this work, we used a Nrl-DsRed reporter to assess the effects of ARS2 KD on rod photoreceptor specification. Although this reporter was specific, as it was expressed with recoverin in the ONL as expected, one of its limitations was low sensitivity. At P8, virtually all postnatally electroporated cells in the ONL are expected to be Nrl-DsRed positive rod photoreceptor cells, as NRL is one of the earliest rod markers [226], yet the observed proportions were much lower, even in control conditions (see Figure 15D, for example). However, ARS2 KD consistently increased the proportion of Nrl-DsRed positive electroporated cells relative to control, as assessed by flow cytometry or confocal microscopy. Conversely, ARS2 KD decreased the proportion of cells expressing Hes1-DsRed or Cralbp-DsRed reporters, suggesting the effect on Nrl-DsRed is not due to a general increase in transcription. We hypothesize that the increased proportion of Nrl-DsRed reporter expression following ARS2 KD represents an increase in rod photoreceptor production, but this needs to be further validated. To verify that ARS2

depleted cells preferentially adopt a rod fate, I would first co-electroporate retinas with shRNA and pCAG-GFP, which uses a strong promoter expressed ubiquitously in the retina [179] and would allow visualization of all electroporated cells. At the desired time points, cell fate could be scored using morphology and localization. In the control condition, I expect GFP expression in rod photoreceptors, bipolar cells, and Müller glia to be in normal proportions (~85% rod photoreceptors, ~10% bipolar cells, ~5% Müller glia) [204]. In contrast, I hypothesize GFP expression would be predominantly limited to rod photoreceptors in the ONL following ARS2 KD, and GFP positive cells adopting a Müller glial morphology would be reduced. Several approaches could be used to corroborate the increase in rod photoreceptor cells in addition to morphology and localization. For example, we could use a different rod photoreceptor reporter such as Rhodopsin-DsRed, as it has a more robust expression relative to Nrl-DsRed [179]. Additionally, the relative expression levels of several rod photoreceptor markers could be assessed using qRT-PCR, such as Rhodopsin, Nrl, and NR2E3 [227].

ARS2 has been implicated as a transcription factor for SOX2, as ARS2 binds to a region within the *Sox2* enhancer in a RNA-independent manner in NSCs, and activates expression of a *Sox2* luciferase reporter [9]. To assess whether ARS2 regulates SOX2 in a similar manner to NSCs, I would first see if it binds to the *Sox2* enhancer region in the context of the retina using chromatin immunoprecipitation (ChIP) with or without RNase treatment. If ARS2 does in fact bind to *Sox2* regulatory DNA, I would next assess whether this has functional relevance in the retina. Using the heterozygous $Sox2^{eGFP}$ knock-in mouse [228] would allow for the comparison between control mice and ARS2 depleted or conditional knockout mice. If using the Ars2 shRNA strategy, I would use a KD vector containing a red fluorescent protein cassette, to visualize electroporated cells and assess $Sox2^{eGFP}$ expression simultaneously. Alternatively, electroporated retinal cells from wildtype mice could be sorted using flow cytometry and the relative levels of Sox2 mRNA could be assessed using qRT-PCR. If there are differences in Sox2 expression following ARS2 KD or conditional knockout, I would assess whether the ARS2-SOX2 pathway is a major contributor to the ARS2 KD phenotype by attempting to rescue the proliferation and cell fate specification defects using electroporation of a Sox2 expression construct [9].

ARS2 has also been shown to regulate the lncRNA TERRA in humans, which are m7G-capped RNAP II transcripts [145,146]. A broader role for ARS2/SERRATE in lncRNA regulation has been shown in plants [150]. *Sox2* is an intronless gene embedded within the lncRNA *SOX2OT* locus [151]. Furthermore, SOX2 expression is positively regulated by SOX2OT [152–154]. Therefore, to gain a more complete understanding of how ARS2 regulates SOX2, the levels of SOX2OT should be assessed. In the same ChIP experiment described above for investigating whether ARS2 binds *Sox2* chromatin in the retina, I would also assess whether ARS2 binds to the *SOX2OT* locus. I would additionally determine whether ARS2 regulates SOX2OT by measuring SOX2OT expression levels either in the ARS2 knockdown or conditional knockout retina. One way to assess SOX2OT expression is through qRT-PCR, and if using the KD electroporation strategy, the electroporated cells would need to be isolated using fluorescence activated cell sorting (FACS). Alternatively, SOX2OT levels could be assessed on retinal tissue sections using *in situ* hybridization.

4.3 ARS2 structure/function

The elucidation of high-resolution structural information requires the purification of your protein(s) of interest. Presently, there is no high-resolution structure of metazoan ARS2. Although the crystal structure of the plant orthologue SE has been solved [53], there are notable differences between SE and mammalian ARS2. For example, I identified a predicted RRM that is conserved within metazoan ARS2 but completely absent from plants, and the orientation of the RRM relative to the ARS2 core is unknown (Figure 23B-D). The overall fold of full-length mammalian ARS2 would be very useful in terms of design and interpretation of biochemical experiments. Therefore, obtaining purified ARS2 protein should be a primary goal for future studies, as it could be used to greatly enhance our understanding of how ARS2 mediates protein-protein and protein-RNA interactions. However, mammalian ARS2 has many unstructured regions, and is unstable when recombinantly expressed in bacteria (Omid Haji-Ghassemi and Stephen Evans, personal communication), so purified ARS2 must be produced using alternative strategies.

I was able to perform ARS2 IP using high salt wash conditions, which removed all associated proteins as visualized by silver staining, but still retained ARS2-bound RNA (Figure 27F). Although this led us the conclusion that the ARS2 ZnF likely directly interacts with RNA, it cannot formally be ruled out that other low abundance proteins or non-specific contaminants present with ARS2 following IP contributed to binding. In addition, the amount of ARS2 protein obtained following IP in high salt wash conditions was estimated to be in the low nanogram range, which precluded my ability to detect whether ARS2 could interact with histone or pri-miRNA probes using an electrophoretic mobility shift assay (EMSA) (data not shown). Andreu-Agullo et al. used *in vitro* translation (IVT) to generate ARS2 and demonstrate its ability to bind a region in the *Sox2* enhancer using EMSA; notably, they used 8 µg of *in vitro* translated ARS2 as input, suggesting a potentially weak interaction with chromatin, and indicating a large amount of purified protein may be necessary to detect certain interactions [9].

To express ARS2 using IVT, I would clone full-length ARS2 as well as the ARS2 mutants into a T7 expression vector and use a commercially available IVT kit (Thermo fisher scientific). I would use a vector containing GST and His tags followed by a protease cleavage site, which would allow for multiple purification steps and subsequent cleavage of the epitope tags from ARS2. Successful IVT of ARS2 and its mutants would allow us to definitively determine which domains directly interact with protein and/or RNA. For example, the ZnF mutants could be incubated with an RNA probe to determine whether this domain interacts with RNA using EMSA. It would also be interesting to see if the ZnF domain could be stably expressed on its own, to test whether the ZnF is sufficient for RNA interaction. In addition, I have shown that the ZnF and RRM are involved in FLASH interaction through the FARB peptide (Figure 29). Using purified FARB peptide and IVT ARS2 and ARS2 mutants would allow us to further refine the region of ARS2 that directly contacts FLASH. Similar to the ZnF, it would be worthwhile to pursue expression and purification of the ARS2 RRM domain on its own, to assess whether it is capable of RNA interaction, and whether it is sufficient for FLASH interaction.

ARS2 regulates a similar set of transcripts as CBP80/20, and forms a stable trimeric complex as CBCA [25,26]. Calero et al. used baculoviruses to express the CBC

in insect cells, and were able to solve the co-crystal structure of CBP80, CBP20 and m7G-capped RNA [229]. Therefore, co-expressing CBP80 and CBP20 may represent a potential strategy to stabilize ARS2, and a co-crystal structure would provide enormous insight. To obtain CBCA, I would co-infect insect cells with baculoviruses encoding epitope tagged ARS2, CBP80 and CBP20 and purify the resulting complex using m7GTP sepharose. I would first assess the ability of purified CBCA to form crystals suitable for X-ray crystallography, as this would provide high-resolution structural information. If CBCA is refractory to crystallization, it is a good candidate for single-particle cryo-electron microscopy (cryo-EM). At a predicted molecular weight of >200 kDa, CBCA is large enough for single-particle cryo-EM [230]. Additionally, the co-crystal structure of CBP80/CBP20 has already been solved, which could be used to dock the CBP80/20 heterodimer within the CBCA cryo-EM images [231]. Furthermore, the ARS2 DUF3546, Mid, ZnF and RRM domains that I modeled could be used for additional docking. The resulting cryo-EM structural model of CBCA would further our understanding of the relative orientation of ARS2 to the CBP80/20 CBC, and would be the first demonstration of the overall fold of mammalian ARS2.

The use of baculovirus to express CBCA in insect cells is not limited to structural studies. Replacing wildtype ARS2 with the mutants I have described in this dissertation into the baculovirus expression constructs represents a strategy to assess protein-protein and protein-RNA interactions mediated by ARS2 in the context of the CBC. For example, we hypothesize that the ARS2-RRM is important for protein-protein interactions, but not protein-RNA interactions, as the ARS2-RRM mutant did not disrupt RNA pulldown (Figure 27G,H and Figure 28A-D), but reduced FARB binding and exacerbated the cell cycle arrest and histone processing phenotype (Figure 25B and Figure 28E-H). Importantly, the ARS2 RRM-Mut did not disrupt binding to the CBC (Figure 25E). Therefore, to assess interactions mediated by the RRM, I would use baculovirus to express either wildtype ARS2 or the RRM-Mut along with CBP80/20, and perform affinity capture mass spectrometry. Comparing the set of proteins pulled down using wildtype ARS2 vs. RRM-Mut can identify ARS2-interacting partners that require the RRM domain. Any hits in this screen can be verified using Western blotting or a yeast two-hybrid approach.

