2021

Silencing Defective 2 is an essential gene required for ribosome biogenesis and the regulation of alternative splicing

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

SILENCING DEFECTIVE 2 IS AN ESSENTIAL GENE REQUIRED FOR RIBOSOME BIOGENESIS AND THE REGULATION OF ALTERNATIVE SPLICING

by

JESS FLORO

B.S., Virginia Polytechnic and State University, 2012

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2021

Approved by

First Reader   _____
                              Rachel Flynn, Ph.D.
                              Assistant Professor of Pharmacology and Experimental Therapeutics

Second Reader_____
                              Neil Ganem, Ph.D.
                              Associate Professor of Pharmacology and Experimental Therapeutics

DEDICATION

I would like to dedicate this work to my parents, who have always supported and

inspired me.

me both as a friend and colleague. Thank you. **Dr. Peter Wu**, thank you for getting me started on the SDE2 project. I enjoyed our experiments, and our jokes that we shared together. **Dr. Himabindu Gali,** I couldn't have asked for a better bay-mate. Thank you for many years of good times, even in the face of terrible data. Your knowledge and work ethic are second to none, and I am lucky to call you my friend. **Jane Lock**, The Saboteur, thank you for so many great memories in the lab. Thank you for listening to me complain about SDE2, and for allowing me to complete construction of The Arch in your bay. I am glad to have undertaken our scientific journeys together. **Abby Metzger,** thank you for being a great trainee. I hope you learned as much from me as I did you, and I look forward to great things on your horizon. **Matt Reiss**, thank you for always bringing a positive attitude to the lab, and for being the sole other RNA researcher on K7. **Lisa Carson**, thank you for always bringing a realistic attitude to the lab, and for your unparalleled anger towards the powers that be. **Victoria Kacprzak and Reyna Stuppard**, thank you for being excellent lab managers. Without you, this would have taken twice as long. To my friends on K7, **Ryan "Bean Curd" Quinton, Marc "Porc" Vittoria, Sang Hee Lim, Kristy Kotynková, Kevin "Splashback" Hua, Amanda "The Bully" Bolgioni, and Rachel Ho,** thank you for the memories, my friends, I will always look back and smile on the good times we had. **Emily**, thank you for always being there for me, and thank you for all the great times we've had so far. I'm lucky to have you and I'm excited for our next chapter together. **Kelly,** thanks for putting up with me and always being a great little sister. To my **Mom** and **Dad**, thank you

for everything. I am truly grateful for your unconditional love, support, and encouragement. I would not be where I am today were it not for both of you.

SILENCING DEFECTIVE 2 IS AN ESSENTIAL GENE REQUIRED FOR RIBOSOME BIOGENESIS AND THE REGULATION OF ALTERNATIVE SPLICING

JESS FLORO

Boston University School of Medicine, 2021

Major Professor: Rachel Flynn, Ph.D., Assistant Professor of Pharmacology and Experimental Therapeutics

ABSTRACT

RNA provides the framework for the assembly of some of the most intricate macromolecular complexes within the cell, including the spliceosome and the mature ribosome. The assembly of these complexes relies on the coordinated association of RNA with hundreds of *trans*-acting protein factors. While some of these *trans*-acting factors are RNA binding proteins (RBPs), others are adaptor proteins, and others still, function as both. Defects in the assembly of these complexes results in a number of human pathologies including neurodegeneration and cancer. Here, we demonstrate that Silencing Defective 2 (SDE2) is both an RNA binding protein and also a *trans*-acting adaptor protein that functions to regulate RNA splicing and ribosome biogenesis. SDE2 depletion leads to widespread changes in alternative splicing, defects in ribosomal biogenesis, and ultimately complete loss of cell viability. Our data highlight SDE2 as a previously uncharacterized essential gene required for the assembly and maturation of some of the most fundamental processes in mammalian cells.

PREFACE

Chapters 2 and 3 are reproduced from a published paper, on which I am first author.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3'SS ....3' Splice Site

5' SS ....5' Splice Site

AA .... Alternative Aceptor

AD ....Alternative Donor

AS .... Alternative Splicing

ASO .... Antisense Oligonucleotide

Bact ....B Active Splicing Complex

BPS .... Branch Point Sequence

BU ....Boston University

CB .... Cajal Body

CBC .... Cap Binding Complex

CBP20 ....Cap Binding Complex Protein 20

CBP80 ....Cap Binding Complex Protein 80

CCHC ....Zinc Finger

cDNA ....complementary DNA

CDS .... Coding Sequence

CE .... Cassette Exon

CHH .... Cartilage Hair Hypoplasia

CHX .... Cycloheximide

CPSF73 ....Cleavage and Polyadenylation Specific Factor 73

CRM1 .... Chromosome Region Maintenance 1

CHAPTER ONE: INTRODUCTION

Section 1: Ribonucleic Acid

Ribonucleic Acid (RNA) is an incredibly versatile molecule. It is a key component of the central dogma of molecular biology – which states that genetic information flows from DNA, through RNA, to proteins. However, beyond the transfer of genetic information from DNA to proteins, RNA functions in multiple distinct processes including reaction catalysis, as structural scaffolding for macromolecular complex assembly, and serves regulatory roles in nearly every pathway in the cell. In light of our ever-increasing understanding and characterization of RNA functions, it has been suggested that the majority of the flow and regulation of genetic information is actually accomplished via RNA, not protein, despite longstanding beliefs otherwise (Mattick and Makunin 2006; Morris and Mattick 2014). The numerous cellular functions of RNA are mirrored by the wide ranging diversity of RNA species that exist within the cell. RNA species can be broadly divided into protein-coding and noncoding RNAs (ncRNAs). This distinction is based on whether said RNA species translate the genetic language of nucleotides into the amino acid language of proteins (protein-coding RNA) or not (ncRNA). Though recent advances in technology have allowed for the identification of nontraditional protein-coding RNAs (Ji et al. 2015) distinct from canonical protein-coding messenger RNAs (mRNAs) (Yeasmin, Yada, and Akimitsu 2018), these are beyond the scope of this work and discussion of the protein-coding RNA species will be limited to mRNA.

Opposite to protein-coding mRNA, are the ncRNA species. ncRNA make up the vast majority of RNA in cells, and their complete identification and characterization is an ongoing endeavor. There is an abundance of distinct ncRNA species, each with a myriad of functions. Here, I will focus discussion on a select few, including small nuclear RNA (snRNA), ribosomal RNA (rRNA), and small nucleolar RNA (snoRNA).

Messenger RNA

mRNA is the main information courier that ferries genetic messages from DNA in the nucleus to cytoplasmic ribosomes where protein synthesis occurs. This journey throughout the cell is mired with multiple processing and surveillance steps to ensure that only high-quality, faithful, and translationally-competent mRNA molecules are translated into protein products. The transcription of mRNA is accomplished by the DNA-dependent RNA Polymerase II (Pol II), and accounts for roughly 5% of the total RNA in a cell. mRNA are transcribed initially as nascent molecules called pre-mRNA. These pre-mRNA contain untranslated regions in both their 5' and 3' termini (5' untranslated region (UTR) and 3' UTR, respectively), as well as an open reading frame (ORF) that contains protein-coding sequences called exons, as well as intervening non-coding sequences called introns. To form a fully mature mRNA, a pre-mRNA must undergo a series of co- and post-transcriptional regulatory processing events. Some of these events include 5' N7-methyl guanosine (m^7G) capping, intron excision via pre-mRNA splicing by the

spliceosome (discussed later), 3' cleavage and polyadenylation, and finally, nuclear export and ribosome engagement. Each of these mRNA processing events are complex and multifactorial, and can require a vast number of *trans*-acting factors, including proteins as well as other RNAs.

The 5' m^{7}G cap has multiple functions including protection of mRNA transcripts from 5' to 3' exonucleases, recruitment of RNA processing and export factors, and ribosome association (Bentley 2014; Huber et al. 1998; Mouaikel et al. 2002). This m^{7}G cap is installed co-transcriptionally by a series of guanyltransferases, triphosphatases, and methyltransferases. At the 3' end of mRNAs lies the polyadenylation signal (PAS) with the canonical sequence AAUAAA. The pre-mRNA transcript is cleaved almost immediately downstream of the PAS (10-30 nucleotides) by the cleavage and polyadenylation specific factor 73 (CPSF73) (Ryan, Calvo, and Manley 2004) nuclease to release it from the chromatin. Polyadenylation (polyA) then occurs post-transcriptionally through the actions of the PolyA Polymerase (PAP) and the PolyA binding protein Nuclear 1 (PABPN1). Dissociation of the PAP is mediated by the accumulation of PABPN1 binding on the polyA tail, which results in a polyA tail numbering roughly 250 nucleotides in most mRNA (Kühn et al. 2009). As mRNA is being capped and polyadenylated, it also associates with a variety of protein factors along its length, forming a messenger ribonucleoprotein (mRNP). Some of these associated proteins mediate export, in particular, the TRanscription-Export (TREX) complex, which recruits the nuclear RNA export factor 1 (NXF1) and nuclear transport factor

2 like export factor 1 (NXT1) mRNA export factors (Heath, Viphakone, and Wilson 2016). Importantly, all of these mRNA processing events, as well as transcription itself, are intimately intertwined. Splicing components are recruited by specific phosphorylations of the C-terminal domain (CTD) of Pol II (Herzel et al. 2017). TREX is loaded on mRNAs in a splicing-dependent manner (Masuda et al. 2005), and in addition to nuclear export, the TREX complex also functions in 3' end processing (Meinel et al. 2013). This interdependency allows multiple levels of regulation to ensure only completely assembled mRNPs are exported into the cytoplasm to engage the ribosome and participate in protein synthesis.

Small Nuclear RNA

Spliceosomal snRNAs interact with proteins to form small nuclear ribonucleoproteins (snRNPs), and play integral roles in pre-mRNA splicing. snRNAs are repetitive elements, in that there are multiple copies of the same gene throughout the genome. The number of functional copies for each snRNA gene varies, ranging from as little as 4 to greater than 15 copies per genome (Didychuk, Butcher, and Brow 2018; Guiro and Murphy 2017). The spliceosomal snRNAs consist of the U1, U2, U4, and U5 snRNAs, and are transcribed by Pol II. The U6 snRNA is transcribed by the DNA-dependent RNA Polymerase III (Pol III), and follows a separate regulatory and maturation process. The Pol II transcribed snRNAs are modified at the onset of transcription with the addition of an (m^7G) cap, similar to protein-coding mRNA. The (m^7G) cap is further recognized and

4

bound by the cap binding complex (CBC) which consists of the cap binding proteins CBP20 and CBP80. However, in contrast to mRNA, snRNA are intronless genes that are nonpolyadenylated. Lacking a PAS, snRNAs are cleaved at a conserved 3' box by the Integrator complex nuclease INTS11 (Wu et al. 2017). Interestingly, INTS11 is paralogous to the mRNA 3' PAS cleavage factor CPSF73. Following capping and transcription termination, snRNAs are exported through the nuclear pore through interaction of the CBC with the phosphorylated adaptor for RNA export (PHAX) and the export receptor Chromosome Region Maintenance 1 (CRM1) (Ohno et al. 2000), to form an snRNA specific export complex. It is likely that this export complex assembly occurs after the snRNA has traversed to the Cajal Body (CB) - a dynamic nonmembranous nuclear organelle. Previous experiments have shown that PHAX depletion causes export defects, accompanied by accumulation of U snRNAs in the CB (Suzuki, Izumi, and Ohno 2010).

Once the snRNA has been exported into the cytoplasm, it undergoes three major processing events before being reimported into the nucleus. First, the snRNA begins the process of snRNP assembly, which is mediated by both general and U-specific factors. snRNP assembly requires the Survival of Motor Neurons (SMN) complex, which facilitates the formation of a ring structure encircling the snRNA, consisting of the Sm proteins (D1, D2, D3, E, F, G, B/B') (Massenet et al. 2002). Sm B/B' dimerizes with Sm D3, and are added as a last step to form the fully assembled Sm ring. The Sm ring is a general snRNP component, and found

5

associated with all Pol II transcribed snRNAs. Second, the newly assembled snRNP undergoes 5' cap hypermethylation of the associated snRNA from an m^7G cap to a 2,2,7 trimethyl guanosine cap (m_3G) (Fischer, Englbrecht, and Chari 2011). This is accomplished by the enzyme trimethyl guanosine synthase (TGS1). TGS1 binds to the Sm ring through interaction with Sm B/B' (Mouaikel et al. 2003), thereby achieving a means of quality control and ensuring only snRNAs with a fully assembled Sm ring are hypermethylated. Third, the snRNA in the assembled snRNP undergoes 3' nucleolytic trimming (Kufel and Grzechnik 2019). As the snRNAs do not contain polyadenylated signals at their 3' end to direct a specific cleavage site following transcription termination, they are transcribed as pre-snRNA molecules containing an extended sequence at their 3' end. This sequence is trimmed by a 3' to 5' exonuclease, the identity of which is still elusive.

Following snRNP assembly in the cytoplasm, the snRNP is imported into the nucleus, to function in pre-mRNA splicing. However, the snRNP is still not yet fully matured, and must undergo further processing upon nuclear import in the CB. CBs are hubs for RNA processing, and it is here that the snRNA undergoes final packaging and folding (Staněk 2017), as well as chemical modifications like pseudouridylation of uridines and 2-O'-methylation of ribose rings, which are required for efficient snRNA functions in splicing. Furthermore, the U1, U2, U4, and U5 snRNPs associate with species-specific protein factors in the CB. The U6 snRNA follows a mechanistically similar pathway, with a few key differences. The U6 snRNA is not reported to be exported to the cytoplasm (Didychuk, Butcher, and

Brow 2018), and consequentially, all its maturation steps take place in the nucleus. On the 5' end, instead of an m^7G cap, the U6 snRNA is modified with a gamma-monomethyl phosphate (g-m-P3) cap (R. Singh and Reddy 1989). On the 3' end, the U6 snRNA contains an oligouridylated tail. The U6 snRNA does not make contact with the Sm proteins, but instead binds a set of U6-snRNA specific LSm (like Sm) proteins (Achsel et al. 1999). The LSm proteins, similar to the Sm proteins for the other U snRNAs, form a ring around the U6 snRNA to form the U6 snRNP. The U6 snRNA undergoes internal chemical modifications in the form of pseudouridylation and 2-O' methylation, though at this time it is unclear if the bulk of these modifications are installed in the CB or another nuclear structure, the nucleolus.

Small Nucleoar RNA

Small nucleolar RNA (snoRNA) are a class of RNA predominantly localized to the nucleolus that range in length from 60-300 nucleotides. Of the hundreds of distinct snoRNAs in the human genome, they can be roughly divided into two groups: box C/D or box H/ACA snoRNAs (Bertrand 2017), based on the eponymous C/D and H/ACA motifs contained within them. Generally, snoRNAs function as guide RNAs for associated enzymes to install chemical modifications on other RNA targets (Ojha, Malla, and Lyons 2020). These targets include both rRNA and snRNA, but emerging evidence also implicates mRNA being modified in a snoRNA dependent manner (Elliott et al. 2019; Beáta E. Jády and Kiss 2001;

Reichow et al. 2007). C/D snoRNAs associate with the small nuclear ribonucleoprotein 13 (SNU13), NOP56, NOP58, and the methyltransferase fibrillarin (FBL) to modify specific RNA nucleotides by 2'-O methylation of their ribose ring. The H/ACA snoRNAs associate with NHP2, GAR1, NOP10, and the pseudouridine synthase Dyskerin (DKC1) to isomerize specific uridines in RNA. This work will focus primarily on box C/D snoRNAs and their biogenesis and function (Ojha, Malla, and Lyons 2020).

The C/D snoRNAs contain highly conserved C box (RUGAUGA, where R is any purine) and D box (CUGA) motifs near their 5' and 3' ends, respectively. They also contain less conserved copies of the C and D boxes located internally, termed C' and D' boxes. Folding of the snoRNA is arranged so that the C and D boxes oppose each other in close physical proximity, creating a hairpin structure. The base of the hairpin is formed through the interaction of the C and D box, forming a bulged structure called a kink-turn, or K-turn. The internal regions of the C/D snoRNA immediately upstream of the D/D' boxes are complementary to the target RNAs, and the target nucleotide to be 2'-O methylated aligns exactly 5 nucleotides upstream of the D/D' box.

A vast majority of snoRNAs are located within the introns of mRNAs and long noncoding RNAs, and are released in a splicing dependent manner (Dieci, Preti, and Montanini 2009). Once the snoRNA-containing intron has been spliced, the intron lariat must be debranched and further processed for snoRNA biogenesis to continue. snoRNA biogenesis is somewhat related to snRNA biogenesis, as

snoRNAs are Pol II transcribed and lack a PAS at their 3' end. However, distinct from snRNAs, snoRNA biogenesis is entirely nuclear. 3' end processing of snoRNAs is still an area of active research, though it is thought that they undergo exonucleolytic trimming by the RNA exosome or other factors (Kufel and Grzechnik 2019; Son, Park, and Kim 2018). Binding of the C/D core components SNU13, NOP56/58, and FBL at the C/D boxes precludes the exosome from trimming past that point, and therefore, these proteins are required for snoRNA stability. In addition to intron-released snoRNA, there are few examples of snoRNAs existing as independent transcriptional units. These independently transcribed snoRNAs include SNORD3 and SNORD118 (Lestrade and Weber 2006), and are absolutely essential for viability. These snoRNAs are not dependent on splicing, and are processed slightly differently, especially at their 5' end. As they are Pol II derived transcripts, they contain a 5' m^7G cap. This cap is either removed or hypermethylated to form a m_3G cap.

Both intron-released and independently transcribed snoRNA are trafficked to the CB before entering the nucleolus to function in rRNA modification (Bertrand 2017). The C/D boxes are known to be sufficient to act as a CB localization signal, and it is in the CB that the C/D snoRNAs undergo final folding and processing. It is believed that the nucleolar-CB phosphoprotein 1 (NOLC1), which shuttles in between the two organelles, is responsible for localizing snoRNA to the nucleolus (Yang et al. 2000). Some snoRNA are retained in the CB due to the presence of additional sequence motifs, and are called small cajal body RNAs (scaRNAs).

9

scaRNAs generally function to enact chemical modifications on snRNAs in the CB, while C/D snoRNAs enact modifications on the rRNA in the nucleolus, though these classifications are not absolute (Deryusheva and Gall 2019).

Beyond chemical modification, many snoRNAs are either orphaned, in that they have no known RNA binding target and no prescribed function, or have functions independent of modification (Falaleeva and Stamm 2013). The most well studied of snoRNAs are the SNORD3 and SNORD118 genes, which function in rRNA cleavage and processing and contain unique sequence motifs beyond the canonical C/D boxes. These snoRNAs are lethal to cells if depleted, in contrast to the majority of 2'-O methylating snoRNAs (Bratkovič, Bozič, and Rogelj 2020). Loss of a single methylation modification on rRNA or snRNA may have translation or splicing defects, respectively, but it is the combined loss of multiple 2'-O methylations that is presumed to cause significant detriment to cell viability in humans.

Section 2: *Trans*-regulation of RNA

The life of an RNA is never a lonely one, as it is accompanied by various protein elements from the time of its transcription until its eventual degradation. *Trans*-regulation of RNA is a pervasive aspect of cellular biology, and is accomplished by both RNA binding proteins (RBPs)-which make direct contact with the RNA, and RNA adaptor proteins which use protein-protein interactions as part of an RNP to exert a regulatory effect on their cognate RNAs. The diversity of

these *trans*-acting protein factors is a direct reflection of the many cellular processes which they influence. However, one commonality shared by a vast number of RBPs and RNA adaptor proteins is that they are absolutely critical for accurate gene expression and cellular viability.

RNA Binding Proteins

There are approximately 1,500 human RBPs which function in a variety of cellular processes and pathways (Gerstberger, Hafner, and Tuschl 2014). RBPs can be added to an RNA at any point in the RNA life cycle, and the proteins associated with a given RNA are dynamically interchanged during RNA maturation, localization, and environmental context. RBP-RNA interaction can occur co-transcriptionally, or post transcriptionally. Adding further complexity, many RBPs can act in concert with other RBPs on a single RNA, or can be added sequentially, as in the case of an initial RBP modifying or remodeling the RNA in some way to allow recognition by an incoming RBP. The intricate interplay between RBPs and RNAs is a defining characteristic of an RNA and often directly responsible for proper function of that RNA. RBPs can act on, or with, their target RNAs. Different RNA species may interact with distinct RBPs. In the case of mRNA, the RBPs cap binding complex protein 20 (CBP20) and cap binding complex protein 80 (CBP80) bind early during transcription and help to stabilize the mRNA to prevent decapping and exonuclease degradation from its 5' end (Gonatopoulos-Pournatzis and Cowling 2014). The CBC20/80 complex also

influences splicing and export of mRNA by recruiting other RBPs involved in both of these processes. In contrast, ribosomal proteins interact not with mRNA but instead with rRNA, but only post-transcriptionally, after ribosomal RNA has undergone multiple processing steps. Like rRNA, snoRNAs associate with specific RBPs, and form RNPs called snoRNPs. These snoRNPs consist of a particular snoRNA in complex with multiple RBPs (McKeegan et al. 2007), and can guide RBP enzymes like the methyltransferase Fibrillarin (FBL) to a target RNA (mediated by snoRNA sequence complementarity) to attach a methyl group to the ribose ring. snRNAs have a large set of RBPs associated with them, and when complexed together, are termed snRNPs. The snRNPs associated with the canonical spliceosome are the U1, U2, U4, U5, and U6 snRNP, and contain an snRNA molecule bound to multiple RBPs and RNA adaptor proteins that collectively function as a vital component of the pre-mRNA splicing reaction. snRNAs bind to both general RBP snRNA factors (Gemins, Sm proteins) (R. Zhang et al. 2011) and complex-specific RBPs like U1-70K (Kondo et al. 2015) for the U1 snRNP and SF3A1 for the U2 snRNP (Martelly et al. 2019), among others.

RBPs can make direct contact with either the nitrogenous base, the sugar ring, or the phosphate backbone of RNA. In this way, RBPs can recognize particular RNA interaction partners through sequence specificity or structural motifs, or they can bind promiscuously to multiple different RNA targets. RBPs achieve this binding through their RNA binding domains (RBDs). Most RBPs have multiple RBDs, which can help increase the specificity, number, and affinity of the

RBP for RNA. These RBDs have an ordered structure, and a majority of them retain evolutionarily conserved amino acid sequence or structure, indicating their importance throughout multiple domains of life. RNA recognition motifs (RRMs), Cold shock domains (CSDs), Zinc Fingers CCHC domains, K homology (KH) domains, S1 motifs, and Sm motifs are some of the most common of the structured canonical RBDs (Lunde, Moore, and Varani 2007). They differ in size (ranging between 30-100 amino acids, roughly) mechanism of RNA contact, and affinity and specificity of binding.

Somewhat recently, it was realized that RBPs existed that lacked canonical RBDs. Instead, these rogue RBPs contained low-complexity intrinsically disordered regions (IDRs) that they used to directly bind RNA. Some reports estimate that at least half of all human RBDs are IDRs (Castello et al. 2016). As the name implies, IDRs contain no definitive structure. However, upon binding of RNA, some IDRs exhibit a rigid conformation that only exists when in complex with their RNA interaction partner. IDRs are enriched for "disorder-promoting" amino acids, including arginine (R), glycine (G), serine (S), and lysine (K), and depleted of aromatic residues. Many IDRs are repetitive, containing multiple iterations of RS and RGG repeats that enhance their overall disordered character. K-rich or R-rich patches are also among the most frequently observed RNA binding surfaces in IDRs. Finally, IDRs lacking all of these repetitive disorder characteristics have also been observed to be RBDs.

It is also understood that IDRs exist in RBPs with multiple ordered RBDs, sometimes inhabiting the intervening spaces between these RBDs (Ottoz and Berchowitz 2020). These intervening IDRs act as a flexible linker to orient the RBPs in a proper binding position. Alternatively, these linker IDRs also serve to increase affinity of the ordered RBP to RNA by mitigating the resultant decrease in entropy upon binding. Because of their prevalence both as bona fide RBDs and for their role in supporting RNA binding of canonical RBDs, IDRs are recognized as a major player in *trans*-regulation of RNA.

RNA adaptor proteins and ribonucleoprotein complexes

The simplest RNP is a single RBP complexed with a single RNA molecule. Quite frequently, however, RNPs are multimeric and consist of multiple RBPs in complex with RNA adaptor proteins. These RNA adaptor proteins reside in the RNP through reliance on protein-protein interactions with either RBPs or other adaptor proteins, and are independent of RNA binding. RNA adaptor proteins can serve a variety of functions within the RNP, including complex stability, recruitment of RNA-modifying enzymes or other RBPs, or they can act as scaffolding components to bridge multiple RNPs together in close proximity.

PHF5A, Gemin2, and NUFIP1 proteins are examples of RNA adaptor proteins. The PHF5A gene encodes a protein that plays an RNA adaptor role in the spliceosomal U2 snRNP complex (Z. Wang et al. 2019). PHF5A interacts with U2 snRNP components independent of RNA and helps bridge U2 snRNP factors

with other spliceosome components to initiate pre-mRNA splicing. The SMN complex consists of the SMN protein, Gemin proteins 2-8, and the Unrip protein. The SMN complex recruits snRNAs through the RBP Gemin5, to mediate association of the snRNA with the Sm proteins, which are present in the mature snRNP. Gemin2 binds and locks the Sm complex conformation in a restrictive state (R. Zhang et al. 2011), where only specific RNAs (snRNAs) can then associate. The eventual association of the snRNA with the Sm complex causes a conformational change, which decreases affinity of the Sm complex for Gemin2 and results in its release from the RNP. In this regard, Gemin2 acts as an RNA adaptor protein that participates in negative cooperativity with target snRNAs to ensure their proper maturation and association. RBPs can function as RNA adaptor proteins in different contexts. For example, nuclear FMRP interacting protein 1 (NUFIP1) is an RBP involved in multiple processes, including pre-snoRNA biogenesis. However, NUFIP1 is predicted to bridge core snoRNA proteins together using protein-protein interactions (Dupuis-Sandoval, Poirier, and Scott 2015), instead of its RNA binding capability.

RNA adaptor proteins represent a dynamic, diverse, and indirect mechanism of *trans*-regulation of RNA. While it is imperative to categorize the direct RNA targets for individual RBPs, it is also equally important to understand the contribution of all members of an RNP, as it is common that these adaptor proteins may modify or regulate RBP characteristics to promote RNA target

binding. The elaborate cooperation between RBPs and RNA adaptor proteins in RNPs is a fundamental aspect of RNA biology.

Section 3: Pre-mRNA Splicing

Protein-coding genes are not transcribed as mature messages. Instead, the nascent pre-mRNA transcript is synthesized as a mosaic, with protein-coding sequences interspersed with noncoding sequences. Nearly all protein-coding genes are transcribed in this manner, with only a few notable exceptions (Grzybowska 2012). To form a mature mRNA from a nascent pre-mRNA, the intervening noncoding sequences must be removed, in a process called pre-mRNA splicing. The process of splicing is highly dynamic and complex and involves a copious assortment of both *cis* characteristics and *trans*-acting factors (De Conti, Baralle, and Buratti 2013; Will and Lührmann 2011). *Cis*-characteristics are contained within the pre-mRNA itself, while *trans*-acting factors include other proteins or RNPs, the most important of which is the multi-megadalton cellular machine called the major spliceosome, hereafter referred to simply as the spliceosome. The spliceosome consists of protein arrays complexed with specific U snRNAs, called U small nuclear ribonucleoproteins (U snRNPs) (Matera and Wang 2014). The coordinated interaction of the U snRNPs, along with a host of splicing auxiliary factors, ensures the timely and efficient splicing of pre-mRNA

transcripts to mature mRNA transcripts, and is crucial to proper regulation of gene expression.

Mechanism of Pre-mRNA Splicing

Pre-mRNA are transcribed as a mosaic, consisting of exonic coding sequences interspaced with intronic noncoding sequences. To form mature mRNA capable of translation by the ribosome, pre-mRNA must undergo the process of splicing, which removes the intronic regions from the transcript and ligates together the exonic regions, thereby ensuring translational capacity.

This process of splicing introns from pre-mRNA is accomplished via a stepwise pathway involving the large RNP called the spliceosome (**Figure 1.1**). The spliceosome is a massive macromolecular machine central to the process of pre-mRNA splicing. It is rivaled in size by the ribosome, with both the spliceosome and the ribosome being multi megadalton molecular complexes. However, unlike the ribosome, which contains a definitive mature structure, the spliceosome exists as a dynamic entity made up of several U snRNP complexes (U snRNPs) (Y. Zhang et al. 2021). The conformation, interaction, consistency, and presence of these U snRNPs changes as the splicing reaction proceeds, lending the spliceosome no definitive structure nor mature form. The dynamic and transient nature of the spliceosome is necessary for the completion of the highly complex splicing reaction. In total, throughout the complete splicing reaction, the

spliceosome forms several distinct complexes throughout the splicing reaction, including the E, A, B, Bact, B*, C, C*, and P complexes (Li et al. 2020).
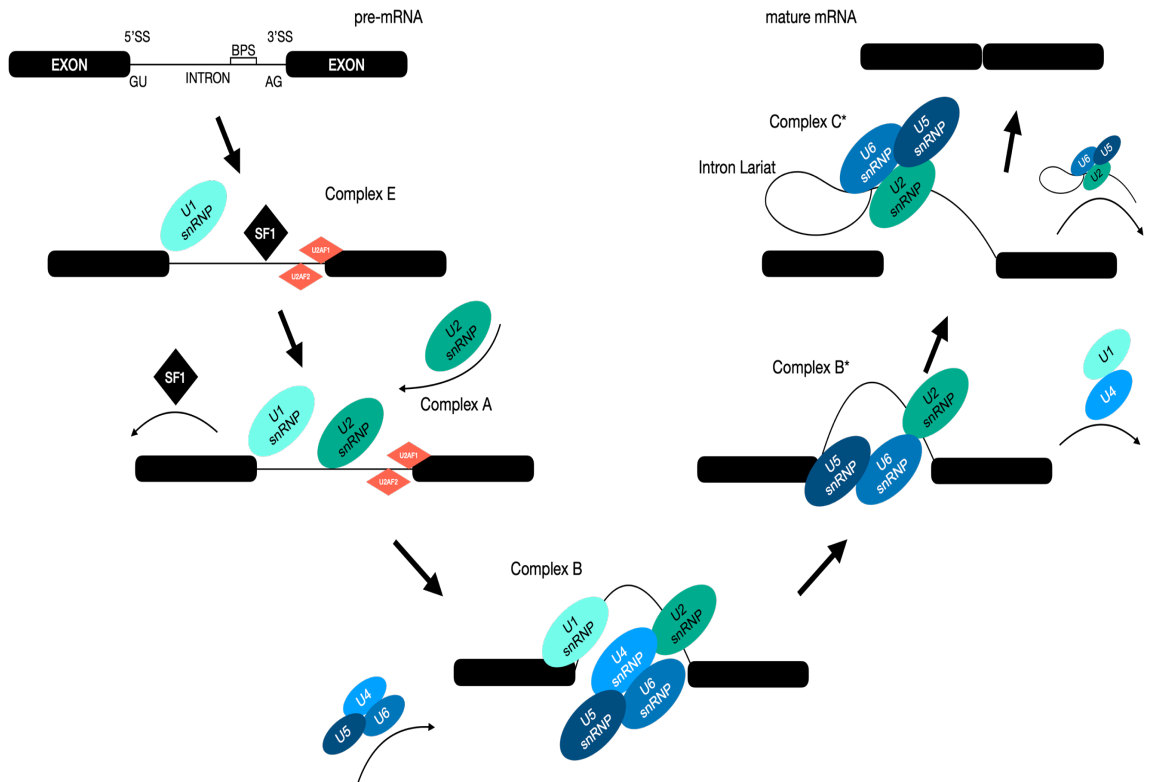
Figure 1.1 Overview of pre-mRNA splicing by the spliceosome. Pre-mRNA contain a GU at the 5'SS and an AG at the 3'SS, as well as a BPS located 19-35 nucleotides upstream of the 3'SS. The spliceosome undergoes a series of rearrangements and factor exchanges, progressing from the E complex to the P complex (not shown here). The two transesterification reactions of pre-mRNA splicing are completed by the B* complex and the C* complex.

Each U snRNP of the spliceosome contains an eponymous snRNA, and as such, the spliceosome consists of the U1, U2, U4, U5, and U6 snRNPs. Each U snRNP contains a set of proteins common to all, as well as U snRNA-specific

proteins that only associate with a single U snRNP. Together, these U snRNPs interact to identify introns and exons and catalyze two sequential transesterification reactions that complete the removal of an intron and the subsequent ligation of adjacent exons.

In humans, pre-mRNA splicing begins with the definition of exons within a nascent pre-mRNA transcript. This is achieved by the spliceosomal recognition of specific sequences within the pre-mRNA transcript, including the 5' splice site (5'SS), the 3' splice site (3'SS), and the branch point sequence (BPS). The 5'SS marks the 5' boundary of an intron, the 3'SS likewise marks the 3' boundary of that same intron. Roughly 15-50 nucleotides upstream of the 3'SS lies a conserved adenosine nucleotide, surrounded by the branch point sequence. The first step in splicing requires formation of the early spliceosome, or E complex. This occurs as the U1 snRNP recognizes and binds the 5'SS and surrounding sequences through complementary base pairing, and the BPS is bound by the auxiliary protein Splicing Factor 1 (SF1) (Matera and Wang 2014). The 3'SS is defined by the U2 auxilliary factor complex (U2AF) heterodimer, which consists of U2AF1 binding to the 3'SS, and U2AF2 binding to a CT-rich region immediately upstream of the 3'SS called the polypyrimidine tract. The next step in pre-mRNA splicing is the conversion of Complex E to the pre-spliceosome, called complex A. Complex A retains U1 snRNP at the 5'SS but exchanges SF1 at the BPS with U2 snRNP, which binds through base pairing similar to the U1 snRNP-5'SS interaction. The

pairing of the BPS with U2 snRNP creates a duplex with the conserved adenosine in the BPS bulged out, in preparation for the next steps of the splicing reaction.

After the A complex is formed, the pre-assembled tri-snRNP, consisting of the U4, U5, and U6 snRNPs, with U4/U6 base paired together , is recruited to the site of splicing to form complex B. The U snRNPs then undergo extensive conformation shifts, disrupting and forming new protein-protein and protein-RNA bonds, as well as the recruitment of multiple auxiliary splicing complexes (RES, NTC, NTR) (Wan, Bai, and Shi 2019). This massive shift in conformation and interaction causes the release of both the U1 and U4 snRNPs, with the U6 snrNP base pairing to both the 5'SS and with U2 snRNP, resulting in the B activated (Bact) complex. The U2-U6 snRNA interaction is thought to be the catalytic center of the spliceosome that is responsible for the transesterification reactions of pre-mRNA splicing (Mefford and Staley 2009), but is inhibited from acting by the presence of the U2 snRNP SF3A and SF3B complexes. Both of these complexes are destabilized by the activity of the DHX16 RNA helicase, after which the B* complex forms and the first transesterification reaction is completed (Gencheva et al. 2010). In this transesterification, the BPS adenosine, bulged out from the BPS-U2 snRNA base pairing, carries out a nucleophilic attack on the 5'SS. This results in a free upstream exon, and a looped intron, called a lariat, attached to the downstream exon, altogether forming spliceosomal complex C. In complex C, the upstream 5' exon is retained in close proximity to the active site of the catalytic center of the U2/U6 interaction by base pairing with U5 snRNA. The C complex

then undergoes further conformational changes eventually culminating in the active C* complex, which completes the second transesterification reaction through nucleophilic attack of the 5' exon on the 3' exon, thereby ligating the two exons together. This forms the final spliceosomal stage, called the P complex, which consists of the ligated exons, the spliceosome, and the still attached intron lariat. The intron lariat is released and the spliceosomal complexes are recycled for future splicing reactions through the helicase activity of DHX15 (Blijlevens, Li, and van Beusechem 2021). Altogether, the sequential assembly of various spliceosome complexes on the pre-mRNA is required for the identification and removal of introns from the pre-mRNA.

Exon Definition

In mammals, introns are generally much larger (>1000 nt (Hong, Scofield, and Lynch 2006)) compared to exons (~150nt (De Conti, Baralle, and Buratti 2013)). The large size of an intron presents a challenge for spliceosomal formation, not only due to the extreme distance over which U snRNPs must interact, but also due to the presence of cryptic splice sites which lie in the intronic sequence separating the canonical 5' and 3'SSs. These cryptic splice sites may contain the invariant GU or AG dinucleotide, but usually have surrounding sequence with lower complementarity to U snRNAs than canonical splice sites. To rectify this, mammalian cells form an exon definition complex (EDC) (De Conti, Baralle, and Buratti 2013). In other words, while the splicing reaction take place across an

intron, the spliceosome forms across an exon. The exon definition complex is defined through the interaction between components of two different spliceosomal A complexes on separate introns. The U2 snRNP, bound at the BPS and near the 3'SS of the upstream intron, also interacts with the U1 snRNP, which is bound at the 5'SS of the downstream intron, forming a cross-exon complex. The switch from this cross-exon complex to a cross-intron complex is still under investigation, though recent work (Li et al. 2020) has suggested that U4/U6/U5 tri-snRNP recruitment forces a conformation change in the EDC that is sterically unfavorable, and induces an interaction between an upstream U1 snRNP at the 5'SS and a downstream U2 snRNP at the 3'SS. In this way, splice sites influence each other. Indeed, it has been observed that a weak 3'SS is more likely to be recognized when followed by a strong 5'SS across the downstream exon. Importantly, while exon definition is the predicted predominant method of spliceosomal formation in mammals (De Conti, Baralle, and Buratti 2013), intron definition may occur when the distances across the intron are minimal. Pre-mRNA splicing utilizing intron definition follows the same mechanistic process as exon-defined splicing, except the U1 snRNP and U2 snRNP bound to specific locations in the same intron recognize and interact with each other, instead of a cross-exon interaction occurring first.

