THE UNIVERSITY OF CHICAGO

TERTIARY STRUCTURE OF INTRONIC PRI-MIR-17-92A REGULATES SPLICING

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

SHABANA MEHTAB SHAIK

CHICAGO, ILLINOIS

JUNE 2017

DEDICATION

To AMMI & BABA

LIST OF CONTENTS

Chapter 3: Tertiary structure of intronic Pri-miR-17-92a regulates splicing

List of figures

List of Tables

Abbreviations

BP   branch point

CTD   carboxy terminal domain

DGCR8   DiGeorge Syndrome Critical Region Gene 8

EJC   exon junction complex

ESE   exonic splicing enhancers

ESS   exonic splicing silencers

hnRNP   heterogenous ribonucleoprotein

ISE   intronic splicing enhancers

ISS   intronic splicing silencers

KSRP   KH-type splicing regulatory protein

LncRNA   long non-coding RNA

m^{6}A   N6-Methyladenosine

miR   microRNA

MPC   Microprocessor complex

nt   nucleotide

PARS   Parallel analysis of RNA structure

piRNA   piwi-interacting RNA

PPT   polypyrimidine tract

pri-miR   primary microRNA

RBP   RNA binding protein

RISC   RNA induced silencing complex

RNA   Ribonucleic Acid

RRM	RNA recognition motif
RT	reverse transcription
SHAPE	Selective 2′ hydroxyl acylation analyzed by primer extension
siRNA	small interfering RNA
snRNP	small nuclear ribonucleoprotein
snoRNAs	small nucleolar RNA
SRE	splicing regulatory element
SS	splice site
ss-ds	single strand-double strand
SR	serine-arginine
TRBP	TAR RNA binding protein
XIST	X-inactive specific transcript
UTR	untranslated region

Acknowledgements

This thesis has taken a long to come out but the work described herein represents a true "trial by fire" and I am delighted to acknowledge the people that have made this journey an exceptional experience. I begin by remembering Late. Prof. Obaid Siddiqui, the vision behind National Center of Biological Sciences-TIFR, Bangalore, India where I started my scientific career for being an inspiration to infinite people including me.

I would like to thank my thesis supervisor Prof. Yamuna Krishnan for giving me the intellectual freedom over the years and for making me think 'outside the box'. I feel fortunate to have been afforded the opportunity to help start her lab at NCBS-TIFR from scratch, witness the evolution of our initial projects and to set an example for new students joining the lab. More importantly, I thank her for encouraging me to move to the University of Chicago that has given me the chance to work for what I believed. It was her passion for research that ignited my own dedication to understanding the influence of RNA structure on its function. Yamuna has been a wonderful mentor and has set a very high bar on how to write and communicate my science. It's because of her support that I have been able to overcome the professional as well as personal challenges. I thank my thesis committee members Prof. Chuan He and Prof. Joe Piccirilli for giving me their time and for evaluating my thesis.

I thank Prof. Joe Piccirilli with whom I collaborated for my *in vitro* work. His prudent guidance, support and input on my work kept me motivated on this daring task. Despite his busy schedule, he always had time to help in troubleshooting an experiment or discussing my research. As an individual I always worked on fixing my weakness but I learnt from him that the key is to amplify one's strength. I am indebted to Joe for the trust he had in me and for the many conversations we had that helped shape this work.

I would like to thank all the past Krishnan group members Drs' Souvik, Saikat, Dhiraj, Suruchi, Sonali, Saheli, Sunaina, Justin and Masood for making it a memorable experience working with them. I would also like to thank Anusuya, Aparna, Raghu, Rajesh, Ramveer, Ramya, Pritha, and Sruthi. It was fun discussing the academic and non-academic matters that has helped me grow as a scientist and as an individual. I would specially thank Saikat my collaborator and friend for his immense support ever since we started working with RNA. I thank Suruchi for all the good times we had organizing and maintaining the radio-active facility for NCBS. I thank all

Synopsis

Chapter 1: Introduction

RNA is a versatile informational molecule. Besides carrying information in the form of its linear primary sequence, RNA can adopt secondary structures by canonical base pairing of complementary nucleotides. These further fold into a compact tertiary structure by invoking non-canonical interactions between sequences that are far apart in one dimension (Hendrix et al 2005; Butcher et al 2011). Secondary and tertiary structured RNA offer a higher level of structural information that influences various steps of gene regulation (Klaff et al 1996; Wan et al. 2011). They enable an RNA to interact with itself, other RNAs, with DNA, with ligands and with proteins Deeper insight into the molecular understanding of RNA function has made it clear that RNA secondary and tertiary structures influence the function of almost all classes of RNAs that include both coding as well as non-coding RNAs (Mortimer et al 2014).