Pir2/Ars2, the orthologue of ARS2 found in *S. pombe*, along with the yeast CBC equivalent Cbc1/Cbc2, have recently been shown to be required for degradation of CUTs by the exosome through interactions with the MTREC complex [85–87], and for heterochromatin formation through multiple pathways (Figure 8) [131]. How Pir2/Ars2 mediates interactions with these various complexes, including Cbc1/Cbc2 is unknown. I was able to refine the mouse ARS2 region required for interaction with the CBC to the unstructured proline-rich C-terminus. Interestingly, there is a Y-X-DLDAP motif at the extreme end of the C-terminus that is conserved in metazoa, *Arabidopsis*, and *S. pombe* (Figure 23D and Figure 30). The unstructured C-terminus is otherwise poorly conserved. Since ARS2/SE/Pir2 functions with their respective cap-binding proteins in animals, plants, and yeast, I hypothesize that the Y-X-DLDAP motif mediates interaction with the cap-binding proteins. To test this, I would generate a deletion mutant for the Y-X-DLDAP motif in *S. pombe*, and probe whether this mutant disrupts binding to Cbc1/Cbc2. Dr. Chris Nelson currently has students working on generating a series of Pir2/Ars2 mutants in *S. pombe*, and these could be used to map the sites used by Pir2/Ars2 to interact with components of the MTREC, Ccr4-Not, CLRC, EMC and RITS complexes, which would greatly enhance our understanding of how Pir2/Ars2 targets transcripts to the exosome for degradation, or for mediating heterochromatin formation.

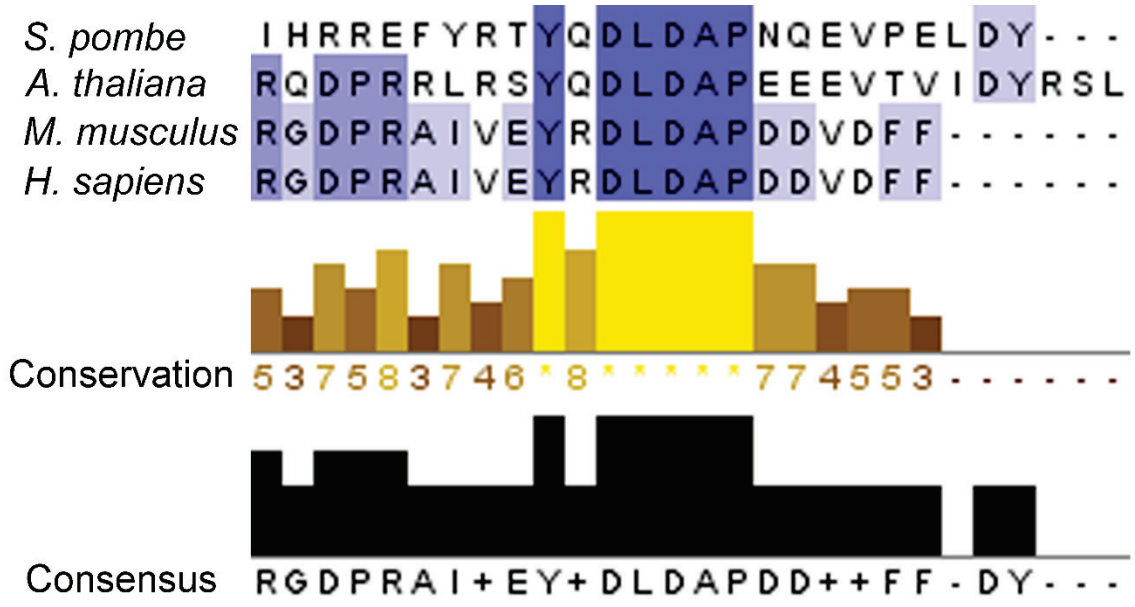

Figure 30 Conserved extreme C-terminus of ARS2

Bioinformatic alignment of the extreme C-terminus of *S. pombe* Pir2/Ars2, *A. thaliana* SE, and *M. musculus, H. sapiens* ARS2 using clustal omega [214]. The Y-X-DLDAP sequence is conserved in all four organisms and represents potential functional importance.

4.4 ARS2 and nonsense-mediated decay

The reported functions of ARS2 are currently restricted to the nucleus, yet evidence is beginning to accumulate in support of a role for ARS2 in the cytoplasm. In *Drosophila*, Ars2 functionally interacts with Dicer2, a cytoplasmic RNase important for viral innate immunity in flies [15]. Moreover, Amanda Carette, a previous honours student, observed co-localization of ARS2 and GW182, a key member in AGO1-mediated mRNA silencing [232], in the cytoplasm of C2C12 mouse myoblast cells (unpublished observations). Remarkably, Jennifer Christie, a previous Master's student in our lab, discovered an alternative isoform of ARS2 called ARS2-X5, which is missing a portion of the N-terminus and starts on Met 228 (unpublished observations). The ARS2-X5 isoform is likely generated from an internal promoter in a conserved region within intron 5 of the *Ars2* gene, as sequencing Ars2-X5 identified a 5'-UTR corresponding to intron 5, and M228 is the first methionine found in exon 6 (Jennifer Christie, unpublished observations). The eGFP-ARS2-ΔN construct (Figure 24A), lacking amino acids 1-227, was generated based on the alternative ARS2-X5 isoform. Notably, this N-terminal region contains a nuclear localization sequence in the unstructured region (Figure 23D), and when deleted, eGFP-ARS2-ΔN is restricted to the cytoplasm (Figure 24A). The existence of a cytoplasmic ARS2 isoform generated from a conserved intron implies it has functions outside of the nucleus.

One of the potential cytoplasmic functions of ARS2 is in the mRNA surveillance process called nonsense-mediated decay (NMD). The CBP80/20 CBC is transiently present on cytoplasmic mRNA during the pioneering round of translation, and plays an important role in NMD [233,234]. NMD ensures aberrant transcripts with a premature termination codon (PTC), located at least 50-55 nt upstream of an exon-exon junction, are efficiently degraded [235]. NMD is stimulated by the retention of an exon junction complex (EJC), which are deposited during splicing and removed during the pioneering round of translation, downstream of the PTC [235]. The main effectors of NMD are the UPF proteins, particularly UPF1 [235]. CBP80 directly interacts with UPF1 and

facilitates formation of the SURF complex (composed of SMG1, UPF1, eRF1, and eRF3) at the PTC, and additionally promotes SMG1-UPF1 binding to UPF2 at the PTC-distal EJC [233,234]. UPF1 is then phosphorylated by SMG1, which ultimately represses translation and recruits mRNA decay factors to degrade the aberrant transcript (Figure 31) [235]. Since ARS2 is a stable component of the nuclear CBC, immunoprecipitates with EJC components [236,237], and is required for mRNA export into the cytoplasm [205], we hypothesized that ARS2 would also play a role in inducing nonsense-mediated decay in the cytoplasm.

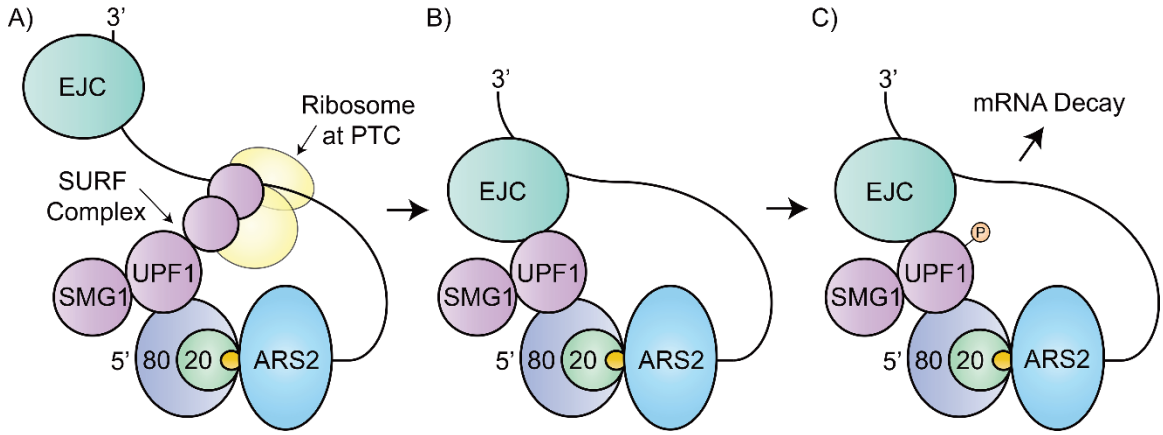

Figure 31 Nonsense-mediated decay

The cap-binding complex is important for nonsense-mediated decay (NMD). A) When a transcript contains a premature termination codon (PTC) at least 50-55 nt upstream of an exon junction complex (EJC), CBP80 will facilitate formation of the SURF complex. Specifically, CBP80 interacts with UPF1 and promotes the association of SMG1-UPF1 with release factors located at the ribosome paused at the PTC. B) CBP80 then promotes the association of SMG1-UPF1, through UPF1, with components of the PTC-distal EJC. C) SMG1 then phosphorylates UPF1, which stimulates recruitment of mRNA decay factors, leading to transcript degradation.