Alternative Splicing

The human genome consists of roughly 20,000 genes, yet the human proteome is over 10 times greater in number. One of the theories put forth to explain this expansion of the human proteome from a limited number of genes is the alternative splicing (AS) of pre-mRNA. That is, the differential inclusion of introns and exons in a single transcript can produce multiple protein isoforms, which can harbor distinct domains or motifs that can alter their subcellular localization, binding partners, or function. Almost 95% of all human genes are predicted to undergo AS (Jiang and Chen 2021). This prevalence underscores its physiological importance to proper tissue function and development, and its deregulation is associated with many genetic diseases. In fact, one third of all disease-causing genetic mutations are thought to result in splicing defects, highlighting an urgent need for further research.

Human splice site sequences are not well conserved, with the exception of an almost invariant GU and AG dinucleotide sequence, as well as a conserved adenosine, located at the 5' and 3'SS and the BPS, respectively (Burset, Seledtsov, and Solovyev 2000). The sequences surrounding these dinculeotides are degenerate, but contributes to the overall splice site strength, with stronger splice sites containing greater complementarity to the U snRNAs that bind them and being more likely to promote intron splicing. Weaker splice sites are less likely to be recognized by the snRNAs of the spliceosomal machinery (Jian, Boerwinkle, and Liu 2014), and therefore are more likely to be associated with retained introns

that are not removed from the pre-mRNA. Beyond splice site strength, there are other *cis*-contained motifs that contribute to increasing or decreasing the likelihood of intron splicing at a specific locus. These cis-motifs are called splicing regulatory elements (SREs), and their specific sequence or structure can be recognized by either individual, or multiple *trans*-acting splicing auxiliary factors (Ramanouskaya and Grinev 2017). These auxiliary factors are either RBPs that bind the SRE directly, or RNA adaptor proteins that function in complex with RBPs to bind the SRE. While the core components of the human spliceosome, namely the U snRNPs, are responsible for the catalysis of the reactions to remove introns from pre-mRNA, it is the SREs and their splicing auxiliary factor interactors that influence the spliceosome to either promote or repress splicing of specific introns, as well as the decision to include or skip exons.

SREs can be located in both intronic and exonic regions, and with their cognate auxiliary factor, can increase or decrease the likelihood of intron splicing. SREs that enhance intron splicing are called exon splicing enhancers (ESEs) if located in an exon, and intron splicing enhancers (ISEs) when located in an intron. Similarly, SREs that repress intron splicing are called exon splicing silencers (ESSs) when located in an exon, and intron splicing silencers (ISSs) when located in an intron. Both the SR protein family and the hnRNP protein family are splicing auxiliary factors and bind SREs to regulate intron exclusion/inclusion (Zhou and Fu 2013), with SR proteins usually binding enhancer elements and promoting splicing, and hnRNP proteins binding silencer elements and repressing splicing,

though there are many exceptions (Shin and Manley 2002; Vu et al. 2013). Adding to the complexity, SREs have a positional bias, effectively enhancing or repressing intron splicing/exon inclusion on the basis of their position relative to the splice sites (Ke and Chasin 2011). For example, NOVA2 is a neuron-specific splicing factor that binds the SRE motif with the sequence YCAY, where Y is any pyrimidine. The NOVA2-bound YCAY motif can act as a splicing enhancer and promote exon inclusion when located within a downstream intron, but can promote exon skipping when located within the exon or immediately upstream of the exon (Ule et al. 2006). A single SRE can also be bound by different auxiliary factors and thus result in differential effects on exon/intron splicing, as is the case for exon P3A of the CHRNA1 gene. CHRNA1 is bound at an ESS by hnRNPL to repress inclusion of exon P3A, but a single nucleotide mutation in this ESS disrupts this binding, instead recruiting the related hnRNPLL factor. hnRNPLL acts as a splicing activator and causes the spliceosome to favor exon P3A inclusion (Rahman et al. 2013). Exon P3A inclusion in the CHRNA1 gene can cause a subtype of congenital myasthenic syndromes, which are marked by high early childhood mortality.

Beyond the myriad of examples listed above there exists a panoply of RBPs that can influence splicing decision through specific SRE interactions. The mechanisms by which SRE recruitment of splicing auxiliary factors influences spliceosomal decision are complex and multifaceted, with both direct and indirect modes of regulation. Most SREs lie adjacent to splice sites or the BPS, and this proximity allows bound auxiliary factors to make direct contact with spliceosomal

components. TIA1 is an RBP and splicing auxiliary factor, which can bind SREs near the 5'SS and recruit the U1 snRNP for 5'SS recognition by protein-protein interaction with the U1 snRNP component SNRPC (Förch et al. 2002). Similarly, but directionally opposite, the RBP polypyrimidine tract binding protein (PTBP1) can inhibit spliceosome formation and repress splicing by binding to the polypyrimidine tract near the BPS and 3'SS, displacing the U2AF complex and preventing U2 snRNP interaction with the BPS (Sharma et al. 2008). These examples show the importance of SRE-bound auxiliary factors and their direct interaction with core spliceosomal components. Indirectly, auxiliary factors can regulate splicing by inducing structural changes, like hnRNPA1 binding in duplicate SREs surrounding an exon, forming a "looped out," structure that promotes exon skipping (Nasim et al. 2002). Finally, though most SREs are in relative proximity to their respective splice sites, it has been shown that long distance interactions can modulate splicing (N. N. Singh et al. 2013).

Alternative splicing relies on the combinatoric inputs from a variety of *cis* and *trans*-acting factors. The cumulative effect of these signals ultimately results in a binary decision to recognize the 5' or 3'SS or the BPS. This recognition is dependent on the strength of the splice sites, the binding of auxiliary factors to SREs, the interaction of the spliceosome with transcription elongation complexes, mainly RNA polymerase II, and the local chromatin and epigenetic environment (see Chapter 4). Following recognition of the splice sites, the spliceosome must form and carry out its transesterification reactions, which are also subject to similar

regulatory constraints. The many layers of regulation regarding the splicing of pre-mRNA serve not only as a means of quality control, but also facilitate the synthesis of diversified mRNA isoforms.

Intron Retention

AS produces a range of different isoforms from a single transcript, and this isoform diversity is achieved through a variety of AS events. AS events can be classified as skipped/cassette exons (CE), mutually exclusive exons (MXE), alternative donor (AD) and alternative acceptor (AA) sites, alternative transcriptional start (TS) and polyadenylation (TE) sites, and finally, intron retention (RI) (**Figure 1.2**). While CE is the most prevalent AS event in humans (Zhiguo, Wang, and Zhou 2013), likely due to exon definition instead of intron definition, RI is more prevalent in lower eukaryotes (Jacob and Smith 2017a), and has more recently been recognized as a critical regulator of gene expression and transcriptomic integrity in humans.

RI has a diverse range of functional consequences for genes, though it is overall inversely correlated with gene expression levels. Classically, RI is known to promote nonsense-mediated decay (NMD) by introduction of a premature termination codon (PTC) into the mRNA transcript (Jacob and Smith 2017a). Briefly, NMD is triggered by ribosomal pausing on termination codons that are physically distanced from the polyA+ tail and polyA+ tail-associated proteins, of an mRNA transcript (Silva et al. 2008). RI that introduce a frameshift or directly

encode a termination codon in the coding sequence (CDS) of the transcript initiate degradation of the PTC-containing mRNA after initial ribosome engagement. Similarly, RI can also induce transcript turnover in the nucleus in the absence of a PTC, in which nucleases of the RNA exosome are required for degradation (Jacob and Smith 2017a). In this way, transcript degradation can be induced by RI in both the nuclear and cytoplasmic components by different surveillance mechanisms.
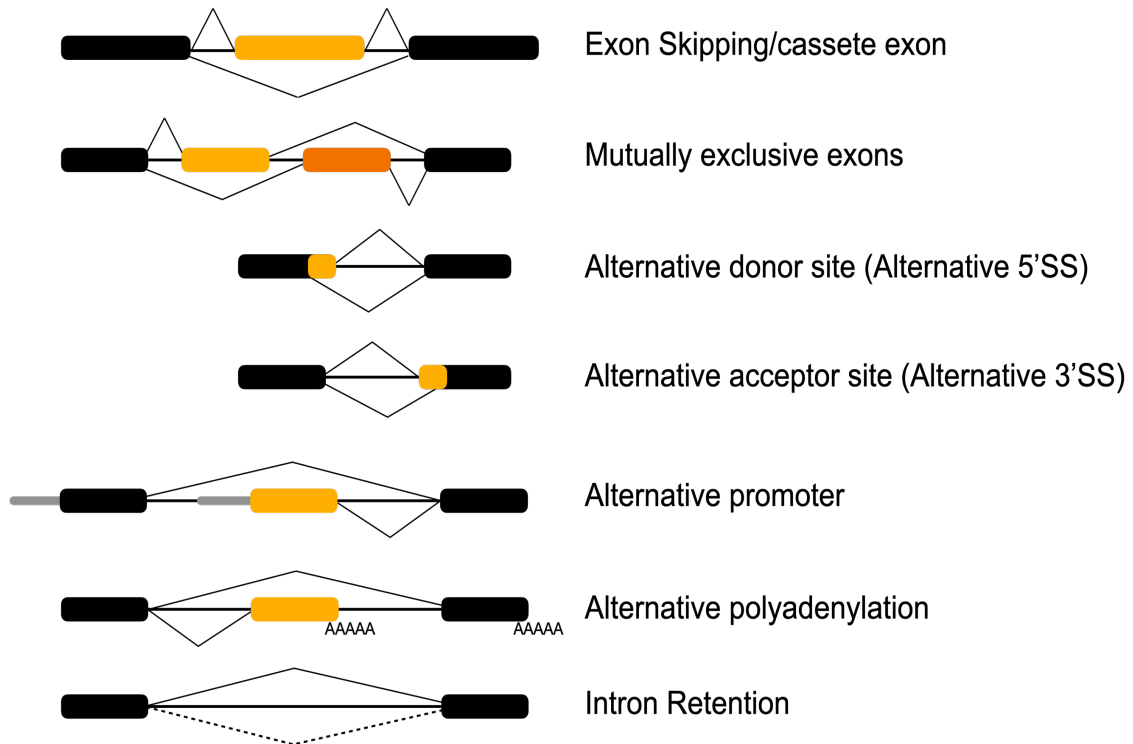
Figure 1.2 Categories of AS. Exons are represented as black, light orange, or dark orange boxes. Introns are represented as horizontal black lines connecting exons. AS patterns are represented by diagonal lines interconnecting different exon or intron elements. Exon skipping (CE), Mutually exclusive exons (MXE), Alternative donor site (AD), Alternative acceptor site (AA), Alternative promoter (TS), Alternative polyadenylation site (TE), Intron Retention (RI).

Though RI is negatively correlated with gene expression, its regulation on the transcriptome is far more nuanced. Gene expression can be regulated by RI through a separate mechanism known as intron detention (ID), in which a single intron in a transcript is retained, causing nuclear retention of the transcript, but avoiding exosome-mediated decay (Boutz, Bhutkar, and Sharp 2015). ID can occur at specific physiological stages or in response to environmental cues. Upon

post-transcriptional splicing of the ID transcripts, they are exported to the cytoplasm and translated in a rapid manner, far quicker than *de novo* transcription and splicing of nascent transcripts could be achieved. This allows the cell to modulate rapid changes in the transcriptome.

Interestingly, introns are still active regulatory effectors post-splicing and release from the transcript. For example, introns can contain entire short RNAs within their sequence, including snoRNAs (Hirose and Steitz 2001) and miRNAs (Lin, Miller, and Ying 2006), which are released upon processing of the intron lariat to perform their respective biological functions. Following   release   from   the transcript, the intron lariat can be processed into a small miRNA to mediate targeted degradation of other mRNA transcripts. snoRNAs are released in a similar manner, and can function in translation fidelity and accuracy by modifying rRNA. Moreover, some introns harbor microRNA (miRNA) binding sites (Xun et al. 2019), which may regulate expression of the host gene, or may confer a sponging effect, sequestering miRNAs from other cellular targets. The efficient and specific splicing of introns therefore enforces another layer of regulation upon gene expression, and their retention can alter the transcriptome in a variety of ways.

Due to their many regulatory roles, introns are a potent mediator of physiological processes like organismal development and cell type specification. RI increases as differentiation proceeds, as shown when comparing murine embryonic stem cells with progressive differentiated stages of glutamatergic neurons (Braunschweig et al. 2014). In the fully differentiated neurons, which

contained the highest levels of RI, the levels of gene expression negatively correlated with the presence of RI, as expected. Importantly, the genes that contained RI in the fully differentiated neurons were enriched for cell cycle-related genes as well as genes involved in non-neuronal biology. Likewise, late-stage human erythroblasts utilize RI as a regulatory mechanism for gene expression, and human granulocyte differentiation and morphology is controlled through orchestrated RI events (Wong et al. 2013).

Splicing Pathologies

The full breadth of splicing-related pathologies is numerous and varied, and its complete annotation is beyond the scope of this work. Genetic mutations in the spliceosomal snRNAs, or any of their core protein components, can cause global splicing errors. Similarly, mutations in the genes of splicing auxiliary factor RBPs or adaptor proteins, as well as mutations in splicing elements (5'SS, 3'SS, BPS) of mRNA can cause mis-splicing. This mis-splicing defect can either be widespread or gene-specific, depending on the mutation location, resulting in cellular malfunction. Indeed, with such an expansive repertoire of factors, it is not surprising that one third of all disease-causing genetic mutations are involved in pre-mRNA splicing (Havens, Duelli, and Hastings 2013). One of these splicing pathologies is discussed briefly below.

Spinal Muscular Atrophy (SMA) is caused by a deletion or loss-of-function mutation in the SMN1 gene, required for snRNA maturation and spliceosomal

assembly. SMA is an autosomal recessive disease that results in the destruction of motor neurons, and is marked by death in infancy or early childhood. The clinical severity of SMA depends on the number of copies of the SMN2 gene, which is identical to SMN1 except for a single nucleotide variation in SMN2, which leads to alternative splicing and exclusion of exon 7. The loss of exon 7 from SMN2 greatly decreases the stability of the resulting protein, making SMN2 an insufficient source of SMN protein (Hua et al. 2007). Currently, one of the therapeutics available to treat SMA is an antisense oligonucleotide (ASO) that targets an ISS in the preceding intron immediately upstream of exon 7 (I. Chen 2019). This site is bound by splicing repressor proteins hnRNPA1/A2 in the absence of the targeted ASO drug, but this binding is inhibited in the presence of the ASO drug, thereby leading to derepression of exon 7 inclusion. Here, both the cause and treatment of SMA rely on understanding the local splicing environment, and it is likely many other diseases can be treated in a similar fashion. Indeed, diseases like cystic fibrosis and Duchenne muscular dystrophy can be caused by splice-altering mutations in the CFTR (Faà et al. 2009) and DMD (ADACHI 2003) gene, respectively.

Section 4: Ribosome Biogenesis

Ribosomes are large macromolecular complexes that are responsible for protein synthesis in all living cells. A single human cell may contain as many as 10 million ribosomes, which consequently make up 50% of cellular dry mass. The

enormous energy requirements of ribosome production underscore their critical function, precise regulation, and overarching governance of cell fate. Ribosomes are comprised of both protein and RNA, and in humans, require action from all of the RNA polymerases (Pol I, II, III) in their biogenesis.

Ribosomal DNA and the Nucleolus

rRNA makes up more than half the mass of the ribosome and as such, has a high transcriptional requirement. Mature rRNA is comprised of four genes: 5.8S, 18S, 28S, and 5S. Three of these genes (5.8S, 18S, 28S) are transcribed from ribosomal DNA (rDNA) which is contained in massive tandemly repeated arrays called nucleolar organizer regions (NORs) on the short arm of the acrocentric chromosomes 13, 14, 15, 21, 22, while the 5S gene is transcribed from an array on chromosome 1. Within these NORs, hundreds of repeated rRNA genes are contained, and often colocalize to form the nucleolus by liquid-liquid phase separation (Lafontaine et al. 2021). The nucleolus is the most prominent subnuclear structure, and is primarily associated with multiple stages of ribosome assembly. It is separated into distinct regions termed the Fibrillar Center (FC), the Dense Fibrillar Component (DFC), and the Granular Component (GC). Transcription of rDNA repeats occurs at the boundary of the FC and the DFC. rDNA in the nucleolus is transcribed by Pol I, but also requires the upstream binding factor (UBF) and the selectivity factor 1 (SL1) complex (Grummt 2003). Interestingly, despite the excessive requirement for rRNA by the ribosomes, up to

half of the rDNA repeats are not transcriptionally active, and may be utilized as templates for repair of actively transcribed rDNA genes, should DNA damage arise (Moss et al. 2019).

Ribosomal RNA

Ribosomes consist of a small and a large subunit that bridge together on an mRNA transcript for translation. Ribosome biogenesis begins with the transcription of rRNA in the nucleolus. The 18S, 5.8S, and 28S gene are transcribed as a single polycistronic transcriptional unit termed 47S precursor. 47S precursor contains both external (5' ETS, 3'ETS) and internal (ITS1, ITS2) transcribed spacers separating and surrounding the rRNA genes. To generate mature rRNAs, 47s precursor interacts with a massive assemblage of *trans*-acting factors that guide the maturation process but are not components of the functional ribosome. These *trans*-acting factors include nucleases, modification enzymes, RNA helicases, snoRNAs, and molecular chaperones, among others. They function in the rRNA maturation process that consists of a series of endonucleolytic cleavages and exonucleolytic trimmings that generate a range of intermediate pre-rRNA, eventually giving rise to the mature forms of the rRNA genes. The rRNA maturation pathway, while generally understood, is still missing specific nucleases for cleavage and trimming events, as well as a complete understanding of the kinetics of each event, and cellular or environmental context for the use of alternative rRNA maturation pathways (Henras et al. 2015). A general rRNA maturation pathway

(Aubert et al. 2018) is shown in (**Figure 1.3**), and described in greater detail here: the Small Subunit Processosome (SSU processome), also called the 90S pre-ribosome, associates with the 47S precursor to mediate the first cleavages occur at sites A' in the 5'ETS and at site 02 in the 3'ETS. This yields the 41S precursor, which is devoid of its most terminal 5'ETS and 3'ETS sequences. Following this, the SSU processosome, which consists of the SNORD3 snoRNP (formerly called U3 snRNP) and a number of other multiprotein complexes including the U three protein (UTP) complexes (Sloan et al. 2014), mediates cleavage of multiple sites in the remaining 5' ETS sequence (A0, 1), through coordinated RNA:RNA interaction of SNORD3 with sequences in the 5'ETS and 18S rRNA. Next, the ribozyme RMRP cleaves the ITS1 and site 2, separating the future ribosomal small subunit (18S) from the future large subunit (5.8S/28S), and each subunit undergoes different processing pathways independent of each other. The sequentiality of site 2 cleavage by RMRP is not absolute, and small subunit and large subunit particles can also be split by cleavage at site E, by the SSU processosome endonuclease UTP24. Further cleavages and trimmings by the nuclear exosome and unknown nucleases results in the biogenesis of the mature 18S rRNA, the RNA component of the small subunit of the ribosome. To generate the mature 5.8S and 28S genes, site 4 is cleaved immediately after the release and separation of the 18S pre-rRNA and splits the 5.8S and 28S pre-rRNA genes from one another. The maturation of 5.8S and 28S requires SNORD118 binding (J. Langhendries et al. 2016), among many other rRNA processing factors. 5.8S

rRNA is generated following several more cleavage and trimming steps, generating the 12S, then the 7S pre-rRNA, eventually culminating in the mature 5.8S rRNA. After separation from the 5.8S pre-rRNA, the remaining 28.5S pre-rRNA undergoes exonucleolytic trimming by the nuclease XRN2 to the mature 28S rRNA (M. Wang and Pestov 2011). Similarly, the Pol III transcribed 5S pre-rRNA undergoes 3' nucleolytic trimming to form the mature 5S rRNA which is eventually incorporated into the large subunit of the ribosome.
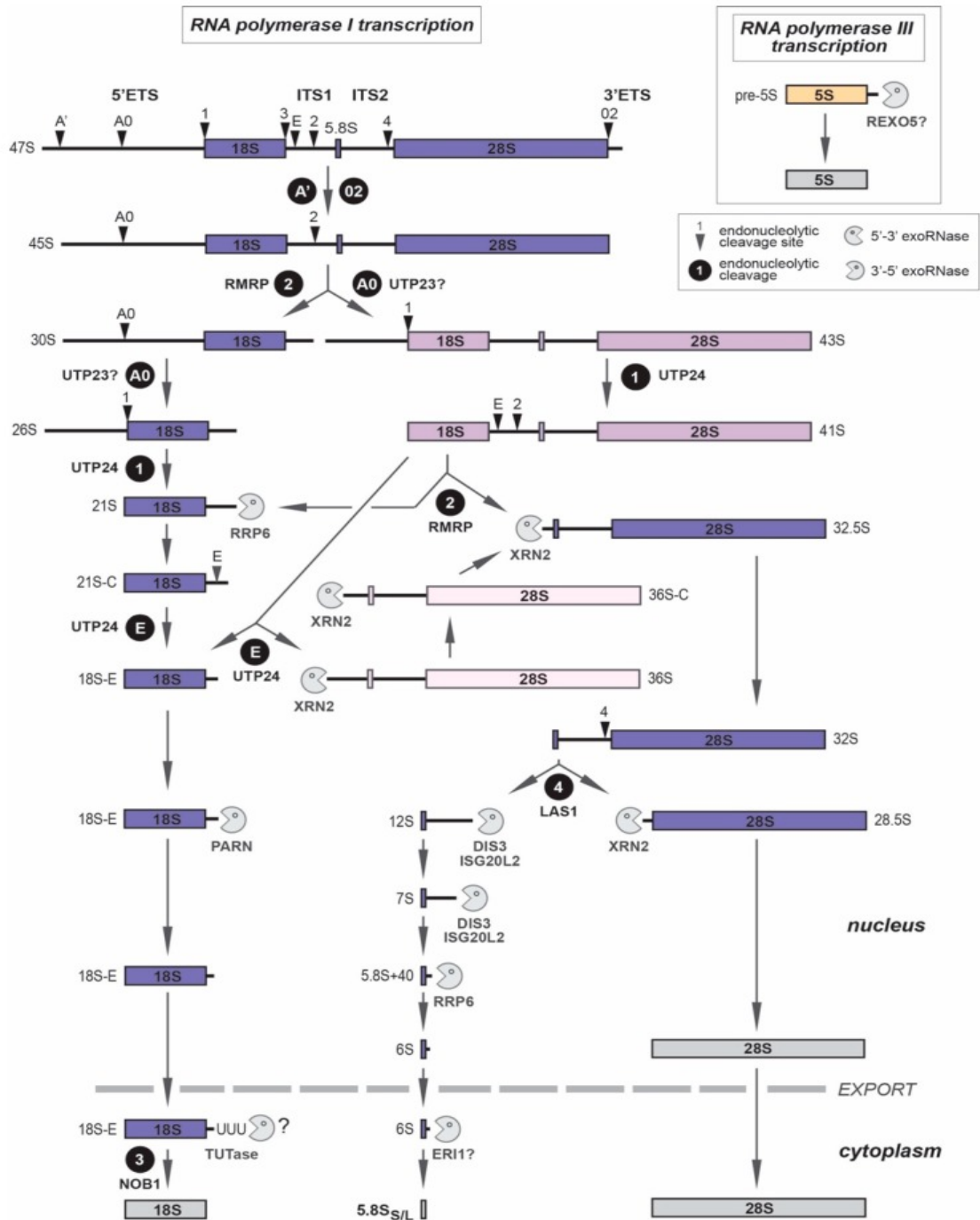
Figure 1.3: Overview of rRNA processing in humans, courtesy of Aubert et al. 2020. rRNA is first transcribed as a 47S polycistronic precursor and eventually matured to the 18S, 5.8S, and 28S rRNAs. Alternative processing pathway choice are shown by split or diagonal arrows.

Ribosome Assembly

The efficient and faithful synthesis of the rRNA genes is paramount in ribosome biogenesis. The ribosome, at its core, is a ribozyme, and derives both its catalytic activity and ability to decode mRNAs from the rRNA (Nissen et al. 2020). Moreover, rRNA acts as a central framework and scaffolding structure for 80 ribosomal proteins (R-proteins), 33 which associate with the SSU (called RPSs), and 47 that associate with the LSU (called RPLs). The R-proteins are critical for ribosome structure, and are imported into the nucleolus and interact with newly transcribed rRNA in stages. These R-proteins associate with different pre-rRNA intermediates, reciprocally influencing structural changes to eventually stabilize the final structures of both the SSU and LSU. Ribosome structure is also dependent on a multitude of snoRNPs. Beyond SNORD3 and 118 which function in rRNA cleavage, the vast majority of both C/D and H/ACA snoRNAs are involved in rRNA modification, with roughly 2% of all nucleotides in rRNA undergoing some form of modification (Sloan et al. 2017). The core 2'-O methylation modifications installed by C/D snoRNAs and pseudouridylations installed by H/ACA snoRNAs function to influence folding and enhance ribosome stability through base stacking and increased hydrogen bonding of rRNA, respectively. Following maturation, modification, folding, and R-protein binding of rRNA, the SSU (40S particle) and the LSU (60S particle) are exported into the cytoplasm. Here, both particles may undergo a final rRNA trimming step before becoming fully mature ribosomal subunits capable of translating mRNAs into protein (Panse and Johnson 2010).

The assembly of ribosomes from the myriad of constituent biomolecules is an energetically expensive process, and as such, is tightly regulated, with many quality control steps. Misprocessed or improperly folded pre-rRNA can be degraded by the exosome (Allmang et al. 2000), or other related nucleases (Zhu et al. 2018) to ensure only faithfully processed and mature rRNAs are incorporated into the ribosome. Defects in the processing or folding of rRNA, whether due to loss of a *trans*-acting factor, R-protein, or snoRNA-mediated cleavage or modification, can have disastrous effects on ribosomal biogenesis, and consequently, the cell or organism.

Ribosomopathies

Given the essentiality of ribosomes for all living cells and organisms, it is not surprising that deficiencies in their biogenesis results in pathologies. In humans, diseases that arise due to abnormal ribosome function or processing are referred to as ribosomopathies. Ribosomopathies are a heterogeneous group of diseases that present with a wide variety of clinical symptoms, generally due to haploinsufficiency of R-protein or rRNA processing genes. Biallelic loss of function mutations are rare in humans, as the complete loss of ribosomal genes is thought to be lethal. Surprisingly, despite the absolute need for ribosomes in all tissues, ribosomopathies exhibit tissue-specific defects and present clinically as bone marrow failure syndromes (Mills and Green 2017). Some of the most common

ribosomopathy diseases and their clinical manifestations and mutant genes are listed below in Table 1.

While the mechanisms behind ribosome dysfunction are varied, the resultant pathophysiology and hypoproliferative state can most likely be attributed to either/both P53-mediated cell cycle arrest/apoptosis and/or altered translation due to reduction of total ribosome number or ribosome substrate preference. P53 stabilization is achieved through the sequestration of E3 ligase MDM2 by free R-proteins RPL5 and RPL11 (Zheng et al. 2015), which accumulate during defective ribosome biogenesis due to inability to incorporate into the maturing ribosome. This stabilization of P53 can prohibit cell proliferation and may be responsible, at least partly, for the hypoproliferative nature of affected tissues in ribosomopathy patients.

Separately, the reduction in functional ribosome number may also play a role in ribosomopathy symptoms. mRNA containing specific sequence or topological features exhibit differing translational efficiencies (Avni, Biberman, and Meyuhas 1996), and reduced ribosome number can disproportionately decrease low translation efficiency mRNAs (Mills and Green 2017). A pertinent example can be found in the observation that GATA1 overexpression rescues cell death in RPS19 haploinsufficient cells from DBA patients(Ludwig et al. 2014). GATA1 is an erythroid lineage transcription factor that contains a highly structured 5' UTR that impedes its translation. The decrease in functional ribosomes due to RPS19 haploinsufficiency renders GATA1 translation inadequate, but this defect can be

mitigated by GATA1 overexpression, in which the 5' UTR of GATA1 was modified to improve its translation efficiency, which is sufficient to partially induce hematopoiesis.

In addition to simple reduction in functional ribosome number, translation may also be altered due to heterogenous ribosome composition, yielding specialized ribosomes. These ribosomes may have a preference for specific subsets of mRNA targets. An example of ribosomal specificity is in XL-DC, in which Dsykerin mutation results in a loss of pseudouridylation of rRNA, and this in turn hampers translation of internal ribosome entry site (IRES)-containing mRNAs, while cap-dependent translation of mRNAs is relatively unaffected(Jack et al. 2011) (Ruggero 2006).

Finally, mutations in the C/D snoRNA SNORD118, responsible for ribosomal LSU biogenesis, cause cerebral microangiopathy leukoencephalopathy with calcification and cysts (LCC) (Jenkinson et al. 2016). LCC is a neurological disorder in which patients undergo continued cerebral degeneration. While data show unequivocally that SNORD118 is the causal factor in LCC, and that SNORD118 depleted cells exhibit proliferation defects, it is unclear if ribosome biogenesis is defective in LCC patients and discussion continues on whether LCC should be classified as a ribosomopathy.

Somewhat paradoxically, many ribosomopathies also confer an increased susceptibility for cancers. How a hypoproliferative disease increases risk for hyperproliferation is still unknown, though several hypotheses have been put

forward. One hypothesis put forth by Kampen et al. (Kampen et al. 2021) suggests that altered translation ratios of tumor suppressor proteins to oncogenes is skewed with dysfunctional ribosome assembly, leading to a pro-oncogenic cellular environment. While most of these cells would exhibit hyproliferation due to general ribosome deficiencies, selective pressure for P53-null cells might allow the escape of a small number of cells that contain the so-called "oncoribosomes." Another hypothesis for the transition from hypo- to hyperproliferation relies on the observation that ribosome function is required for maintaining proper levels of reactive oxygen species (ROS), and ribosome deficient cells have higher levels of ROS (Sulima, Kampen, and De Keersmaecker 2019). Ribosome defects keep the cell in a hypoproliferative state, yet the cells exhibit a high mutagenicity due to ROS-mediated DNA damage. Eventually, ROS-driven mutations lead to a cellular environment that favors proliferation, while still maintaining a high mutational rate, thereby increasing cancer risk.

Altogether, these data show the necessity of efficient and faithful ribosome biogenesis. Defects in any step of this process have extreme consequences for cells or organisms. The continued study and complete elucidation of this incredibly complex process will reap great rewards in terms of patient outcomes.

Table 1. List of ribosomopathies. *The precise molecular defect in ribosome biogenesis has not been elucidated here.

Disease Name	Clinical Presentation	Gene(s) mutated(Aspesi and Ellis 2019)	Defect in Ribosome Biogenesis	Prevalence
5q-Myelodysplastic Syndrome	Extreme fatigue, macrocytic anemia, hypolobulated megakaryocytes	RPS14	SSU biogenesis*	1 in 130,000
Aplasia Cutis Congenita	Skin defects	BMS1	SSU biogenesis	1 in 10,000
Bowen-Conradi Syndrome	Bone marrow failure, growth retardation, skeletal deformities, death in infancy	EMG1	SSU pseudouridine methylation	unknown
Cartilage Hair Hypoplasia (CHH)	Dwarfism, hypotrichosis (hair growth abnormalities), immunodeficiency	RMRP	LSU and SSU biogenesis	unknown
Diamond Blackfan Anemia (DBA)	Bone marrow failure, physical abnormalities (skeletal, cardiac, genitourinary)	RPS(7,10,15A,17,19, 24, 26, 27, 28, 29, RPL(5, 11, 15, 18, 26, 27, 31, 35, 35A), TSR2	LSU and SSU biogenesis	1 in 200,000
North America Indian Cirrhosis (NAIC)	Liver cirrhosis	UTP4	SSU biogenesis	Only found in children of Ojibway-Cree descent
Schwachman-Diamond Syndrome (SDS)	Bone marrow failure, pancreatic insufficiency	SBDS,	LSU biogenesis	1 in 80,000
Treacher Collins Syndrome (TCS)	Craniofacial abnormalities	TCOF1, POLR1C, POLR1D	Loss of rRNA transcription	1 in 50,000
X-linked Dyskeratosis Congenita (XL-DC)	Bone marrow failure, oral leukoplakia, hyperpigmentation of skin, nail dystrophy, pulmonary fibrosis	DKC1 (other genes mutated but involved in telomere maintenance)	Loss of rRNA pseudouridylation	1 in 1,000,000

Section 5: Silencing Defective 2

Silencing Defective 2 (SDE2) is a protein-coding gene of unknown function in humans. It was first characterized in 2011 in *Schizosaccharomyces pombe* as a telomere maintenance factor (Sugioka-Sugiyama and Sugiyama 2011a), its presence recruited histone deacetylases to ensure transcriptional silencing of telomeres, thereby ensuring genomic stability. Since this initial discovery, subsequent research in yeast has suggested SDE2 is predominantly involved in pre-mRNA splicing. Thakran et al. showed that SDE2 is required for the splicing of a small subset of specific introns in *Schizosaccharomyces pombe*, though many molecular details regarding its involvement with the spliceosome remain unclear (Thakran et al. 2018). SDE2 has been detected in human spliceosome complexes (Bessonov et al. 2008), though its appearance is not entirely uniform, as spliceosomal complexes have been isolated that failed to detect SDE2. This may indicate that SDE2 is a splicing auxiliary factor that either facilitates transient interactions with spliceosomal complexes, or associates with the spliceosome only in specific contexts, similar to the splicing requirement of SDE2 at specific introns in yeast. SDE2 involvement in human pre-mRNA splicing was further strengthened when a partial crystal structure of an exon ligation complex (a late stage spliceosomal complex) was obtained, containing within it SDE2 and several other associated splicing factors (Fica et al. 2019b).

SDE2 is a 451 amino acid protein predicted to contain an N-terminal ubiquitin-like (UBL) motif, a C-terminal nucleic acid binding SAP domain (named after similar structural motifs in the SAF-A/B, Acinus, and PIAS proteins) (Aravind and Koonin 2000), and a central region predicted to harbor a large IDR. Similar to other UBL-containing proteins, the UBL of SDE2 may undergo a sequence specific cleavage event at a diglycine motif in the N-terminus, though this has never been demonstrated with endogenous SDE2. Interestingly, UBL-containing proteins have been implicated in a wide range of cellular processes, especially splicing and ribosome biogenesis (Chanarat and Mishra 2018). Both IDRs and SAP domains are known to bind RNA and nucleic acids, respectively, suggesting that SDE2 may function as an RNA binding protein. These observations led us to investigate the role of SDE2 as both an RBP and its role in pre-mRNA splicing in humans.

CHAPTER TWO: SDE2 REGULATES ALTERNATIVE SPLICING

Section 1: Introduction

Pre-mRNA splicing consists of the removal of introns and the ligation of neighboring exons to create a mature RNA transcript. This process relies on the intricate coordination between both *cis*-elements located on the transcript itself and *trans*-acting ribonucleoprotein factors. The most frequently described cis-elements include the 5' and 3' splice sites, the branch-point sequence, and the splicing regulatory elements (SREs) which together serve as a platform for the association of the trans-acting ribonucleoprotein factors that make up the human spliceosome (Will and Lu 2011). The human spliceosome is a dynamic complex that in totality consists of five small nuclear ribonucleoproteins (snRNPs), U1, U2, U4, U5, and U6 and approximately 300 protein factors (Jurica and Moore 2003). Once assembled, the spliceosome functions to regulate two sequential transesterification reactions that lead to the cleavage of pre-mRNA at the exon-intron boundaries, removal of the intervening intron, and ligation of the adjacent exons (Jurica and Moore 2003; Will and Lu 2011). In this way, pre-mRNA is processed to mature mRNA, exported to the cytoplasm, and incorporated into active ribosomes for translation. Efficient and accurate splicing of mRNA transcripts is critical for cellular function, and defects in even a single factor or step in the splicing reaction can lead to significant changes in the cellular transcriptome in the form of differential, or alternative splicing.

Alternative splicing (AS) gives rise to unique mRNA transcripts through the differential inclusion of exons and/or introns into the mature RNA transcript. AS not only promotes the formation of unique mature RNA isoforms from the expression of a single gene, but it also increases protein diversity, and likely functions to refine and coordinate complex biological processes within the cell (Graveley 2001; Maniatis and Tasic 2002). The physiological importance of AS is underscored by the fact that over 90 percent of human genes undergo some form of AS (Pan et al. 2008; E. T. Wang et al. 2008) However, genetic mutations in RNA-splicing factors or cis-elements contribute to widespread changes in the splicing profile within cell and have been linked to a number of human diseases including muscular dystrophy, neurodegeneration, and cancer, suggesting that defects in the regulation of AS can also be pathogenic (Ahn and Kunkel 1993; Dvinge et al. 2016; Lagier-Tourenne, Polymenidou, and Cleveland 2010; Xiong et al. 2015). To date, the major AS events include cassette exons, mutually exclusive exons, alternative 5′ splice sites, alternative 3′ splice sites, intron retention, alternative 3′ terminal exons, and alternative 5′ exons (Breitbart 1987). These AS events may arise individually within a transcript, or concomitant with other AS events generating more complex isoforms. Yet, physiologically, how each AS isoform functions within the cell, whether each AS event is processed by the same spliceosomal machinery, and what cellular cues regulate AS remains unclear.