The presence of spliceosomal introns is one of the most defining features of eukaryotic genomes. Intronic sequences in the genome was initially thought to impose substantial burden on the host and regarded as 'junk'. Pre-mRNAs that are transcribed with these intronic RNAs are removed and the exonic RNAs that encode for proteins are joined to form mature mRNAs catalyzed by a dynamic ribonucleoprotein complex called Spliceosome. Several studies revealed many families of non-coding RNAs like small nucleolar RNAs (snoRNAs), microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), small interfering RNAs (siRNAs), and various long non-coding RNAs (lncRNAs) that are preferentially associated with introns showing that genes autoregulate their expression by hosting relevant ncRNAs (Rearick et al. 2011). At the heart of the multifaceted regulatory potential of RNA lies its property to fold into intricate shapes (Tinoco and Bustamante 1999; Schroeder et al. 2002; Wan et al. 2014; Onoa and Tinoco 2004). RNA secondary

and tertiary structures are shown to influence chromatin remodeling, transcription, splicing, cellular localization, translation and turnover of the RNA (Mortimer et al. 2014; Klaff et al. 1996; McManus and Graveley 2011).

In chapter 1, I briefly outline 'RNA structure based gene regulation', emphasizing on two examples (i) Riboswitches that are ribo-regulators of prokaryotic gene regulation (ii) long non-coding RNAs (lncRNAs) that are now emerging as regulators of eukaryotic gene expression. In my thesis, I have shown the influence of tertiary structure of an intronic RNA element on two different steps of gene regulation. In the first study, I demonstrate that a polycistronic primary microRNA transcript pri-miR-17-92a called as 'OncomiR-1' folds into a higher order tertiary structure that acts as a kinetic barrier to autoregulate its pre-miRNA processing (Chapter 2). My second study addresses the potential of this tertiary structured intronic pri-miR-17-92a to regulate splicing of its host transcript (Chapter 3). I have described the background for (i) microRNA mediated gene regulation and the mechanisms that regulate microRNA biogenesis and (ii) mechanism of pre-mRNA splicing and its regulation in this chapter (Chapter 1).

Chapter 2: Pri-miR-17-92a folds into tertiary structure and autoregulates its processing

MicroRNAs (miRNAs) are a large family of small noncoding RNAs that have emerged as key post-transcriptional regulators of gene expression. In mammals, miRNAs are implicated in regulating the activity of ~50% of all protein-coding genes either by degradation or repression of translation of specific target mRNAs (Ambros 2004; Chen and Rajewsky 2007). They have been implicated in numerous biological processes including cellular differentiation, cell proliferation, apoptosis, metabolism, immunity and synaptic plasticity (Bushati and Cohen 2007). Involvement of miRNAs across wide range of important biological phenomena requires stringent control over their expression.

MiRNAs are formed from primary microRNA (pri-miRNA) transcript that are first cropped to release hairpin-shaped precursors (pre-miRNAs) in the nucleus by the microprocessor complex (MPC), comprising Drosha and partner DGCR8 (Han et al 2004; Kim, 2005). The pre-miRNAs are exported to the cytoplasm where they are processed by Dicer to form mature microRNA (miRNA) that target their cognate mRNA via base-paring leading to mRNA cleavage or translational repression Thus, miRNAs control protein levels by controlling concentrations of translationally competent mRNA directly without affecting its transcription (Bartel et al 2005). This indicates that there is a need of stringent control over miRNA expression and function. Several studies have uncovered a wide range of post transcriptional mechanisms which regulate miRNA biogenesis and activity (Krol et al. 2010). Investigations on pri-miRNA have largely focused on transcripts incorporating a single pre-miRNA stem loop which is processed sequentially into pre-miRNA and mature miRNA (Davis et al. 2008). However, in vertebrates, it is known that 30% of miRNAs are found as part of a polycistronic cluster i.e. they are present in quick succession on the same pri-miR transcript (Megraw et al. 2007). Though they are co-transcribed as a single transcript, the component miRNAs of the cluster are present in different abundances in a given tissue (Tang et al. 2008; Thomson et al. 2006).

The mechanisms by which such differential expression of miRNAs is achieved was unclear when I started my thesis. Thus, we considered a possibility where the primary miRNA cluster might self-orchestrate the binding of accessory proteins associated with processing different microRNA domains resulting in their differential regulation. We chose the intronic miRNA cluster called pri-miR-17-92a which encompasses ~800 nucleotides of a non-coding RNA called *C13orf25* (Ota et al 2004) present on human chromosome 13 to address such differential