One of the properties of essential NMD components is the ability to induce decay when tethered to a transcript sufficiently downstream of a termination codon [238,239]. Therefore, tethering assays are widely used as a readout for NMD. For example, a firefly luciferase reporter containing 5× boxb sites in its 3'-UTR is recognized and bound by the 22 amino acid RNA-binding domain of the bateriophage λ antiterminator protein (λN) when co-expressed [240]. λN on its own does not have the ability to induce decay of the boxb containing luciferase; however, tethering a λN-ARS2 fusion protein to boxb luciferase decreased reporter activity, suggesting it may be involved in NMD (Appendix

A – Supplementary Information Figure 33). Due to ARS2's role in mRNA surveillance through the exosome [26,81], I wanted to further assess whether this decrease in reporter activity was specifically due to a disruption in NMD. Since UPF1 is the primary effector of NMD [235], depleting UPF1 should decrease the efficiency of NMD and therefore rescue the boxb luciferase reporter decrease seen following ARS2 KD. As expected, tethering λN-ARS2 to the reporter following UPF1 KD partially rescued luciferase activity (Appendix A – Supplementary Information Figure 33), supporting the hypothesis that ARS2 is involved in NMD.

NMD also regulates several endogenous transcripts, some of which are well characterized in HeLa cells [241–243]. To corroborate the tethering assay data implicating ARS2 in NMD, I would next assess whether ARS2 KD affects the levels of endogenous NMD targets in HeLa cells using qRT-PCR. To further characterize how ARS2 contributes to NMD, I would determine the protein-protein interactions mediated between ARS2 and NMD pathway components. I would use the strategy employed by Andersen et al., who performed affinity capture coupled with mass spectrometry in triplicate to show interactions between CBCA and multiple NEXT components, as well as approximate the stoichiometry of subcomplexes [26]. An obvious first candidate to test would be UPF1, as UPF1 KD partially rescued the decrease in reporter activity following ARS2 KD, and CBP80 has been shown to directly interact with UPF1 and would act as a positive control in UPF1 pulldowns [233,234]. I would also pulldown other SURF complex members, such as SMG1, or EJC members, such as UPF2, to assess ARS2 interactions. If ARS2 is identified in these screens, interactions would be verified using Western blotting. Additionally, purified ARS2 protein from either *in vitro* translation or insect cell expression could be used in conjunction with purified ARS2-interacting proteins to assess whether any observed interactions are direct.

The cytoplasmic localization of ARS2-X5 makes it a strong candidate for being involved in NMD. To test this, I would clone a λN-ARS2-X5 construct and use it in the already established boxb luciferase assay. It would also be informative to probe ARS2-X5 interacting partners by performing affinity capture mass spectrometry. An expression vector with ARS2-X5 tagged with a small epitope tag could be used to enrich for the alternative isoform. If ARS2-X5 is involved in NMD, I would expect members of SURF

or EJC to be identified. I would verify potential interactors using Western blotting. Due to the unique 5'-UTR on Ars2-X5, it is possible to generate isoform-specific siRNA, which would potentially enable KD of Ars2-X5 while leaving full-length Ars2 mRNA intact. After confirming the efficacy of the siRNA using either Western blotting or qRT-PCR, I would assess whether Ars2-X5 KD affects the decay of endogenous NMD targets in HeLa cells. Overall, NMD represents a novel ARS2 function that can be easily pursued using the tools and expertise available in our lab.

References

1. Rossman TG, Wang Z. Expression cloning for arsenite-resistance resulted in isolation of tumor-suppressor fau cDNA: possible involvement of the ubiquitin system in arsenic carcinogenesis. Carcinogenesis 1999;20:311–6.

2. Gruber JJ, Zatechka DS, Sabin LR, Yong J, Lum JJ, Kong M, et al. Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. Cell 2009;138:328–39.

3. Lobbes D, Rallapalli G, Schmidt DD, Martin C, Clarke J. SERRATE: a new player on the plant microRNA scene. EMBO Rep. 2006;7:1052–8.

4. Oh SW, Kingsley T, Shin HH, Zheng Z, Chen HW, Chen X, et al. A P-element insertion screen identified mutations in 455 novel essential genes in Drosophila. Genetics 2003;163:195–201.

5. Golling G, Amsterdam A, Sun Z, Antonelli M, Maldonado E, Chen W, et al. Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. Nat. Genet. 2002;31:135–40.

6. Amsterdam A, Nissen RM, Sun Z, Swindell EC, Farrington S, Hopkins N. Identification of 315 genes essential for early zebrafish development. Proc. Natl. Acad. Sci. U. S. A. 2004;101:12792–7.

7. Wilson MD, Wang D, Wagner R, Breyssens H, Gertsenstein M, Lobe C, et al. ARS2 is a conserved eukaryotic gene essential for early mammalian development. Mol. Cell. Biol. 2008;28:1503–14.

8. Kim D-U, Hayles J, Kim D, Wood V, Park H-O, Won M, et al. Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat. Biotechnol. 2010;28:617–23.

9. Andreu-Agullo C, Maurin T, Thompson CB, Lai EC. Ars2 maintains neural stem-cell identity through direct transcriptional activation of Sox2. Nature 2012;481:195–8.

10. Carrasco-Garcia E, Arrizabalaga O, Serrano M, Lovell-Badge R, Matheu A. Increased gene dosage of Ink4/Arf and p53 delays age-associated central nervous system functional decline. Aging Cell 2015;14:710–4.

11. He Q, Cai L, Shuai L, Li D, Wang C, Liu Y, et al. Ars2 is overexpressed in human cholangiocarcinomas and its depletion increases PTEN and PDCD4 by decreasing microRNA-21. Mol. Carcinog. 2013;52:286–96.

12. He Q, Huang Y, Cai L, Zhang S, Zhang C. Expression and prognostic value of Ars2 in hepatocellular carcinoma. Int. J. Clin. Oncol. 2014;19:880–8.

13. Cui L, Gao C, Zhang R-D, Jiao Y, Li W-J, Zhao X-X, et al. Low expressions of ARS2 and CASP8AP2 predict relapse and poor prognosis in pediatric acute lymphoblastic leukemia patients treated on China CCLG-ALL 2008 protocol. Leuk. Res. 2015;39:115–23.

14. Yang L, Liu Z, Lu F, Dong A, Huang H. SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. Plant J. 2006;47:841–50.

15. Sabin LR, Zhou R, Gruber JJ, Lukinova N, Bambina S, Berman A, et al. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell 2009;138:340–51.

16. Kiriyama M, Kobayashi Y, Saito M, Ishikawa F, Yonehara S. Interaction of FLASH with arsenite resistance protein 2 is involved in cell cycle progression at S phase. Mol. Cell. Biol. 2009;29:4729–41.

17. Lewis JD, Izaurralde E. The role of the cap structure in RNA processing and nuclear export. Eur. J. Biochem. FEBS 1997;247:461–9.

18. Furuichi Y, LaFiandra A, Shatkin AJ. 5'-Terminal structure and mRNA stability. Nature 1977;266:235–9.

19. Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW. A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell 1994;78:657–68.

20. Visa N, Izaurralde E, Ferreira J, Daneholt B, Mattaj IW. A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. J. Cell Biol. 1996;133:5–14.

21. Flaherty SM, Fortes P, Izaurralde E, Mattaj IW, Gilmartin GM. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc. Natl. Acad. Sci. U. S. A. 1997;94:11893–8.

22. Narita T, Yung TMC, Yamamoto J, Tsuboi Y, Tanabe H, Tanaka K, et al. NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs. Mol. Cell 2007;26:349–65.

23. Laubinger S, Sachsenberg T, Zeller G, Busch W, Lohmann JU, Rätsch G, et al. 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. 2008;105:8795–800.

24. Gruber JJ, Olejniczak SH, Yong J, La Rocca G, Dreyfuss G, Thompson CB. Ars2 promotes proper replication-dependent histone mRNA 3' end formation. Mol. Cell 2012;45:87–98.

25. Hallais M, Pontvianne F, Andersen PR, Clerici M, Lener D, Benbahouche NEH, et al. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. Nat. Struct. Mol. Biol. 2013;20:1358–66.

26. Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, Verheggen C, et al. The human cap-binding complex is functionally connected to the nuclear RNA exosome. Nat. Struct. Mol. Biol. 2013;20:1367–76.

27. O'Sullivan C, Christie J, Pienaar M, Gambling J, Nickerson PEB, Alford SC, et al. Mutagenesis of ARS2 Domains To Assess Possible Roles in Cell Cycle Progression and MicroRNA and Replication-Dependent Histone mRNA Biogenesis. Mol. Cell. Biol. 2015;35:3753–67.

28. Clarke JH, Tack D, Findlay K, Van Montagu M, Van Lijsebettens M. The SERRATE locus controls the formation of the early juvenile leaves and phase length in Arabidopsis. Plant J. 1999;20:493–501.

29. Prigge MJ, Wagner DR. The arabidopsis serrate gene encodes a zinc-finger protein required for normal shoot development. Plant Cell 2001;13:1263–79.

30. Grigg SP, Canales C, Hay A, Tsiantis M. SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. Nature 2005;437:1022–6.

31. Bezerra IC, Michaels SD, Schomburg FM, Amasino RM. Lesions in the mRNA cap-binding gene ABA HYPERSENSITIVE 1 suppress FRIGIDA-mediated delayed flowering in Arabidopsis. Plant J. 2004;40:112–9.