Retained introns (RI) are the most common type of AS event in plants, fungi, and unicellular eukaryotes, however the prevalence and relevance of RI in

mammals is only just beginning to emerge (Grützmann et al. 2014; Ner-Gaon et al. 2004; Sebé-Pedrós et al. 2013). Transcripts containing RI events can be retained in the nucleus where they are either stored for future splicing or degraded (Boutz, Bhutkar, and Sharp 2015; Braunschweig et al. 2014). Alternatively, RI containing transcripts can be exported to the cytoplasm where they are incorporated into active ribosomes for translation or are actively degraded by various RNA decay mechanisms (Ge and Porse 2014). The fate of transcripts containing retained introns is diverse and dependent not only on the position of the AS event within the transcript, but also on the context in which the AS event occurs. Introns retained within the untranslated region (UTR) of the transcript can affect overall transcript stability and/or translational efficiency. Intron retention events within the open reading frame of a transcript often lead to the incorporation of a premature termination codon (PTC) in the mRNA. Incorporation of a PTC can either promote degradation of the transcript through the nonsense-mediated decay (NMD) pathway or, if not degraded, the RI containing transcript can be translated and result in the expression of a truncated or misfolded protein product (Ge and Porse 2014). As with other AS events, RI events have been demonstrated to be physiologically important in organismal development, but also pathological in the development of human disease (Boutz, Bhutkar, and Sharp 2015; Braunschweig et al. 2014; Jacob and Smith 2017b; Wong et al. 2016).

RNA splicing is an essential process, and more than one-third of all disease-causing genetic mutations occur in RNA-splicing or RNA-splicing related genes

(Havens, Duelli, and Hastings 2013). In light of this, fully defining the catalogue of proteins and cis-elements required for pre-mRNA splicing is essential to our understanding of both basic biological processes and complex human diseases. Silencing defective 2 (SDE2) was originally identified in Schizosaccharomyces pombe and has since been linked to several cellular process in eukaryotes including heterochromatin formation, telomere silencing, DNA replication, and mRNA processing (Fica et al. 2019a; Jo et al. 2016; Rageul et al. 2019; Sugioka-Sugiyama and Sugiyama 2011b; Thakran et al. 2018). Here, we show SDE2 plays an important role in maintaining transcriptomic stability in mammalian cells, as loss of SDE2 results in widespread AS. Intron retention events are particularly increased upon SDE2 knockdown, and are defined by distinct characteristics in established *cis*-elements. This work identifies SDE2 as a previously uncharacterized factor that is required for efficient pre-mRNA splicing and proper gene expression.

Section 2: Materials and Methods

Cell Lines

All cell lines were submitted for Short Tandem Repeat (STR) analysis by ATCC and certificates of authentication can be provided upon request. HeLa, U2OS, and 293FT cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (10% FBS, 1% penicillin/streptomycin). U2OS cells were engineered to

stably express H2B-mCherry and were a gift from Dr. Neil J. Ganem. RPE cells were cultured in DMEM/F12 (10% FBS, 1% penicillin/streptomycin). Cell culture media and supplements were obtained from Gibco Invitrogen and all plasticware came from Corning (Corning, NY). All cells were maintained at 37°C in a humidified incubator at 5% CO_2.

Transfections and siRNA

Cells were seeded at 5 X 10^4 cells per well in a 6 well plate and reverse transfected with ON-TARGETplus siRNA (Dharmacon). Cells were transfected with either 100nM (Non-targeting control, siSDE2-2) or 20nM (siSDE2-1) siRNA using Lipofectamine RNAiMax diluted in Opti-MEM according to the manufacturer's instructions. The next day, the siRNA solution was removed from cells and replaced with fresh media. Cells were collected for various downstream applications after 3 days, or collected, subjected to a second reverse transfection and plated, and collected after another 2 days (5 days total from the time of the initial reverse transfection). siRNA target sequences: siSDE2-1 (CUACUAAAUCUCAAACAGAdTdT),siSDE2-2 (GGAAGCUUGUAGAACCCA-AdTdT), ON-TARGETplus Non-targeting siRNA #1 (UGGUUUACAUGUCGA-CUAA).

Antibodies and Plasmids

The following antibodies and plasmids were used where indicated. SF3B1 (Bethyl Laboratories A300-997A), SDE2 (Bethyl Laboratories A302-098A and A302-099A), GAPDH (Santa Cruz Biotechnologies sc-47724), U2AF1 (Bethyl Laboratories A302-079A, Histone 4 (Active Motif #39269), (Cell Signaling Technology 2855S), Cactin (Bethyl Laboratories A303-349A), Tubulin (Cell Signaling Technology 2125S).

Western Blotting

Western blots were performed using standard protocols. Briefly, cells were collected by trypsinization and washed with ice-cold 1XPBS. Samples were then lysed in 2X sample buffer and sonicated in a water bath at 4°C for 5 minutes (20-second pulse on/30-second pulse off at 100% amplitude), then boiled at 95°C for 10 minutes. Soluble protein lysates were then analyzed by western blot using standard SDS-PAGE techniques and transferred onto PVDF membranes. Membranes were blocked in TBS-T (1X TBS, 0.1% Tween-20) containing 5% milk for one hour and then incubated overnight at 4°C with the appropriate primary antibodies. Following overnight incubation with primary antibodies, membranes were washed 3X in TBS-T for 10 minutes each, incubated with peroxidase conjugated secondary antibodies for 1 hour, then washed 3X in TBS-T for 10 minutes each and visualized using enhanced chemiluminescence reagents from

BioRad and Thermo Fisher. Zyagen Human Tissue Western Blot was probed with the indicated antibodies.

Cellular Fractionation

Cells were collected by trypsinization and lysed in CSK buffer (10mM HEPES pH 7.5, 10mM KCl, 2mM MgCl$_2$, 300mM sucrose, 10% glycerol, 0.1% Triton X-100) for 5 minutes on ice and then centrifuged for 5 minutes at 1,400 × g at 4°C. The supernatant was collected and analyzed as the cytoplasmic fraction. The cell pellet was washed in CSK buffer without Triton X-100, then centrifuged for 5 minutes at 1,400 × g at 4°C and supernatant discarded. The remaining cell pellet was then collected and analyzed as the nuclear fraction.

Immunoprecipitation

5 µg of each antibody was incubated with Dynabeads Protein A (Invitrogen 1001D) suspended in NETN buffer (150mM NaCl, 20mM TRIS pH 8.0, 0.5mM EDTA, 0.5% NP-40) supplemented with protease inhibitor cocktail (Sigma-Aldrich P8340) at a 1:100 ratio for one hour at room temperature with gentle rotation. Cells were lysed in NETN buffer and sonicated in a water bath at 4°C for 15 minutes (20-second pulse on/40-second pulse off at 50% amplitude). Lysate was centrifuged at 12,000 x g for three minutes at 4°C and supernatant was collected. Supernatant was added to Dynabeads-antibody mixture and allowed to incubate for 16 hours at 4°C with gentle rotation. After incubation, beads were collected on

a magnet and supernatant was discarded. Beads were further washed 3X with NETN buffer supplemented 1:100 with protease inhibitor complex. After final wash, beads were collected on a magnet, wash buffer was discarded, and beads were resuspended in standard western blot sample buffer.

RT-PCR Amplification & Gel Electrophoresis

Cell pellets for each cell line or siRNA condition were collected from actively growing cells. Genomic DNA (gDNA) and total cellular RNA were then extracted from the cell pellets using the QIAamp DNA Mini Kit and RNeasy Mini Kit protocols according to manufacturer's instructions. For RT-PCR experiments, 0.5 µg of total cellular RNA was converted to complementary DNA (cDNA) using the SuperScript IV Reverse Transcriptase protocol (Life Technologies #18090010) in a volume of 20 µl according to manufacturer's instructions. gDNA or cDNA samples underwent three-step PCR amplification using Phusion High-Fidelity PCR Master Mix according to manufacturer's instructions. PCR reaction parameters were as follows: [1] 0:30 at 98°C; [2] 0:10 at 98°C; [3] 0:30 at annealing temperature (58.5°C for ARSA and RFNG, 57.4°C for PITPNM1 and SPATA20, 64.2°C for CUL7, and 55°C for FBF1); [4] 0:10 at 72°C; [5] Repeat steps 2–4 for 31 cycles; [6] 10:00 at 72°C [7] Hold at 12°C. PCR products were then resolved using agarose gel electrophoresis, stained using GelRed, and visualized using a BioRad ChemiDoc XRS+ imaging system. Sequences of primers are described in Table 1 (Appendix A).

Computational Alternative Splicing Profiling

Total RNA was extracted from three biological replicates of HeLa cells with SDE2 knockdown on day 3 and day 5, using Qiagen RNeasy Kit (Qiagen 74106). Samples were submitted to the BU Microarray and Sequencing Core for library preparation and polyadenylated RNA selection using Kapa RNA HyperPrep kit with Riboerase, and sequenced yielding 2x75 bp paired-end read datasets. Read library quality was assessed using FastQC (Andrews 2015) and MultiQC (Ewels et al. 2016) packages. Illumina adapters were removed and leading and trailing low quality bases (below quality 30) were trimmed using Trimmomatic (Bolger, Lohse, and Usadel 2014). Reads which were less than 36 bases long after these steps were dropped. Three tools were engaged in the AS profiling process: Whippet, IRFinder and rMATS.

Whippet

Annotation (GTF) only index was built using GENCODE v27 annotations (Harrow et al. 2012) by Whippet (Sterne-Weiler et al. 2018). Trimmed reads were then aligned directly to the contiguous splice graphs built from the hg38 human reference genome, and Percent Spliced In (PSI) of each AS event was quantified. After pooling the three PSI results for each biological sample, the comparison between siSDE2-1 and control, as well as siSDE2-2 and control were implemented. For both of these comparisons, three filters were applied. The

absolute difference between PSI in knockdown and in control should be greater than 0.1. The probability of this event happening as identified by Whippet algorithm should be greater than 0.9, and the PSI in the knockdown condition should be greater than 0.1. The filtered results from the two comparisons were then intersected to find the significant events.

IRFinder

The IRFinder (Middleton et al. 2017) reference was built from the GENCODE v27 annotations and hg38 human reference genome, prior to aligning trimmed reads to it in FASTQ mode. The resulted unsorted BAM files from the three replicates were then concatenated and pooled together for each biological sample, producing quantification of the intron retention (RI) events for siSDE2-1, siSDE-2 and control. Next, the comparisons between siSDE2-1 and control, as well as siSDE2-2 and control were analyzed using the provided script for small amounts of replicates. For both of the results, two filters were applied. Events marked as known exons or minor isoforms were removed, and events with FDR less than 0.05 were selected. The filtered results from the two comparisons were then intersected to find the significant events.

rMATS

Mapping the trimmed reads to the reference genome was completed via STAR (Dobin et al. 2013) using the same annotation and reference genome files

as noted above. The produced BAM files were then used as input for pooling and contrasting through the rMATS (Shen et al. 2014) algorithm between knockdown and control, generating quantitative results of the five AS categories. Only FDR values less than 0.05 were considered for further analysis.

The three sets (Whippet, IRFinder, rMATS) of overlapped results were subsequently combined to create an aggregated view with all of the AS events, after removing duplicated ones that were identified by more than one tool. The PSI in knockdown was calculated as the average of that in siSDE2-1 and siSDE2-2, and delta PSI was computed as the difference between PSI knockdown and PSI control.

Genome-wide Intron Identification

Only genes that have expression in at least one sample(s) were chosen for further analysis. Next, for each gene we retrieved the coordinates of the unique introns from all of the transcripts according to the GENCODE v27 annotations. Finally, only the introns with a length greater than 30nt were kept to eliminate inclusion of any false positive candidates that were actually indels.

Characterization of RI Events: Comparison between ENCODE data and SDE2 data

The IRFinder paper accessed RNA sequencing data from 8 shRNA knockdown samples in ENCODE (Dunham et al. 2012), and included a

presentation of the number of both significantly increased and decreased RI events under nominal P values less than 0.05 for each of the experiments. The same processing was performed on the SDE2 RNA sequencing data, but only the increased RI events were selected to show in the figure, as well as those from the ENCODE data. The individual siRNAs against SDE2, the intersection of the two siRNAs, and the union of the two siRNA were included in the final result. The raw numbers were divided by the total number of all introns expressed in the genome to compute a percentage of RI.

Characterization of RI Events: Location of the retained introns

Two groups were involved in the comparison, the significant (Sig) events and the non-significant (Non-sig) ones. After removing events recognized as known exons and minor isoforms in both the siSDE2-1 VS control and siSDE2-2 VS control, the Sig events were selected as the intersected events that have FDR less than 0.05, while the Non-sig events were selected as the intersected events that have FDR greater than or equal to 0.05. The locations of the RI events were computed as a fraction of the middle point of the RI in proportion to the length of the gene, with 0 representing the 5' splice site and 1 representing the 3' splice site.

Characterization of RI Events: Length of the retained introns

Three groups were included in the comparison: Sig, Non-sig and Non-AS. The last group includes all the introns in the genome except for the events in the Sig and Non-sig groups. Length is calculated as the difference between the start and end coordinates. The pairwise statistical comparison was done using the Mann-Whitney U test.

Characterization of RI Events: GC content of the retained introns

Four groups were included in the comparison, with the Non-AS events divided into Non-AS-short and Non-AS-long. The discriminating threshold is the 75th percentile of the length of the Sig events, which is 358nt. Percentage of GC was calculated in the retrieved sequences and statistical comparison was done using the Mann-Whitney U test.

Characterization of RI Events: Splice site score of the retained introns

Utilizing the online portal for the MaxEntScan (G. Yeo and Burge 2004), which is based on the maximum entropy modeling of short sequence motifs, scores were computed with sequences at both the 5' and 3' splice sites. At 5' splice site, each sequence must be 9 bases long (3 bases in the exon and 6 bases in the intron), while at the 3' splice site, each sequence must be 23 bases long (20 bases in the intron and 3 bases in the exon). An additional filter was added to remove any sites that appeared to be a splice site, but were not validated when scrutinizing

the sequences. The satisfying score threshold was determined in such a way that at least 50% of the sequences falling into the ± 0.2 range of the chosen score should have the correct splicing signal at the correct position, which led to -7.2 for the 5' splice site and 1.4 for the 3' splice site. The statistical comparison was done using the Mann-Whitney U test.

Code Availability

All code written to implement the analyses in this paper are available at https://bitbucket.org/bucab/sde2/.

Section 3: Results

Previous studies have identified SDE2 in spliceosomal complexes, suggesting SDE2 may function in pre-mRNA splicing. In fact, structural databases highlight a region of homology between SDE2 and the human splicing factor SF3A3, a canonical component of the U2 snRNP. In addition, a partial crystal structure of SDE2 was recently resolved as part of the spliceosomal P complex bound to precursor mRNA and the U2/U6 snRNAs, suggesting a role for SDE2 in 3' splice site recognition (Fica et al. 2019a). Therefore, we asked whether SDE2 associated with the U2 snRNP complex. Using reciprocal immunoprecipitation, we detected an interaction between SDE2, SF3B1, and U2AF1, confirming that SDE2 associates with the U2 snRNP complex (**Figure 2.1A**). Moreover, SDE2 protein

interactions extended beyond the U2 snRNP, to proteins critical for 3' splice site docking including Cactin (**Figure 2.1A**), fueling the hypothesis that SDE2 was involved in the regulation of pre-mRNA splicing. Finally, we demonstrate that SDE2 expression, while prevalent in all tissue tested, is specifically enriched in brain, liver, and lung (**Figure 2.1B**), and these tissues are associated with increased and distinctive patterns of AS (G. Yeo et al. 2004). Notably, the SDE2 expression pattern in human tissue mimics that of the splicing factor and core U2 snRNP component SF3B1, consistent with the idea that SDE2 is a component of the U2 snRNP.

To determine whether loss of SDE2 leads to defects in pre-mRNA splicing, we treated cultured HeLa cells with one of two unique siRNAs to deplete SDE2 protein and extracted RNA at 72 hours (3 days) and 120 hours (5 days) in preparation for high throughput polyA+ mRNA sequencing. We identified transcriptome-wide differential alternative splicing (dAS) events using a custom bioinformatic pipeline (Appendix B, **Figure AB1**). Briefly, an AS event is defined as a genomic locus (e.g. an intron) that shows evidence of differential inclusion. For example, an intron involved in an RI event might have evidence of being included in only 30% of transcripts that originate from that locus. Such an event would have a "percent spliced in" (PSI) of 0.3. To identify changes in AS, we identified AS events that show significantly different PSI values, or delta PSI, between conditions. When considering the above example, if in one condition the RI event has a PSI of 0.3, while in another the PSI is 0.9, the delta PSI is 0.9

- 0.3 =0.6. This delta PSI is large, and thus we infer that there is dAS at this locus. Our pipeline uses multiple published methods for detecting dAS events and identifies consensus results from both siRNAs to arrive at the set of dAS events that are likely the most robust (see Methods for more details).

At the 72-hour timepoint, we found that there was a significant increase in AS with 2,431 dAS events affecting 1,474 unique genes with an FDR < 0.05 (**Figure 2.1C-D**). This increased to 3,577 dAS events consisting of 2,288 unique genes with an FDR < 0.05 at the 120-hour timepoint (**Figure 2.2A-B**). Of the
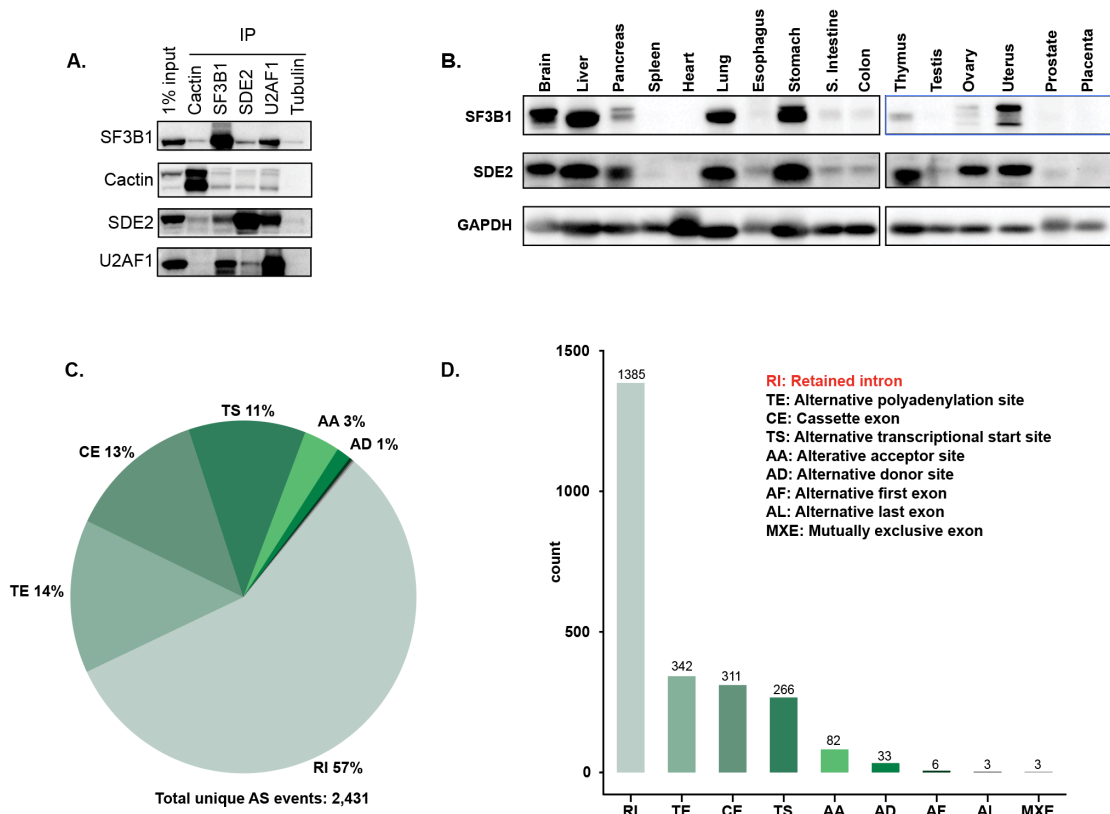
Figure 2.1. SDE2 is a U2 snRNP component required for pre-mRNA splicing. (A) Reciprocal immunoprecipitation of SDE2, SF3B1, U2AF1, and Cactin, with a 1% input from whole cell extract in HeLa cells. Tubulin is used as a negative IP control. (B) Western blot of SF3B1, SDE2, and GAPDH on a membrane containing 16 human tissue samples. (C) Overview of the AS events identified using our custom pipeline after 3 days of SDE2 depletion via siRNA. Pie chart represents the percentage of each AS type. The total number of significant AS events is annotated at the bottom. (D) Bar plot demonstrating the absolute number of each AS type.

2,431 AS events detected 3 days after knockdown, RI were the most common

AS type, consisting of 1,385 events, or 57% of all AS events (**Figure 2.1C**). Also

enriched were alternate terminal exon (TE 14%), cassette exons (CE12.7%),

and alternative transcriptional start site (TS, 11%). The distribution of dAS event types was similar at day 5 albeit with an increase in CE (**Figure 2.2A-B**).

A.

B.
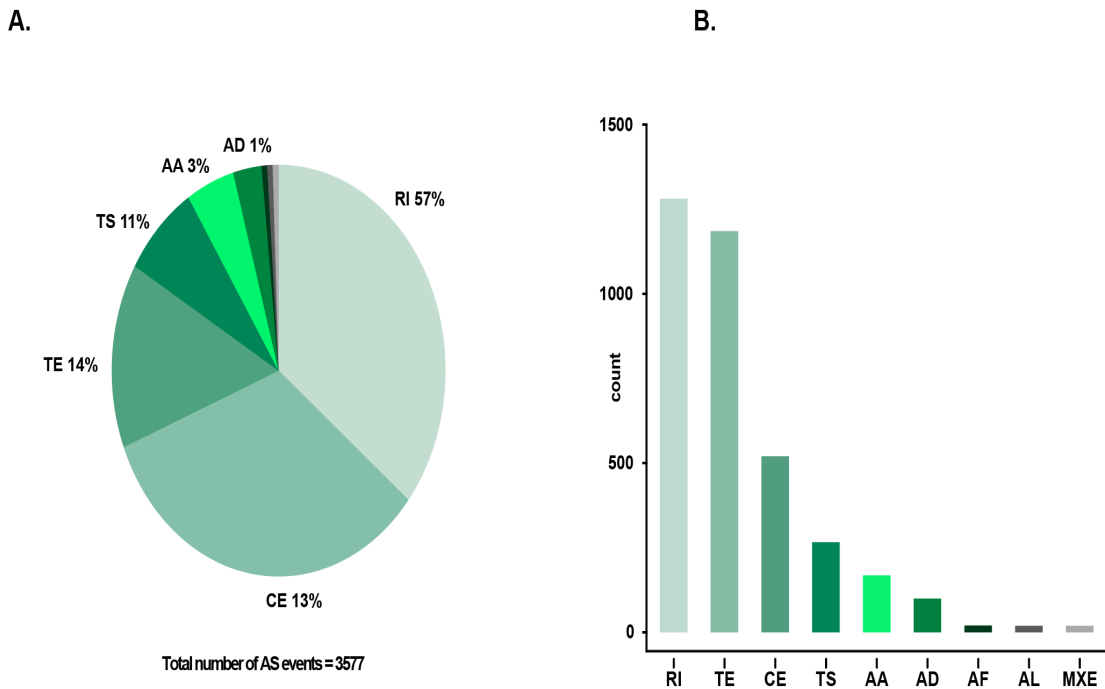
Figure 2.2. AS profile after 120 hours of SDE2 depletion. (A) Pie chart depicting the distribution of AS event types identified 120 hours following SDE2 knockdown. Event types are Retained Intron (RI), Terminal Exon (TE), Casette Exon (CE, aka exon skipping), Alternative Transcription Start (TS), Alternative Acceptor (AA, aka alternative 3' intron splice site), Alternative Donor (AD, aka alternative 5' intron splice site), Alternative First Exon (AF), Alternative Last Exon (AL), and Mutually Exclusive Exons (MXE). (B) Number of events by event type 120 hours after SDE2 knockdown.

Given that RI were by far the most abundant AS event following SDE2 depletion, we chose ten candidate genes (ADAMTSL4, ARSA, CUL7, DGKQ, FBF1, KIFC2, PARP10, PITPNM1, RHBDF1, and SPATA20), each with at least one significant RI dAS event identified by our pipeline, to validate using RT-PCR (**Figure 2.3A-B and Figure 2.4A**). In addition, we also selected one event from the gene RFNG, predicted to demonstrate intron retention, but that was not significantly different between control cells and SDE2 knockdown cells (no change in delta PSI). We then visualized each of these ten loci in the integrated genome viewer (IGV) to confirm the predictions from our pipeline (**Figure 2.3B**). As predicted by our analysis, ADAMTSL4, ARSA, CUL7, DGKQ, FBF1, KIFC2, PARP10, PITPNM1, RHBDF1, and SPATA20 all demonstrated robust and significant increases in intron retention at the predicted loci by RT-PCR following knockdown of SDE2 (**Figure 2.4B, D**). In contrast, RFNG, which served as a negative control, had no change in intron retention between control and SDE2 knockdown cells (**Figure 2.4B, D**). To ensure that the AS events were not cell line specific, we also analyzed the RI events in all ten of our candidate genes and RFNG in U2OS cells following SDE2 knockdown (**Figure 2.4C, E**). Consistent with the data in HeLa cells, loss of SDE2 in U2OS also led to a significant increase in intron retention in nine out of ten candidate genes. The tenth candidate gene, PARP10, displayed increased intron retention, but was not significantly changed with SDE2 depletion, while intron retention in RFNG remained unchanged.
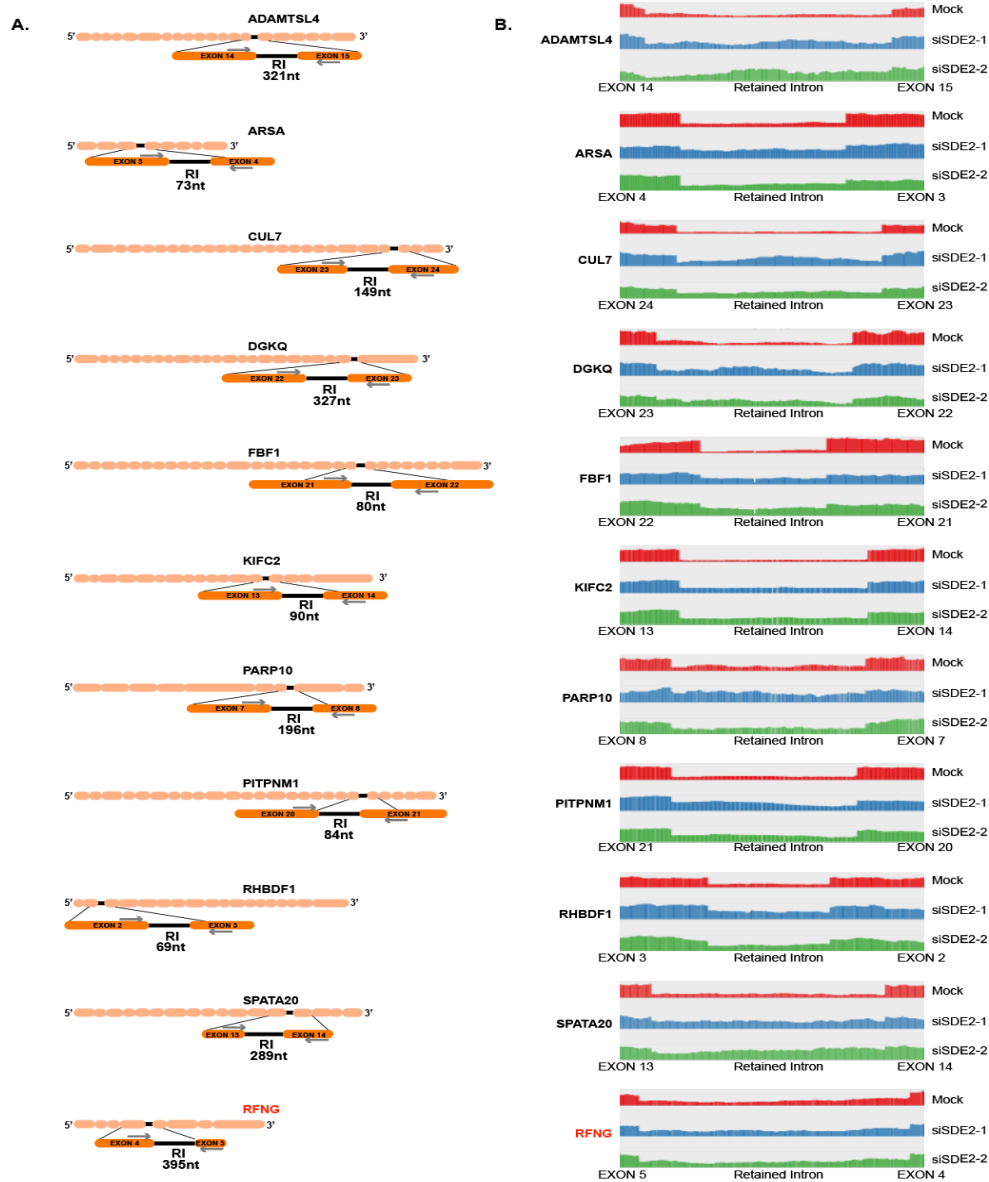
Figure 2.3. Location and coverage of validated RI events. (A) Diagrams depicting position of RI event loci in the corresponding gene and PCR amplification region. Exons are shown as peach or orange colored boxes, retained introns are shown as thick black line, PCR primers are shown as gray arrows. (B) Sashimi plots of mRNA-Seq reads aligned to dAS RI event loci selected for validation. Each validation target corresponds to a different set of genomic coordinates flanking the RI event. The height of each colored segment is proportional to the number of reads that align to that location in the locus. The greater abundance of reads in the intronic region for the ten validation targets (ADAMTSL4, ARSA, CUL7, DGKQ, FBF1, KIFC2, PARP10, PITPNM1, RHBDF1, SPATA20) illustrates the increased dAS of each event. The negative control RFNG shows no difference in intron inclusion.
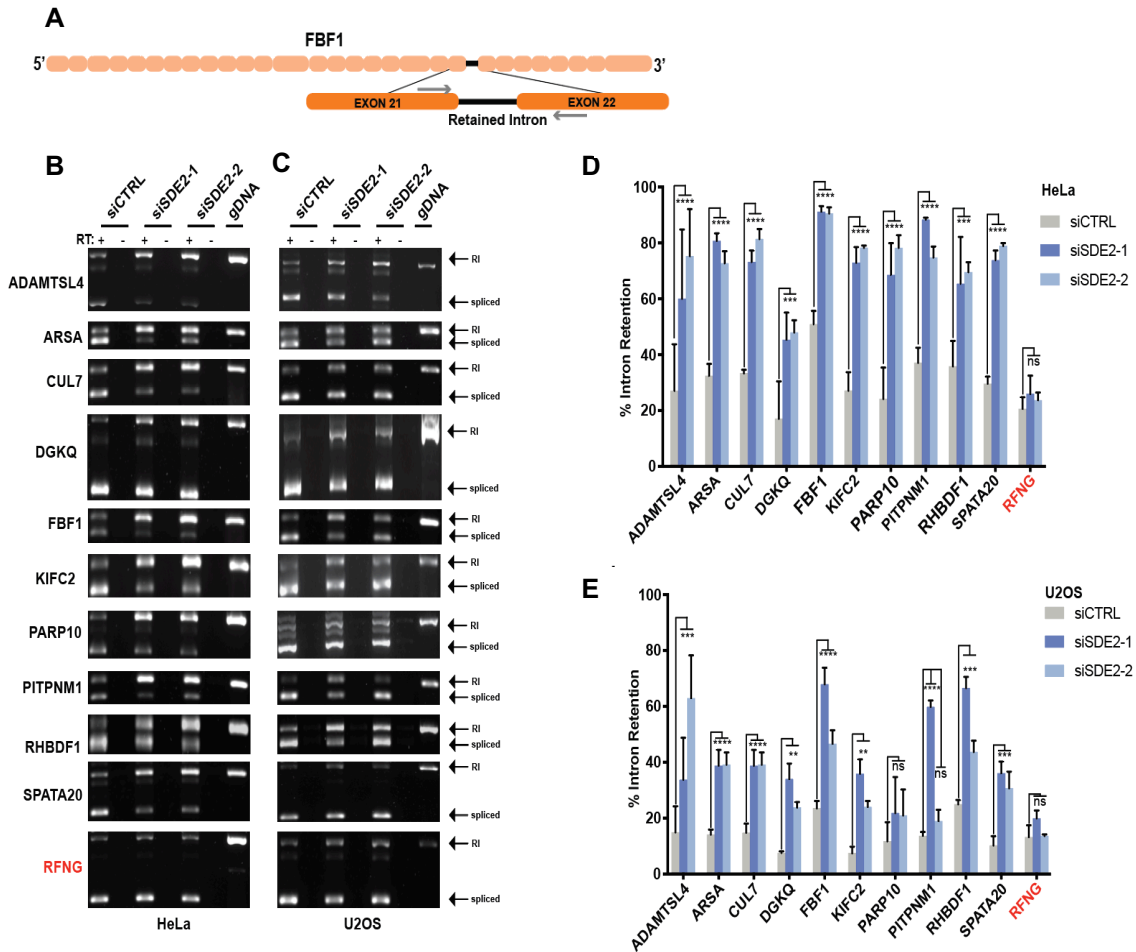
Figure 2.4. SDE2 depletion causes increased RI. (A) Schematic representation of one gene containing an RI event identified in our analysis pipeline; exons are depicted by peach boxes, the retained intron is depicted with a thick black line, and the exons flanking the RI event are depicted by larger orange colored boxes. Location of primers used to amplify RI events are shown as gray arrows. (B-C) DNA gels depicting intron retention events following RT-PCR. RNA was extracted from siCTRL and siSDE2 cells and target genes were amplified by RT-PCR in HeLa cells (B,D) and U2OS cells (C,E) with (+RT) and without (-RT) reverse transcriptase. RI events and spliced events are labeled. Graphs show mean percent intron retention in each condition relative to total transcript (intron retained + spliced) +/- SD. RFNG does not demonstrate a dAS RI event and serves as a negative control. *n=3* for all quantifications and experiments, *P*-values denoted by comparing siCTRL vs siSDE2-1 or siSDE2-2 by two-way ANOVA followed by Dunnet's multiple comparison test, $P \leq 0.01$ denoted by **, $P \leq 0.001$ denoted by ***, and $P \leq 0.0001$ denoted by ****, ns = not significant. For RI events with different levels of *P*-values between siSDE2-1 and siSDE2-2, the least significant value is shown.
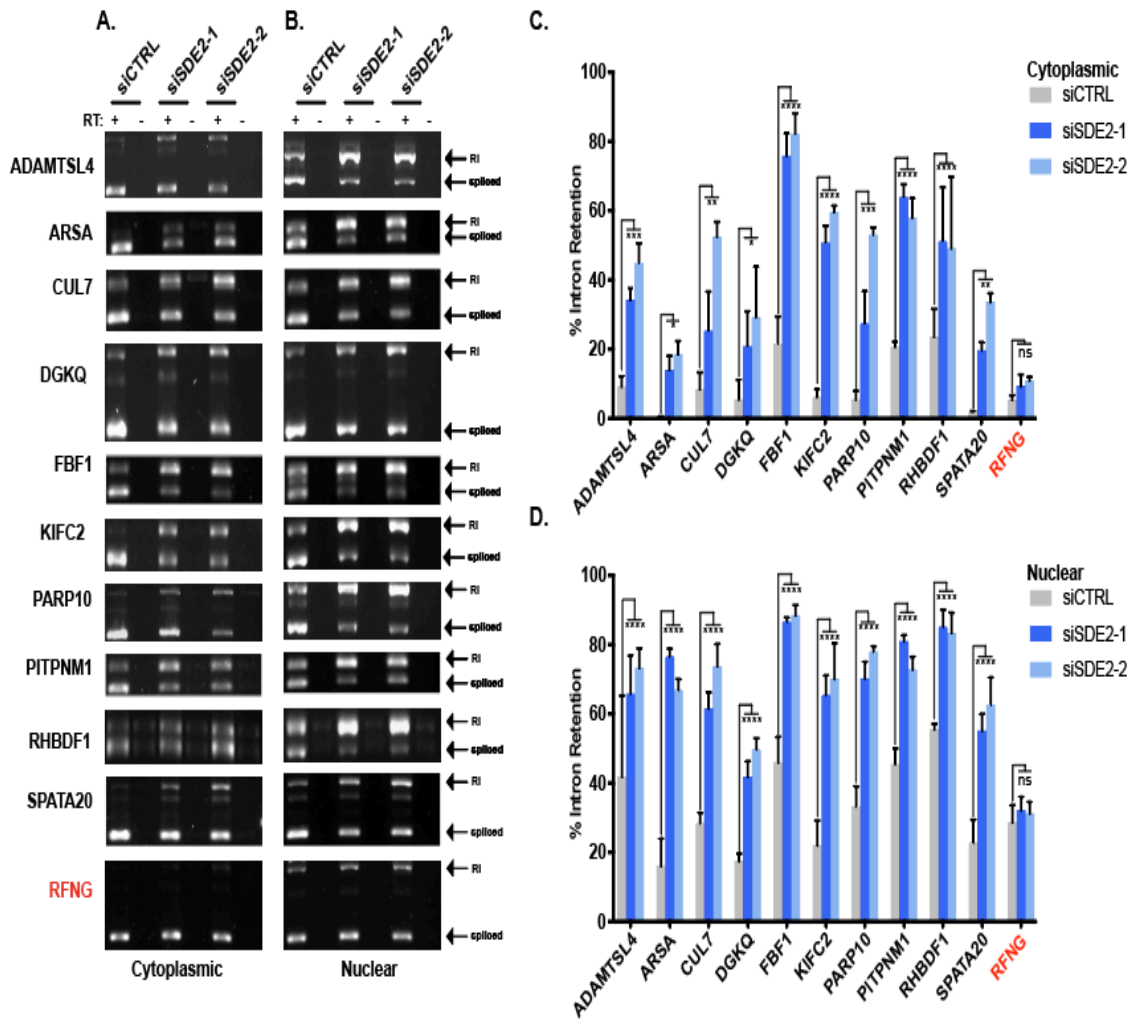
Figure 2.5. SDE2 depletion does not affect mRNA export. (A-B) DNA gels of intron retention events in the corresponding conditions from cytoplasmic and nuclear fractions of HeLa cells. (C-D) Graphs show mean percent intron retention in each condition relative to total transcript (intron retained + spliced) +/- SD. n=3 for all quantifications and experiments, P-values denoted by comparing siCTRL vs siSDE2-1 or siSDE2-2 by two-way ANOVA followed by Dunnet's multiple comparison test, P ≤ 0.01 denoted by **, P ≤ 0.001 denoted by ***, and P ≤ 0.0001 denoted by ****, ns = not significant. For RI events with different levels of P-values between siSDE2-1 and siSDE2-2, the least significant value is shown.