regulation. Upon processing this cluster produces six individual mature miRNAs (miR-17, miR-18a, miR-19a, miR-20a, miR-19b, and miR-92a) (Mendell et al. 2008) **(Figure 1)**.
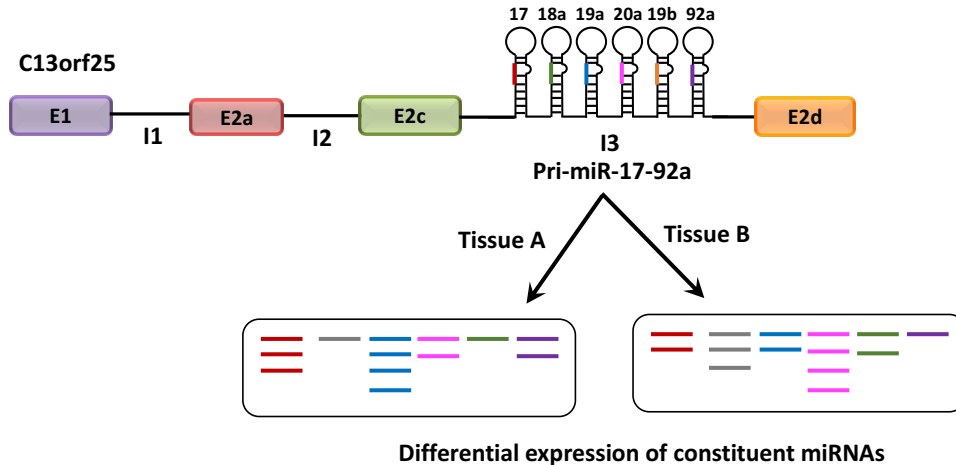
Figure 1: Schematic representation of genomic locus of pri-miR-17-92a cluster. The cluster resides in intron 3 (I3) of the gene *C13orf25* and contains six stem loop structures which gives six individual microRNAs (17, 18a, 19a, 20a, 19b, 92a). It undergoes differential processing to yield component miRNAs in different amounts in different tissues (A & B).

Formation of a higher order structure by the pri-miR-17-92a cluster could create a suboptimal display of recognition sites for microprocessor complex. Structural studies indicate Drosha-DGCR8 binding requires a single strand double strand junction (ss-ds) (Han et al., 2006). Thus, sequestration of such a recognition site could arise from tertiary structure formation from the helices of this cluster thereby masking key ss-ds junctions. This hypothesis was supported by a study that revealed that recruitment of hnRNP A1 remodeled the local structure of this pri-miRNA. Chakraborty *et al* have shown that the pri-miR-17-92a folds into a tertiary structure (Chakraborty et al. 2012) and I wanted to investigate the influence of this tertiary structure of the pri-miR-17-92a on its miRNA processing. To do this, a shuffled or swapped pri-miRNA (pri-miR-20a-19b-92a-17-18a-19a) also refered as pri-miR-20a-19a was made by disrupting conserved

structured elements, such as bases 388-480 i.e., the inter-pre miR region between pre-miR-19a and pre-miR-20a. The disruption of such structured inter pre-miRNA elements, such as the region between pre-miR-19a and pre-miR-20a, in the shuffled transcript pri-miR-20a-19a resulted in its impaired tertiary structure.

Overexpression of native pri-miR and shuffled pri-miR showed that although native pri-miR-17-92a was present at reasonable levels, shuffled pri-miR-20a-19a level was below detection limit, indicating that *in cellulis* equilibrium levels of native and shuffled transcript were quite different. *In cellulis* processing of the native and shuffled transcripts to its constituent pre-miRs revealed that both the transcripts had altered processing efficiencies with shuffled pri-miR showing elevated levels of pre-miRs compared to native. This was further confirmed by *in vitro* processing studies wherein the tertiary structured native pri-miR-17-92a undergoes slow microprocessing compared to the shuffled transcript. Thus, a mere shuffling of discrete, pre-miR containing, hairpin domains is sufficient to alter the relative abundance of the processed pre-miRNAs. Using *in cellulis* and *in vitro* processing studies we have shown that structural differences between the native transcript and the shuffled transcript impacts their processing, which is beyond a simplistic secondary structure mode.

This suggests a model where tertiary structure formation by a primary miRNA transcript imposes a kinetic barrier to its processing, where the conformation adopted is transparent to the MPC leading to an inhibition imposed at its earliest stage of its processing into pre-miRs **(Figure 2)**.
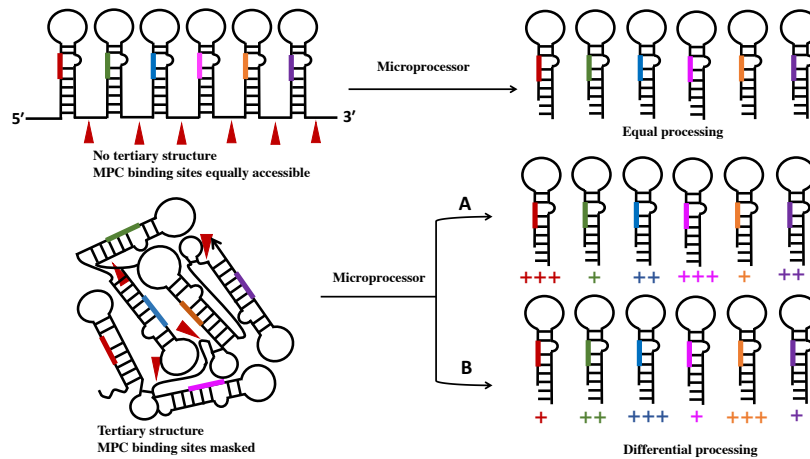
Figure 2: Schematic showing the differential processing of pri-miR-17-92a cluster. In the absence of a tertiary structure the pri-miR would undergo equal processing to give equimolar amounts of individual pre-miRs (colors represent the six individual miRs) due to equal accessibility of the microprocessor (shown by arrowheads). Folding of the pri-miRNA into a distinct tertiary structure would allow differential processing in different tissues (A, B) by masking the recognition sites of the microprocessor to give different levels of its component pre-miRs.