32. Papp I, Mur LA, Dalmadi A, Dulai S, Koncz C. A mutation in the Cap Binding Protein 20 gene confers drought tolerance to Arabidopsis. Plant Mol. Biol. 2004;55:679–86.

33. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136:215–33.

34. Lee Y, Kim M, Han J, Yeom K-H, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23:4051–60.

35. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA 2004;10:1957–66.

36. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. 2014;15:509–24.

37. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 2010;466:835–40.

38. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat. Rev. Genet. 2011;12:99–110.

39. Djuranovic S, Nahvi A, Green R. A Parsimonious Model for Gene Regulation by miRNAs. Science 2011;331:550–3.

40. Ren G, Yu B. Post-transcriptional control of miRNA abundance in Arabidopsis. Plant Signal. Behav. 2012;7:1443–6.

41. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, et al. Methylation as a Crucial Step in Plant microRNA Biogenesis. Science 2005;307:932–5.

42. Li J, Yang Z, Yu B, Liu J, Chen X. Methylation Protects miRNAs and siRNAs from a 3′-End Uridylation Activity in Arabidopsis. Curr. Biol. 2005;15:1501–7.

43. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS. Nuclear processing and export of microRNAs in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 2005;102:3691–6.

44. Baumberger N, Baulcombe DC. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. U. S. A. 2005;102:11928–33.

45. Li S, Liu L, Zhuang X, Yu Y, Liu X, Cui X, et al. MicroRNAs Inhibit the Translation of Target mRNAs on the Endoplasmic Reticulum in Arabidopsis. Cell 2013;153:562–74.

46. Wang L, Song X, Gu L, Li X, Cao S, Chu C, et al. NOT2 Proteins Promote Polymerase II–Dependent Transcription and Interact with Multiple MicroRNA Biogenesis Factors in Arabidopsis. Plant Cell 2013;25:715–27.

47. Iwata Y, Takahashi M, Fedoroff NV, Hamdan SM. Dissecting the interactions of SERRATE with RNA and DICER-LIKE 1 in Arabidopsis microRNA precursor processing. Nucleic Acids Res. 2013;41:9129–40.

48. Raczynska KD, Stepien A, Kierzkowski D, Kalak M, Bajczyk M, McNicol J, et al. The SERRATE protein is involved in alternative splicing in Arabidopsis thaliana. Nucleic Acids Res. 2014;42:1224–44.

49. Kim S, Yang J-Y, Xu J, Jang I-C, Prigge MJ, Chua N-H. Two Cap-Binding Proteins CBP20 and CBP80 are Involved in Processing Primary MicroRNAs. Plant Cell Physiol. 2008;49:1634–44.

50. Fang Y, Spector DL. Identification of Nuclear Dicing Bodies Containing Proteins for MicroRNA Biogenesis in Living Arabidopsis Plants. Curr. Biol. 2007;17:818–23.

51. Fujioka Y, Utsumi M, Ohba Y, Watanabe Y. Location of a Possible miRNA Processing Site in SmD3/SmB Nuclear Bodies in Arabidopsis. Plant Cell Physiol. 2007;48:1243–53.

52. Dong Z, Han MH, Fedoroff N. The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc. Natl. Acad. Sci. U. S. A. 2008;105:9970–5.

53. Machida S, Chen HY, Yuan YA. Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. Nucleic Acids Res. 2011;39:7828–36.

54. Ren G, Xie M, Dou Y, Zhang S, Zhang C, Yu B. Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 2012;109:12817–21.

55. Baranauskė S, Mickutė M, Plotnikova A, Finke A, Venclovas Č, Klimašauskas S, et al. Functional mapping of the plant small RNA methyltransferase: HEN1 physically interacts with HYL1 and DICER-LIKE 1 proteins. Nucleic Acids Res. 2015;43:2802–12.

56. Kwon SC, Nguyen TA, Choi Y-G, Jo MH, Hohng S, Kim VN, et al. Structure of Human DROSHA. Cell 2016;164:81–90.

57. Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. Processing of primary microRNAs by the Microprocessor complex. Nature 2004;432:231–5.

58. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, et al. Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. Cell 2004;117:69–81.

59. Tsutsumi A, Kawamata T, Izumi N, Seitz H, Tomari Y. Recognition of the pre-miRNA structure by Drosophila Dicer-1. Nat. Struct. Mol. Biol. 2011;18:1153–8.

60. Okamura K, Ishizuka A, Siomi H, Siomi MC. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev. 2004;18:1655–66.

61. Liu Q, Rand TA, Kalidas S, Du F, Kim H-E, Smith DP, et al. R2D2, a Bridge Between the Initiation and Effector Steps of the Drosophila RNAi Pathway. Science 2003;301:1921–5.

62. Iwasaki S, Sasaki HM, Sakaguchi Y, Suzuki T, Tadakuma H, Tomari Y. Defining fundamental steps in the assembly of the Drosophila RNAi enzyme complex. Nature 2015;521:533–6.

63. DeLisle AJ, Graves RA, Marzluff WF, Johnson LF. Regulation of histone mRNA production and stability in serum-stimulated mouse 3T6 fibroblasts. Mol. Cell. Biol. 1983;3:1920–9.

64. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. 2008;9:843–54.

65. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3′ processing endonuclease. RNA 2004;10:565–73.

66. Dominski Z, Yang X, Marzluff WF. The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell 2005;123:37–48.

67. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, et al. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature 2006;444:953–6.

68. Wang ZF, Whitfield ML, Ingledue TC, Dominski Z, Marzluff WF. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. Genes Dev. 1996;10:3028–40.

69. Dominski Z, Yang X, Kaygun H, Dadlez M, Marzluff WF. A 3' exonuclease that specifically interacts with the 3' end of histone mRNA. Mol. Cell 2003;12:295–305.

70. Tan D, Marzluff WF, Dominski Z, Tong L. Structure of Histone mRNA Stem-Loop, Human Stem-Loop Binding Protein, and 3′hExo Ternary Complex. Science 2013;339:318–21.

71. Scharl EC, Steitz JA. The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. EMBO J. 1994;13:2432–40.

72. Yang X, Torres MP, Marzluff WF, Dominski Z. Three proteins of the U7-specific Sm ring function as the molecular ruler to determine the site of 3'-end processing in mammalian histone pre-mRNA. Mol. Cell. Biol. 2009;29:4045–56.

73. Sabath I, Skrajna A, Yang X, Dadlez M, Marzluff WF, Dominski Z. 3′-End processing of histone pre-mRNAs in Drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. RNA 2013;19:1726–44.

74. Sullivan E, Santiago C, Parker ED, Dominski Z, Yang X, Lanzotti DJ, et al. Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. Genes Dev. 2001;15:173–87.

75. Yang XC, Burch BD, Yan Y, Marzluff WF, Dominski Z. FLASH, a proapoptotic protein involved in activation of caspase-8, is essential for 3' end processing of histone pre-mRNAs. Mol. Cell 2009;36:267–78.

76. Burch BD, Godfrey AC, Gasdaska PY, Salzler HR, Duronio RJ, Marzluff WF, et al. Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila. RNA 2011;17:1132–47.

77. Yang XC, Xu B, Sabath I, Kunduru L, Burch BD, Marzluff WF, et al. 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. 2011;31:1492–502.

78. Yang X-C, Sabath I, Dębski J, Kaus-Drobek M, Dadlez M, Marzluff WF, et al. 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. 2013;33:28–37.

79. Barcaroli D, Bongiorno-Borbone L, Terrinoni A, Hofmann TG, Rossi M, Knight RA, et al. FLASH is required for histone transcription and S-phase progression. Proc. Natl. Acad. Sci. U. S. A. 2006;103:14808–12.

80. Nguyen VT, Kiss T, Michels AA, Bensaude O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature 2001;414:322–5.

81. Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, Lykke-Andersen S, et al. Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. Mol. Cell 2011;43:624–37.

82. Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G, Jensen TH. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. Cell Rep. 2015;10:178–92.

83. Chlebowski A, Lubas M, Jensen TH, Dziembowski A. RNA decay machines: The exosome. Biochim. Biophys. Acta BBA - Gene Regul. Mech. 2013;1829:552–60.

84. Schneider C, Tollervey D. Threading the barrel of the RNA exosome. Trends Biochem. Sci. 2013;38:485–93.

85. Lee NN, Chalamcharla VR, Reyes-Turcu F, Mehta S, Zofall M, Balachandran V, et al. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. Cell 2013;155:1061–74.

86. Zhou Y, Zhu J, Schermann G, Ohle C, Bendrin K, Sugioka-Sugiyama R, et al. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. Nat. Commun. 2015;6:7050.

87. Egan ED, Braun CR, Gygi SP, Moazed D. Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. RNA 2014;20:867–81.

88. Sugiyama T, Sugioka-Sugiyama R. Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast: Meiotic mRNA degradation by Red1 in mitotic cells. EMBO J. 2011;30:1027–39.

89. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, et al. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 1997;385:357–61.

90. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates III JR, et al. Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. Mol. Cell 2009;33:365–76.

91. Xiang K, Tong L, Manley JL. Delineating the Structural Blueprint of the Pre-mRNA 3′-End Processing Machinery. Mol. Cell. Biol. 2014;34:1894–910.

92. de Vries H, Rüegsegger U, Hübner W, Friedlein A, Langen H, Keller W. Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. 2000;19:5895–904.

93. Ntini E, Järvelin AI, Bornholdt J, Chen Y, Boyd M, Jørgensen M, et al. Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. Nat. Struct. Mol. Biol. 2013;20:923–8.

94. Yamamoto J, Hagiwara Y, Chiba K, Isobe T, Narita T, Handa H, et al. DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes. Nat. Commun. 2014;5:4263.