Previous studies have demonstrated that transcripts with RIs can exhibit defects in nuclear mRNA export (Boutz, Bhutkar, and Sharp 2015; Braunschweig et al. 2014).   Therefore, we analyzed the cellular localization of 10 of the transcripts containing an increase in RI events by RT-PCR. Following SDE2 knockdown, we fractionated HeLa cells into cytoplasmic and nuclear fractions and analyzed ADAMTSL4, ARSA, CUL7, DGKQ, FBF1, KIFC2, PARP10, PITPNM1, RHBDF1, SPATA20, and RFNG RI events by RT-PCR. Consistently, we found that all transcripts containing increased RI events and RFNG were localized in both the cytoplasm and the nucleus suggesting there was not a universal defect in mRNA export (**Figure 2.5A-D**).

Our data demonstrated that SDE2 functions to regulate pre-mRNA splicing, yet it was unclear how SDE2 activity compared to other established splicing factors. Using publicly available RNA sequencing data from the ENCODE Project, we asked whether the percent of dAS RI events following SDE2 knockdown was similar to the percent of RI events following the knockdown of other known splicing factors including TIA1, SRSF1, U2AF2, PCBP1, PCBP2, PTBP1, SRSF7, and FUS (Dunham et al. 2012). To do this, we calculated the number of RI events in each knockdown condition using a custom pipeline and divided that by the total number of introns within the genome to calculate the percent of RI each condition (Appendix B, **Figures AB1 A-B, AB2 A**). We calculated the percent RI for each SDE2 siRNA individually, for the RI events shared between both SDE2 siRNAs (intersection), and for the RI events in both SDE2 siRNAs combined (union).

Regardless of which way we analyzed the SDE2 knockdown condition, the data demonstrated that between 0.5-1.95% of introns within the genome demonstrate an increase in retention in the absence of SDE2. This is consistent with the percent RI following knockdown of any of the known splicing factors using the data from the ENCODE Project, highlighting a role for SDE2 in the regulation of AS (**Figure 2.6A**).

To determine whether a specific subset of introns was differentially affected by loss of SDE2, we analyzed the cis-characteristics of the RI events. Initially, we noted that several of the transcripts demonstrated dAS RI events located towards the 3' end of the transcript (**Figure 2.3A**). Therefore, we asked whether this 3' bias was a consistent feature amongst all of the significant RI-containing transcripts. To address this, we computed the position of each RI event relative to the gene body length of each gene for both significant and non-significant dAS RI events (Appendix B, **Figure AB3 B**). While 8 of our 10 analyzed transcripts (ADAMTSL4, CUL7, DGKQ, FBF1, KIFC2, PARP10, PITPNM1 and SPATA20) all demonstrated an RI event located near the 3' end of the transcript, the position of the remainder of the dAS RI events was uniform across the gene body, suggesting that loss of SDE2 does not lead to a significant positional bias in splicing efficiency (**Figure 2.6B**).
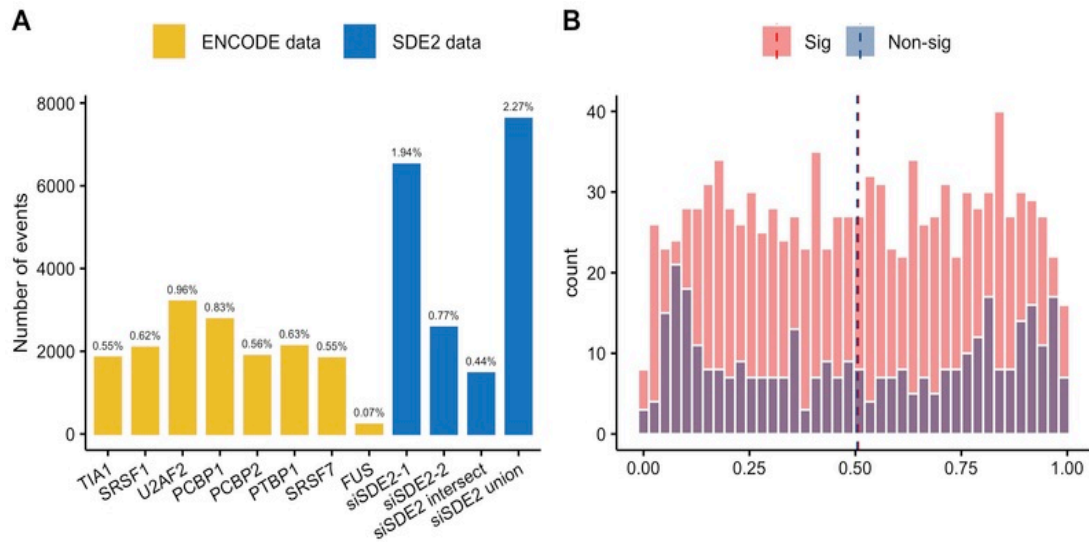
Figure 2.6. Comparison of total RI events between splicing factor depletions and position of SDE2-mediated RI events. (A) Comparison of the number and percentage of increased significant RI events (P < 0.05) in knockdown between ENCODE data and SDE2 data. The percentages represent the fraction of introns in the whole genome affected by the knockdown of the corresponding splicing factor. (B) Histogram of the Location of the RIs between significant (Sig) and non-significant (Non-sig) events. On the X-axis, '0' corresponds to the 5' end of the transcript and the '1' corresponds to the 3' end of the transcript. Position of the RI is determined using the center of the RI. Dashed lines mark the median of the fraction in each group.

In addition to the perceived positional bias, we also noticed that the introns of all 10 RI events we had initially validated were relatively short, between 69-327 nucleotides in length (**Figure 2.3A**). Therefore, we asked whether the introns amongst the significant RI events were overall shorter in length than the introns of the non-significant RI events, as shorter introns could affect spliceosomal assembly. We found that in the absence of SDE2, the significant RI events were on average shorter in length than the non-significant RI events.

70

Specifically, the significant RI events had a median length of 136 nucleotides whereas the non-significant RI events had a median length of 1,398 nucleotides. However, the non-significant group of RI events had almost half of the number of events as the significant RI events (1,265 significant events vs 717 non-significant events), raising the possibility that the unequal distribution of events may have skewed the analysis. To account for this possibility, we included a third group of introns for comparison. This group consists of all introns from expressed genes in our RNA-sequencing analysis that were not alternatively spliced (i.e. not identified in our AS pipeline) and consists of 333,400 introns (referred to as 'non-AS') (Appendix B, **Figure AB2 A-B, AB3 C**). Even when compared to this much larger group, the introns in the significant RI events group were still significantly shorter than all non-AS introns combined (significant RI events-136 nucleotides vs non-AS events-1,438 nucleotides) (**Figure 2.7A**). Short introns have often been associated with differential GC content and weaker splice site strength (Braunschweig et al. 2014; Galante et al. 2004; Sakabe and de Souza 2007). Therefore, we asked whether the introns in the significant RI events displayed changes in the GC content as compared to either the non-significant RI events or the non-AS introns. As with our earlier analysis, we considered the possibility that the analysis of the GC content in the significantly retained introns may be skewed simply because the introns were on average significantly shorter than both of the other comparison groups. Therefore, we split the non-AS intron group into non-AS short and non-AS long
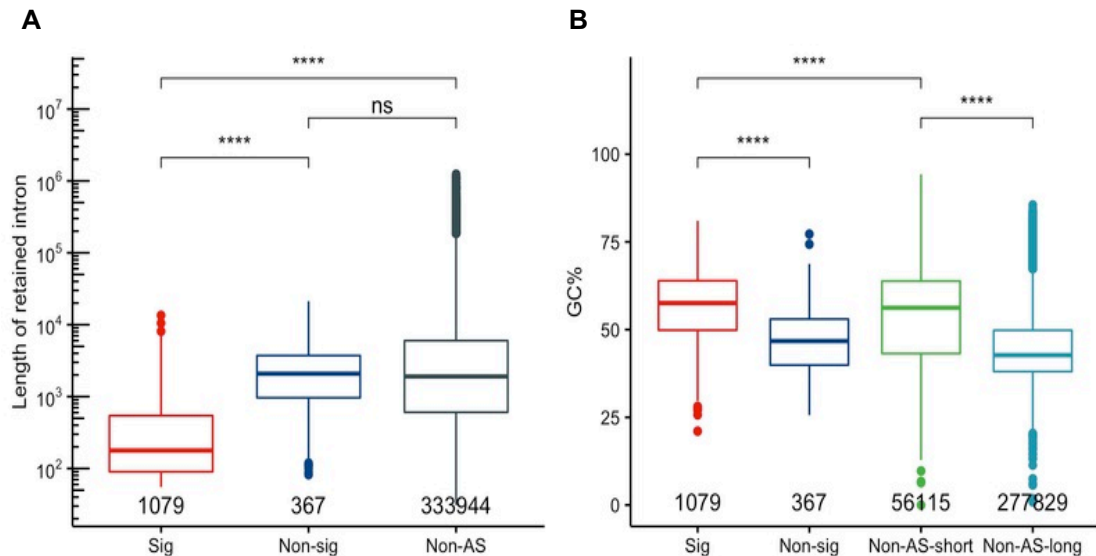
Figure 2.7. Characterization of RI events following SDE2 depletion. (A) Boxplot of the Length of the RIs among Sig, Non-sig, and Non-AS group. The number of events in each group is annotated. Length is log10 transformed on y-axis. (ns: not significant, * p<0.05, **p<0.01, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test). (B) Boxplot of the GC content of the RIs among Sig, Non-sig, Non-AS-short and Non-AS-long groups. The number of events in each group is annotated. (**** P < 0.0001, Mann-Whitney U test).

introns (Appendix B, **Figure AB2 A-B, AB3 D**). Here, 'short' and 'long' were determined using the box and whisker plot from the significant RI events where 'short' was defined as intron lengths less than or equal to the 75th percentile (≤400nt) and 'long' was defined as lengths greater than the 75th percentile (>400nt). Notably, the significant RI events had a GC content of 62% whereas the non-significant, non-AS short, and non-AS long introns had GC contents of 50%, 56%, and 42% respectively (**Figure 2.7B**). These data highlight a high GC content as a defining feature of the significant RI events. To determine whether

Figure 2.8 Boxplot of the splice site scores on both the 5' (ss5) and 3' (ss3) end of the RIs among the four groups. The plot is only showing the distribution of positive scores, but the statistical test included all of the data. (ns: not significant, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test).

the shorter length and higher GC content affected the strength of the 5' and/or 3' splice site we calculated the maximum entropy (MaxENT) score (G. Yeo and Burge 2004) of each splice site within each intron. A higher MaxENT score is indicative of a stronger splice site. The significant RI events have significantly lower MaxENT scores as compared to non-significant RI events (MaxENT = 8.05 vs 8.55). Moreover, while the MaxENT scores of the significant RI events were lower than the non-AS short introns at the 3' splice site (MaxENT = 8.05 vs. 8.25), there was no difference in MaxENT scores at the 5' splice site (MaxENT = 8.05 vs 8.1) (**Figure 2.8,** Appendix B**, Figure AB3 E**), suggesting that SDE2 may be especially critical for 3' splice site definition.

Taken together, our data define the subset of introns that are retained in the absence of SDE2 as short, GC-rich, and containing weak 3' splice sites. All

of these analyses were done using the set of RI events that were shared between each SDE2 siRNA. However, the overall results were identical if we combined all of the RI events between both siRNAs (union) (**Figure 2.9 A-E**). In addition, the characteristics of these RI events were consistent at 120 hours after knockdown, albeit with an increase in total events by this time (**Figures 2.10 A-E and 2.11 A-E**).
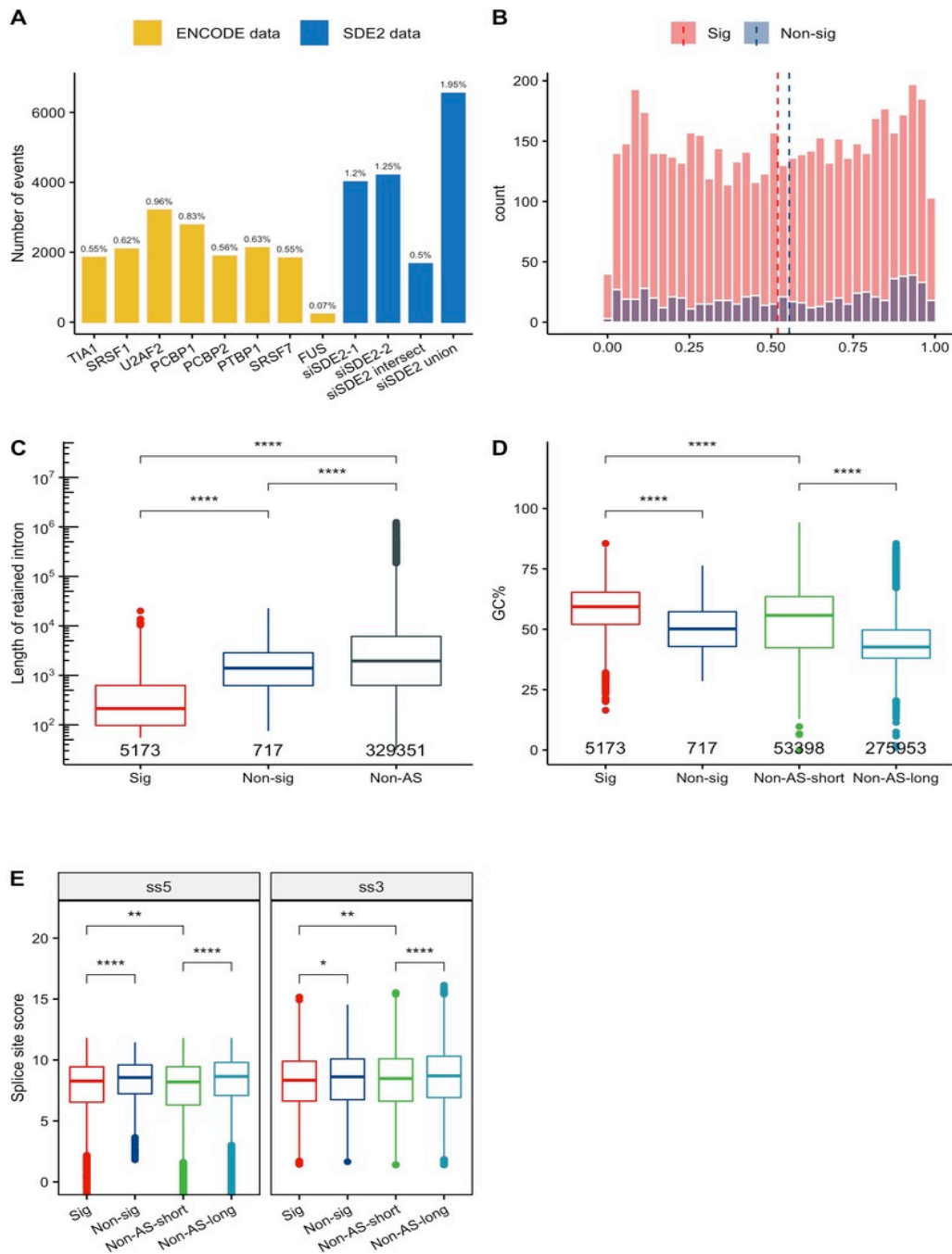
Figure 2.9. Characterization of the significant retained intron events using the union of the results of the two individual siRNAs for SDE2 at the 72h timepoint. (A) Comparison of the number and percentage of increased significant RIevents (P < 0.05) in knockdown between ENCODE data and SDE2 data. The percentages represent the fraction of introns in the whole genome affected by the SDE2 knockdown.

75

Figure 2.9 (continued). (B) Histogram of the location of the retained introns between significant (Sig) and non-significant (Non-sig) events. Non-sig events are represented by the intersected pool. On the X-axis '0' corresponds to the 5' end of the gene and the '1' corresponds to the 3' end of the gene. Position of the retained intron is determined using the center of the retained intron. Dashed lines mark the median of the fraction in each group. (C) Boxplot of the Length of the retained introns among Sig, Non-sig and Non-AS group. The number of events in each group is annotated. Length is log10 transformed on y-axis. (ns: not significant, * p<0.05, **p<0.01, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test). (D) Boxplot of the GC content of the retained introns among Sig, Non-sig, Non-AS-short and Non-AS-long groups. The number of events in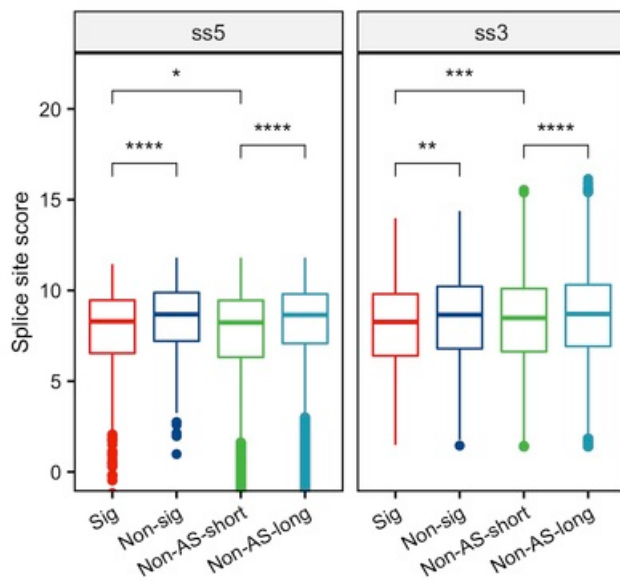 each group is annotated. (**** P < 0.0001, Mann-Whitney U test). (E) Boxplot of the splice site scores on both the 5' and 3' end of the retained introns among the four groups. The plot is only showing the distribution of positive scores, but the statistical test included all of the data. (ns: not significant, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test).

Figure 2.10. Characterization of the retained intron events with significant events defined as the intersected results of the two individual siRNAs for SDE2 at the 120h timepoint. (A) Comparison of the number and percentage of increasedsignificant RI events (P < 0.05) in knockdown between ENCODE data and SDE2 data.

77

Figure 2.10 continued. The percentages represent the fraction of introns in the whole genome affected by the SDE2 knockdown. (B) Histogram of the location of the retained introns between significant (Sig) and non-significant (Non-sig) events. Non-sig events are represented by the intersected pool. On the X-axis '0' corresponds to the 5' end of the gene and the '1' corresponds to the 3' end of the gene. Position of the retained intron is determined using the center of the retained intron. Dashed lines mark the median of the fraction in each group. (C) Boxplot of the Length of the retained introns among Sig, Non-sig and Non-AS group. The number of events in each group is annotated. Length is log10 transformed on y-axis. (ns: not significant, * p<0.05, **p<0.01, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test). (D) Boxplot of the GC content of the retained introns among Sig, Non-sig, Non-AS-short and Non-AS-long groups. The number of events in each group is annotated. (**** P < 0.0001, Mann-Whitney U test). (E) Boxplot of the splice site scores on both the 5' and 3' end of the retained introns among the four groups. The plot is only showing the distribution of positive scores, but the statistical test included all of the data. (ns: not significant, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test).

Figure 2.11. Characterization of the retained intron events with significant events using the union of the results of the two individual siRNAs for SDE2 at the 120h timepoint. (A) Comparison of the number and percentage of increased significant RI events (P < 0.05) in knockdown between ENCODE data and SDE2 data.

Figure 2.11 continued. The percentages represent thefraction of introns in the whole genome affected by the SDE2 knockdown. (B) Histogram of the location of the retained introns between significant (Sig) and non-significant (Non-sig) events. Non-sig events are represented by the intersected pool. On the X-axis '0' corresponds to the 5' end of the gene and the '1' corresponds to the 3' end of the gene. Position of the retained intron is determined using the center of the retained intron. Dashed lines mark the median of the fraction in each group. (C) Boxplot of the Length of the retained introns among Sig, Non-sig and Non-AS group. The number of events in each group is annotated. Length is log10 transformed on y-axis. (ns: not significant, * p<0.05, **p<0.01, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test). (D) Boxplot of the GC content of the retained introns among Sig, Non-sig, Non-AS-short and Non-AS-long groups. The number of events in each group is annotated. (**** P < 0.0001, Mann-Whitney U test). (E) Boxplot of the splice site scores on both the 5' and 3' end of the retained introns among the four groups. The plot is only showing the distribution of positive scores, but the statistical test included all of the data. (ns: not significant, ***P < 0.001, ****P < 0.0001, Mann-Whitney U test).
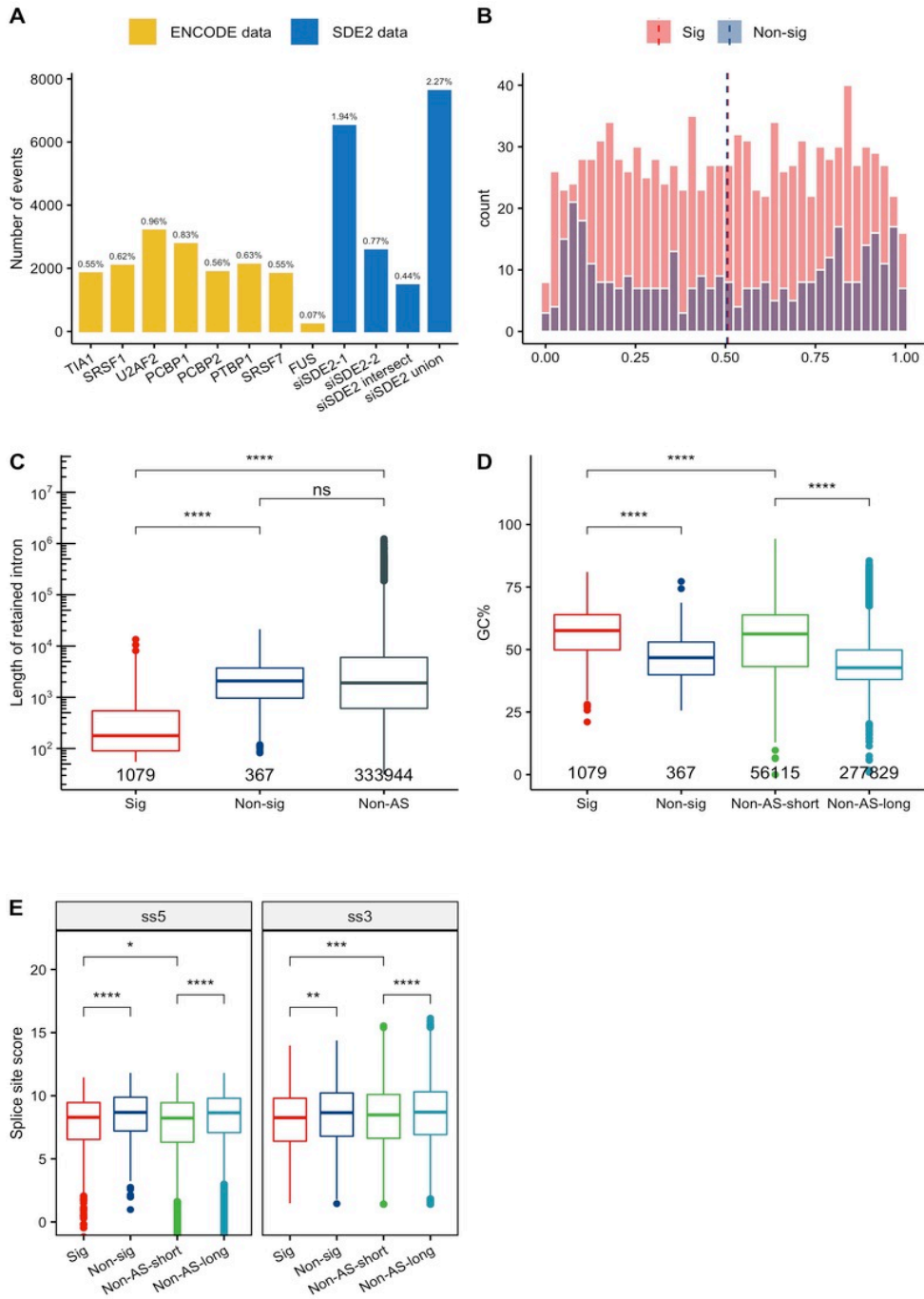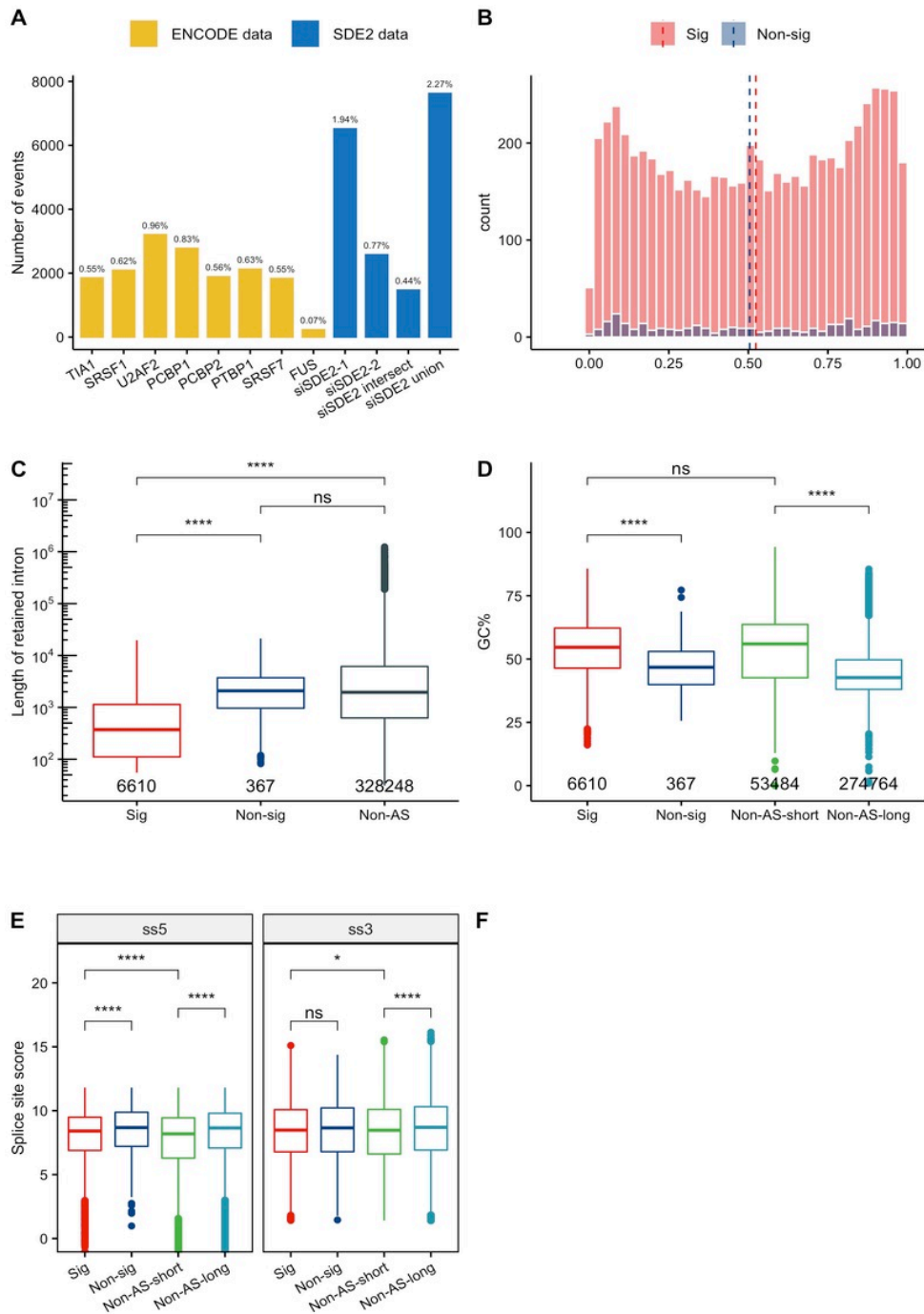
Section 4: Discussion

Pre-mRNA splicing is a highly dynamic process that involves the transient association of hundreds of protein and RNA factors with chromatin to ensure efficient pre-mRNA processing. The transient nature of the interaction of each binding protein with RNA makes the complete isolation and characterization of the human spliceosome a challenging task. Therefore, there are likely many proteins and RNAs that function as auxiliary factors throughout the process that have yet to be identified. Given that mutations in splicing factors and cis-regulatory elements are strongly associated with human disease, fully defining the catalogue of proteins responsible for splicing may lead to the discovery of genes implicated in disease pathology. Here, we demonstrate that SDE2 is required for efficient pre-mRNA splicing. SDE2 depletion leads to a widespread increase in AS, especially RI, highlighting the requirement of SDE2 for proper gene expression in mammalian cells.

We have found that SDE2 interacts with the U2 snRNP complex and, likewise, others have identified SDE2 in biochemical purifications of human and yeast spliceosomal complexes (Bayne et al. 2008; Bessonov et al. 2008; W. Chen et al. 2014; Fica et al. 2019a; Thakran et al. 2018). SDE2 is not a constitutive component across all spliceosomal purifications, suggesting that SDE2 may be a dynamic factor with unique functions in pre-mRNA processing. We demonstrate that loss of SDE2 led to an increase in dAS, with RI being the most significantly

enriched dAS event type. Moreover, our bioinformatics pipeline has allowed us to capture both novel RI events as well as annotated RI events. One of the defining characteristics of the significant dAS RI events is that they are shorter in length than introns that were properly spliced, suggesting that SDE2 may be a *trans*-acting splicing factor that functions to define splice site boundaries at short introns. Given SDE2's interaction with components of the U2 snRNP, SDE2 function may be especially critical to define the boundaries of short introns that have high GC content and weak splice sites. Although the RI events we identified following loss of SDE2 were shorter in length, they were not always the smallest introns within a given AS transcript. Likewise, we identified short introns that were retained, but were not dAS following loss of SDE2. These data suggest that the short introns that are not retained, or short RI events that did not demonstrate dAS, have strong splice sites and lower GC content and are therefore efficiently spliced during mRNA processing in the absence of SDE2. Alternatively, there may simply be additional defining features for the significantly retained introns as compared to the non-significantly retained introns that are yet to be uncovered. Our analyses here have focused on RI events, however, as computational tools continue to expand, our ability to detect AS events may improve. As a result, we may uncover novel AS events in the future, that are significantly enriched in the absence of SDE2 beyond RI events.

The predicted protein domains contained within SDE2 served as an initial indicator for SDE2's functions. In addition to the IDR and SAP domains, SDE2 also

contains an N-terminal ubiquitin-like domain (UBL) (Jo et al. 2016).   Proteins containing UBL domains are often enzymatically cleaved, generating a ubiquitin peptide that is then covalently linked to a substrate protein to regulate various cellular pathways, including RNA processing and ribosome biogenesis (Chanarat S 2018; Shcherbik and Pestov 2010).   In fission yeast, the UBL domain of Sde2 is cleaved by the deubiquitinating enzyme paralogs Ubp5 and Ubp15, creating an N-terminal UBL fragment and a C-terminal fragment (Thakran et al. 2018). Likewise, in mammalian cells, SDE2 has been purported to be cleaved in response to DNA damage, resulting in both an N-terminal UBL fragment and a C-terminal SAP fragment (Jo et al. 2016).   While the exact function of the liberated UBL domain has not been explored, recent studies highlight a role for the C-terminal fragment in pre-mRNA splicing.   Specifically, SDE2 was identified in a crystal structure of the human spliceosomal P complex associated with a precursor mRNA substrate. These studies suggested that cleavage of the UBL may be required to allow SDE2 to associate with spliceosomal complex proteins (Fica et al. 2019a). Our experiments were performed in the absence of exogenous DNA damage, which limited our ability to detect this cleavage event. However, our data support a role for SDE2 in pre-mRNA splicing through an interaction with both the U2 snRNP complex and the auxiliary factor reported in the resolved crystal structure, Cactin.

Here, we demonstrate that loss of SDE2 leads to significant defects in pre-mRNA processing, leading to an increase in AS and ultimately, cell death. While

we describe RIs as the most significant event detected following SDE2 knockdown, our RNA sequencing analysis was performed on polyA+ purified RNA. Therefore, if defects in splicing led to transcripts that were inherently unstable and preferentially degraded, we would be unable to detect them in our analysis. Likewise, the purification of spliceosomal complexes has placed SDE2 in various complexes throughout the splicing process. If SDE2 were also to function in late stages of the splicing reaction, it is plausible that defects in the transesterification reactions would lead to the formation of RNA fragments that might not be recovered in our polyA+ purified RNA. Thus, we may be underestimating the effects of loss of SDE2 on the transcriptome. Regardless, loss of SDE2 led to significant and widespread increases in dAS, revealing that SDE2 is a previously uncharacterized human splicing factor

CHAPTER THREE: SDE2 IS REQUIRED FOR RIBOSOME BIOGENESIS

Section 1: Introduction

Mammalian cells contain thousands of RNA-binding proteins (RBPs) often fulfilling critical roles in a multitude of cellular processes from transcription to translation. These RBPs harbor structural domains that directly interact with their cognate RNA targets, including RNA recognition motifs (RRMs), K-Homology domains (KH), cold shock domains (CSD), or Zinc finger CCHC domain(Lunde, Moore, and Varani 2007). However, more recent biochemical purifications of RNA have identified RBPs containing no known RNA binding domains (Castello et al. 2016; C. He et al. 2016). These findings not only highlight unexplored protein interfaces supporting RNA binding, but may also indicate novel biological functions in RNA biology. Thus, although the involvement of RBPs across cellular processes is ubiquitous, there remain many enigmatic RBPs that are yet to be functionally defined (Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume AM, Schwarzl T, Curk T, Foehr S, Huber W, Krijgsveld J 2015). Identifying these RBPs and their functions provide an opportunity to further delineate the most complex and energy consuming pathways in the cell: ribosome biogenesis.

Ribosome biogenesis relies on 80 different ribosomal proteins (r-proteins), hundreds of *trans*-acting factors, and four non-polyadenylated ribosomal RNAs (rRNAs) 18S, 5.8S, 28S, and 5S. Of these rRNAs 18S, 5.8S, and 28s are transcribed as a single 47S polycistronic precursor transcript from the tandem-

repeat ribosomal DNA (rDNA) arrays by RNA polymerase I (Pol I). This 47S precursor undergoes a series of cleavage events and nucleolytic trimmings to remove external (5'ETS, 3'ETS) and internal (ITS1, ITS2) transcribed spacers from the transcript to generate the mature rRNAs (18S, 5.8S, and 28S) (Baßler and Hurt 2019). The rRNA maturation process relies heavily on two major classes of small nucleolar RNAs (snoRNAs), the H/ACA snoRNAs and the C/D snoRNAs. snoRNAs are bound by additional proteins to form small nucleolar ribonucleoprotein (snoRNP) complexes which function to regulate the chemical modification, folding, and/or cleavage of rRNA (Ojha, Malla, and Lyons 2020). Mature rRNAs are critical for ribosome structure, with 18S rRNA essential for the formation of the small ribosomal subunit (SSU) and the 5.8S, 28S, and 5S essential for formation of the large ribosomal subunit (LSU). Together, these rRNAs provide the framework for the assembly of the ribosomal proteins, and ultimately the formation of a fully functional and translationally competent ribosome. Thus, rRNA maturation is a requisite step in ribosome biogenesis that is highly dependent on an array of RBPs and snoRNAs

Ribosome biogenesis is an essential process within the cell that requires the function of hundreds of unique RBPs and trans-acting protein factors. Approximately 25% of all known RBPs are mutated in various human diseases including neurological pathologies and cancer (Gebauer F, Schwarzl T, Valcárcel J 2021), and ribosome-related diseases, called ribosomopathies, often present with severe clinical phenotypes, including early death. Therefore, fully defining the

catalogue of RBPs required for fundamental pathways like ribosome biogenesis is essential to our understanding of both basic biological processes and complex human diseases. Silencing defective 2 (SDE2) was originally identified in *Schizosaccharomyces* pombe and has since been linked to several cellular processes in eukaryotes including heterochromatin formation, telomere silencing, DNA replication, and mRNA processing (Fica et al. 2019a; Jo et al. 2016; Rageul et al. 2019; Sugioka-Sugiyama and Sugiyama 2011b; Thakran et al. 2018). Here, we identify SDE2 as a previously uncharacterized RBP, with critical roles in both rRNA processing and ribosome biogenesis. Moreover, we demonstrate that concomitant with increased AS, loss of SDE2 function causes defects in translation, and ultimately results in the complete loss of cellular viability.

Section 2: Materials and Methods

Cell Lines

All cell lines were submitted for Short Tandem Repeat (STR) analysis by ATCC and certificates of authentication can be provided upon request. HeLa, U2OS, and 293FT cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (10% FBS, 1% penicillin/streptomycin). U2OS cells were engineered to stably express H2B-mCherry and were a gift from Dr. Neil J. Ganem. RPE cells were cultured in DMEM/F12 (10% FBS, 1% penicillin/streptomycin). Cell culture media and supplements were obtained from Gibco Invitrogen and all plasticware

87

came from Corning (Corning, NY). All cells were maintained at 37°C in a humidified incubator at 5% CO_2.

Transfections and siRNA

Cells were seeded at 5 X 10^4 cells per well in a 6 well plate and reverse transfected with ON-TARGETplus siRNA (Dharmacon). Cells were transfected with either 100nM (Non-targeting control, siSDE2-2) or 20nM (siSDE2-1) siRNA using Lipofectamine RNAiMax diluted in Opti-MEM according to the manufacturer's instructions. The next day, the siRNA solution was removed from cells and replaced with fresh media. Cells were collected for various downstream applications after 3 days, or collected, subjected to a second reverse transfection and plated, and collected after another 2 days (5 days total from the time of the initial reverse transfection). siRNA target sequences: siSDE2-1 (CUACUAAAUCUCAAACAGAdTdT), siSDE2-2 (GGAAGCUUGUAGAACCCAAdTdT), ON-TARGETplus Non-targeting siRNA #1 (UGGUUUACAUGUCGACUAA).