95. Baillat D, Hakimi M-A, Näär AM, Shilatifard A, Cooch N, Shiekhattar R. Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing, Associates with the C-Terminal Repeat of RNA Polymerase II. Cell 2005;123:265–76.

96. Izaurralde E, Lewis J, Gamberi C, Jarmolowski A, McGuigan C, Mattaj IW. A cap-binding protein complex mediating U snRNA export. Nature 1995;376:709–12.

97. Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW. PHAX, a Mediator of U snRNA Nuclear Export Whose Activity Is Regulated by Phosphorylation. Cell 2000;101:187–98.

98. Mourão A, Varrot A, Mackereth CD, Cusack S, Sattler M. Structure and RNA recognition by the snRNA and snoRNA transport factor PHAX. RNA 2010;16:1205–16.

99. Kitao S, Segref A, Kast J, Wilm M, Mattaj IW, Ohno M. A Compartmentalized Phosphorylation/Dephosphorylation System That Regulates U snRNA Export from the Nucleus. Mol. Cell. Biol. 2008;28:487–97.

100. Boulon S, Verheggen C, Jady BE, Girard C, Pescia C, Paul C, et al. PHAX and CRM1 Are Required Sequentially to Transport U3 snoRNA to Nucleoli. Mol. Cell 2004;16:777–87.

101. McCloskey A, Taniguchi I, Shinmyozu K, Ohno M. hnRNP C Tetramer Measures RNA Length to Classify RNA Polymerase II Transcripts for Export. Science 2012;335:1643–6.

102. Masuyama K, Taniguchi I, Kataoka N, Ohno M. RNA length defines RNA export pathway. Genes Dev. 2004;18:2074–85.

103. Zhou Z, Luo M, Straesser K, Katahira J, Hurt E, Reed R. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature 2000;407:401–5.

104. Luo M-J, Zhou Z, Magni K, Christoforides C, Rappsilber J, Mann M, et al. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. Nature 2001;413:644–7.

105. Rodrigues JP, Rode M, Gatfield D, Blencowe BJ, Carmo-Fonseca M, Izaurralde E. REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. Proc. Natl. Acad. Sci. 2001;98:1030–5.

106. Lei H, Dias AP, Reed R. Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex. Proc. Natl. Acad. Sci. 2011;108:17985–90.

107. Gromadzka AM, Steckelberg A-L, Singh KK, Hofmann K, Gehring NH. A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. Nucleic Acids Res. 2016;gkw009.

108. Braun IC, Herold A, Rode M, Conti E, Izaurralde E. Overexpression of TAP/p15 Heterodimers Bypasses Nuclear Retention and Stimulates Nuclear mRNA Export. J. Biol. Chem. 2001;276:20536–43.

109. Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R. Recruitment of the human TREX complex to mRNA during splicing. Genes Dev. 2005;19:1512–7.

110. Dufu K, Livingstone MJ, Seebacher J, Gygi SP, Wilson SA, Reed R. ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. Genes Dev. 2010;24:2043–53.

111. Chi B, Wang Q, Wu G, Tan M, Wang L, Shi M, et al. Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. Nucleic Acids Res. 2013;41:1294–306.

112. Dias AP, Dufu K, Lei H, Reed R. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. Nat. Commun. 2010;1:97.

113. Le Hir H. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J. 2001;20:4987–97.

114. Viphakone N, Hautbergue GM, Walsh M, Chang C-T, Holland A, Folco EG, et al. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat. Commun. 2012;3:1006.

115. Hautbergue GM, Hung M-L, Golovanov AP, Lian L-Y, Wilson SA. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. Proc. Natl. Acad. Sci. 2008;105:5154–9.

116. Fribourg S, Braun IC, Izaurralde E, Conti E. Structural Basis for the Recognition of a Nucleoporin FG Repeat by the NTF2-like Domain of the TAP/p15 mRNA Nuclear Export Factor. Mol. Cell 2001;8:645–56.

117. Cheng H, Dufu K, Lee C-S, Hsu JL, Dias A, Reed R. Human mRNA Export Machinery Recruited to the 5′ End of mRNA. Cell 2006;127:1389–400.

118. Nojima T, Hirose T, Kimura H, Hagiwara M. The Interaction between Cap-binding Complex and RNA Export Factor Is Required for Intronless mRNA Export. J. Biol. Chem. 2007;282:15645–51.

119. Chi B, Wang K, Du Y, Gui B, Chang X, Wang L, et al. A Sub-Element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. Nucleic Acids Res. 2014;42:7305–18.

120. Gebhardt A, Habjan M, Benda C, Meiler A, Haas DA, Hein MY, et al. mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. Nat. Commun. 2015;6:8192.

121. Will CL, Lührmann R. Spliceosome Structure and Function. Cold Spring Harb. Perspect. Biol. 2011;3:a003707.

122. Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat. Rev. Mol. Cell Biol. 2013;14:153–65.

123. Matera AG, Wang Z. A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. 2014;15:108–21.

124. Raczynska KD, Simpson CG, Ciesiolka A, Szewc L, Lewandowska D, McNicol J, et al. Involvement of the nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana. Nucleic Acids Res. 2010;38:265–78.

125. Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev. 1996;10:1683–98.

126. O'Mullane L, Eperon IC. The Pre-mRNA 5′ Cap Determines Whether U6 Small Nuclear RNA Succeeds U1 Small Nuclear Ribonucleoprotein Particle at 5′ Splice Sites. Mol. Cell. Biol. 1998;18:7510–20.

127. Lenasi T, Peterlin BM, Barboric M. Cap-binding Protein Complex Links Pre-mRNA Capping to Transcription Elongation and Alternative Splicing through Positive Transcription Elongation Factor b (P-TEFb). J. Biol. Chem. 2011;286:22758–68.

128. Pabis M, Neufeld N, Steiner MC, Bojic T, Shav-Tal Y, Neugebauer KM. The nuclear cap-binding complex interacts with the U4/U6·U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. RNA 2013;19:1054–63.

129. Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. Nature 2002;419:182–5.

130. Rappsilber J, Ryder U, Lamond AI, Mann M. Large-Scale Proteomic Analysis of the Human Spliceosome. Genome Res. 2002;12:1231–45.

131. Sugiyama T, Thillainadesan G, Chalamcharla VR, Meng Z, Balachandran V, Dhakshnamoorthy J, et al. Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly. Mol. Cell 2016;61:747–59.

132. Martienssen R, Moazed D. RNAi and heterochromatin assembly. Cold Spring Harb. Perspect. Biol. 2015;7:a019323.

133. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SIS, et al. RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. Science 2004;303:672–6.

134. Schalch T, Job G, Shanker S, Partridge JF, Joshua-Tor L. The Chp1–Tas3 core is a multifunctional platform critical for gene silencing by RITS. Nat. Struct. Mol. Biol. 2011;18:1351–7.

135. Jia S, Kobayashi R, Grewal SIS. Ubiquitin ligase component Cul4 associates with Clr4 histone methyltransferase to assemble heterochromatin. Nat. Cell Biol. 2005;7:1007–13.

136. Zhang K, Mosch K, Fischle W, Grewal SIS. Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. Nat. Struct. Mol. Biol. 2008;15:381–8.

137. Ukleja M, Cuellar J, Siwaszek A, Kasprzak JM, Czarnocki-Cieciura M, Bujnicki JM, et al. The architecture of the Schizosaccharomyces pombe CCR4-NOT complex. Nat. Commun. 2016;7:10433.

138. Cotobal C, Rodríguez-López M, Duncan C, Hasan A, Yamashita A, Yamamoto M, et al. Role of Ccr4-Not complex in heterochromatin formation at meiotic genes and subtelomeres in fission yeast. Epigenetics Chromatin 2015;8:28.

139. Mersman DP, Du H-N, Fingerman IM, South PF, Briggs SD. Polyubiquitination of the demethylase Jhd2 controls histone methylation and gene expression. Genes Dev. 2009;23:951–62.

140. Morgunova V, Akulenko N, Radion E, Olovnikov I, Abramov Y, Olenina LV, et al. Telomeric repeat silencing in germ cells is essential for early development in Drosophila. Nucleic Acids Res. 2015;43:8762–73.

141. Czech B, Preall JB, McGinn J, Hannon GJ. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol. Cell 2013;50:749–61.

142. Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, et al. Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature 2006;442:45–50.

143. Zofall M, Yamanaka S, Reyes-Turcu FE, Zhang K, Rubin C, Grewal SIS. RNA Elimination Machinery Targeting Meiotic mRNAs Promotes Facultative Heterochromatin Formation. Science 2012;335:96–100.

144. Hiriart E, Vavasseur A, Touat-Todeschini L, Yamashita A, Gilquin B, Lambert E, et al. Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast: Targeting RITS to meiotic RNAs and genes. EMBO J. 2012;31:2296–308.

145. Scheibe M, Arnoult N, Kappei D, Buchholz F, Decottignies A, Butter F, et al. Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. Genome Res. 2013;23:2149–57.

146. Rippe K, Luke B. TERRA and the state of the telomere. Nat. Struct. Mol. Biol. 2015;22:853–8.

147. Porro A, Feuerhahn S, Delafontaine J, Riethman H, Rougemont J, Lingner J. Functional characterization of the TERRA transcriptome at damaged telomeres. Nat. Commun. 2014;5:5379.

148. Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. Mol. Cell 2009;35:403–13.

149. Arnoult N, Van Beneden A, Decottignies A. Telomere length regulates TERRA levels through increased trimethylation of telomeric H3K9 and HP1α. Nat. Struct. Mol. Biol. 2012;19:948–56.

150. Liu J, Jung C, Xu J, Wang H, Deng S, Bernad L, et al. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. Plant Cell 2012;24:4333–45.

151. Fantes J, Ragge NK, Lynch S-A, McGill NI, Collin JRO, Howard-Peebles PN, et al. Mutations in SOX2 cause anophthalmia. Nat. Genet. 2003;33:462–3.

152. Shahryari A, Rafiee MR, Fouani Y, Oliae NA, Samaei NM, Shafiee M, et al. Two Novel Splice Variants of SOX2OT, SOX2OT-S1, and SOX2OT-S2 are Coupregulated with SOX2 and OCT4 in Esophageal Squamous Cell Carcinoma. STEM CELLS 2014;32:126–34.

153. Askarian-Amiri ME, Seyfoddin V, Smart CE, Wang J, Kim JE, Hansji H, et al. Emerging Role of Long Non-Coding RNA SOX2OT in SOX2 Regulation in Breast Cancer. PLOS ONE 2014;9:e102140.

154. Shahryari A, Jazi MS, Samaei NM, Mowla SJ. Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis. RNA 2015;196.

155. Matheu A, Maraver A, Klatt P, Flores I, Garcia-Cao I, Borras C, et al. Delayed ageing through damage protection by the Arf/p53 pathway. Nature 2007;448:375–9.

156. Young RW. Cell differentiation in the retina of the mouse. Anat. Rec. 1985;212:199–205.

157. Agathocleous M, Harris WA. From Progenitors to Differentiated Cells in the Vertebrate Retina. Annu. Rev. Cell Dev. Biol. 2009;25:45–69.

158. Alexiades MR, Cepko C. Quantitative analysis of proliferation and cell cycle length during development of the rat retina. Dev. Dyn. 1996;205:293–307.

159. Das G, Choi Y, Sicinski P, Levine EM. Cyclin D1 fine-tunes the neurogenic output of embryonic retinal progenitor cells. Neural Develop. 2009;4:15.

160. Ohnuma S, Hopper S, Wang KC, Philpott A, Harris WA. Co-ordinating retinal histogenesis: early cell cycle exit enhances early cell fate determination in the Xenopus retina. Development 2002;129:2435–46.

161. Dyer MA, Cepko CL. p27Kip1 and p57Kip2 Regulate Proliferation in Distinct Retinal Progenitor Cell Populations. J. Neurosci. 2001;21:4259–71.

162. Cepko CL, Austin CP, Yang X, Alexiades M, Ezzeddine D. Cell fate determination in the vertebrate retina. Proc. Natl. Acad. Sci. U. S. A. 1996;93:589–95.

163. Taverna E, Huttner WB. Neural Progenitor Nuclei IN Motion. Neuron 2010;67:906–14.

164. Leung L, Klopper AV, Grill SW, Harris WA, Norden C. Apical migration of nuclei during G2 is a prerequisite for all nuclear motion in zebrafish neuroepithelia. Development 2011;138:5003–13.

165. Baye LM, Link BA. Interkinetic Nuclear Migration and the Selection of Neurogenic Cell Divisions during Vertebrate Retinogenesis. J. Neurosci. 2007;27:10143–52.

166. Cisneros E, Latasa MJ, García-Flores M, Frade JM. Instability of Notch1 and Delta1 mRNAs and reduced Notch activity in vertebrate neuroepithelial cells undergoing S-phase. Mol. Cell. Neurosci. 2008;37:820–31.

167. Del Bene F, Wehman AM, Link BA, Baier H. Regulation of Neurogenesis by Interkinetic Nuclear Migration through an Apical-Basal Notch Gradient. Cell 2008;134:1055–65.

168. Nickerson PE, Ronellenfitch KM, Csuzdi NF, Boyd JD, Howard PL, Delaney KR, et al. Live imaging and analysis of postnatal mouse retinal development. BMC Dev. Biol. 2013;13:24.

169. Walsh K, Perlman H. Cell cycle exit upon myogenic differentiation. Curr. Opin. Genet. Dev. 1997;7:597–602.

170. Yaffe D, Saxel O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature 1977;270:725–7.

171. Silberstein L, Webster SG, Travis M, Blau HM. Developmental progression of myosin gene expression in cultured muscle cells. Cell 1986;46:1075–81.

172. Rando TA, Blau HM. Primary mouse myoblast purification, characterization, and transplantation for cell-mediated gene therapy. J. Cell Biol. 1994;125:1275–87.

173. Ge Y, Chen J. MicroRNAs in skeletal myogenesis. Cell Cycle 2011;10:441–8.

174. Seok HY, Tatsuguchi M, Callis TE, He A, Pu WT, Wang DZ. miR-155 inhibits expression of the MEF2A protein to repress skeletal muscle differentiation. J. Biol. Chem. 2011;286:35339–46.

175. Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E, et al. miR-24 Inhibits Cell Proliferation by Targeting E2F2, MYC, and Other Cell-Cycle Genes via Binding to "Seedless" 3′UTR MicroRNA Recognition Elements. Mol. Cell 2009;35:610–25.

176. Sun Q, Zhang Y, Yang G, Chen X, Zhang Y, Cao G, et al. Transforming growth factor-beta-regulated miR-24 promotes skeletal muscle differentiation. Nucleic Acids Res. 2008;36:2690–9.

177. Jadhav AP, Mason HA, Cepko CL. Notch 1 inhibits photoreceptor production in the developing mammalian retina. Development 2006;133:913–23.

178. Mizeracka K, DeMaso CR, Cepko CL. Notch1 is required in newly postmitotic cells to inhibit the rod photoreceptor fate. Development 2013;140:3188–97.

179. Matsuda T, Cepko CL. Controlled expression of transgenes introduced by in vivo electroporation. Proc. Natl. Acad. Sci. U. S. A. 2007;104:1027–32.

180. Matsuda T, Cepko CL. Electroporation and RNA interference in the rodent retina in vivo and in vitro. Proc. Natl. Acad. Sci. U. S. A. 2004;101:16–22.

181. Kato H, Sakai T, Tamura K, Minoguchi S, Shirayoshi Y, Hamada Y, et al. Functional conservation of mouse Notch receptor family members. FEBS Lett. 1996;395:221–4.

182. Young RW. Cell proliferation during postnatal development of the retina in the mouse. Brain Res. 1985;353:229–39.

183. Solovei I, Kreysing M, Lanctôt C, Kösem S, Peichl L, Cremer T, et al. Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. Cell 2009;137:356–68.

184. Bassett EA, Wallace VA. Cell fate determination in the vertebrate retina. Trends Neurosci. 2012;35:565–73.

185. Mears AJ, Kondo M, Swain PK, Takada Y, Bush RA, Saunders TL, et al. Nrl is required for rod photoreceptor development. Nat. Genet. 2001;29:447–52.

186. Oh ECT, Khan N, Novelli E, Khanna H, Strettoi E, Swaroop A. Transformation of cone precursors to functional rod photoreceptors by bZIP transcription factor NRL. Proc. Natl. Acad. Sci. 2007;104:1679–84.

187. Dizhoor AM, Ray S, Kumar S, Niemi G, Spencer M, Brolley D, et al. Recoverin: a calcium sensitive activator of retinal rod guanylate cyclase. Science 1991;251:915–8.

188. Furukawa T, Mukherjee S, Bao Z-Z, Morrow EM, Cepko CL. rax, Hes1, and notch1 Promote the Formation of Müller Glia by Postnatal Retinal Progenitor Cells. Neuron 2000;26:383–94.

189. Bunt-Milam AH, Saari JC. Immunocytochemical localization of two retinoid-binding proteins in vertebrate retina. J. Cell Biol. 1983;97:703–12.

190. Miyoshi H, Takahashi M, Gage FH, Verma IM. Stable and efficient gene transfer into the retina using an HIV-based lentiviral vector. Proc. Natl. Acad. Sci. 1997;94:10319–23.

191. Klein RL, Meyer EM, Peel AL, Zolotukhin S, Meyers C, Muzyczka N, et al. Neuron-Specific Transduction in the Rat Septohippocampal or Nigrostriatal Pathway by Recombinant Adeno-associated Virus Vectors. Exp. Neurol. 1998;150:183–94.

192. Haeseleer F, Sokal I, Verlinde CLMJ, Erdjument-Bromage H, Tempst P, Pronin AN, et al. Five Members of a Novel Ca2+-binding Protein (CABP) Subfamily with Similarity to Calmodulin. J. Biol. Chem. 2000;275:1247–60.

193. Iso T, Kedes L, Hamamori Y. HES and HERP families: Multiple effectors of the notch signaling pathway. J. Cell. Physiol. 2003;194:237–55.

194. Burch BD, Godfrey AC, Gasdaska PY, Salzler HR, Duronio RJ, Marzluff WF, et al. Interaction between FLASH and Lsm11 is essential for histone pre-mRNA processing in vivo in Drosophila. RNA N. Y. N 2011;17:1132–47.

195. Dominski Z, Zheng LX, Sanchez R, Marzluff WF. Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. Mol. Cell. Biol. 1999;19:3561–70.

196. Imai F, Yoshizawa A, Matsuzaki A, Oguri E, Araragi M, Nishiwaki Y, et al. Stem-loop binding protein is required for retinal cell proliferation, neurogenesis, and intraretinal axon pathfinding in zebrafish. Dev. Biol. 2014;394:94–109.