Antibodies and Plasmids

The following antibodies and plasmids were used where indicated. SDE2 (Bethyl Laboratories A302-098A and A302-099A), GAPDH (Santa Cruz Biotechnologies sc-47724), U2AF1 (Bethyl Laboratories A302-079A), Puromycin

(EMD Millipore MABE343), Histone 4 (Active Motif #39269), eIF2 Cell Signaling Technology 9722S), Phospho- eIF2 Ser51 Cell Signaling Technology 9721S), 4E-BP1 (Cell Signaling Technology 9644S), Phospho-4E-BP1-Thr37/46 (Cell Signaling Technology 2855S), FBL (Abcam ab5821), Tubulin (Cell Signaling Technology 2125S), p53 (Santa Cruz Biotechnology sc-126), p21 (Santa Cruz Biotechnology sc-6246) . The following plasmids were used: pLKO.1.

Western Blotting

Western blots were performed using standard protocols. Briefly, cells were collected by trypsinization and washed with ice-cold 1XPBS. Samples were then lysed in 2X sample buffer and sonicated in a water bath at 4°C for 5 minutes (20-second pulse on/30-second pulse off at 100% amplitude), then boiled at 95°C for 10 minutes. Soluble protein lysates were then analyzed by western blot using standard SDS-PAGE techniques and transferred onto PVDF membranes. Membranes were blocked in TBS-T (1X TBS, 0.1% Tween-20) containing 5% milk for one hour and then incubated overnight at 4°C with the appropriate primary antibodies. Following overnight incubation with primary antibodies, membranes were washed 3X in TBS-T for 10 minutes each, incubated with peroxidase conjugated secondary antibodies for 1 hour, then washed 3X in TBS-T for 10 minutes each and visualized using enhanced chemiluminescence reagents from BioRad and Thermo Fisher.

UV Crosslinking and Immunoprecipitation

CLIP was performed as previously described with the following modifications (Huppertz et al. 2014). Briefly, cells were UV crosslinked at 254 nm at 150 mJ/cm^2. Immediately after crosslinking, cells were collected via cell lifter and pelleted. Pellet was resuspended in 1 mL lysis buffer (50mM Tris-HCl pH 7.4, 100mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1:1000 SUPERase-IN) and vortexed, then placed on ice. 4 μL of Turbo DNase and 10 μL of RNase I diluted 1:500 were added to the mixture, and the solution was incubated in a thermomixer for 3 minutes at 37°C and 1100 rpm. Following this, 5 μL of SUPERase-IN was added and the solution was placed on ice for 3 minutes. Next, the solution was centrifuged for 10 minutes at 14,000 rpm at 4°C in a tabletop centrifuge. The resulting supernatant was then immunoprecipitated by incubation with beads for 120 minutes at 4°C. Beads were previously prepared by washing in lysis buffer 3X, then incubating with the appropriate antibody for 60 minutes at RT. Here, we used 3 μg antibody and 30 μL Dynabeads A per condition. Beads were removed from solution by magnet, and washed 2X with a high salt wash (50mM Tris-HCl ph7.4, 1M NaCl, 1mM EDTA, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate), followed by washing 3X with PNK buffer (20mM Tris-HCl ph7.4, 10mM MgCl$_2$, 0.2% Tween-20). After removing the last wash buffer, beads were incubated for 5 minutes with 4 μL of a radiolabeling mixture consisting of 0.2 μL T4 PNK and 0.4 μL 10X PNK Buffer, 0.4 μL gamma-32-ATP, 2.8 μL H$_2$O, and 0.2 μL SUPERase-IN. The beads were separated from the radiolabeling mixture

by magnet, and the mixture was discarded. Beads were washed once more with PNK buffer. Beads were then resuspended in 30 μL Laemmli's sample buffer and boiled for 5 minutes. After boiling, samples were again placed on a magnet, and the supernatant was subjected to western blotting following standard procedures. After transfer of gel to a PVDF membrane using the iBlot 2 gel transfer device, membrane was wrapped in plastic wrap and set into a light-protected radiographic film cassette. Image was developed using a GE Healthcare Typhoon FLA 7000.

eCLIP Protocol and Analysis

eCLIP was performed to the manufacturer's instructions with the eCLIP Kit (Eclipse Bioinnovations #ECEK8-0001) using antibodies for SDE2 and U2AF1 listed above. eCLIP-Seq short read datasets for SDE2 and U2AF IPs and their corresponding inputs were first quality and adapter trimmed using trimmomatic (Bolger, Lohse, and Usadel 2014) and then assessed for quality using FastQC (Andrews 2015) and MultiQC (Ewels et al. 2016) packages. Trimmed libraries were analyzed using the published analytical pipeline strategy described in (Van Nostrand et al. 2020). Briefly, PCR duplicate reads were collapsed by Unique Molecular Identifier (UMI) using the umi-tools package (Ewels et al. 2016). A database of repetitive elements was constructed included human RepBase sequences (Bao, Kojima, and Kohany 2015), snoRNA sequences from snoDB (Bouchard-Bourelle et al. 2020), tRNA sequences from GtRNAdb (Chan and Lowe 2016), snRNA sequences extracted from the GRCh38 reference human genome

using the GENCODE v27 annotation coordinates for genes with biotype 'snRNA', and the rRNA sequences with NCBI accessions U13369.1 and NR_046235.3. Reads were aligned against the repetitive element database using STAR and reads were separated based on whether they align against this database. Reads that did not align to repetitive elements were then aligned against the GRCh38 reference genome using STAR and analyzed with the Yeo lab clipper pipeline (Clipper n.d.) to characterize SDE2 and U2AF binding events to the non-repetitive genome.

Reads mapping to the repetitive element database were further analyzed to identify which classes and families of repetitive elements were enriched in SDE2 and U2AF IP over input. First, the Yeo lab repetitive element analysis pipeline (Repetitive-Element-Mapping n.d.) was employed to provide family level quantification of different repetitive element classes. Because this default pipeline only provides summary information on the family level, a custom analysis pipeline was developed in parallel to further characterize more detailed binding event information on a per-repetitive-sequence basis. Briefly, the reads aligned with STAR against the repetitive element database allowing a read to have 1000 multimaps were counted for each repetitive element sequence. Library size normalized counts were computed by dividing the number of aligned reads for each repetitive element by the total number of reads in each dataset. Repetitive sequence level log2 fold change was computed as the log2 ratio of normalized counts of IP vs corresponding input. Family (e.g. snoRNAs) and subfamily (e.g.

SNORD3 A-D) log2 fold change was computed similarly by first summing normalized counts for each family or subfamily member and then taking the log2 ratio of IP vs input. Finally, the STAR aligned reads were consolidated into per-base pileups for each repetitive element sequence, where only the 5' base position of each alignment was counted. This strategy yielded a base-resolution binding profile for each repetitive element that is not influenced by the repetitive nature of the sequence database. The code that implements this analysis pipeline is available at https://bitbucket.org/bucab/sde2/src/master/analysis/24_eclip.

Phenol Toluol Extraction (PTex)

Phenol Toluol extraction of RNA was performed as previously described (Urdaneta et al. 2019). Briefly, cells were UV crosslinked at 254 nm at 150 mJ/cm^2 and immediately collected with cell lifters. Cells were counted, and 6 X 10^6 per condition were centrifuged and resuspended in 1 mL PBS. From here, 400 μL of cells were collected for input, and the rest were mixed with 300 μL phenol, 300 μL toluene, and 200 μL 1,3-bromochloropropane (BCP). The resulting solution was shaken in a thermomixer for 1 minute at 21°C and centrifuged at 17,500 x g for 3 minutes at 4°C. The aqueous phase was collected and mixed with 5.85M guanidine isothiocyanate, 31.1mM sodium citrate, 25.6mM N-lauroyl-sarcosine, and 1% 2-mercaptoethanol and mixed well. 600 μL phenol and 200 μL BCP were added to this solution, and it was mixed and centrifuged as before. 3/4 of the resulting aqueous phase and 3/4 of the organic phase were removed and

discarded; the remaining interphase was mixed with 200 μL 100% ethanol and 400 μL water, followed by 400 μL phenol and 200 μL BCP. The resulting solution was mixed and centrifuged as before. Again, 3/4 of the aqueous phase and 3/4 of the organic phase were removed from the solution after centrifugation. The remaining interphase was mixed with 4.5 mL of ethanol and precipitated at -80°C overnight. The next day, tubes with ethanol and the remaining interphase were centrifuged at 17,500 x g for 30 minutes at 4°C. Ethanol was carefully decanted and tubes dried for 5 minutes in the chemical fume hood, followed by resuspension in 2X Laemmli sample buffer and subsequent western blotting. To increase detection by western blot, each condition was run in duplicate (6 X 10^6 cells per tube, 12 X 10^6 cells total) and combined in the last step by resuspension in 2X Laemmli sample buffer to be run in the same well by western blot.

Immunoprecipitation

5 μg of each antibody was incubated with Dynabeads Protein A (Invitrogen 1001D) suspended in NETN buffer (150mM NaCl, 20mM TRIS pH 8.0, 0.5mM EDTA, 0.5% NP-40) supplemented with protease inhibitor cocktail (Sigma-Aldrich P8340) at a 1:100 ratio for one hour at room temperature with gentle rotation. Cells were lysed in NETN buffer and sonicated in a water bath at 4°C for 15 minutes (20-second pulse on/40-second pulse off at 50% amplitude). Lysate was centrifuged at 12,000 x g for three minutes at 4°C and supernatant was collected. Supernatant was added to Dynabeads-antibody mixture and allowed to incubate

94

for 16 hours at 4°C with gentle rotation. After incubation, beads were collected on a magnet and supernatant was discarded. Beads were further washed 3X with NETN buffer supplemented 1:100 with protease inhibitor complex. After final wash, beads were collected on a magnet, wash buffer was discarded, and beads were resuspended in standard western blot sample buffer.

Lentiviral Transduction

293 FT cells were seeded at a density of 5×10^5 cells per well and transfected using Fugene 6 Transfection Reagent (Promega #E2691). Standard lentiviral packaging plasmids (0.5 μg pMD2.G, 1.5 μg psPAX2) and 2 μg of indicated plasmid DNA (pLKO.1 empty vector or shSDE2-30, Sigma Aldrich TRCN0000370430) were incubated with Fugene 6 diluted in Opti-MEM according to the manufacturer's instructions. Transfection media was removed after 8 hours and replaced with fresh media. Cells were allowed to proliferate for 48 hours before supernatant was collected and filtered using 0.45 μm filters (Corning #431220). Viral supernatants were then directly used to infect target cells. RPE cells were seeded at 7×10^4 cells per well one day prior to 293 FT viral media collection. RPE media was removed and replaced with filtered viral media (either undiluted or diluted 1:10 with DMEM/F12) from 293 FT cells. After 24 hours, media was removed from RPE cells and replaced with fresh media. After a subsequent 24 hours, 7 μg/mL puromycin was added to media for 72 hours. After 72 hours, media with puromycin was removed and replaced with fresh media, and the RPE cells

95

were allowed to proliferate for 7 days before being counted and processed for downstream application.

RNA electrophoresis and Northern Blotting

5 μg of RNA in RNA loading buffer [70% v/v formamide (Fisher Scientific BP228), 7.4% v/v formaldehyde (Fisher Scientific BP531), 20mM HEPES/5mM sodium acetate/1mM EDTA, pH to 7.0] was heated at 85°C for 10 minutes, followed by a 10 minute incubation on ice. RNA samples were then separated on a denaturing agarose gel (1.2% agarose for high molecular weight RNA species or 2.4% agarose for low molecular weight RNA species, 7% v/v formaldehyde, in HEPES/sodium acetate/EDTA buffer) for either 6 hours (2.4% agarose gel) or 16 hours (1.2% agarose gel) at 55V. Following electrophoresis, the 1.2% agarose gel was washed with water, followed by 50mM NaOH/10mM NaCl for 10 minutes, then 2.5X TBE buffer for 10 minutes, then 10X SSC (1.5M NaCl,150mM SSC) buffer. The 2.4% agarose gel was washed with water, then 10X SSC buffer for 10 minutes. After incubation with 10X SSC buffer, agarose gels were transferred to positively-charged nylon membranes (GE Healthcare RPN 203S) in 10X SSC buffer by capillary action. After transfer, membranes were pre-hybridized for 1 hour at 55°C in hybridization buffer (Invitrogen AM8669). Membranes were then incubated with DIG-labeled probe (see below) overnight at 55°C. Sequences of the probes are described in Table S1.

96

DIG-Labelling

Following transfer and UV crosslinking, membranes were incubated with Dig-labeled DNA probes. DNA probes were Dig-labeled following manufacturer's protocol from Sigma-Aldrich (03353575910) and detected with Sigma-Aldrich Anti-Digoxigenin-AP (11093274910) and Sigma-Aldrich chemiluminescent substrate CDP-Star (11685627001). Images were captured and visualized using a BioRad ChemiDoc XRS+ imaging system.

Immunofluorescence (IF) Imaging

HeLa cells were grown on glass cover slips (Corning Cat No. 2845-22). Upon processing for IF, cells were washed in PBS, pre-extracted (in 100mM NaCl, 300mM sucrose, 3mM MgCl$_2$, 10mM PIPES pH 7.0, 0.1% Triton X-100), washed with PBS, then fixed with 4% paraformaldehyde in PBS for 10 minutes at RT. The cells were then washed again with PBS and permeabilized with 0.1% Triton X-100 for 10 minutes at RT. Following this step, cells were washed with PBS and incubated with blocking buffer (0.5% BSA, 0.2% fish gelatin) for one hour at RT. The cells were then incubated with FBL antibody diluted 1:250 in blocking buffer overnight in a humidified chamber at 4°C. The following day, cells were washed with PBS and incubated with secondary antibody (Abcam ab6564) diluted in blocking buffer for 45 minutes in a humidified chamber at RT. The cells were then washed in PBS and subjected to a final wash in 2X SSC with DAPI (Sigma-Aldrich D9542) for 10 minutes, followed by Vectashield (Vector Laboratories H-1000-10)

treatment before mounting slides and imaging with a Zeiss LSM 710 confocal microscope. Images were visualized with FIJI Image software package.

Population Doubling Assays

Cells were seeded at 5×10^4 cells per well and reverse transfected with the appropriate siRNAs. Cells were allowed to proliferate for 3 days. At this time, cells were collected and counted, then 5×10^4 cells per well were re-seeded into a new plate and subjected to a second reverse transfection with the appropriate siRNAs. After another 3 days, cells were collected and counted, then used for downstream applications. To determine population doubling, we used the formula: Pop. Doubling = $\log(N_{final}/N_{initial})/\log(2)$ where N_{final} is the cell count after 3 days and $N_{initial}$ is 5×10^4 cells.

Crystal Violet Stain

Cells were washed twice with cold PBS, then fixed for 15 minutes with 100% methanol at -20°C. Methanol was removed from cells and replaced with 10% crystal violet stain, 25% methanol for 60 minutes. Crystal violet stain was removed from cells and cells were washed with water 6X and left to dry overnight. The next day, pictures were taken of the cells.

SUnSET Assays

Cells were reverse transfected with the appropriate siRNAs as described above. 60 minutes before collection, cells were pulsed with 10 μg/mL of puromycin. For our cycloheximide control, we added 10 μg/mL cycloheximide (Sigma Aldrich C7698) 10 minutes before addition of puromycin (70 minutes before collection of cells). Cells were then subject to standard western blotting procedures and puromycin incorporation was detected by the anti-puromycin antibody listed above.

Polysome Profiling

10-50% sucrose gradients (in polysome gradient buffer: 20mM Tris, pH 7.5, 0.125M NaCl, 10mM $MgCl_2$) were poured one day before use and kept at 4°C. HeLa cells were grown to 80% confluence, then on the day of collection, were treated with 100ug/ml cycloheximide for 10 minutes, then washed twice in PBS with 100ug/ml cycloheximide, before being collected. 6 X 10^6 cells were collected in each condition, and lysed in polysome gradient buffer supplemented with 2mM DTT, 100ug/ml cycloheximide, 2% NP-40, and 1X EDTA-free protease inhibitor (Millipore-Sigma 11873580001). Cells were centrifuged for 10 minutes at 16,000 X g to collect the cytoplasmic fraction, and supernatant was extracted and loaded to top of sucrose gradients. Gradients and lysates were centrifuged at 100,000 X g for 2 hours at 4°C in an ultracentrifuge (Beckman Coulter) with an SW41 Ti rotor. Following centrifugation, samples were fractionated into 12 fractions of roughly 1ml

volume each, using a BioLogic DuoFlow Chromatography System with continuous measurement of the absorbance at 254nm.

Live Cell Imaging

U2OS cells stably expressing H2B-mCherry were grown on glass-bottom 12-well tissue culture dishes (MatTek) and transfected with either scrambled or SDE2-1 siRNAs as previously described. Cells were imaged on a Nikon TE2000-E2 inverted microscope equipped with the Nikon Perfect Focus system beginning 24-48 hrs following transfection. The microscope was enclosed within a temperature- and CO_2-controlled environment that maintained an atmosphere of 37°C and 5% humidified CO_2. Fluorescence images were captured from a single focal plane every 10 minutes for 4 days with a 10X 0.5 NA Plan Fluor objective. All captured images were analyzed using NIS-Elements software.

Section 3: Results

The C-terminus of SDE2 was initially demonstrated to share a region of homology with both the splicing factor SF3A3/SF3A60, and also the SAF-A/B, Acinus and PIAS (SAP) domain (**Figure 3.1A**) (Jo et al. 2016; Sugioka-Sugiyama and Sugiyama 2011b). SF3A3 is the third subunit of the trimeric SF3A complex which together with the SF3B complex binds the U2 snRNA to facilitate spliceosomal assembly (Tanackovic and Krämer 2005). SAP domains are found in diverse nuclear and cytoplasmic proteins and function to mediate nucleic acid binding (Aravind and Koonin 2000; C. He et al. 2016; Iida, Kawaguchi, and Nakayama 2006). These regions of homology provided early evidence that SDE2 may function to regulate RNA processing. At the cellular level SDE2 is distributed across both the nucleus and the cytoplasm. This pattern of localization for SDE2 was consistent across several established cell lines, including HeLa, U2OS, 293 FT, and RPE-1 (**Figure 3.1B**). To determine whether SDE2 directly interacted with RNA, we performed UV crosslinking and immunoprecipitation (CLIP) analysis. Here, we either left cells untreated or irradiated them with UV prior to lysis and immunoprecipitation using IgG or SDE2 specific antibodies. The precipitated protein-RNA complexes were subjected to 5′ labelling using ^{32}P-γ-ATP and then visualized by autoradiography following SDS-PAGE separation and transfer to a PVDF membrane. Neither the uncrosslinked samples, nor the IgG control antibody precipitated RNA. However, SDE2 immunoprecipitation

demonstrated a signal by autoradiography, suggesting that SDE2 directly interacts with RNA (**Figure 3.1C**). To further confirm the interaction between SDE2 and RNA, we also performed a Phenol Toluol extraction (PTex) analysis (Urdaneta et al. 2019). While CLIP protocols require immunoprecipitation of the protein of interest, PTex allows for the separation of RNA, proteins, and cross-linked protein-RNA complexes in biphasic extractions. This eliminates the risk of non-specific interactions associated with antibodies in the CLIP experiments, providing an additional and potentially less biased approach to assessing protein-RNA interactions. Following crosslinking and biphasic extractions, we confirmed that the splicing factor U2AF1 was highly enriched in the ribonucleoprotein fraction while the histone protein H4 was undetectable, thus confirming the purity of the phases in the extraction process. Consistent with our CLIP experiments, we demonstrate that SDE2 is highly enriched in the fraction containing the crosslinked ribonucleoprotein complexes suggesting that like U2AF1, SDE2 is an RBP (**Figure 3.1D**).

To gain better mechanistic insight into SDE2 function and identify unique RNA targets, we performed enhanced CLIP (eCLIP) in HeLa cells (Van Nostrand et al. 2016). Using SDE2 and U2AF1 antibodies, we performed each eCLIP reaction in duplicate with corresponding size-matched inputs as controls. Using a previously published bioinformatic pipeline (Clipper n.d.; Lovci et al. 2013; G. W. Yeo et al. 2009) for analysis we found that approximately half of all usable

reads from our U2AF1 immunoprecipitations (~44%) uniquely mapped to the

genome and largely corresponded to mRNA elements (**Figure 3.1E-F**).

A.

SDE2 Protein - 451 AA

N- UBL C-Terminal Domain -C

SF3a60 homology; AA: 385-449

Telomere stability domain; AA 385-444

SAP Domain; AA 395-430

B.

	HeLa			U2OS			293 FT			RPE		
	WCL	Cyto.	Nuc.	WCL	Cyto.	Nuc.	WCL	Cyto.	Nuc.	WCL	Cyto.	Nuc.

SDE2
GAPDH
H4

C. -UV +UV

SDE2 IP / IgG IP / SDE2 IP / IgG IP

D. Input PTex

UV: + - + -

SDE2
U2AF1
H4

E.

% usable counts

■ % repetive counts
■ % unique counts

SDE2 IP-1, SDE2 IP-2, SDE2 input-1, SDE2 input-2, U2AF1 IP-1, U2AF1 IP-2, U2AF1 input-1, U2AF1 input-2

F.

Normalized Counts vs Log₂ Fold Change

● SDE2
● U2AF1

unique proxintron
unique 3' UTR
unique 5' UTR
unique CDS
unique intergenic
unique noncoding proxintron

unique CDS
unique 5' UTR
unique 3' UTR
unique intergenic
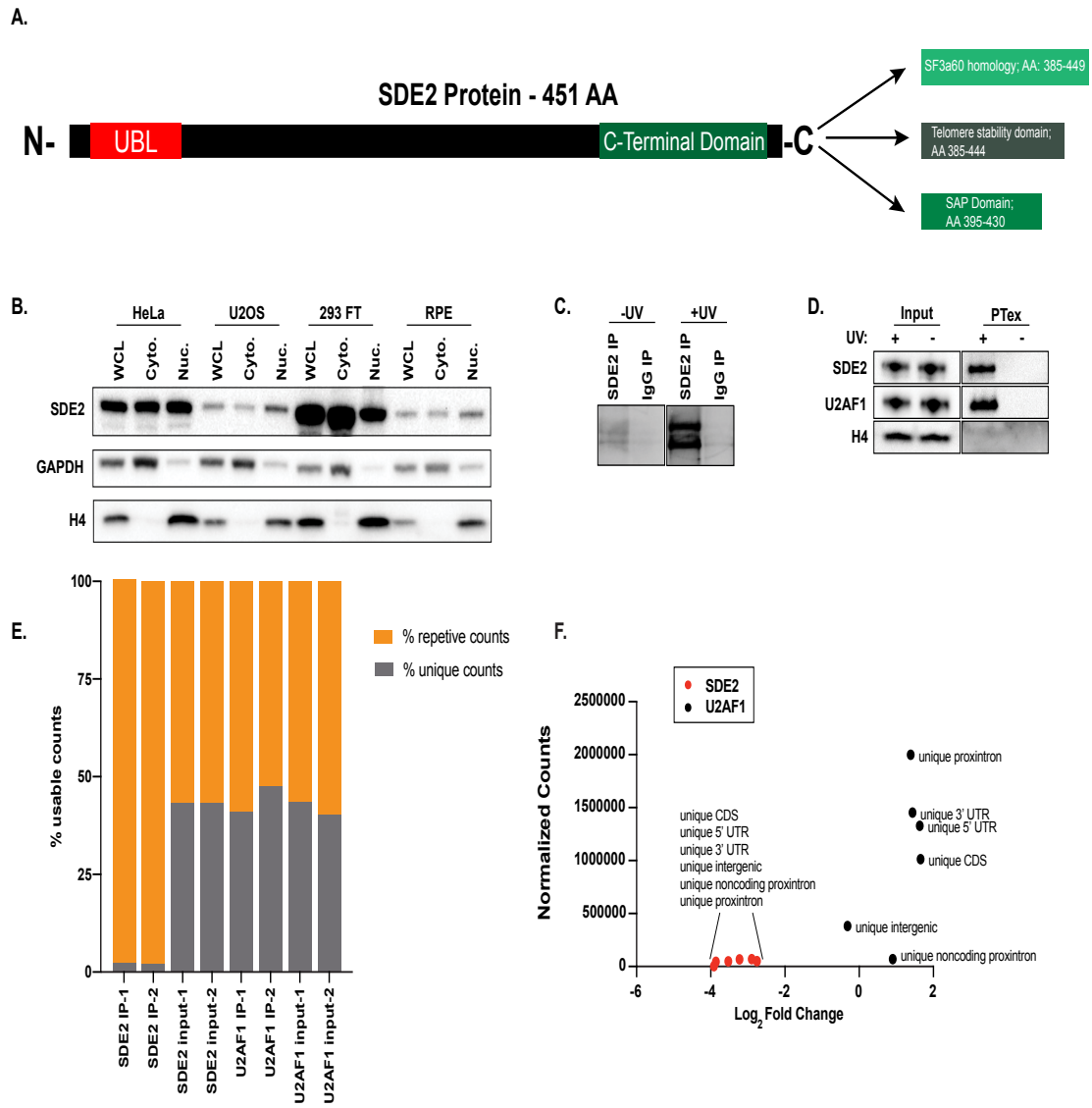unique noncoding proxintron
unique proxintron

Figure 3.1

Figure 3.1 (previous page). SDE2 is an RBP associated with repetitive ncRNAs. (A) SDE2 protein domain map. N-terminal Ubiquitin-like motif (UBL) is shown along with a putative C-terminal domain. Arrows indicate different possible motifs, as this sequence has been identified differentilally by different groups. (B) Western blot of SDE2, GAPDH, and H4 following cellular fractionation of HeLa, U2OS, 293 FT, and RPE-1 cell lines. GAPDH is used as a marker for the cytoplasmic fraction and H4 is used as a marker for the nuclear fraction. (C) Radiographic image of ^{32}P labelled RNA from CLIP of SDE2 and IgG in HeLa cells, in both uncrosslinked (-UV) and crosslinked (+UV) conditions. (D) Western blot of PTex samples in HeLa cells from both uncrosslinked (-UV) and crosslinked (+UV) conditions. U2AF1 is used as a positive control and DNA-binding protein H4 is used as a negative control. (E) eCLIP usable reads were identified as either "unique," or "repetitive," and were plotted as percentages of total usable reads for both IPs and inputs for SDE2 and U2AF1. (F) All (unique and repetitive) elements were filtered for information content >100. Unique non-repetitive element counts were averaged between IP replicates and plotted at the family level for both SDE2 and U2AF1.

These RNA sequences were specifically enriched for proximal introns consistent with previously published data and validating the quality of our own eCLIP profiles (Nostrand et al. 2017). In contrast, only ~2% of reads in the SDE2 eCLIP uniquely mapped to the genome, with the vast majority (~98%) of all usable reads mapping to repetitive elements (**Figure 3.1E-F**). Using a custom bioinformatic pipeline to analyze repetitive elements at the gene level, we found that SDE2 immunoprecipitations were enriched for non-coding RNAs (ncRNAs) including ribosomal RNA (rRNA), transfer RNA (tRNA), and small nucleolar RNAs (snoRNA) (**Figure 3.2** and Appendix C **Figure AC1**). Notably, SDE2's interaction with snoRNAs was not universal across the entire family, but almost entirely restricted to C/D box snoRNAs.

There are over 350 C/D snoRNAs in the human genome ranging in length from 60-300nt (Jorjani et al. 2016; Kiss 2002). C/D snoRNAs are defined by two conserved sequence motifs, a 5' C box (RUGAUGA) and a 3' D box (CUGA). In addition, some C/D snoRNAs also contain more centrally located degenerate C/D box motifs referred to as C'/D' box. The C/D snoRNAs are bound by four core proteins including Small Nuclear Ribonucleoprotein 13 (SNU13), NOP56 ribonucleoprotein (NOP56), NOP58 ribonucleoprotein (NOP58), and fibrillarin (FBL) to form active snoRNP complexes. While most C/D snoRNAs function as guide RNAs to promote the 2'-O-methylation of rRNA by FBL, others do not chemically modify rRNA. Instead, they function to direct the cleavage and processing of pre-rRNA to promote rRNA maturation (Ojha, Malla, and Lyons 2020). We identified 85 C/D snoRNAs by SDE2 eCLIP, of these, the SNORD3 (A, B, C, and D) family and SNORD118 were highly enriched. These C/D snoRNA are not only abundant in the cell, but have been demonstrated to regulate the cleavage and processing of rRNA. Although the SNORD3 and SNORD118 families share little sequence conservation, they both maintain the canonical C box and D box motifs (J.-L. Langhendries et al. 2016). We found that SDE2 directly interacts with nucleotides 135-151 of SNORD3, and between nucleotides 45 to 52 of SNORD118 (**Figure 3.3A**).
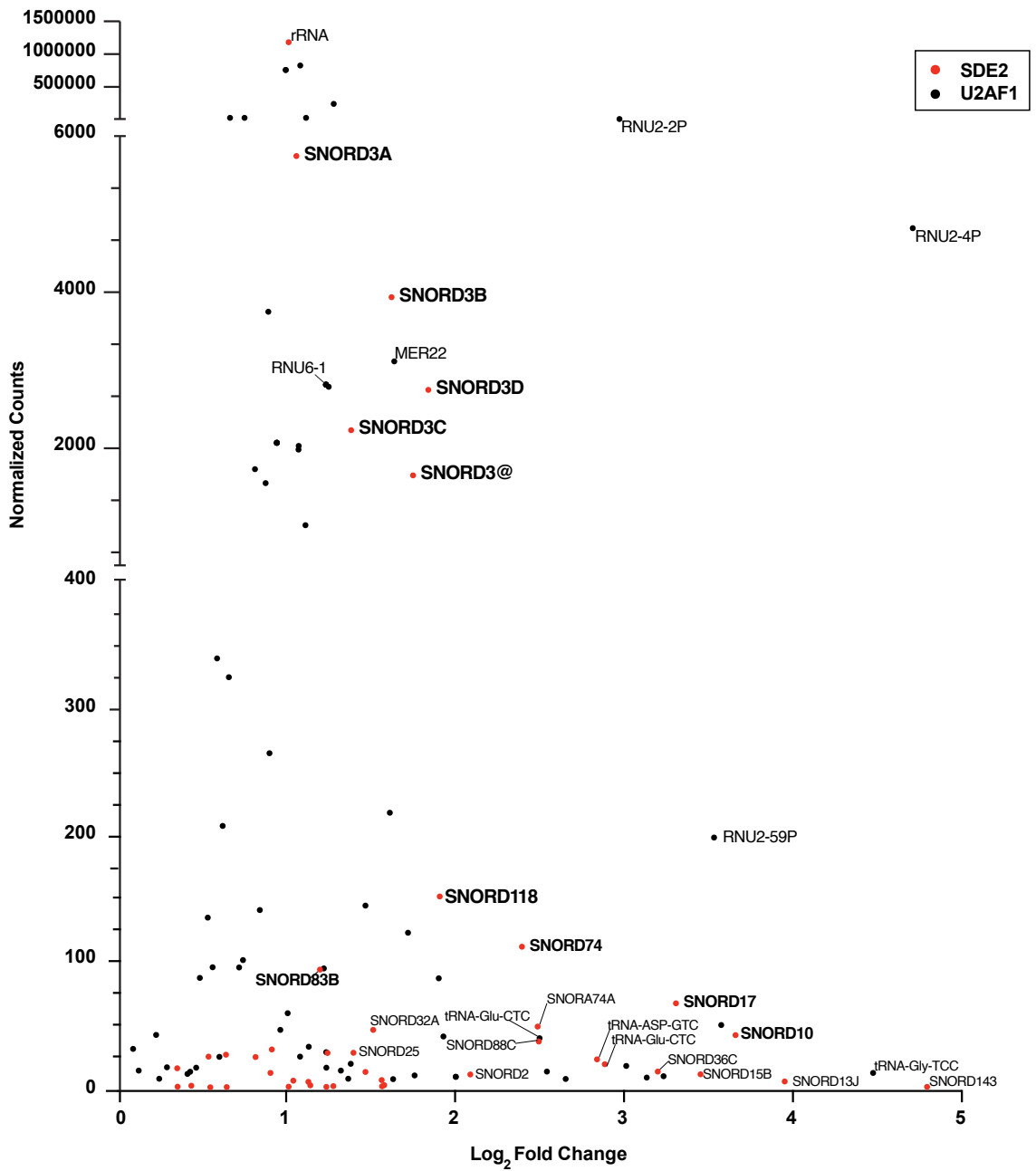
Figure 3.2

Figure 3.2 (previous page). SDE2 interacts predominantly with C/D snoRNAs. Using a custom bioinformatic pipeline, repetitive elements from SDE2 and U2AF1 eCLIP were identified at the gene level by filtering for raw counts >100 and Log_2 fold change >0 in duplicate IP conditions relative to respective inputs. Normalized counts were calculated and averaged between replicates and plotted against the average Log_2 fold change between replicates. Two genes from SDE2 and four genes from U2AF1 were excluded in (F) because their data points were outside the axis limits. The omission did not alter the overall analysis or interpretation of the data.

Remarkably, these SDE2 binding sites are located immediately adjacent to the 5' end of the C box in both SNORD3 and SNORD118. Given that SDE2 directly interacted with C/D snoRNAs, we hypothesized that SDE2 would also associate with one of the core C/D snoRNA associated proteins, FBL. Therefore, we performed reciprocal immunoprecipitations with SDE2 and FBL antibodies and analyzed interactions by western blot (**Figure 3.3 B**). These reciprocal immunoprecipitations demonstrated robust interaction between FBL and SDE2, further supporting our eCLIP analysis.
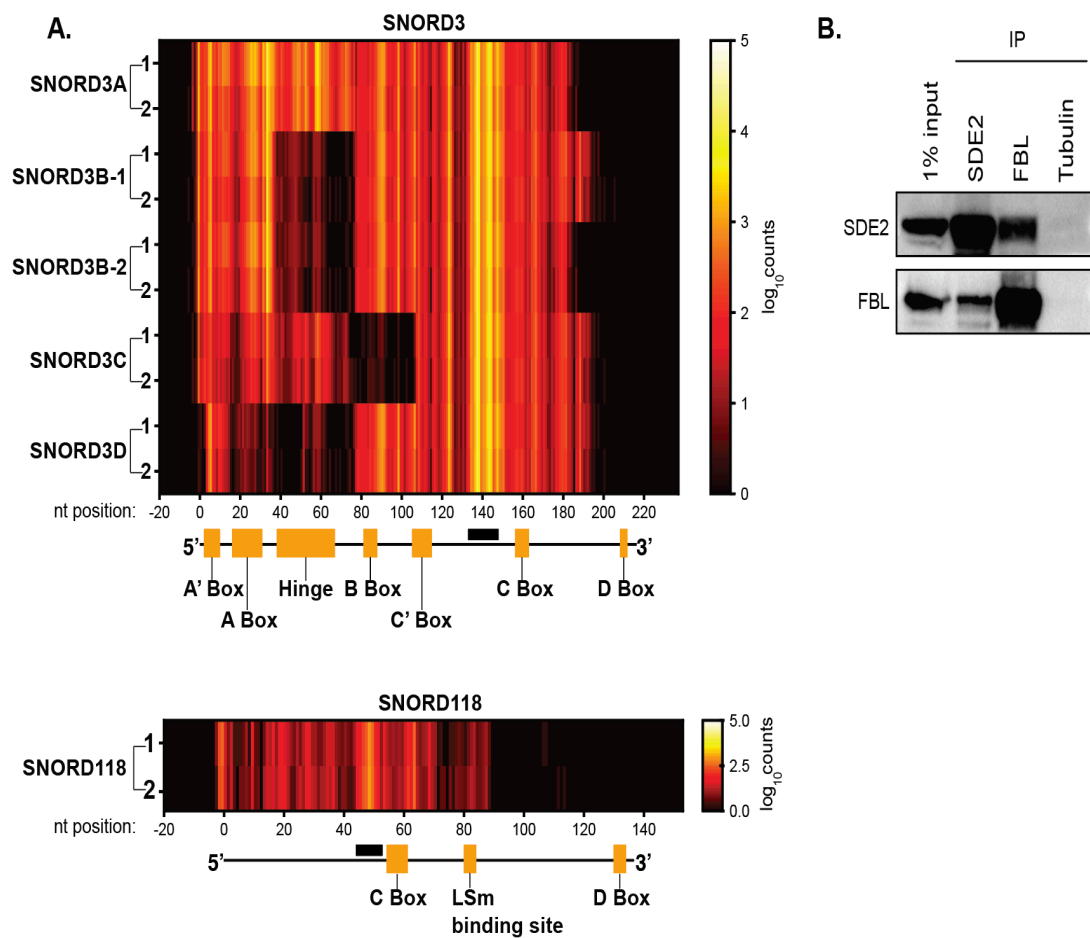
Figure 3.3. SDE2 binds adjacent to the C box in multiple C/D snoRNA family members. (A) Heat map corresponding to read counts from SDE2 eCLIP replicates 1 and 2 for the SNORD3 family (top panel) and SNORD118 (bottom panel). Nucleotide positions are labeled on the horizontal axis with 0 = 5' start site of each gene. Below each heatmap is a diagram showing annotated domains of each respective snoRNA. Black boxes indicate SDE2 binding regions on each transcript. (B) Reciprocal immunoprecipitation of SDE2 and FBL with a 1% input from whole cell extract in HeLa cells. Tubulin is used as a negative IP control.

Previous studies have demonstrated that SNORD3 functions to direct the cleavage of the A0 and 1 cleavage sites within the 5' ETS of the 47S rRNA precursor to promote rRNA maturation (J.-L. Langhendries et al. 2016).

109
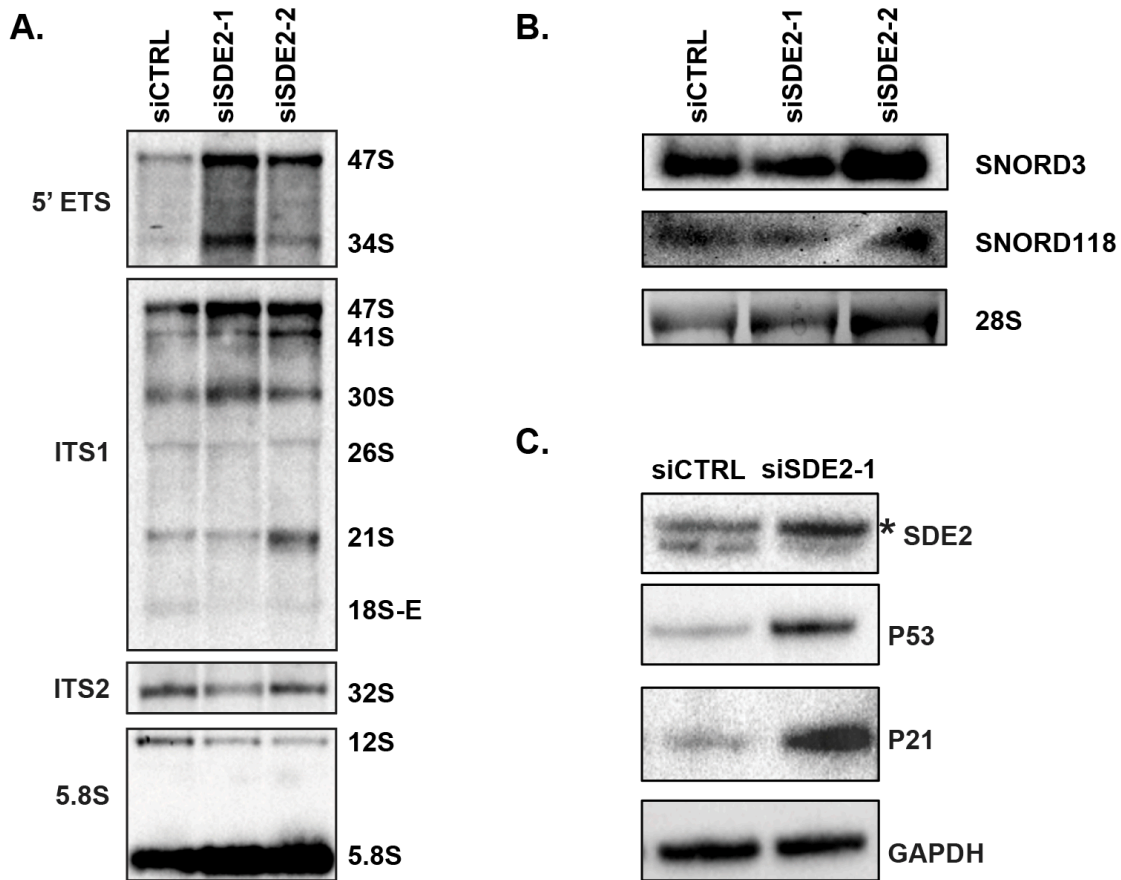
Figure 3.4. SDE2 regulates rRNA processing. (A) Northern blot of SDE2-depleted or control HeLa cells, incubated with the indicated probes in separate northern blots. n = 3 replicates. (B) Northern Blot for SNORD3 and SNORD118. 28S, detected by EtBr, serves as a loading control. (C) Western blot of RPE-1 cells depleted of SDE2. * indicates a background band immediately above the SDE2 band.

Consequently, loss of SNORD3, or SNORD3 interacting factors, leads to defects in rRNA cleavage and the accumulation of the 47S and 34S rRNA precursors. Here, we demonstrate that similar to loss of SNORD3, depletion of SDE2 also leads to defects in rRNA cleavage within the 5' ETS region resulting in the

accumulation of both the 47S and 34S rRNA precursors by northern blot (**Figure 3.4A**, top panel). SNORD118 functions to promote rRNA processing within the 5' ETS and also the internal transcribed spacer regions ITS1 and ITS2 to ensure processing of the 5.8S and 28S precursors. Loss of SNORD118 leads to an accumulation of the 47S precursor and decrease of both the 32S and the 12S rRNA precursors (J.-L. Langhendries et al. 2016). Consistent with defects in SNORD118, knockdown of SDE2 led to an increase in 47S and a decrease in the 32S and 12S precursors by northern blot (**Figure 3.4A**, bottom two panels). Notably, SNORD3 and SNORD118 RNA levels are unchanged following SDE2 depletion demonstrating that SDE2 does not regulate snoRNA stability (**Figure 3.4B**). Taken together, these data demonstrate that SDE2 is critical for pre-rRNA processing conceivably by regulating SNORD3 and SNORD118 RNA function.

Defects in rRNA processing shift the stoichiometry between rRNA and r-proteins leading to an excess of unbound, or free, r-proteins. These free r-proteins are then available to bind MDM2 and inhibit p53 degradation activating a ribosomal stress response pathway that leads to cell cycle arrest (Golomb, Volarevic, and Oren 2014). Given that loss of SDE2 led to defects in rRNA processing, we asked whether these defects were sufficient to release r-proteins and trigger p53 stabilization. Here, we silenced SDE2 in RPE-1 cells, a non-transformed, immortalized, p53 WT cell line, and analyzed p53 by western blot. After knockdown of SDE2, we observed p53 stabilization and downstream p21 activation (**Figure 3.4C**) in the absence of DNA damage (Jo et al. 2016), indicative of a defect in

ribosome biogenesis. Many of the r-proteins that stabilize p53 and function in ribosome biogenesis are also required for maintaining the structural integrity of the nucleolus. The nucleolus forms around the rDNA arrays and is composed of a fibrillar center (FC), a granular component (GC), and a dense fibrillar component (DFC) (Pederson 2011). FBL is enriched in the DFC and is considered a marker of nucleolar integrity (J.-S. He et al. 2018; Hernandez-Verdun D, Roussel P, Thiry M, Sirri V 2010). We observed defects in nucleolar integrity as measured by FBL immunostaining in SDE2 depleted HeLa cells (**Figure 3.5**), further supporting the conclusion that ribosome biogenesis is defective.

Defects in ribosome biogenesis and/or maturation can be visualized by changes in the sedimentation of the ribosomes with mRNA using polysome profiles. Following polysome analysis, we show that loss of SDE2 leads to a decrease in the abundance of the 40S ribosomal subunit, and to a lesser extent the 60S subunit. In addition, we also observe that loss of SDE2 leads to an accumulation of 80S monosomes and a relative decrease in higher molecular weight polysomes resulting in a 2-fold decrease in the polysome/monosome ratio (**Figure 3.6A**).
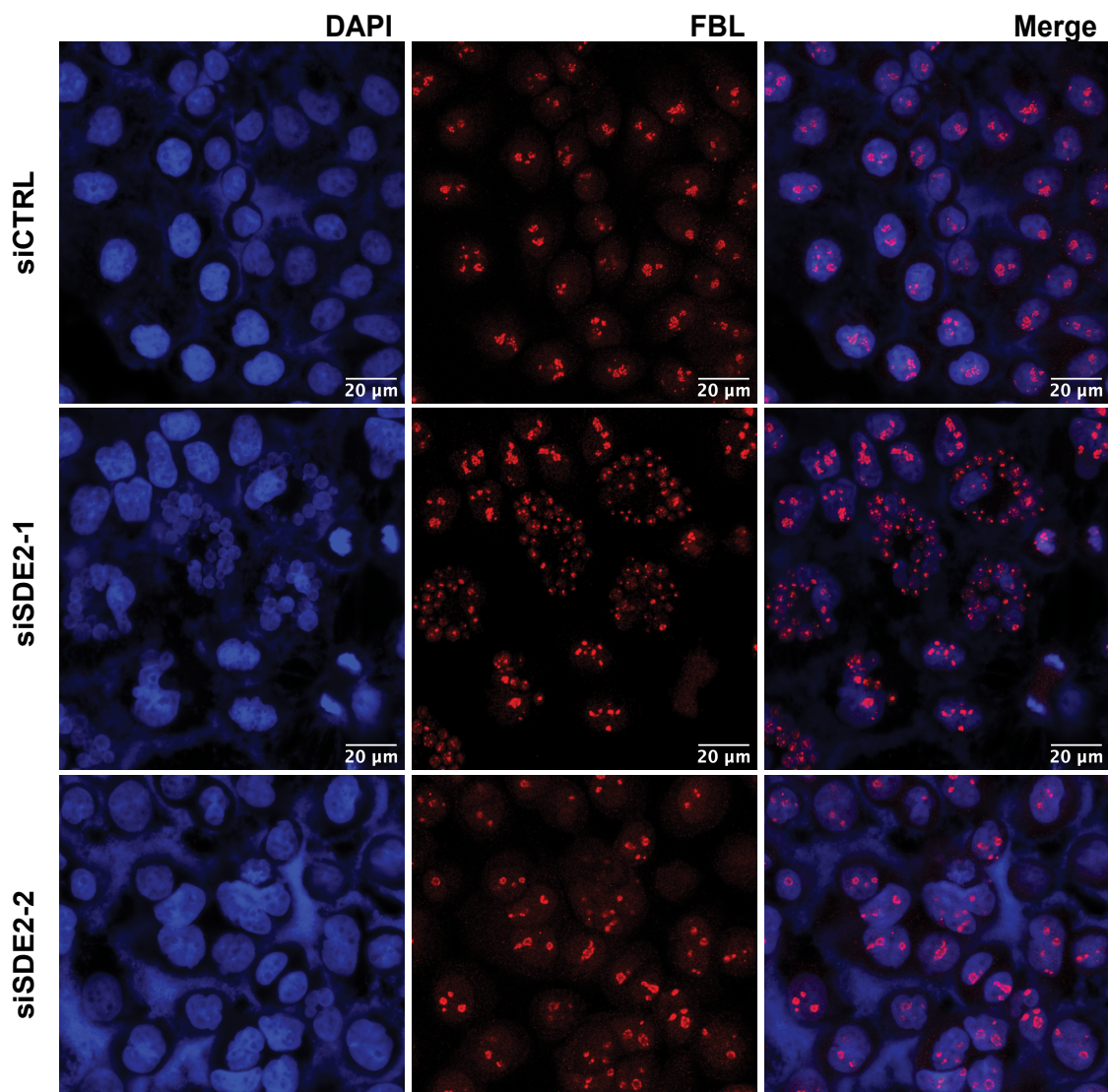
Figure 3.5. SDE2 depletion compromises nucleolar integrity. Immunostaining of FBL in SDE2-depleted HeLa cells (red). Scale bar 20 μm.

We reasoned that these defects in ribosome biogenesis and/or maturation may drive a concomitant loss in global protein translation. Therefore, we asked whether loss of SDE2 led to defects in protein synthesis using SUrface seNSing of Translation, or SunSET assays (Schmidt et al. 2009). Following transfection with siRNA for SDE2, we allowed cells to proliferate for either 3 or 5 days before pulsing cells with puromycin. Following the incubation with puromycin, we collected cells and analyzed whole cell lysates for puromycin incorporation by western blot using a puromycin-specific antibody. As expected, cells transfected with a control siRNA demonstrated a strong pattern of active protein synthesis by western blot. However, loss of SDE2 led to a significant and progressive decrease in protein synthesis over the course of 5 days (**Figure 3.6B-C**). This defect in translation was not simply due to the activation of the integrated stress response that drives translational repression via the eIF2α signaling pathway, nor was it caused by disruptions in mTOR signaling (**Figure 3.6D-E**). Together, our data suggest SDE2 plays a critical role in maintaining ribosome biogenesis and global protein synthesis in mammalian cells through regulation of snoRNA-dependent rRNA cleavage and the maturation of functional ribosomes.
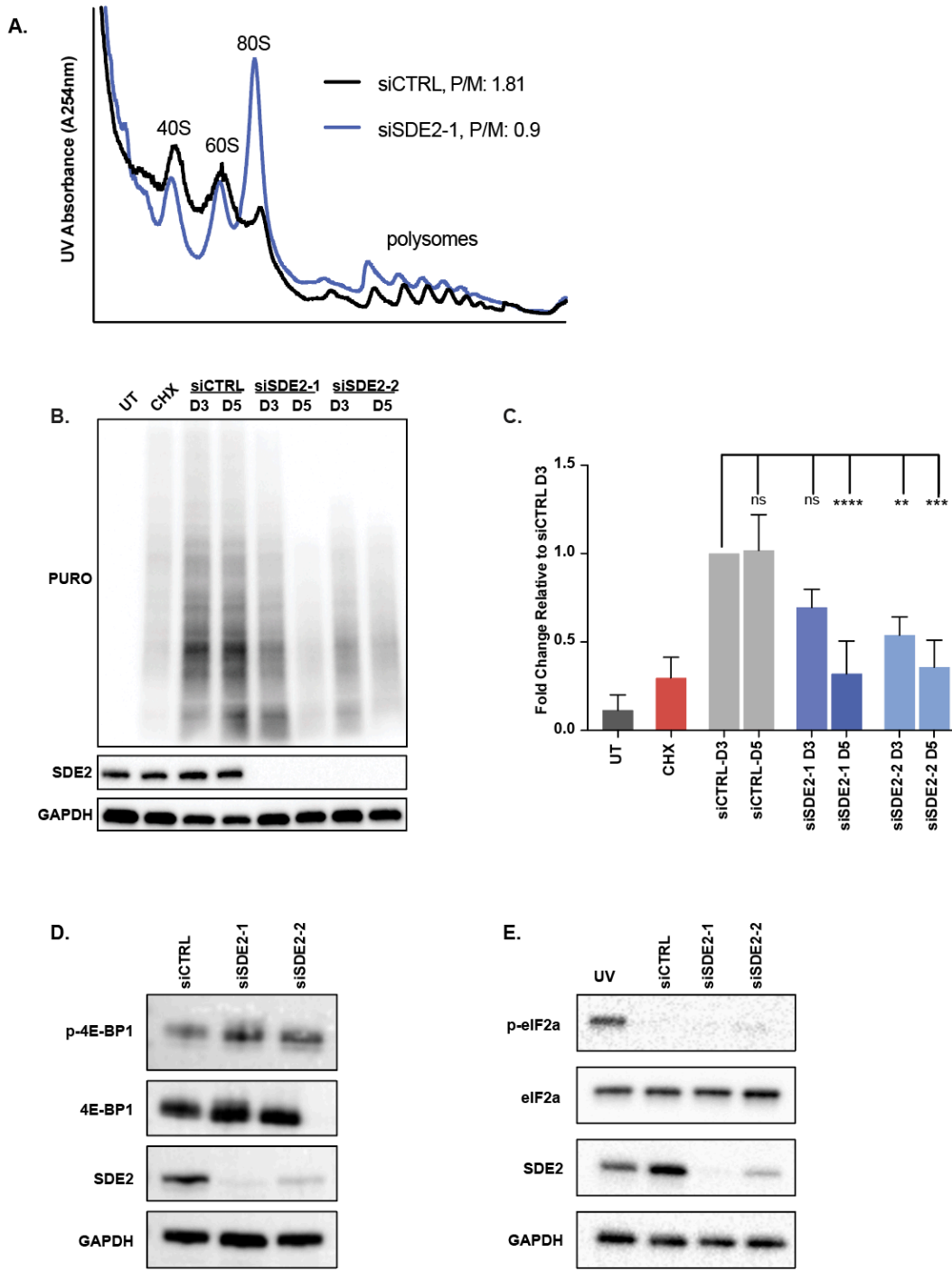
Figure 3.6

Figure 3.6 (previous page). SDE2 depletion causes defects in ribosome biogenesis and protein synthesis. (A) Polysome profiling from HeLa cells treated with siCTRL or siSDE2 for 5 days. 40S, 60S, and 80S peaks reflect SSU, LSU, and assembled ribosomes, respectively. Polysome to monosome (P/M) ratio is calculated using the area under the curve for the first four polysome peaks divided by the area under the curve for the 80S peak. (B) Western blot of SUnSET assay measuring global translation in HeLa cells 3 or 5 days following SDE2 knockdown. Protein synthesis was measured by the incorporation of puromycin, and puromycin was detected using a puromycin specific antibody. UT, cells not treated with puromycin; CHX, cells treated with puromycin and cycloheximide. Representative image shown from experiments performed in triplicate. (C) Quantification of experiments performed in (B). Graph shows mean puromycin intensity for each lane relative to the mean puromycin intensity for Day 3 siCTRL +/- SD. The intensity of the puromycin in the Day 3 siCTRL lane was normalized to 1.0. *P*-values calculated by two-way ANOVA comparing mean puromycin intensity of each condition to Day 3 siCTRL, followed by Dunnet's multiple comparisons test, with $P \leq 0.01$ denoted by **, $P \leq 0.001$ denoted by ***, and $P \leq 0.0001$ denoted by ****. (D) HeLa cells were transfected with control or SDE2 siRNA and analyzed for changes in 4E-BP1 phosphorylation by Western blot. (E) HeLa cells were transfected with control or SDE2 siRNA and analyzed for changes in eIF2a phosphorylation by Western blot. Untransfected HeLa cells treated with 50J/m UV light were used as a positive control for eIF2a phosphorylation.

Our functional characterization of SDE2 has demonstrated that SDE2 is an RBP with critical functions in regulating specific snoRNA-mediated cleavage in rRNA maturation, and that it also associates with the U2 snRNP to regulate the AS of a specific subset of introns. Given the essentiality of these pathways in cell viability, we used live cell imaging to monitor the fate of U2OS cells depleted of SDE2. We found that loss of SDE2 led to an increase in the fraction of cells that arrested in the cell cycle, as well as an increase in the number of interphase or mitotic cells that underwent cell death, relative to controls (**Figure 3.7A-C**). These

116

defects were observed in almost half (47.8%) of all cells depleted of SDE2 (**Figure 3.7D**). The remaining SDE2 depleted cells that were able to proceed to mitosis exhibited mitotic slippage or prolonged mitosis (**Figure 3.7E-H**). Unsurprisingly, cells that underwent abnormal mitosis did not survive multiple rounds of cell division in culture.

Following the results of our live cell imaging, we asked whether cells would survive long-term following SDE2 depletion. Not surprisingly, knockdown of SDE2 using two different siRNAs in HeLa cells led to significant defects in cellular proliferation and culminated in cell death six days after SDE2 knockdown (**Figure 3.8A**). Likewise, we used shRNA to knock down SDE2 in RPE-1 cell lines and analyzed viability by clonogenic survival. While cells infected with empty vector continued to proliferate, the cells infected with shSDE2 demonstrated significant defects in proliferation and increased cell death 12 days after shRNA knockdown (**Figure 3.8B-C**).
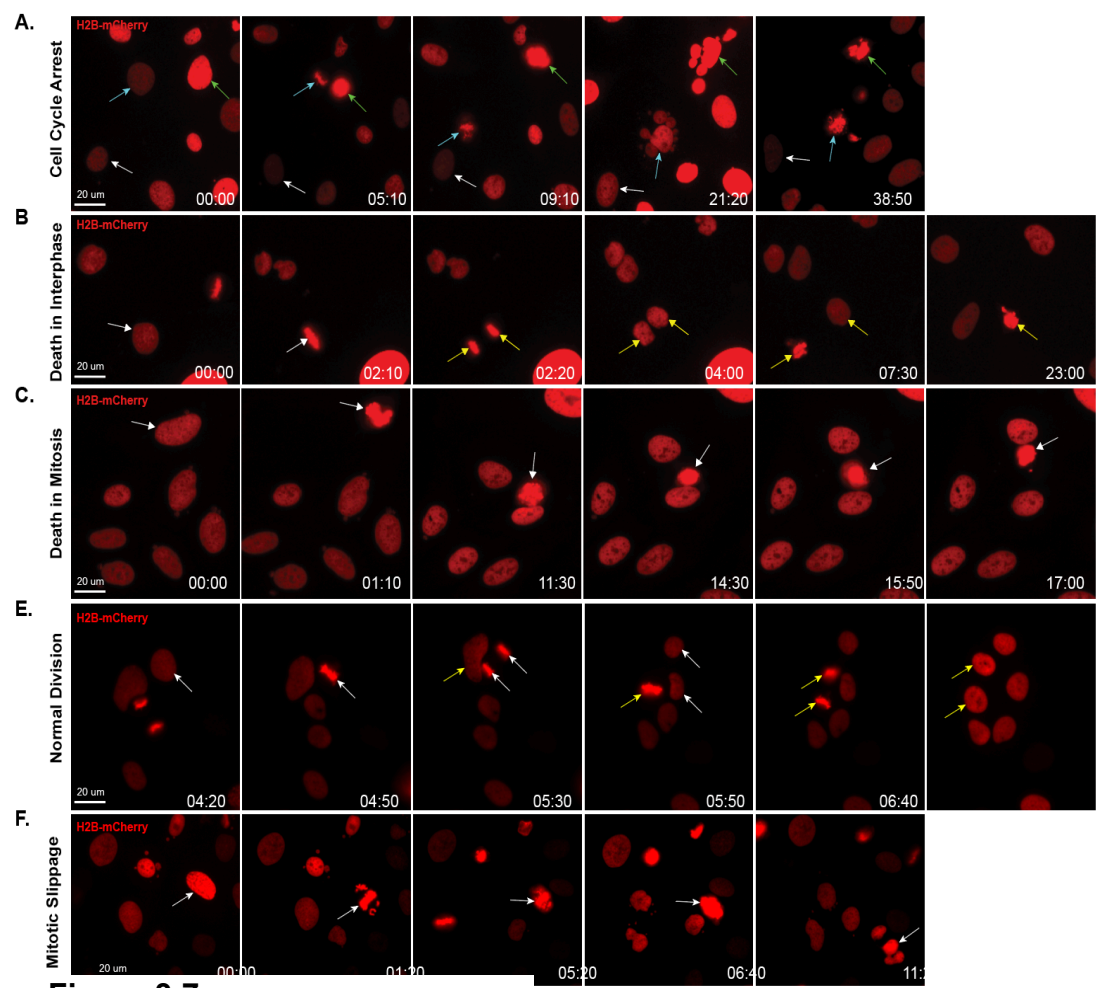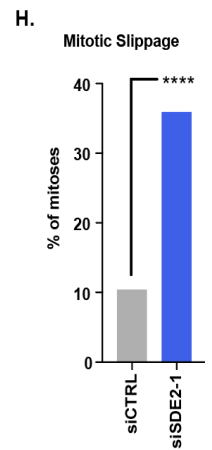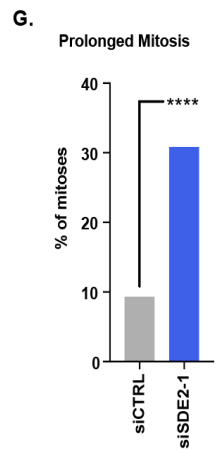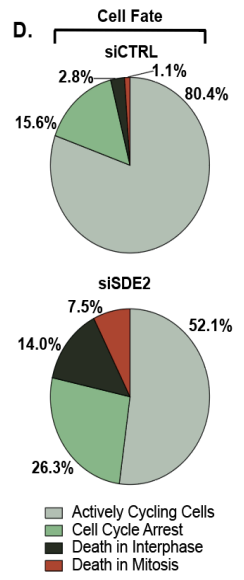
Figure 3.7

Figure 3.7 (previous page). Live cell imaging reveals cell fate following SDE2 depltion. (A-C) Representative still images from live cell imaging experiments of U2OS cells stably expressing H2B-mCherry transfected with siSDE2-1. (A) Representative image of a cell undergoing cell cycle arrest (White arrow), defined by the absence of mitosis for >36 hours. Teal and green arrows denote cells that undergo catastrophic mitosis and subsequent cellular death. (B) Representative image of cells undergoing death in interphase. White arrows mark a cell that enters mitosis and divides, forming two daughter cells (yellow arrows) that both exhibit cellular death in interphase. (C) Representative image of a cell undergoing mitotic cell death (white arrow). Time, hr:min. (D) Pie charts representing the fate of cells following control or siSDE2-1 knockdown, n = 179 cells in the siCTRL condition and 186 cells in the siSDE2-1 condition. (E) Representative still images from a live cell imaging experiment of U2OS cells stably expressing H2B-mCherry transfected with control siRNA. White arrows and yellow arrows track single cells from interphase through successful completion of mitosis. (F) Still images of U2OS cells transfected with siSDE2-1 siRNA. A cell (marked by a white arrow) is shown undergoing both prolonged mitosis and mitotic slippage. Time, hr: min. (G) Quantification of mitotic duration. Mitoses lasting longer than 90 minutes were scored as having prolonged mitoses, and mitoses lasting less than 90 minutes scored as normal duration mitoses. The graph shows the percent of prolonged mitoses relative to all mitoses in each condition. n = 216 cells in the siCTRL condition and n=166 cells for siSDE2-1 condition. (H) Quantification of mitotic slippage. Graph shows the percent of mitoses that result in a mitotic slippage event relative to all mitoses for each condition. n = 479 cells in the siCTRL condition and n=295 cells in the siSDE2-1 condition. For live cell imaging experiments, P-values were calculated from total cell counts over two (prolonged mitosis) or three (mitotic slippage) independent experiments by Fischer's exact test and P ≤ 0.0001 is denoted by ****.

Finally, the recent advances in CRISPR technology and the release of the DepMap through the Broad Institute allowed us to ask whether SDE2 was essential for viability in all cancer cells (Depmap Broad 2019). In support of our analysis in HeLa and U2OS cells, DepMap confirmed that SDE2 was essential for cellular viability in 100% of the 625 cell lines analyzed following CRISPR knockout. Although this loss of viability following CRISPR knockout was consistent with other genes known to regulate pre-mRNA splicing and ribosomal biogenesis, essentiality
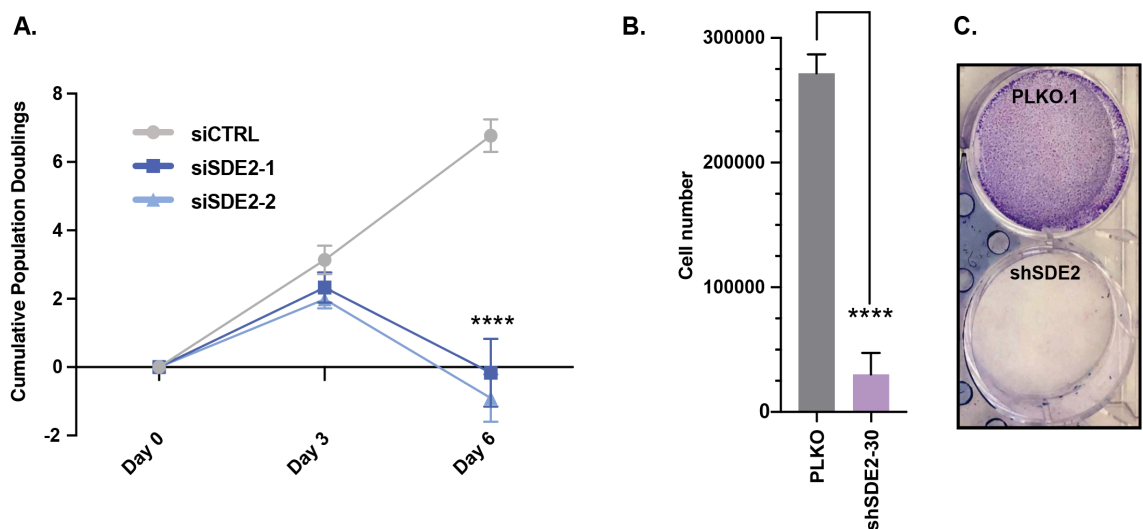
Figure 3.8. SDE2 is essential for cell viability. (A) Population doubling assay in HeLa cells. Cells were counted at Day 0, 3, and 6 and reverse transfected with corresponding siRNAs at Day 0 and Day 3. n=3 independent experiments with duplicate technical replicates for each condition in each independent experiment. Graph shows cumulative population doubling in each condition at the indicated time +/- SD. *P*-values calculated by two-way ANOVA followed by Dunnet's multiple comparisons test, with $P \leq 0.0001$ denoted by ****. (B,C) RPE-1 cells expressing shSDE2 were collected after 12 days, counted, and plated again. The following day, cells were stained with crystal violet or collected for western blot (n=3 independent experiments). The graph shows mean cell number +/- SD. *P*-value calculated by unpaired t-test with $P \leq 0.0001$ denoted by ****.

is not a universal characteristic for all splicing or ribosome assembly factors, highlighting the critical nature of SDE2 function in pre-mRNA processing and ribosome biogenesis (**Figure 3.9**).
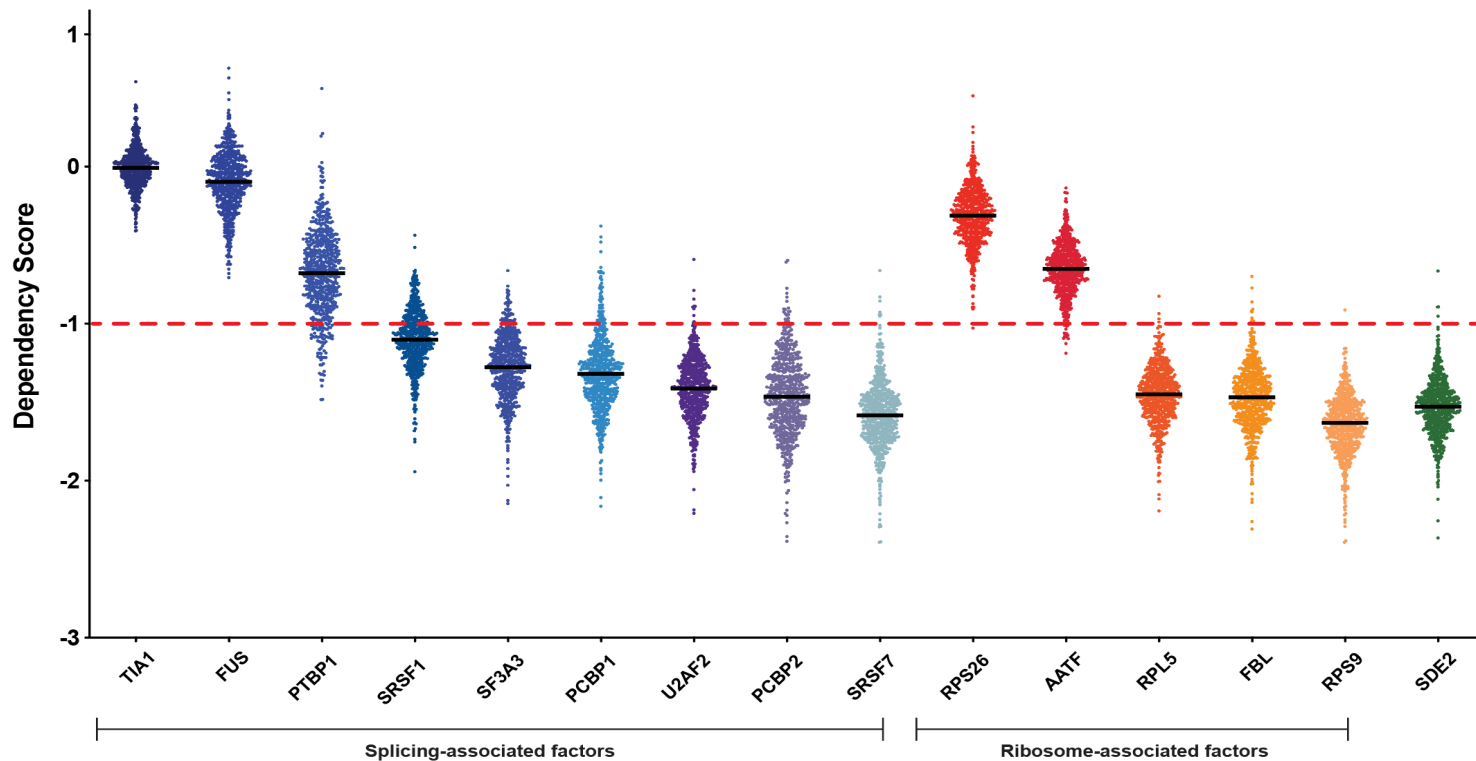
Figure 3.9. SDE2 is a pan-essential gene. Data from The Broad Institute's Project Achilles dataset *DepMap 19Q3 Public*. The graph shows dependency scores for each gene listed on the *x*-axis in 625 different cancer cell lines. Each cell line is represented by a single point, and all cell lines are represented for each listed gene. The average dependency score of an essential gene is -1, denoted by the red dashed horizontal line drawn at *y* = -1. For the indicated genes, the black horizontal line in front of the dots is the average mean dependency score between all cell lines.

Section 4: Discussion

The complex and dynamic nature of ribosome biogenesis has precluded the identification and characterization of an ell-encompassing inventory of every critical RBP. Therefore, there are likely many proteins and RNA moieties that function as *trans*-acting factors throughout these processes that have yet to be defined. Here, we demonstrate that SDE2 is an RNA binding protein that directly interacts with C/D snoRNAs and is required for efficient rRNA processing. SDE2 depletion leads to misprocessed rRNA intermediates, defects in ribosome subunit biogenesis, and global loss of translation. These deficits overwhelm the cell and result in the complete loss of cell viability, highlighting the essentiality of SDE2 in mammalian cells.

Recent biochemical purifications of RNA have identified over one thousand RBPs and approximately half lack the classical RNA binding domains (i.e. RRM, KH, CSD, or CCHC). Instead, these RBPs contain classical DNA binding domains such as the SAP domain, while others contain intrinsically disordered regions (IDRs) (Castello et al. 2016; C. He et al. 2016). Similar to these RBPs, SDE2 not only contains both an IDR and a SAP domain, but also demonstrates direct interactions with RNA. Using eCLIP, we identified that SDE2 directly interacts with the C/D box subfamily of snoRNAs. C/D snoRNAs are defined by two highly conserved C and D box motifs and two less conserved C' and D' box motifs.

Structurally, snoRNAs fold to ensure that both the C and D (and C'/D') core motifs oppose one another, creating a hairpin structure. The C/D and C'/D' box motifs are bound by the core C/D snoRNP factors NOP56, NOP58, SNU13, and FBL to regulate a number of functions in snoRNA biogenesis including the trimming of snoRNA from within introns of pre-mRNA, localization of snoRNA to the nucleolus, and methylation of the snoRNA cap (Terns and Terns 2002). Conceivably, the SAP and/or IDR of SDE2 directly interacts with snoRNA near the C box motif to promote snoRNA biogenesis.

Within the C/D box class of snoRNA, the SNORD3 family is structurally unique, containing A, A', Hinge, B, C, C', and D sequence motifs. In contrast to other C/D box snoRNAs, SNORD3 family members fold into a hairpin structure where the C' box opposes the D box and the C box opposes the B box. As a result, SNORD3 maintains both a C'/D box motif and a unique B/C box motif. The B/C box is critical for interactions with not only the core C/D snoRNP factors, but also the SNORD3 specific factor U3-55K (Terns and Terns 2002). Together, these proteins function to regulate both SNORD3 stability and the assembly of the small ribosomal subunit (Hughes 1996; Terns and Terns 2002). In contrast, SNORD118 contains only the canonical C/D box motif, in which the C box directly opposes the D box (Tyc and Steitz 1989). SDE2 interacted with both SNORD3 and SNORD118 adjacent to and immediately 5' of each C box. Although the C box sequences are identical, SNORD3 and SNORD118 lack sequence conservation within the region surrounding the C/D box motifs. This suggests that the folding and/or structure of

the C box flanking regions play a role in mediating SDE2 interaction, similar to the interaction between the B/C box of SNORD3 with U3-55K (Lukowiak et al. 2000). SDE2's interaction with the 5' C box flanking regions suggests that SDE2 binds near these motifs and functions to regulate C/D snoRNP complexes to coordinate rRNA processing events including, cleavage, methylation, folding, and/or localization ultimately, providing the framework for assembly and maturation of the ribosome.

Our polysome profiles provided additional evidence to support a role for SDE2 in ribosome biogenesis and/or ribosome maturation. Although the exact mechanism of SDE2 function in translation remains unclear, we provide several possibilities that are not necessarily exhaustive, nor mutually exclusive. First, SDE2 interacts broadly with C/D snoRNAs, and also with the methyltransferase FBL. FBL is the catalytic core of the C/D snoRNP complexes and functions to regulate 2'-O-methylation of a range of RNA species including rRNA, mRNA, and snRNA (Elliott et al. 2019; B E Jády and Kiss 2001; Karijolich and Yu 2010; Kiss-László et al. 1996). Defects in 2'-O-methylation are known to impair ribosome biogenesis and protein translation (Elliott et al. 2019). Therefore, it's tempting to speculate that SDE2 may promote unique 2'-O-methylation events on rRNA and/or mRNA to regulate protein synthesis. RNA methyl-sequencing approaches are rapidly expanding and will undoubtedly provide an opportunity for future studies investigating the genome-wide methylation events regulated by SDE2. Second, SDE2 depleted cells could have defects in the kinetics of ribosomal subunit

assembly or produce suboptimal subunits that, while competent for initiation, may be defective for ribosomal elongation. Finally, monosomes are known to be enriched on misspliced mRNAs targeted for nonsense-mediated decay (NMD) (Heyer and Moore 2016). Given that SDE2 depleted cells harbor an increase in RI events, and many of these RI events contain NMD-triggering premature termination codons, it stands to reason that these RI events may sequester functional ribosomes and competitively inhibit translation. Future studies investigating SDE2 function in ribosome biogenesis and/or maturation will aim to provide additional insight towards these possible mechanisms.

Ribosome biogenesis and pre-mRNA splicing are essential processes in mammalian cells, and as such, many of the factors within these pathways are essential genes. However, some (FUS, TIA1, and RPS26) are not required for viability in the vast majority of cancer cells, highlighting a critical function for SDE2 (Depmap Broad 2019). Though impaired ribosome biogenesis and pathologic AS are linked to human disease, it is unclear on a molecular level the exact mechanism, or mechanisms, by which they lead to loss of cell viability. Future studies will likely shed light on these mechanisms and provide insight to the highly complex regulation of these essential processes.