197. Taranova OV, Magness ST, Fagan BM, Wu Y, Surzenko N, Hutton SR, et al. SOX2 is a dose-dependent regulator of retinal neural progenitor competence. Genes Dev. 2006;20:1187–202.

198. Lin Y, Ouchi Y, Satoh S, Watanabe S. Sox2 Plays a Role in the Induction of Amacrine and Müller Glial Cells in Mouse Retinal Progenitor Cells. Investig. Opthalmology Vis. Sci. 2009;50:68.

199. Bhatia B, Singhal S, Tadman DN, Khaw PT, Limb GA. SOX2 Is Required for Adult Human Müller Stem Cell Survival and Maintenance of Progenicity In Vitro. Investig. Opthalmology Vis. Sci. 2011;52:136.

200. Surzenko N, Crowl T, Bachleda A, Langer L, Pevny L. SOX2 maintains the quiescent progenitor cell state of postnatal retinal Muller glia. Dev. Camb. Engl. 2013;140:1445–56.

201. Bachleda AR, Pevny LH, Weiss ER. Sox2-Deficient Müller Glia Disrupt the Structural and Functional Maturation of the Mammalian Retina. Invest. Ophthalmol. Vis. Sci. 2016;57:1488–99.

202. Ringuette R, Atkins M, Lagali PS, Bassett EA, Campbell C, Mazerolle C, et al. A Notch-Gli2 axis sustains Hedgehog responsiveness of neural progenitors and Müller glia. Dev. Biol. 2016;411:85–100.

203. Li H, Collado M, Villasante A, Matheu A, Lynch CJ, Cañamero M, et al. p27Kip1 Directly Represses Sox2 during Embryonic Stem Cell Differentiation. Cell Stem Cell 2012;11:845–52.

204. Jeon C-J, Strettoi E, Masland RH. The Major Cell Populations of the Mouse Retina. J. Neurosci. 1998;18:8936–46.

205. Chi B, Wang K, Du Y, Gui B, Chang X, Wang L, et al. A Sub-Element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. Nucleic Acids Res. 2014;42:7305–18.

206. Müller-McNicoll M, Neugebauer KM. Good cap/bad cap: how the cap-binding complex determines RNA fate. Nat. Struct. Mol. Biol. 2014;21:9–12.

207. Wagner EJ, Burch BD, Godfrey AC, Salzler HR, Duronio RJ, Marzluff WF. A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing. Mol. Cell 2007;28:692–9.

208. Terry NHA, White RA. Flow cytometry after bromodeoxyuridine labeling to measure S and G2+M phase durations plus doubling times in vitro and in vivo. Nat. Protoc. 2006;1:859–69.

209. Veitia RA. Exploring the molecular etiology of dominant-negative mutations. Plant Cell 2007;19:3843–51.

210. Wang Y, Baskerville S, Shenoy A, Babiarz JE, Baehner L, Blelloch R. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nat. Genet. 2008;40:1478–83.

211. Pacheco TR, Moita LF, Gomes AQ, Hacohen N, Carmo-Fonseca M. RNA interference knockdown of hU2AF35 impairs cell cycle progression and modulates alternative splicing of Cdc25 transcripts. Mol. Biol. Cell 2006;17:4187–99.

212. Pawellek A, McElroy S, Samatov T, Mitchell L, Woodland A, Ryder U, et al. Identification of small molecule inhibitors of pre-mRNA splicing. J. Biol. Chem. 2014;289:34683–98.

213. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 2009;4:363–71.

214. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 2011;7:539.

215. Cerdà-Costa N, Bonet J, Fernández MR, Avilés FX, Oliva B, Villegas S. Prediction of a new class of RNA recognition motif. J. Mol. Model. 2011;17:1863–75.

216. Allain FHT, Gubser CC, Howe PWA, Nagai K, Neuhaus D, Varani G. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. Nature 1996;380:646–50.

217. Cléry A, Blatter M, Allain FHT. RNA recognition motifs: boring? Not quite. Curr. Opin. Struct. Biol. 2008;18:290–8.

218. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 2012;40:D261–70.

219. Amanchy R, Zhong J, Molina H, Chaerkady R, Iwahori A, Kalume DE, et al. Identification of c-Src tyrosine kinase substrates using mass spectrometry and peptide microarrays. J. Proteome Res. 2008;7:3900–10.

220. Fribourg S, Gatfield D, Izaurralde E, Conti E. A novel mode of RBD-protein recognition in the Y14-Mago complex. Nat. Struct. Biol. 2003;10:433–9.

221. Kadlec J, Izaurralde E, Cusack S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. Nat. Struct. Mol. Biol. 2004;11:330–7.

222. Zarbin M. Cell-Based Therapy for Degenerative Retinal Disease. Trends Mol. Med. 2016;22:115–34.

223. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell 2013;154:1370–9.

224. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. Nat. Protoc. 2014;9:1956–68.

225. Klimova L, Lachova J, Machon O, Sedlacek R, Kozmik Z. Generation of mRx-Cre Transgenic Mouse Line for Efficient Conditional Gene Deletion in Early Retinal Progenitors. PLOS ONE 2013;8:e63029.

226. Akimoto M, Cheng H, Zhu D, Brzezinski JA, Khanna R, Filippova E, et al. Targeting of GFP to newborn rods by Nrl promoter and temporal expression profiling of flow-sorted photoreceptors. Proc. Natl. Acad. Sci. U. S. A. 2006;103:3890–5.

227. Cheng H, Khanna H, Oh ECT, Hicks D, Mitton KP, Swaroop A. Photoreceptor-specific nuclear receptor NR2E3 functions as a transcriptional activator in rod photoreceptors. Hum. Mol. Genet. 2004;13:1563–75.

228. Arnold K, Sarkar A, Yram MA, Polo JM, Bronson R, Sengupta S, et al. Sox2+ Adult Stem and Progenitor Cells Are Important for Tissue Regeneration and Survival of Mice. Cell Stem Cell 2011;9:317–29.

229. Calero G, Wilson KF, Ly T, Rios-Steiner JL, Clardy JC, Cerione RA. Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. Nat. Struct. Mol. Biol. 2002;9:912–7.

230. Cheng Y. Single-Particle Cryo-EM at Crystallographic Resolution. Cell 2015;161:450–7.

231. Nogales E, Scheres SHW. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. Mol. Cell 2015;58:677–89.

232. Zielezinski A, Karlowski WM. Early origin and adaptive evolution of the GW182 protein family, the key component of RNA silencing in animals. RNA Biol. 2015;12:761–70.

233. Hosoda N, Kim YK, Lejeune F, Maquat LE. CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. Nat. Struct. Mol. Biol. 2005;12:893–901.

234. Hwang J, Sato H, Tang Y, Matsuda D, Maquat LE. UPF1 Association with the Cap-Binding Protein, CBP80, Promotes Nonsense-Mediated mRNA Decay at Two Distinct Steps. Mol. Cell 2010;39:396–409.

235. Kervestin S, Jacobson A. NMD: a multifaceted response to premature translational termination. Nat. Rev. Mol. Cell Biol. 2012;13:700–12.

236. Tange TØ, Shibuya T, Jurica MS, Moore MJ. Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. RNA 2005;11:1869–83.

237. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, et al. Large-scale mapping of human protein–protein interactions by mass spectrometry. Mol. Syst. Biol. [Internet] 2007 [cited 2016 May 30];3. Available from: http://msb.embopress.org/cgi/doi/10.1038/msb4100134

238. Lykke-Andersen J, Shu M-D, Steitz JA. Human Upf Proteins Target an mRNA for Nonsense-Mediated Decay When Bound Downstream of a Termination Codon. Cell 2000;103:1121–31.

239. Gehring NH, Hentze MW, Kulozik AE. Tethering assays to investigate nonsense-mediated mRNA decay activating proteins. Methods Enzymol. 2008;448:467–82.

240. Baron-Benhamou J, Gehring NH, Kulozik AE, Hentze MW. Using the lambdaN peptide to tether proteins to RNAs. Methods Mol. Biol. Clifton NJ 2004;257:135–54.

241. Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat. Genet. 2004;36:1073–8.

242. Wittmann J, Hol EM, Jäck H-M. hUPF2 Silencing Identifies Physiologic Substrates of Mammalian Nonsense-Mediated mRNA Decay. Mol. Cell. Biol. 2006;26:1272–87.

243. Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Mühlemann O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. RNA 2011;17:2108–18.