CHAPTER 4: FUTURE DIRECTIONS

SDE2 is an essential gene with roles in both pre-mRNA splicing and ribosomal biogenesis and maturation. The critical nature of SDE2 regulation on these pathways is borne out by the fact that all human cells require SDE2 for viability, and in addition, loss-of-function mutations in SDE2 in other metazoans have proven incompatible with survival. We have provided direct evidence of SDE2's involvement in both pre-mRNA splicing and ribosomal maturation, however, the exact mechanism by which SDE2 exerts its critical molecular function to these processes still eludes characterization. Future work, proposed below, will be necessary to further our understanding not only of SDE2 and its functions, but also to fully elucidate the molecular underpinnings that guide these most vital cellular pathways.

Section 1: SDE2's Role in pre-mRNA Splicing

Despite the expansive accumulated knowledge from decades of research regarding both the *cis* and *trans*-acting factors involved in pre-mRNA splicing, we are still unable to accurately articulate a 'splicing code'-- a set of rules that allow us to determine intron, exon, and splice site inclusion in a given context or tissue type. The enormous complexity underlying the decision-making process in splice site choice is due to multivalent regulatory factors that can act either synergistically or in opposition with one another. We have demonstrated that SDE2 is required

for proper splicing of a specific subset of introns in human cells, but the exact modality by which SDE2 exerts this regulation upon the splicing mechanism is unknown. Future work will undoubtedly uncover the mechanism by which SDE2 functions in pre-mRNA splicing.

There are many avenues through which SDE2 may regulate intron retention, though the RNA binding profile of SDE2 may offer up some clues as to which is most likely. Our eCLIP data suggested that SDE2 interacted broadly with C/D snoRNAs, most of which function as guide RNAs for the methyltransferase FBL to install 2'-O methylations on the ribose sugar of various RNAs. Spliceosomal snRNAs undergo extensive modification, and 2'-O methylations especially are required for the splicing reaction to occur. The U2 snRNA is the most modified snRNA out of all the spliceosomal RNAs (Morais, Adachi, and Yu 2021), and requires four separate 2'-O methylations at its extreme 5' terminus to function efficiently in splicing (Dönmez, Hartmuth, and Lührmann 2004). This is especially intriguing because SDE2 interacts with FBL, as well as protein components of the U2 snRNP. Additionally, some of the 2'-O methylation sites in U2 snRNA are unaccounted for in terms of snoRNA guide, raising the possibility that SDE2 may link C/D snoRNAs to the U2 snRNA to promote 2'-O methylation, and subsequently, proper pre-mRNA splicing. However, most snRNA modifications are deposited late in their biogenesis, usually in the CB. It remains to be seen if SDE2 localizes to this membraneless nuclear structure. As all commercial SDE2 antibodies have failed to detect SDE2 in immunocytochemistry experiments,

tagged constructs of SDE2 may allow us to visualize SDE2 and CB components colocalizing, though this will require further optimization as current tagging strategies have been unsuccessful.

The partial crystal structure of SDE2 was identified in complex with the spliceosome in human cells (Fica et al. 2019b). This was accomplished by the use of a synthetic intron called MINX, which contains strong splice sites and exhibits a short length, as well as being MS2 tagged. In this crystal structure, as well as in splicing experiments performed in yeast, SDE2 associated with the auxiliary splicing factor Cactin (Cay1 in yeast) (Thakran et al. 2018). The interaction of SDE2 with Cactin may result in the timely delivery of Cactin to the spliceosome, thereby stabilizing its interaction with pre-mRNA across specific introns, and ensuring efficient and complete splicing. It is unclear if loss of SDE2 would result in loss of Cactin in the spliceosome, though Thakran et. al showed that SDE2 and Cactin act in the same splicing pathway in yeast, as individual deletions of SDE2 or Cactin resulted in similar introns being retained, and co-depletion of SDE2 and Cactin did not prove to have a synergistic effect on number or content of RI. In humans, it may be that SDE2 recruits Cactin to specific spliceosomes formed through intron definition. Depletion of SDE2 followed by PTEX experiments to detect Cactin in the RBP fraction would allow us to investigate if Cactin is dependent solely on SDE2 for its RNA-binding function. However, like most RBPs, it is unlikely that Cactin relies on a single protein to recruit it into splicing factors, and any specific loss of RNA-interaction by Cactin due to SDE2 depletion would

go undetected, at least by PTEX. Instead, spliceosomal pulldowns using MS2 minigene constructs, similar to the MINX synthetic intron that was used to crystallize SDE2 in the spliceosome in humans, might allow detection of SDE2-mediated Cactin exclusion from specific spliceosomes. Here, instead of using MS2-tagged MINX pre-mRNA, an SDE2-depletion candidate RI (ARSA, CUL7, FBF1, etc.) minigene construct, consisting of a single intron flanked by an exon on each could be introduced into cells. MS2-tagging allows for the isolation of these pre-mRNAs, as well as the spliceosomal complexes that form over them, followed by detection of specific splicing factors predicted to be stabilized by SDE2, like Cactin.

Our bioinformatic analysis of the SDE2-depleted RI indicated defining *cis* characteristics of the introns, but failed to identify any common sequence motifs. One of these *cis* defined characteristics was an extreme shortness in length of the intron. This likely alludes to SDE2 influencing splicing in an intron-defined manner, whereas most splicing in humans is accomplished through exon-definition, due to the relative shortness of most exons compared to introns. Though exons and introns are often represented as individual 'units' of a transcript, exon definition models clearly indicate that the splicing machinery recognize exons using cues from surrounding exons/introns (De Conti, Baralle, and Buratti 2013). Intron-defined pre-mRNA splicing allows the formation of the spliceosome across a single intron, but it has been observed that short introns, that presumably utilize intron-defined spliceosomes (Sakabe and de Souza 2007), are influential in determining

splicing efficiency of flanking exons (Sterner, Carlo, and Berget 1996). It is possible that flanking exons containing additional ESEs that reinforce spliceosome assembly across the intron may further contribute to splicing efficiency and spliceosome formation. Future analysis of the flanking exons of our RI events may uncover conserved sequence-specific binding sites for specific RBPs that are delivered, stabilized, or interact in some manner with SDE2 to establish proper splicing of these short introns.

RI was not the only category of dAS affected by SDE2 depletion. Future work examining the full dAS profile of SDE2 depleted cells may yield additional clarity of the mechanism, or mechanisms by which SDE2 regulates pre-mRNA splicing. We identified both exon skipping and alternative polyadenylation sites as being the two most increased AS categories besides RI. It would be interesting to analyze the splice site strength in the flanking exons/introns surrounding exon skipping events to determine if SDE2 is required for efficient splicing at these splice site junctions with lower MaxENT splice site strength scores, regardless of length or GC content. We also noticed an increase in the polyadenylation of histone mRNAs (**Figure 4.1**). Histone mRNAs are unique in that they do not canonically contain polyadenylated tails (Krieg and Melton 1999), which are replaced by a conserved sequence that forms a stem loop structure and is bound by the stem loop binding protein (SLBP) (Dominski et al. 1999). In this case, the choice of polyadenylation site was not altered inasmuch as a noncanonical installation of a poly(A) tail on the histone mRNA was observed. Polyadenylation of histones has

131

been noted to affect cells deleteriously (D. Chen et al. 2020), thereby leading to another possible route SDE2 disrupts transcriptomic integrity.
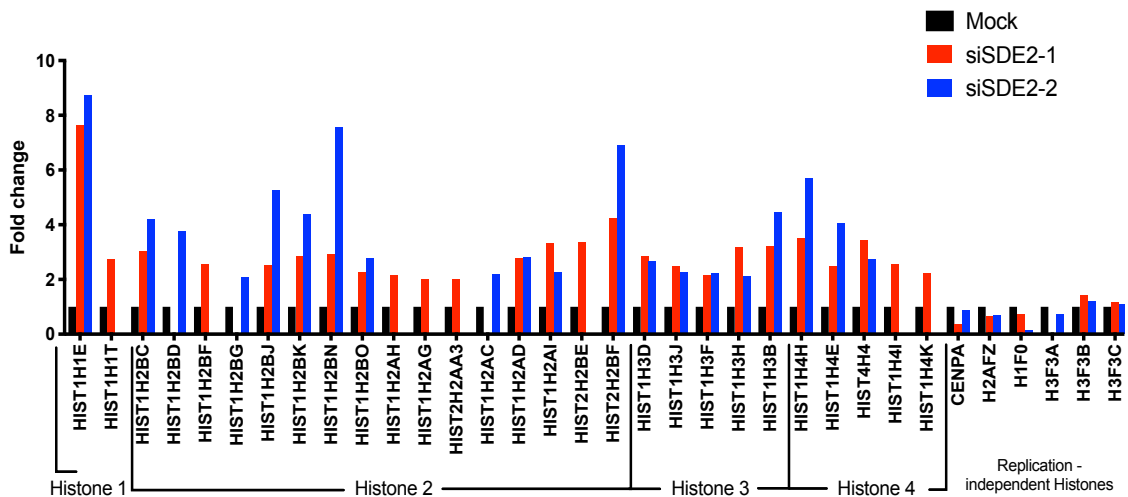
Figure 4.1. Histone mRNA display increased polyadenylation following SDE2 depletion. Fold change of histone mRNA read counts in poly(A) RNA sequencing of Hela cells.

Pre-mRNA splicing is regulated by multiple modalities, including the chromatin landscape that transcription and co-transcriptional splicing occur within (Allemand et al. 2016). Interestingly, one group showed that SDE2 associates with DNA and DNA replication factors (Luo et al. 2020). Our own data indicate that SDE2 associates with Histone 4. In Schizosaccharomyces pombe, where loss of SDE2 leads to both retention of specific introns (of short length) as well as increased telomeric transcription (Sugioka-Sugiyama and Sugiyama 2011b),

(Thakran et al. 2018). Loss of telomere silencing via SDE2 was linked to decreased recruitment of the histone deacetylase complex SHREC, indicating that at least in yeast, SDE2 may serve as a chromatin modifying agent. Whether this chromatin modification role exists in humans, and if it has any effect on splicing fidelity, is an intriguing question.

Section 2: SDE2's Role in Ribosome Biogenesis and Maturation

Similarly to SDE2's role in pre-mRNA splicing, our data are provocative in that they implicate a crucial role for SDE2 in ribosome maturation and translation, but further work is needed to define the precise molecular function SDE2 contributes to these processes. Additionally, SDE2 is an essential gene, and it's depletion results in death in cells and in metazoans like *C. elegans* (Piano et al. 2002) and *Bos taurus taurus* (Holstein cattle) (Fritz et al. 2018). What then, is the pathology status of humans with heterozygous loss-of-function mutations in the SDE2 gene? It is clear that the relatively recent first characterization of this gene has only scratched the surface in determining the necessary roles SDE2 plays in cells and organisms.

One of the limitations of eCLIP is its inability to distinguish RNA target affinity from RNA target abundance. In most cases where RBPs bind specific mRNAs, this does not affect overall results, as mRNAs are transcribed from unique loci and contain only two copies throughout the human genome, with mRNA

accounting for less than 5% of total RNA within a cell. However, eCLIP is not optimized for RBPs that bind repetitive elements, which can be transcribed from four to hundreds of different genomic loci, and make up the remaining 95% of cellular RNA. Repetitive elements differ wildly in transcript abundance, with certain transcripts displaying expression levels that are orders of magnitude greater than other repetitive elements, even those from similar gene families. For example, SNORD3 is expressed at approximately 200,000 transcripts per cell, while SNORD13/14 is expressed only around 20,000 transcripts per cell (Falaleeva et al. 2017), and other SNORDS have yet lower levels of expression. The issue of abundance becomes apparent when RBPs recognize motifs that are common in multiple differentially expressed repetitive elements, such as C or D boxes in C/D snoRNAs. Here, the most highly expressed snoRNAs containing a common RBD would be overrepresented in an eCLIP dataset. Our SDE2 eCLIP suggested that SDE2 bound predominantly to the region flanking the C box in both SNORD3 and SNORD118, which are two of the most highly expressed C/D snoRNAs in cells. However, beyond SNORD3 and SNORD118, we detected a broad association with multiple C/D snoRNAs, which may have been underrepresented in our eCLIP due to their relatively low abundance when compared to SNORD3 and SNORD118. To that end, future investigation of SDE2's interaction with other C/D snoRNAs to determine the identification of the specific binding location within the transcript will allow us to grasp the full extent of the SDE2 RNA 'interactome.' If C box flanking regions are a common motif amongst C/D snoRNAs in our eCLIP dataset, it is

likely that SDE2 is a ubiquitous regulator of C/D snoRNAs, and implicates SDE2 in canonical C/D snoRNA 2'-O methylation. Global 2'-O methyl sequencing would uncover differential modification in the absence of SDE2, and loss of this modification is proposed to have many deleterious effects, including suboptimal ribosomal subunit biogenesis likely due to disrupted rRNA folding (Dennis et al. 2015; Sloan et al. 2017), as well as translation inhibition.

We observed a striking increase in 80S monosomes in our polysome profiling following SDE2 depletion. Interestingly, 80S monosomes have been observed to be enriched in NMD-triggering RNAs (Heyer and Moore 2016), which can cause stalling of the ribosome in the initial pioneer round of translation. The increased RI observed simultaneously with translation inhibition in the absence of SDE2 may account for the enrichment in 80S monosomes, possibly by overwhelming the ribosome recycling machinery (Serdar et al. 2020) and causing an accumulation of stalled monosomes unproductively engaged with an mRNA transcript. This can easily be tested using material collected from our polysome profiling experiment. RNA from the monosome fractions of both siCTRL and siSDE2 cells can be extracted and the presence of RI containing transcripts can be compared between conditions, as well as between monosome and polysome fractions within the same sample. If indeed RI-containing PTCs are responsible for monosome accumulation due to ribosomal stalling, then the ratio of RI to fully spliced transcripts should be highly increased in the monosome fraction, while decreased in the polysome fraction of siSDE2 treated cells. It is of note that many

splicing factors shown to increase RI following their depletion also incur translation deficits in cells (**Figure 4.2**) this could be indicative of a broader mechanism by which increased AS, in particular RI, impedes translation. It should be noted that the degree of translation repression corresponds, at least qualitatively, to the lethality of the silenced splicing factor. For example, PRPF8 is a pan-essential gene required for viability, and near complete loss of translation is observed following siRNA treatment. In contrast to this, TIA1 induces RI, but is a nonessential gene, and its silencing shows only moderate translation defects by SUnSET assay (**Figure 4.2**).

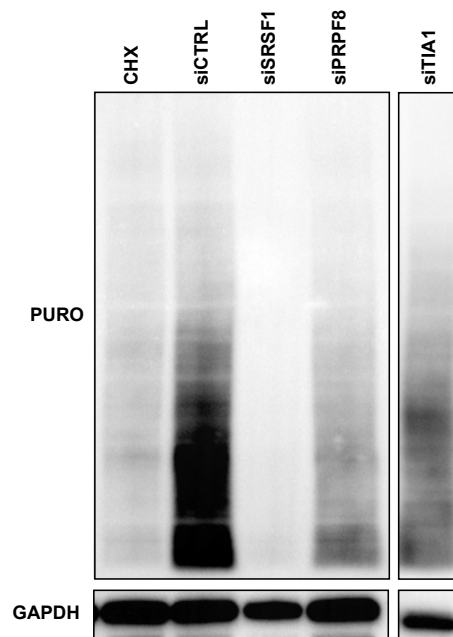**Figure 4.2. Knockdown of splicing components causes translation defects.** Western blot of SUnSET assay measuring global translation in HeLa cells 3 after treatment with siRNA against different spliceosome factors. Protein synthesis was measured by the incorporation of puromycin, and puromycin was detected using a puromycin specific antibody. CHX, cells treated with puromycin and cycloheximide.

Finally, certain cases of ribosomopathies have no mutation in genes known to be involved in ribosome biogenesis/assembly/maturation or translation. This is especially true in DBA, where 35% of cases have no known genetic cause (Engidaye, Melku, and Enawgaw 2019). As we are the first to implicate SDE2's function in the regulation of ribosomes and translation, it would be especially interesting, and could prove clinically useful, to sequence these patients to look at the mutational status of SDE2. While homozygous loss of SDE2 is likely lethal in humans, similar to homozygous loss of many ribosomal protein or biogenesis genes, heterozygous loss of SDE2 may result in manifestation of DBA or other ribosomopathies.

Section 3: Conclusions

Fundamental processes such as pre-mRNA splicing and ribosome biogenesis are imperative for cellular function. As such, these processes are under complex multifactorial and multilayered regulation to ensure their timely and efficient operation. Misregulation of these processes at any single step can result in a wide range of pathologies in humans, as well as resulting in cell death. While the complexity of both pre-mRNA splicing and ribosome biogenesis belies their importance, it also obstructs our complete understanding of these processes and hinders our identification of the full inventory of participating factors.

Here, we identify SDE2 as a gene involved in both ribosome biogenesis, as well as pre-mRNA splicing. We discovered that SDE2 is a bona fide RBP, with a strong preference for repetitive RNA elements. Most strikingly, SDE2 associates with a large subset of C/D snoRNAs, canonically implicated in modifying multiple RNA species with a 2'-O methylation of specific ribose rings. SDE2 also associated with C/D snoRNAs SNORD3 and SNORD118, which are required for rRNA processing and ribosomal biogenesis. We show that in the absence of SDE2, cells exhibit rRNA processing defects, decreased ribosome subunit biogenesis, nucleolar fragmentation, and impaired translation. These data are consistent with the requirement of SDE2 as a ribosome biogenesis factor. Moreover, we strengthened the idea that SDE2 is a member of the human spliceosome. We show that SDE2 interacts with the U2 snRNP, as well as additional spliceosomal components and auxiliary factors. In the absence of SDE2, proper splicing patterns are perturbed, resulting in widespread increases in AS, especially RI. These RI are defined by shortness in length, high GC content, and weak splice sites. The malfunction of the fundamental pathways of both pre-mRNA splicing and ribosome biogenesis due to SDE2 results in cell death in all cells tested, showing that SDE2 is an essential gene.

There likely exist other yet-uncharacterized genes with functions in both splicing and ribosome biogenesis similar to SDE2. The identification of all factors involved in these processes will allow us to probe these pathways for therapeutic vulnerabilities in human disease.

APPENDIX A

TABLE 2

| Primer/Probe name	Primer/Probe sequence
ADAMTSL4 FWD	TGAGTCGGGAGAGGAACT
ADAMTSL4 REV	GGCAAGACTGGGTGATGTT
ARSA FWD	CAGAACCTGACCTGCTTCC
ARSA REV	GATGACCAGCGTCTCTTCAA
CUL7 FWD	TGAGCGAGGCTCACAGA
CUL7 REV	TTTGCTCGTGAAGGTCCAG
DGKQ FWD	ATTGCCCAGGGTTCCTACTT
DGKQ REV	GCCTCCAGACCACCTGAAA
FBF1 FWD	AGCCTGCACGAGTTGTC
FBF1 REV	CCGGCTCTGCTCATTCA
KIFC2 FWD	TCTTCAGAGAGCTGGAACCT
KIFC2 REV	CTCCACCATGCTGAGTGTC
PARP10 FWD	CAGACAGTACAGGTGGTGAC
PARP10 REV	CCTCCACATCCTGCTCAAA
PITPNM1 FWD	GGCACCGAAGTCACCAATA
PITPNM1 REV	CTGCCCATGATGGAGACG
RFNG FWD	TTTGCCACGTGGATGATGA
RFNG REV	CCATGGGCTCATCTTGAGG
RHBDF1 FWD	GAAGCCACCCTGGCTAAA
RHBDF1 REV	CGGATGGTCTGTGTGATGG
SPATA20 FWD	GCTTGATGGTGTCAGGCTAT
SPATA20 REV	AAGAGCTTGTCCTGTGTGTC
5' ETS	CGGAGGCCCAACCTCTCCGACGACAGGTCGCCAGAGGACAGCGTGTCAGC
ITS1	CCTCGCCCTCCGGGCTCCGGGCTCCGTTAATGATC
ITS2	CTGCGAGGGAACCCCCAGCCGCGCA
5.8S	CAATGTGTCCTGCAATTCAC
SNORD3	AACGATCATCAATGGCTGACGGCAGTTGCAGCCAAG
SNORD118	GTTCTAATCTGCCCTCCGGAGGAGGAACAG

139
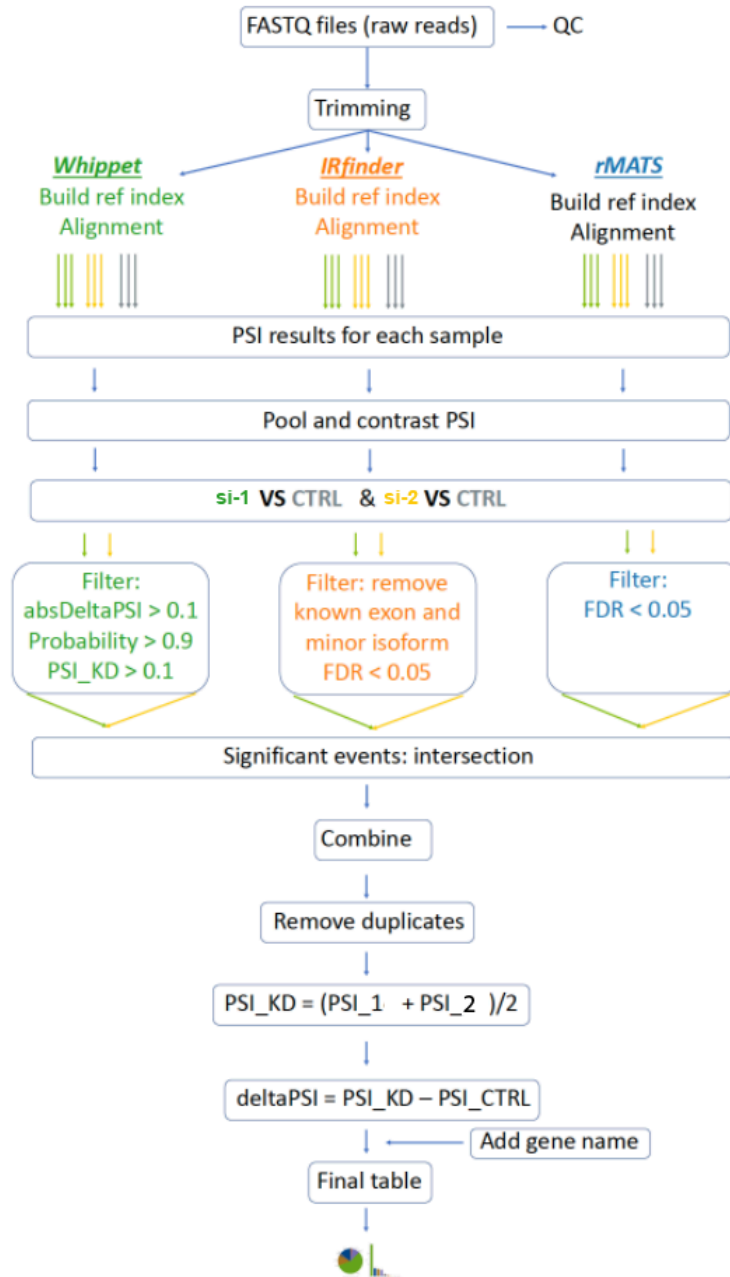
Figure AB1. AS analysis and target validation by IGV. Conceptual diagram of dAS analytical pipeline, starting from raw mRNA-Seq reads through to final dAS event table.
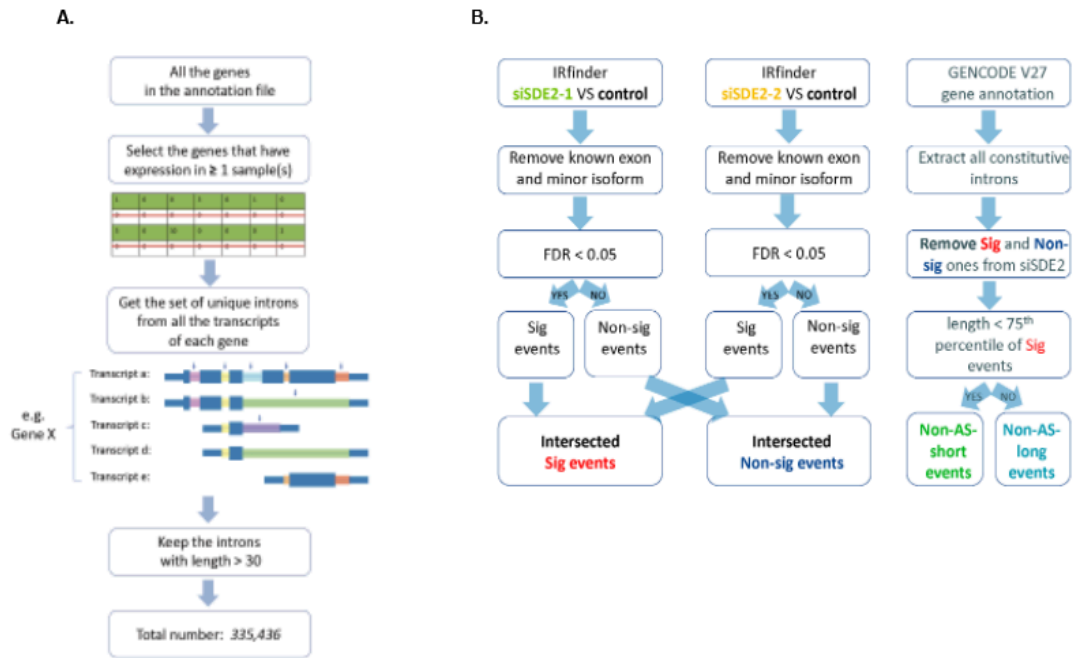
Figure AB2. Identification of significantly retained introns and non-significantly retained introns amongst all introns in the genome. (A) Conceptual diagram of pipeline to identify all constitutive introns in the genome as defined by the GENCODE gene annotation. (B) Pipeline identifying significant and non-significant introns from the SDE2 knockout experiments compared with all constitutive introns from A. Only RI events found as significant in both replicates are considered Sig events; all other events identified by IRfinder are either significant in only one replicate, or significant in neither, indicating they are putative real dAS events that did not reach statistical significance. All introns not included in Sig events or Non-sig events groups are retained as Non-AS RI events and retained for comparison.
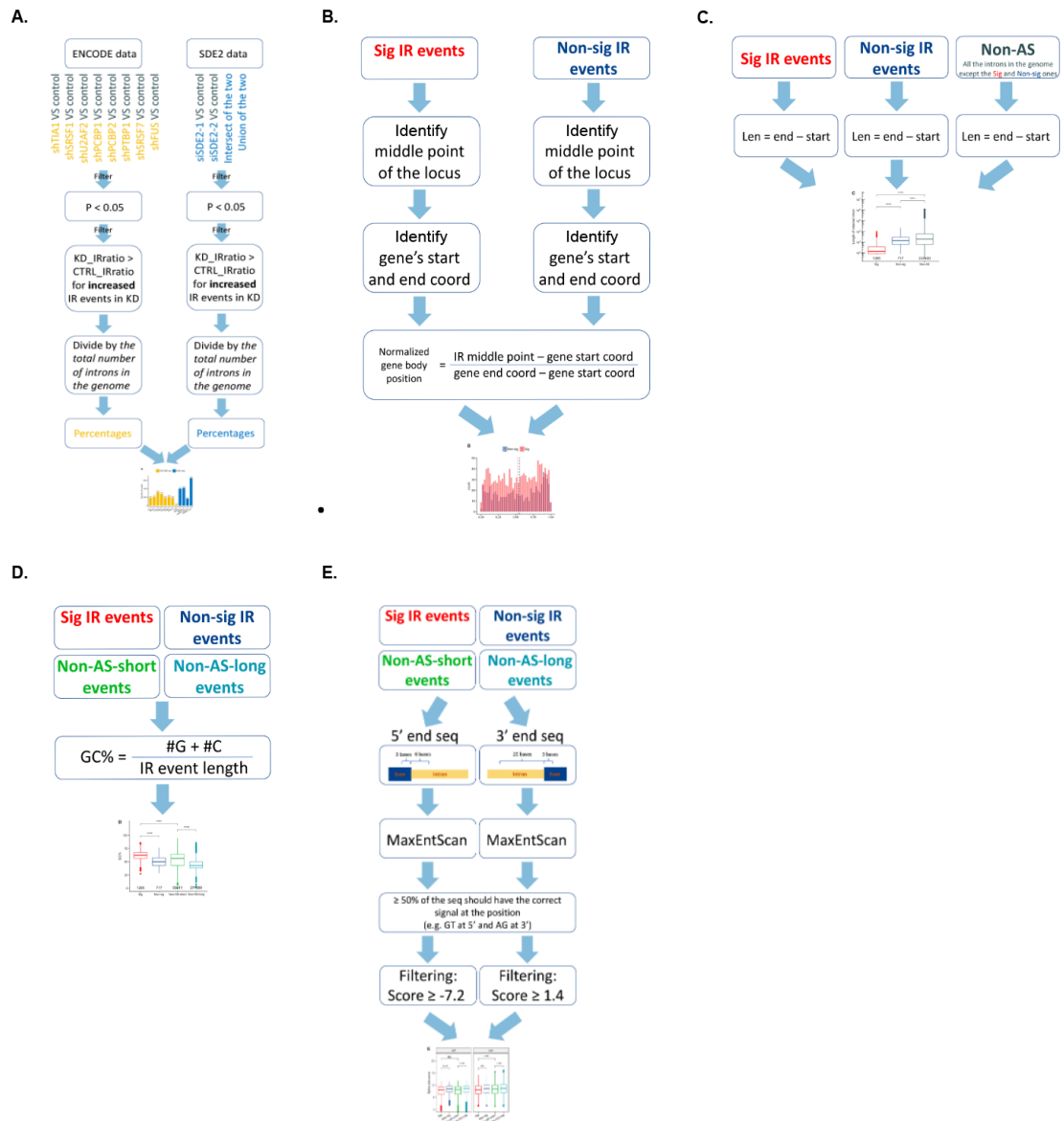
Figure AB3. Schematic representation of workflow and calculations for RI event characterization. (A) Workflow for computing the percentage of significantly increased RI events in whole intron populations with ENCODE and SDE2 data respectively. (B) Workflow for computing the RI event's normalized gene body position in Sig and Non-sig groups. (C) Workflow for computing the length of the retained intron in Sig, Non-sig and Non-AS groups. (D) Workflow for computing the GC content of the IR events in all Sig, Non-sig, Non-AS-short, and Non-AS-long groups. (E) Workflow for computing and filtering the 5' and 3' splice site scores of the 4 groups from (D).

142

Repetitive Element (RE) Analysis Pipeline

Figure AC1. Repetitive Element (RE) Analysis pipeline. Schematic of repetetive element bioinformatic pipeline for eCLIP.

BIBLIOGRAPHY

Achsel, Tilmann et al. 1999. "A Doughnut-Shaped Heteromer of Human Sm-like Proteins Binds to the 3'-End of U6 SnRNA, Thereby Facilitating U4/U6 Duplex Formation in Vitro." *EMBO Journal* 18(20): 5789–5802.
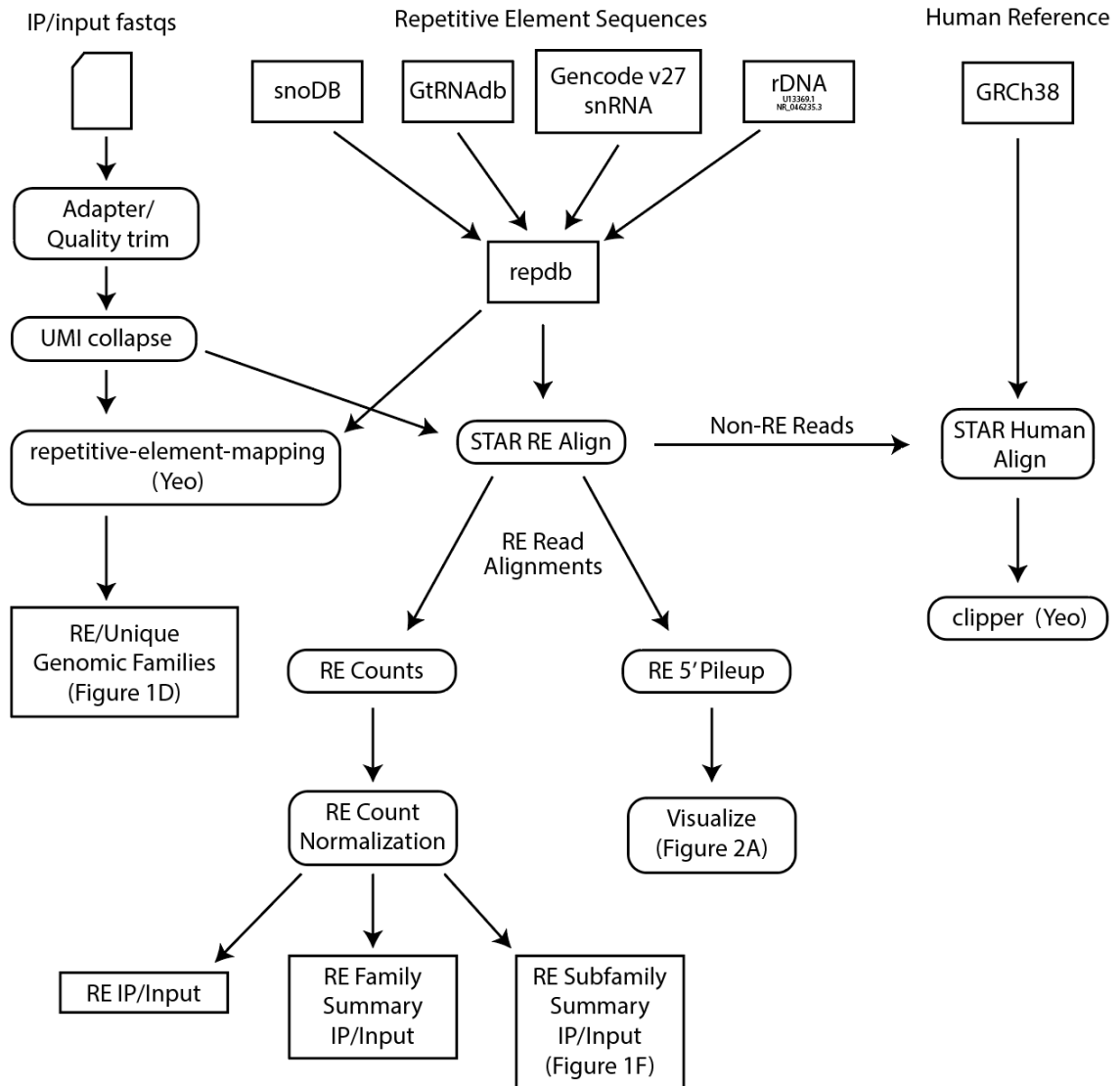
ADACHI, K. 2003. "Heterogous Dystrophin MRNA Produced by a Novel Splice Acceptor Site Mutation in Intermediate Dystrophinopathy." *Pediatric Research* 53(1): 125–31.

Ahn, Andrew H., and Louis M. Kunkel. 1993. "The Structural and Functional Diversity of Dystrophin." *Nature Genetics*.

Allemand, Eric et al. 2016. "A Broad Set of Chromatin Factors Influences Splicing." *PLoS Genetics* 12(9): 1–21.

Allmang, Christine, Philip Mitchell, Elisabeth Petfalski, and David Tollervey. 2000. "Degradation of Ribosomal RNA Precursors by the Exosome." *Nucleic Acids Research* 28(8): 1684–91.

Andrews, Simon. 2015. "FASTQC A Quality Control Tool for High Throughput Sequence Data." *Babraham Institute*.

Aravind, L., and Eugene V. Koonin. 2000. "SAP - A Putative DNA-Binding Motif Involved in Chromosomal Organization." *Trends in Biochemical Sciences* 25(3): 112–14.

Aspesi, Anna, and Steven R. Ellis. 2019. "Rare Ribosomopathies: Insights into Mechanisms of Cancer." *Nature Reviews Cancer* 19(4): 228–38. http://dx.doi.org/10.1038/s41568-019-0105-0.

Aubert, Maxime, Marie Françoise O'donohue, Simon Lebaron, and Pierre Emmanuel Gleizes. 2018. "Pre-Ribosomal RNA Processing in Human Cells: From Mechanisms to Congenital Diseases." *Biomolecules* 8(4).

Avni, Dror, Yael Biberman, and Oded Meyuhas. 1996. "The 5′ Terminal Oligopyrimidine Tract Confers Translational Control on Top Mrnas in a Cell Type-and Sequence Context-Dependent Manner." *Nucleic Acids Research* 25(5): 995–1001.

Bao, Weidong, Kenji K Kojima, and Oleksiy Kohany. 2015. "Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes." *Mobile DNA* 6(1): 11.

Baßler, Jochen, and Ed Hurt. 2019. "Eukaryotic Ribosome Assembly." *Annual*

Review of Biochemistry 88(1): 281–306.

Bayne, Elizabeth H. et al. 2008. "Splicing Factors Facilitate RNAi-Directed Silencing in Fission Yeast." *Science*.

Beckmann BM, Horos R, Fischer B, Castello A, Eichelbaum K, Alleaume AM, Schwarzl T, Curk T, Foehr S, Huber W, Krijgsveld J, Hentze MW. 2015. "The RNA-Binding Proteomes from Yeast to Man Harbour Conserved EnigmRBPs." *Nat Commun* 6(10127).

Bentley, David L. 2014. "Coupling MRNA Processing with Transcription in Time and Space." *Nature Reviews Genetics* 15(3): 163–75.

Bertrand, Edouard. 2017. "Assembly and Traf Fi Cking of Box C / D and H / ACA SnoRNPs." 14(6): 680–92.

Bessonov, Sergey et al. 2008. "Isolation of an Active Step I Spliceosome and Composition of Its RNP Core." *Nature* 452(7189): 846–50.

Blijlevens, Maxime, Jing Li, and Victor W. van Beusechem. 2021. "Biology of the Mrna Splicing Machinery and Its Dysregulation in Cancer Providing Therapeutic Opportunities." *International Journal of Molecular Sciences* 22(10).

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics*.

Bouchard-Bourelle, Philia et al. 2020. "SnoDB: An Interactive Database of Human SnoRNA Sequences, Abundance and Interactions." *Nucleic acids research* 48(D1): D220–25.

Boutz, Paul L., Arjun Bhutkar, and Phillip A. Sharp. 2015. "Detained Introns Are a Novel, Widespread Class of Post-Transcriptionally Spliced Introns." *Genes and Development* 29(1): 63–80.

Bratkovič, Tomaz, Janja Bozič, and Boris Rogelj. 2020. "Functional Diversity of Small Nucleolar RNAs." *Nucleic Acids Research* 48(4): 1627–51.

Braunschweig, Ulrich et al. 2014. "Widespread Intron Retention in Mammals Functionally Tunes Transcriptomes." *Genome Research* 24(11): 1774–86.

Breitbart, R. 1987. "Alternative Splicing: A Ubiquitous Mechanism For The Generation Of Multiple Protein Isoforms From Single Genes." *Annual Review of Biochemistry*.

Burset, M., I. A. Seledtsov, and V. V. Solovyev. 2000. "Analysis of Canonical and

Non-Canonical Splice Sites in Mammalian Genomes." *Nucleic Acids Research* 28(21): 4364–75.

Castello, Alfredo et al. 2016. "Comprehensive Identification of RNA-Binding Domains in Human Cells." *Molecular Cell* 63(4): 696–710. http://dx.doi.org/10.1016/j.molcel.2016.06.029.

Chan, Patricia P, and Todd M Lowe. 2016. "GtRNAdb 2.0: An Expanded Database of Transfer RNA Genes Identified in Complete and Draft Genomes." *Nucleic acids research* 44(D1): D184–89.

Chanarat S, Mishra SK. 2018. "Emerging Roles of Ubiquitin-like Proteins in Pre-MRNA Splicing." *Trends Biochem Sci.* 43(11): 896–907.

Chanarat, Sittinan, and Shravan Kumar Mishra. 2018. "Emerging Roles of Ubiquitin-like Proteins in Pre-MRNA Splicing." *Trends in Biochemical Sciences* 43(11): 896–907. https://doi.org/10.1016/j.tibs.2018.09.001.

Chen, Danqi et al. 2020. "Polyadenylation of Histone H3.1 MRNA Promotes Cell Transformation by Displacing H3.3 from Gene Regulatory Elements." *iScience* 23(9): 101518. http://dx.doi.org/10.1016/j.isci.2020.101518.

Chen, Inês. 2019. "An Antisense Oligonucleotide Splicing Modulator to Treat Spinal Muscular Atrophy." *Nature Research 2020*.

Chen, Weijun et al. 2014. "Endogenous U2•U5•U6 SnRNA Complexes in S. Pombe Are Intron Lariat Spliceosomes." *RNA* 20(3): 308–20.

Clipper. "Clipper. n.d. Github."

De Conti, Laura, Marco Baralle, and Emanuele Buratti. 2013. "Exon and Intron Definition in Pre-MRNA Splicing." *Wiley Interdisciplinary Reviews: RNA* 4(1): 49–60.

Dennis, Patrick P. et al. 2015. "C/D Box SRNA-Guided 2'-O-Methylation Patterns of Archaeal RRNA Molecules." *BMC Genomics* 16(1): 1–12. http://dx.doi.org/10.1186/s12864-015-1839-z.

Depmap Broad. 2019. "DepMap Achilles 19Q1 Public." *figshare. Fileset.*

Deryusheva, Svetlana, and Joseph G. Gall. 2019. "Scarnas and SnoRNAs: Are They Limited to Specific Classes of Substrate RNAs?" *Rna* 25(1): 17–22.

Didychuk, Allison L., Samuel E. Butcher, and David A. Brow. 2018. "The Life of U6 Small Nuclear RNA, from Cradle to Grave." *Rna* 24(4): 437–60.

Dieci, Giorgio, Milena Preti, and Barbara Montanini. 2009. "Eukaryotic SnoRNAs: A Paradigm for Gene Expression Flexibility." *Genomics* 94(2): 83–88. http://dx.doi.org/10.1016/j.ygeno.2009.05.002.

Dobin, Alexander et al. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." *Bioinformatics*.

Dominski, Zbigniew, Lian-Xing Zheng, Ricardo Sanchez, and William F. Marzluff. 1999. "Stem-Loop Binding Protein Facilitates 3′-End Formation by Stabilizing U7 SnRNP Binding to Histone Pre-MRNA." *Molecular and Cellular Biology* 19(5): 3561–70.

Dönmez, Gizem, Klaus Hartmuth, and Reinhard Lührmann. 2004. "Modified Nucleotides at the 5′ End of Human U2 SnRNA Are Required for Spliceosomal E-Complex Formation." *Rna* 10(12): 1925–33.

Dunham, Ian et al. 2012. "An Integrated Encyclopedia of DNA Elements in the Human Genome." *Nature*.

Dupuis-Sandoval, Fabien, Mikaël Poirier, and Michelle S. Scott. 2015. "The Emerging Landscape of Small Nucleolar RNAs in Cell Biology." *Wiley Interdisciplinary Reviews: RNA* 6(4): 381–97.

Dvinge, Heidi, Eunhee Kim, Omar Abdel-Wahab, and Robert K. Bradley. 2016. "RNA Splicing Factors as Oncoproteins and Tumour Suppressors." *Nature Reviews Cancer*.

Elliott, Brittany A. et al. 2019. "Modification of Messenger RNA by 2′-O-Methylation Regulates Gene Expression in Vivo." *Nature Communications* 10(1): 1–9. http://dx.doi.org/10.1038/s41467-019-11375-7.

Engidaye, Getabalew, Mulugeta Melku, and Bamlaku Enawgaw. 2019. "Diamond Blackfan Anemia: Genetics, Pathogenesis, Diagnosis and Treatment." *Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine* 30(1): 67–81.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics*.

Faà, Valeria et al. 2009. "Characterization of a Disease-Associated Mutation Affecting a Putative Splicing Regulatory Element in Intron 6b of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene." *Journal of Biological Chemistry* 284(44): 30024–31.

Falaleeva, Marina, and Stefan Stamm. 2013. "Processing of SnoRNAs as a New Source of Regulatory Non-Coding RNAs: SnoRNA Fragments Form a New Class of Functional RNAs." *BioEssays* 35(1): 46–54.

Falaleeva, Marina, Justin R. Welden, Marilyn J. Duncan, and Stefan Stamm. 2017. "C/D-Box SnoRNAs Form Methylating and Non-Methylating Ribonucleoprotein Complexes: Old Dogs Show New Tricks." *BioEssays* 39(6): 1–28.

Fica, Sebastian M et al. 2019a. "A Human Postcatalytic Spliceosome Structure Reveals Essential Roles of Metazoan Factors for Exon Ligation." *Science*: eaaw5569.

———. 2019b. "R ES E A RC H." 714(February): 710–14.

Fischer, Utz, Clemens Englbrecht, and Ashwin Chari. 2011. "Biogenesis of Spliceosomal Small Nuclear Ribonucleoproteins." *Wiley Interdisciplinary Reviews: RNA* 2(5): 718–31.

Förch, Patrik et al. 2002. "The Splicing Regulator TIA-1 Interacts with U1-C to Promote U1 SnRNP Recruitment to 5′ Splice Sites." *EMBO Journal* 21(24): 6882–92.

Fritz, Sébastien et al. 2018. "An Initiator Codon Mutation in SDE2 Causes Recessive Embryonic Lethality in Holstein Cattle." *Journal of Dairy Science* 101(7): 6220–31. http://dx.doi.org/10.3168/jds.2017-14119.

Galante, Pedro Alexandre Favoretto, Noboru Jo Sakabe, Natanja Kirschbaum-Slager, and Sandro José De Souza. 2004. "Detection and Evaluation of Intron Retention Events in the Human Transcriptome." *RNA* 10(5): 757–65.

Ge, Ying, and Bo T. Porse. 2014. "The Functional Consequences of Intron Retention: Alternative Splicing Coupled to NMD as a Regulator of Gene Expression." *BioEssays*.

Gebauer F, Schwarzl T, Valcárcel J, Hentze MW. 2021. "RNA-Binding Proteins in Human Genetic Disease." *Nat Rev Genet.* 22((3)): 185–98.

Gencheva, Marieta, Mitsuo Kato, Alain N.S. Newo, and Ren Jang Lin. 2010. "Contribution of DEAH-Box Protein DHX16 in Human Pre-MRNA Splicing." *Biochemical Journal* 429(1): 25–32.

Gerstberger, Stefanie, Markus Hafner, and Thomas Tuschl. 2014. "A Census of Human RNA-Binding Proteins." *Nature Reviews Genetics* 15(12): 829–45.

Golomb, Lior, Sinisa Volarevic, and Moshe Oren. 2014. "P53 and Ribosome Biogenesis Stress: The Essentials." *FEBS Letters* 588(16): 2571–79.

Gonatopoulos-Pournatzis, Thomas, and Victoria H. Cowling. 2014. "Cap-Binding Complex (CBC)." *Biochemical Journal* 457(2): 231–42.

Graveley, Brenton R. 2001. "Alternative Splicing: Increasing Diversity in the Proteomic World." *Trends in Genetics*.

Grummt, Ingrid. 2003. "Life on a Planet of Its Own: Regulation of RNA Polymerase I Transcription in the Nucleolus." *Genes and Development* 17(14): 1691–1702.

Grützmann, Konrad et al. 2014. "Fungal Alternative Splicing Is Associated with Multicellular Complexity and Virulence: A Genome-Wide Multi-Species Study." *DNA Research*.

Grzybowska, Ewa A. 2012. "Human Intronless Genes: Functional Groups, Associated Diseases, Evolution, and MRNA Processing in Absence of Splicing." *Biochemical and Biophysical Research Communications* 424(1): 1–6. http://dx.doi.org/10.1016/j.bbrc.2012.06.092.

Guiro, Joana, and Shona Murphy. 2017. "Regulation of Expression of Human RNA Polymerase II-Transcribed SnRNA Genes." *Open Biology* 7(6): 3–11.

Harrow, Jennifer et al. 2012. "GENCODE: The Reference Human Genome Annotation for the ENCODE Project." *Genome Research* 22(9): 1760–74.

Havens, Mallory A., Dominik M. Duelli, and Michelle L. Hastings. 2013. "Targeting RNA Splicing for Disease Therapy." *Wiley Interdisciplinary Reviews: RNA* 4(3): 247–66.

He, Chongsheng et al. 2016. "High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells." *Molecular Cell*.

He, Jin-Shu et al. 2018. "High-Content Imaging Approaches to Quantitate Stress-Induced Changes in Nucleolar Morphology." *ASSAY and Drug Development Technologies* 16(6): 320–32.

Heath, Catherine G., Nicolas Viphakone, and Stuart A. Wilson. 2016. "The Role of TREX in Gene Expression and Disease." *Biochemical Journal* 473(19): 2911–35.

Henras, Anthony K. et al. 2015. "An Overview of Pre-Ribosomal RNA Processing in Eukaryotes." *Wiley Interdisciplinary Reviews: RNA* 6(2): 225–42.

Hernandez-Verdun D, Roussel P, Thiry M, Sirri V, Lafontaine DL. 2010. "The Nucleolus: Structure/Function Relationship in RNA Metabolism." *Wiley Interdiscip Rev RNA* 1(3): 415–31.

Herzel, Lydia, Diana S.M. Ottoz, Tara Alpert, and Karla M. Neugebauer. 2017. "Splicing and Transcription Touch Base: Co-Transcriptional Spliceosome Assembly and Function." *Nature Reviews Molecular Cell Biology* 18(10): 637–50. http://dx.doi.org/10.1038/nrm.2017.63.

Heyer, Erin E., and Melissa J. Moore. 2016. "Redefining the Translational Status of 80S Monosomes." *Cell* 164(4): 757–69.

Hirose, T., and J. A. Steitz. 2001. "Position within the Host Intron Is Critical for Efficient Processing of Box C/D SnoRNAs in Mammalian Cells." *Proceedings of the National Academy of Sciences of the United States of America* 98(23): 12914–19.

Hong, Xin, Douglas G. Scofield, and Michael Lynch. 2006. "Intron Size, Abundance, and Distribution within Untranslated Regions of Genes." *Molecular Biology and Evolution* 23(12): 2392–2404.

Hua, Yimin et al. 2007. "Enhancement of SMN2 Exon 7 Inclusion by Antisense Oligonucleotides Targeting the Exon." *PLoS Biology* 5(4): 729–44.

Huber, Jochen et al. 1998. "Snurportin1, an M3G-Cap-Specific Nuclear Import Receptor with a Novel Domain Structure." *EMBO Journal* 17(14): 4114–26.

Hughes, John M X. 1996. "Functional Base-Pairing Interaction Between Highly Conserved Elements of U3 Small Nucleolar RNA and the Small Ribosomal Subunit RNA." *Journal of Molecular Biology* 259(4): 645–54.

Huppertz, Ina et al. 2014. "ICLIP: Protein-RNA Interactions at Nucleotide Resolution." *Methods*.

Iida, Tetsushi, Rika Kawaguchi, and Jun ichi Nakayama. 2006. "Conserved Ribonuclease, Eri1, Negatively Regulates Heterochromatin Assembly in Fission Yeast." *Current Biology* 16(14): 1459–64.

Jack, Karen et al. 2011. "RRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells." *Molecular Cell* 44(4): 660–66. http://dx.doi.org/10.1016/j.molcel.2011.09.017.

Jacob, Aishwarya G., and Christopher W.J. Smith. 2017a. "Intron Retention as a Component of Regulated Gene Expression Programs." *Human Genetics* 136(9): 1043–57.

———. 2017b. "Intron Retention as a Component of Regulated Gene Expression Programs." *Human Genetics* 136(9): 1043–57.

Jády, B E, and T Kiss. 2001. "A Small Nucleolar Guide RNA Functions Both in 2'-O-Ribose Methylation and Pseudouridylation of the U5 Spliceosomal RNA." *The EMBO journal* 20(3): 541–51.

Jády, Beáta E., and Tamás Kiss. 2001. "A Small Nucleolar Guide RNA Functions Both in 2′-O-Ribose Methylation and Pseudouridylation of the U5 Spliceosomal RNA." *EMBO Journal* 20(3): 541–51.

Jenkinson, Emma M. et al. 2016. "Mutations in SNORD118 Cause the Cerebral Microangiopathy Leukoencephalopathy with Calcifications and Cysts." *Nature Genetics* 48(10): 1185–92.

Ji, Zhe, Ruisheng Song, Aviv Regev, and Kevin Struhl. 2015. "Many LncRNAs, 5′UTRs, and Pseudogenes Are Translated and Some Are Likely to Express Functional Proteins." *eLife* 4(DECEMBER2015): 1–21.

Jian, Xueqiu, Eric Boerwinkle, and Xiaoming Liu. 2014. "In Silico Tools for Splicing Defect Prediction: A Survey from the Viewpoint of End Users." *Genetics in Medicine* 16(7): 497–503.

Jiang, Wei, and Liang Chen. 2021. "Alternative Splicing: Human Disease and Quantitative Analysis from High-Throughput Sequencing." *Computational and Structural Biotechnology Journal* 19: 183–95. https://doi.org/10.1016/j.csbj.2020.12.009.

Jo, Ukhyun et al. 2016. "PCNA-Dependent Cleavage and Degradation of SDE2 Regulates Response to Replication Stress" ed. Kristijan Ramadan. *PLOS Genetics* 12(12): e1006465. http://www.ncbi.nlm.nih.gov/pubmed/27906959 (July 22, 2017).

Jorjani, Hadi et al. 2016. "An Updated Human SnoRNAome." *Nucleic acids research* 44(11): 5068–82.

Jurica, Melissa S., and Melissa J. Moore. 2003. "Pre-MRNA Splicing: Awash in a Sea of Proteins." *Molecular Cell*.

Kampen, Kim R., Sergey O. Sulima, Stijn Vereecke, and Kim de Keersmaecker. 2021. "Hallmarks of Ribosomopathies." *Nucleic Acids Research* 48(3): 1013–28.

Karijolich, John, and Yi-Tao Yu. 2010. "Spliceosomal SnRNA Modifications and Their Function." *RNA Biology* 7(2): 192–204.

Ke, Shengdong, and Lawrence A. Chasin. 2011. "Context-Dependent Splicing Regulation: Exon Definition, Co-Occurring Motif Pairs and Tissue Specificity." *RNA Biology* 8(3): 384–88.

Kiss-László, Zsuzsanna et al. 1996. "Site-Specific Ribose Methylation of Preribosomal RNA: A Novel Function for Small Nucleolar RNAs." *Cell* 85(7): 1077–88.

Kiss, T. 2002. "Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions." *Cell* 109(2): 145–48.

Kondo, Yasushi, Chris Oubridge, Anne Marie M. van Roon, and Kiyoshi Nagai. 2015. "Crystal Structure of Human U1 SnRNP, a Small Nuclear Ribonucleoprotein Particle, Reveals the Mechanism of 5' Splice Site Recognition." *eLife* 4: 1–19.

Krieg, P A, and D A Melton. 1999. "Formation of the 3' End of Histone MRNA." *Gene* 239(1): 203–6.

Kufel, Joanna, and Pawel Grzechnik. 2019. "Small Nucleolar RNAs Tell a Different Tale." *Trends in Genetics* 35(2): 104–17. https://doi.org/10.1016/j.tig.2018.11.005.

Kühn, Uwe et al. 2009. "Poly(A) Tail Length Is Controlled by the Nuclear Poly(A)-Binding Protein Regulating the Interaction between Poly(A) Polymerase and the Cleavage and Polyadenylation Specificity Factor." *Journal of Biological Chemistry* 284(34): 22803–14.

Lafontaine, Denis L.J., Joshua A. Riback, Rümeyza Bascetin, and Clifford P. Brangwynne. 2021. "The Nucleolus as a Multiphase Liquid Condensate." *Nature Reviews Molecular Cell Biology* 22(3): 165–82. http://dx.doi.org/10.1038/s41580-020-0272-6.

Lagier-Tourenne, Clotilde, Magdalini Polymenidou, and Don W. Cleveland. 2010. "TDP-43 and FUS/TLS: Emerging Roles in RNA Processing and Neurodegeneration." *Human Molecular Genetics*.

Langhendries, Jean-louis et al. 2016. "#79 Langhendries Oncotarget 2016 SnoRNA U3 and U8 Required for Pre-RRNA Processing." 7(37).

Langhendries, Jean-Louis et al. 2016. "The Human Box C/D SnoRNAs U3 and U8 Are Required for Pre-RRNA Processing and Tumorigenesis." *Oncotarget* 7(37): 59519–34.

Lestrade, Laurent, and Michel J. Weber. 2006. "SnoRNA-LBME-Db, a

Comprehensive Database of Human H/ACA and C/D Box SnoRNAs." *Nucleic acids research* 34(Database issue): 158–62.

Li, Xueni et al. 2020. "Splicing." 573(7774): 375–80.

Lin, Shi Lung, Joseph D. Miller, and Shao Yao Ying. 2006. "Intronic MicroRNA (MiRNA)." *Journal of Biomedicine and Biotechnology* 2006: 1–13.

Lovci, Michael T et al. 2013. "Rbfox Proteins Regulate Alternative MRNA Splicing through Evolutionarily Conserved RNA Bridges." *Nature structural & molecular biology* 20(12): 1434–42.

Ludwig, Leif S. et al. 2014. "Altered Translation of GATA1 in Diamond-Blackfan Anemia." *Nature Medicine* 20(7): 748–53.

Lukowiak, A A et al. 2000. "Interaction of the U3-55k Protein with U3 SnoRNA Is Mediated by the Box B/C Motif of U3 and the WD Repeats of U3-55k." *Nucleic acids research* 28(18): 3462–71.

Lunde, Bradley M., Claire Moore, and Gabriele Varani. 2007. "RNA-Binding Proteins: Modular Design for Efficient Function." *Nature Reviews Molecular Cell Biology* 8(6): 479–90.

Luo, Ang, Yao Gong, Hyungjin Kim, and Yue Chen. 2020. "Proteome Dynamics Analysis Identifies Functional Roles of SDE2 and Hypoxia in DNA Damage Response in Prostate Cancer Cells." *NAR Cancer* 2(2): 1–12.

Maniatis, Tom, and Bosiljka Tasic. 2002. "Alternative Pre-MRNA Splicing and Proteome Expansion in Metazoans." *Nature*.

Martelly, William et al. 2019. "Identification of a Noncanonical RNA Binding Domain in the U2 SnRNP Protein SF3A1." *Rna* 25(11): 1509–21.

Massenet, Séverine et al. 2002. "The SMN Complex Is Associated with SnRNPs throughout Their Cytoplasmic Assembly Pathway." *Molecular and Cellular Biology* 22(18): 6533–41.

Masuda, Seiji et al. 2005. "Recruitment of the Human TREX Complex to MRNA during Splicing." *Genes and Development* 19(13): 1512–17.

Matera, A. Gregory, and Zefeng Wang. 2014. "Erratum: A Day in the Life of the Spliceosome (Nature Reviews Molecular Cell Biology (2014) 15 (108-122))." *Nature Reviews Molecular Cell Biology* 15(4): 294.

Mattick, John S., and Igor V. Makunin. 2006. "Non-Coding RNA." *Human molecular genetics* 15 Spec No(1): 17–29.

McKeegan, Kenneth Scott et al. 2007. "A Dynamic Scaffold of Pre-SnoRNP Factors Facilitates Human Box C/D SnoRNP Assembly." *Molecular and Cellular Biology* 27(19): 6782–93.

Mefford, Melissa A., and Jonathan P. Staley. 2009. "Evidence That U2/U6 Helix I Promotes Both Catalytic Steps of Pre-MRNA Splicing and Rearranges in between These Steps." *Rna* 15(7): 1386–97.

Meinel, Dominik M. et al. 2013. "Recruitment of TREX to the Transcription Machinery by Its Direct Binding to the Phospho-CTD of RNA Polymerase II." *PLoS Genetics* 9(11).

Middleton, Robert et al. 2017. "IRFinder: Assessing the Impact of Intron Retention on Mammalian Gene Expression." *Genome Biology* 18(1): 1–11.

Mills, Eric W., and Rachel Green. 2017. "Ribosomopathies: There's Strength in Numbers." *Science* 358(6363).

Morais, Pedro, Hironori Adachi, and Yi Tao Yu. 2021. "Spliceosomal SnRNA Epitranscriptomics." *Frontiers in Genetics* 12(March).

Morris, Kevin V., and John S. Mattick. 2014. "The Rise of Regulatory RNA." *Nature Reviews Genetics* 15(6): 423–37.

Moss, Tom, Jean Clement Mars, Michel G. Tremblay, and Marianne Sabourin-Felix. 2019. "The Chromatin Landscape of the Ribosomal RNA Genes in Mouse and Human." *Chromosome Research* 27(1–2): 31–40.

Mouaikel, John et al. 2002. "Hypermethylation of the Cap Structure of Both Yeast SnRNAs and SnoRNAs Requires a Conserved Methyltransferase That Is Localized to the Nucleolus." *Molecular Cell* 9(4): 891–901.

———. 2003. "Interaction between the Small-Nuclear-RNA Cap Hypermethylase and the Spinal Muscular Atrophy Protein, Survival of Motor Neuron." *EMBO Reports* 4(6): 616–22.

Nasim, Faiz Ul Hassan, Stephen Hutchison, Mélanie Cordeau, and Benoit Chabot. 2002. "High-Affinity HnRNP A1 Binding Sites and Duplex-Forming Inverted Repeats Have Similar Effects on 5′ Splice Site Selection in Support of a Common Looping out and Repression Mechanism." *Rna* 8(8): 1078–89.

Ner-Gaon, Hadas et al. 2004. "Intron Retention Is a Major Phenomenon in Alternative Splicing in Arabidopsis." *Plant Journal*.

Nissen, Poul et al. 2020. "The Structural Basis of Ribosome Activity in Peptide

Bond Synthesis." *Structural Insights into Gene Expression and Protein Synthesis* 289(August): 501–11.

Van Nostrand, Eric L et al. 2016. "Robust Transcriptome-Wide Discovery of RNA-Binding Protein Binding Sites with Enhanced CLIP (ECLIP)." *Nature methods* 13(6): 508–14.

Nostrand, Eric L Van et al. 2017. "A Large-Scale Binding and Functional Map of Human RNA Binding Proteins Correspondence and Requests for Materials Should Be Addressed to Brenton Graveley (Graveley@uchc.Edu), Chris Burge (Cburge@mit.Edu), Xiang-Dong Fu (Xdfu@ucsd.Edu),." : 1–74.

Van Nostrand, Eric L et al. 2020. "A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins." *Nature* 583(7818): 711–19.

Ohno, Mutsuhito et al. 2000. "PHAX, a Mediator of U SnRNA Nuclear Export Whose Activity Is Regulated by Phosphorylation." *Cell* 101(2): 187–98.

Ojha, Sandeep, Sulochan Malla, and Shawn M. Lyons. 2020. "SnoRNPs: Functions in Ribosome Biogenesis." *Biomolecules* 10(5): 1–28.

Ottoz, Diana S.M., and Luke E. Berchowitz. 2020. "The Role of Disorder in RNA Binding Affinity and Specificity: The Role of Disorder in RNA Binding." *Open Biology* 10(12).

Pan, Qun et al. 2008. "Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing." *Nature Genetics*.

Panse, Vikram Govind, and Arlen W. Johnson. 2010. "Maturation of Eukaryotic Ribosomes: Acquisition of Functionality." *Trends in Biochemical Sciences* 35(5): 260–66.

Pederson, Thoru. 2011. "The Nucleolus." *Cold Spring Harbor perspectives in biology* 3(3): a000638.

Piano, Fabio et al. 2002. "PianoF02CurrBiol.Pdf." 12(02): 1959–64.

Rageul, Julie et al. 2019. "Conditional Degradation of SDE2 by the Arg/N-End Rule Pathway Regulates Stress Response at Replication Forks." *Nucleic acids research*.

Rahman, Mohammad Alinoor et al. 2013. "HnRNP L and HnRNP LL Antagonistically Modulate PTB-Mediated Splicing Suppression of CHRNA1 Pre-MRNA." *Scientific Reports* 3: 1–11.

Ramanouskaya, Tatsiana V., and Vasily V. Grinev. 2017. "The Determinants of

Alternative RNA Splicing in Human Cells." *Molecular Genetics and Genomics* 292(6): 1175–95.

Reichow, Steve L., Tomoko Hamma, Adrian R. Ferré-D'Amaré, and Gabriele Varani. 2007. "The Structure and Function of Small Nucleolar Ribonucleoproteins." *Nucleic Acids Research* 35(5): 1452–64.

Repetitive-Element-Mapping. "Repetitive-Element-Mapping."

Ruggero, Davide. 1881. "American Association for the Advancement of Science." *Science* os-2(57): 341–42.

Ryan, Kevin, Olga Calvo, and James L. Manley. 2004. "Evidence That Polyadenylation Factor CPSF-73 Is the MRNA 3′ Processing Endonuclease." *Rna* 10(4): 565–73.

Sakabe, Noboru Jo, and Sandro José de Souza. 2007. "Sequence Features Responsible for Intron Retention in Human." *BMC Genomics* 8(59).

Schmidt, Enrico K., Giovanna Clavarino, Maurizio Ceppi, and Philippe Pierre. 2009. "SUnSET, a Nonradioactive Method to Monitor Protein Synthesis." *Nature Methods* 6(4): 275–77.

Sebé-Pedrós, Arnau et al. 2013. "Regulated Aggregative Multicellularity in a Close Unicellular Relative of Metazoa." *eLife*.

Serdar, Lucas D. et al. 2020. "Inhibition of Post-Termination Ribosome Recycling at Premature Termination Codons in UPF1 Atpase Mutants." *eLife* 9: 1–21.

Sharma, Shalini et al. 2008. "Polypyrimidine Tract Binding Protein Controls the Transition from Exon Definition to an Intron Defined Spliceosome." *Nature Structural and Molecular Biology* 15(2): 183–91.

Shcherbik, Natalia, and Dimitri G Pestov. 2010. "Ubiquitin and Ubiquitin-Like Proteins in the Nucleolus: Multitasking Tools for a Ribosome Factory." *Genes & Cancer* 1(7): 681–89.

Shen, Shihao et al. 2014. "RMATS: Robust and Flexible Detection of Differential Alternative Splicing from Replicate RNA-Seq Data." *Proceedings of the National Academy of Sciences of the United States of America*.

Shin, Chanseok, and James L. Manley. 2002. "The SR Protein SRp38 Represses Splicing in M Phase Cells." *Cell* 111(3): 407–17.

Silva, Ana Luísa et al. 2008. "Proximity of the Poly(A)-Binding Protein to a Premature Termination Codon Inhibits Mammalian Nonsense-Mediated

MRNA Decay." *Rna* 14(3): 563–76.

Singh, Natalia N. et al. 2013. "An Intronic Structure Enabled by a Long-Distance Interaction Serves as a Novel Target for Splicing Correction in Spinal Muscular Atrophy." *Nucleic Acids Research* 41(17): 8144–65.

Singh, R., and M. Reddy. 1989. "γ-Monomethyl Phosphate: A Cap Structure in Spliceosomal U6 Smal Nuclear RNA." *Proceedings of the National Academy of Sciences of the United States of America* 86(21): 8280–83.

Sloan, Katherine E. et al. 2017. "Tuning the Ribosome: The Influence of RRNA Modification on Eukaryotic Ribosome Biogenesis and Function." *RNA Biology* 14(9): 1138–52.

Sloan, Katherine E., Markus T. Bohnsack, Claudia Schneider, and Nicholas J. Watkins. 2014. "The Roles of SSU Processome Components and Surveillance Factors in the Initial Processing of Human Ribosomal RNA." *Rna* 20(4): 540–50.

Son, Ahyeon, Jong Eun Park, and V. Narry Kim. 2018. "PARN and TOE1 Constitute a 3′ End Maturation Module for Nuclear Non-Coding RNAs." *Cell Reports* 23(3): 888–98. https://doi.org/10.1016/j.celrep.2018.03.089.

Staněk, David. 2017. "Cajal Bodies and SnRNPs - Friends with Benefits." *RNA Biology* 14(6): 671–79.

Sterne-Weiler, Timothy et al. 2018. "Efficient and Accurate Quantitative Profiling of Alternative Splicing Patterns of Any Complexity on a Laptop." *Molecular Cell*.

Sterner, Deborah A., Troy Carlo, and Susan M. Berget. 1996. "Architectural Limits on Split Genes." *Proceedings of the National Academy of Sciences of the United States of America* 93(26): 15081–85.

Sugioka-Sugiyama, Rie, and Tomoyasu Sugiyama. 2011a. "Sde2: A Novel Nuclear Protein Essential for Telomeric Silencing and Genomic Stability in Schizosaccharomyces Pombe." *Biochemical and Biophysical Research Communications* 406(3): 444–48.

———. 2011b. "Sde2: A Novel Nuclear Protein Essential for Telomeric Silencing and Genomic Stability in Schizosaccharomyces Pombe." *Biochemical and Biophysical Research Communications* 406(3): 444–48. http://linkinghub.elsevier.com/retrieve/pii/S0006291X11002609 (July 22, 2017).

Sulima, Sergey, Kim Kampen, and Kim De Keersmaecker. 2019. "Cancer Biogenesis in Ribosomopathies." *Cells* 8(3): 229.

Suzuki, Tatsuya, Hiroto Izumi, and Mutsuhito Ohno. 2010. "Cajal Body Surveillance of U SnRNA Export Complex Assembly." *Journal of Cell Biology* 190(4): 603–12.

Tanackovic, Goranka, and Angela Krämer. 2005. "Human Splicing Factor SF3a, but Not SF1, Is Essential for Pre-MRNA Splicing in Vivo." *Molecular Biology of the Cell* 16(3): 1366–77.

Terns, Michael P, and Rebecca M Terns. 2002. "Small Nucleolar RNAs: Versatile Trans-Acting Molecules of Ancient Evolutionary Origin." *Gene expression* 10(1–2): 17–39.

Thakran, Poonam et al. 2018. " Sde2 Is an Intron-specific Pre- MRNA Splicing Regulator Activated by Ubiquitin-like Processing ." *The EMBO Journal* 37(1): 89–101.

Tyc, K, and J A Steitz. 1989. "U3, U8 and U13 Comprise a New Class of Mammalian SnRNPs Localized in the Cell Nucleolus." *The EMBO journal* 8(10): 3113–19.

Ule, Jernej et al. 2006. "An RNA Map Predicting Nova-Dependent Splicing Regulation." *Nature* 444(7119): 580–86.

Urdaneta, Erika C. et al. 2019. "Purification of Cross-Linked RNA-Protein Complexes by Phenol-Toluol Extraction." *Nature Communications* 10(1): 990.

Vu, Ngoc T. et al. 2013. "HnRNP U Enhances Caspase-9 Splicing and Is Modulated by AKT-Dependent Phosphorylation of HnRNP L." *Journal of Biological Chemistry* 288(12): 8575–84. http://dx.doi.org/10.1074/jbc.M112.443333.

Wan, Ruixue, Rui Bai, and Yigong Shi. 2019. "Molecular Choreography of Pre-MRNA Splicing by the Spliceosome." *Current Opinion in Structural Biology* 59: 124–33. https://doi.org/10.1016/j.sbi.2019.07.010.

Wang, Eric T. et al. 2008. "Alternative Isoform Regulation in Human Tissue Transcriptomes." *Nature*.

Wang, Minshi, and Dimitri G. Pestov. 2011. "5′-End Surveillance by Xrn2 Acts as a Shared Mechanism for Mammalian Pre-RRNA Maturation and Decay." *Nucleic Acids Research* 39(5): 1811–22.

Wang, Zhe et al. 2019. "Acetylation of PHF5A Modulates Stress Responses and Colorectal Carcinogenesis through Alternative Splicing-Mediated Upregulation of KDM3A." *Molecular Cell* 74(6): 1250-1263.e6. https://doi.org/10.1016/j.molcel.2019.04.009.

Will, Cindy L., and Reinhard Lührmann. 2011. "Spliceosome Structure and Function." *Cold Spring Harbor Perspectives in Biology* 3(7): 1–2.

Will, Cindy L, and Reinhard Lu. 2011. "Spliceosome Structure and Function." *Cold Spring Harb Perspect Biol.* 3(7): 1–23.

Wong, Justin J.L. et al. 2013. "XOrchestrated Intron Retention Regulates Normal Granulocyte Differentiation." *Cell* 154(3): 583–95. http://dx.doi.org/10.1016/j.cell.2013.06.052.

Wong, Justin J.L., Amy Y.M. Au, William Ritchie, and John E.J. Rasko. 2016. "Intron Retention in MRNA: No Longer Nonsense: Known and Putative Roles of Intron Retention in Normal and Disease Biology." *BioEssays*.

Wu, Yixuan et al. 2017. "Molecular Basis for the Interaction between Integrator Subunits IntS9 and IntS11 and Its Functional Importance." *Proceedings of the National Academy of Sciences of the United States of America* 114(17): 4394–99.

Xiong, Hui Y. et al. 2015. "The Human Splicing Code Reveals New Insights into the Genetic Determinants of Disease." *Science*.

Xun, Yu et al. 2019. "Purification and Identification of MiRNA Target Sites in Genome Using DNA Affinity Precipitation." *Frontiers in Genetics* 10(September): 1–12.

Yang, Yunfeng et al. 2000. "Conserved Composition of Mammalian Box H/ACA and Box C/D Small Nucleolar Ribonucleoprotein Particles and Their Interaction with the Common Factor Nopp140." *Molecular Biology of the Cell* 11(2): 567–77.

Yeasmin, Fouzia, Tetsushi Yada, and Nobuyoshi Akimitsu. 2018. "Micropeptides Encoded in Transcripts Previously Identified as Long Noncoding RNAs: A New Chapter in Transcriptomics and Proteomics." *Frontiers in Genetics* 9(APR): 1–10.

Yeo, Gene, and Christopher B. Burge. 2004. "Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals." In *Journal of Computational Biology*,.

Yeo, Gene, Dirk Holste, Gabriel Kreiman, and Christopher B. Burge. 2004. "Variation in Alternative Splicing across Human Tissues." *Genome biology* 5(R74).

Yeo, Gene W et al. 2009. "An RNA Code for the FOX2 Splicing Regulator Revealed by Mapping RNA-Protein Interactions in Stem Cells." *Nature structural & molecular biology* 16(2): 130–37.

Zhang, Rundong et al. 2011. "Structure of a Key Intermediate of the SMN Complex Reveals Gemin2's Crucial Function in SnRNP Assembly." *Cell* 146(3): 384–95.

Zhang, Yuanjiao, Jinjun Qian, Chunyan Gu, and Ye Yang. 2021. "Alternative Splicing and Cancer: A Systematic Review." *Signal Transduction and Targeted Therapy* 6(1).

Zheng, Jiangge et al. 2015. "Structure of Human MDM2 Complexed with RPL11 Reveals the Molecular Basis of P53 Activation." *Genes and Development* 29(14): 1524–34.

Zhiguo, E., Lei Wang, and Jianhua Zhou. 2013. "Splicing and Alternative Splicing in Rice and Humans." *BMB Reports* 46(9): 439–47.

Zhou, Zhihong, and Xiang-Dong Fu. 2013. "Regulation of Splicing by SR Proteins." *Chromosoma* 122(3): 191–207.

Zhu, Chengming et al. 2018. "Erroneous Ribosomal RNAs Promote the Generation of Antisense Ribosomal SiRNA." *Proceedings of the National Academy of Sciences of the United States of America* 115(40): 10082–87.

CURRICULUM VITAE