Appendix A – Supplementary Information

Table 1 shRNA and siRNA targeting sequences, and qRT-PCR primers

Ars2 shRNA 78	ATGCAGCTGTCATTAAGATGGAAGGTGGC
Ars2 shRNA 79	CAGGCTGAGAATGACAGTTCCAACGATGA
Ars2 shRNA 80	GCAAGGATAAGTGGCTATGTCCTCTCAGT
Ars2 shRNA 81	CCAGGCTTTATGCGAGTGGCACTGTCAGA
FLASH shRNA	AACATTGTGCCAATAATGTCTGGTCACGT
Drosha shRNA	GTTCATTGAGCGGAAATACAGACAAGAGT
Lsm11 shRNA	GGATTACCAGCAGGTATTCACTCGGCACA
Ars2 siRNA 1	CCCGTCGTGTCCGCAACATAA
Ars2 siRNA 2	CCGAAGAAGCACTTAAAGAAA
Ars2 siRNA 3	CCCAGCTTTGCCTGAGATCAA
UPF1 siRNA	AAGAUGCAGUUCCGCUCCAdTdT
HIST3H2A-Tot For	5’-AAGCTCGTGCAAAAGCGAAG-3’
HIST3H2A-Tot Rev	5’-TATTCTAGCACAGCCGCCAG-3’
HIST3H2A-NC For	5’-GCTACTGCCCAAGAAGACCG-3’
HIST3H2A-NC Rev	5’-GTAACTGGCGACAGGCTCAG-3’
HIST1H2BE-NC For	5’-ACACCAGCTCCAAGTGAGTT-3’
HIST1H2BE-NC Rev	5’-AGGTGGTACTGCGGGTACT-3’
GAPDH For	5’-ACGACCCCTTCATTGACCTC-3’
GAPDH Rev	5’-GTCTCGCTCCTGGAAGATGG-3’

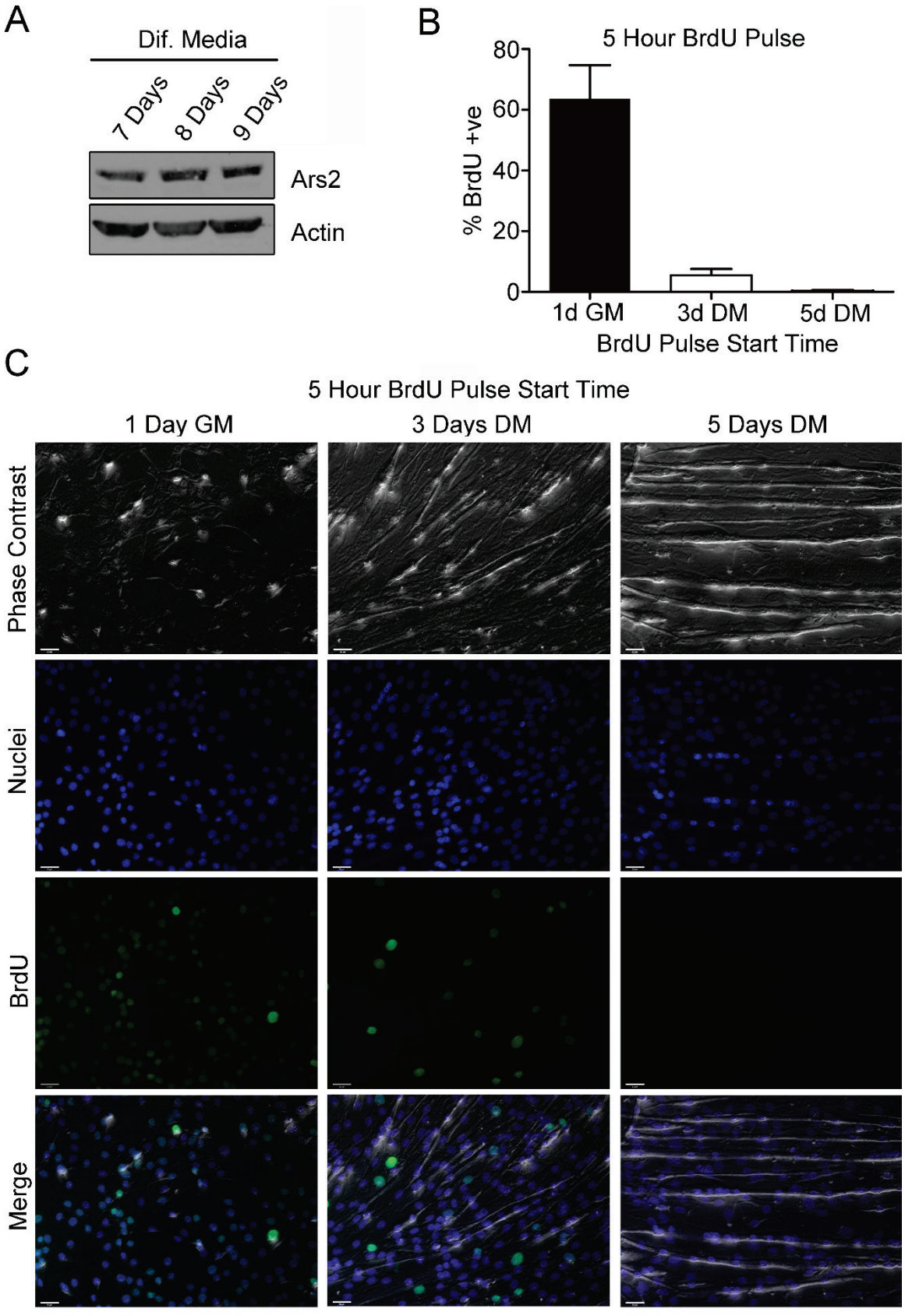
A
Dif. Media
7 Days
8 Days
9 Days
Ars2
Actin
B
5 Hour BrdU Pulse
% BrdU +ve
80
60
40
20
0
1d GM
3d DM
5d DM
BrdU Pulse Start Time
C
5 Hour BrdU Pulse Start Time
1 Day GM
3 Days DM
5 Days DM
Phase Contrast
Nuclei
BrdU
Merge

Figure 32 ARS2 is expressed in post-mitotic C2C12 myotubes

A) C2C12 myoblasts were incubated in Differentiation Media (DM) for 7, 8 and 9 days and endogenous ARS2 protein levels were measured using Western blotting. Actin was used as a loading control. B) and C) A 5 hour BrdU pulse was given to C2C12 myoblasts that had been incubated in either growth media (GM) for 1 day (1d), DM for 3 days (3d) or DM for 5 days (5d). After the BrdU pulse, cells were fixed and labeled with a BrdU antibody conjugated to AF-488 and nuclei were stained using Hoechst. The proportion of BrdU +ve nuclei is quantified in B). After 5 days of being induced to differentiate, most nuclei were in multinucleated myotubes and were BrdU –ve, suggesting these are post-mitotic myotubes.

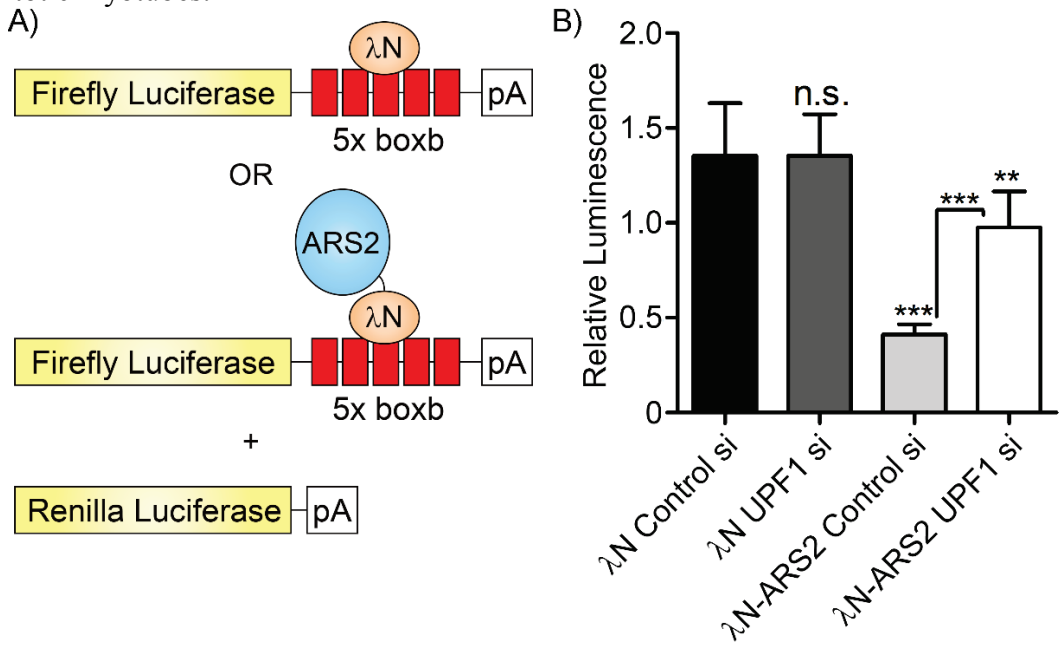

Figure 33 ARS2 decreases nonsense-mediated decay luciferase reporter in UPF1-dependent manner

A) Schematic of nonsense-mediated decay (NMD) luciferase reporter and tethering. The NMD firefly luciferase reporter contains 5× boxb sites in its 3'-UTR (NMD-firefly), which is bound by the 22 amino acid RNA-binding domain of the bacteriophage λ antiterminator protein (λN) when co-expressed. A fusion λN-ARS2 protein was also generated, which is tethered to the NMD-firefly luciferase reporter when co-expressed. If λN-ARS2 recruits NMD factors and induces decay, a decrease in luciferase activity is observed. Firefly luciferase activity was normalized to a Renilla luciferase construct lacking 5× boxb sites. Jennifer Christie cloned the NMD firefly luciferase reporter, as well as λN and λN-ARS2. B) HeLa cells were transfected with either control siRNA (si) or UPF1 si for two consecutive days, and on day 3 were co-transfected with NMD-firefly and Renilla luciferase and either λN or λN-ARS2. Firefly luminescence was normalized to Renilla luminescence.

Appendix B – Additional Publications

1. Alford SC*, **O'Sullivan C***, Obst J, Christie J, Howard PL. Conditional protein splicing of α-sarcin in live cells. Mol. Biosyst. 2014;10:831-7.

2. Alford SC*, **O'Sullivan C***, Howard PL. Conditional toxin splicing using a split intein system. Meth. Mol. Biol. Accepted.

3. Meanwell MW, **O'Sullivan C**, Howard PL, Fyles TM. Branched-chain and dendritic lipids for nanoparticles. Can. J. Chem. In preparation.

*These authors contributed equally to this work.