Chapter 3: Tertiary structure of intronic Pri-mir-17-92a regulates splicing of its host transcript.

The presence of introns is a defining feature of eukaryotic genes that undergo splicing within the nucleus to remove introns and join exons in a process catalyzed by a multi-mega Dalton ribonucleoprotein complex called the 'Spliceosome'. Pre-mRNA splicing is a process that demands remarkable accuracy, and it is intriguing to contemplate the mechanisms that enable this accuracy that can be either constitutive or alternative (Green 1991). During constitutive exon splicing, a dynamic process involving the spliceosome joins the 5′ and 3′ splice sites (SS) that define the exon-intron boundary through the interplay of small nuclear ribonucleoproteins (snRNPs) and other associated proteins with the pre-mRNA (De Conti et al. 2013). The accuracy of splicing is defined by consensus splice sites and regulatory elements such as exon and intron splicing enhancers (ESE, ISE) and silencers (ESS, ISS) (Liu et al. 2010; Wang and Burge 2008). These regulatory elements are recognized by members of the SR and hnRNP family of proteins

Alternative splicing is an important mechanism for regulating gene expression. It expands the coding capacity of a single gene to produce different proteins with distinct functions (Stamm et al. 2005; Matlin et al. 2005; Chen and Manley 2009; Kelemen et al. 2013).

There is a wealth of information on regulatory signatures within introns that dictate splicing in terms of *cis*-elements such as primary sequence, and *trans* elements such as proteins. Yet it is not known whether the next level of information coded by RNA, i.e., its tertiary structure, can regulate splicing. Hence, I wanted to investigate if the tertiary structure of intronic polycistronic microRNA cluster, pri-miR-17-92a can influence the splicing of its host transcript. The tertiary structured pri-miR 17-92a is located in the third intron of its endogenous non-coding transcript known as *C13orf25*, which is alternatively spliced.

I reasoned that microRNAs would be particularly relevant as potential regulators of splicing in that 80% of the human microRNA loci have been reported to be located within the intronic regions of coding or non-coding transcription units and are transcribed along with their host genes (Rodriguez et al. 2004). Studies on processing of intronic miRNAs have shown that splicing is not a prerequisite for miRNA processing and that unspliced introns can also serve as the substrate for the microprocessor complex (Kim and Kim 2007). RNA splice site choice is likely to be regulated kinetically. Recognition of splice sites could be masked by proteins or presented in a different order, all of which can guide the splicing outcome. This is revealed by the influence of kinetics of cleavage on splicing. Insertion of fast or slow self-cleaving ribozymes within the intron of β-globin transcript show that fast cleavage leads to impaired splicing due to inhibition of co-transcriptional assembly of the spliceosome, while slow cleavage allows for efficient assembly, leading to proper exon tethering that results in effective pre-mRNA splicing (Fong et al. 2009). Micro processing of miR-211 located in intron6 of melastatin gene promoted the splicing of exon6-

exon7 by a mechanism that required cleavage by Drosha. Based on their studies a feed-forward regulation between miRNA processing and splicing was proposed, whereby 5′ SS recognition by the U1 complex promotes miRNA processing of intronic miR-211 by Drosha and this cropping of miR-211 promotes splicing at its host melastatin intron 6 (Janas et al. 2011).

These studies clearly revealed the crosstalk between the microprocessor and spliceosome that resulted in the production of two different functional RNAs: (i) regulatory miRNAs and (ii) spliced mRNA from a single pre-mRNA transcript. However, the contribution of tertiary structure to influence the splicing needed investigation.

First, using a constitutively spliced b-globin reporter gene with a TS-pri-miR engineered into its intron, I was able to study its effect on splicing. RT-PCR analysis of *in cellulis* spliced products clearly revealed that the introduction of tertiary structured pri-miR impeded the removal of introns from its host transcript. This impedance was more pronounced on the upstream intron compared to the intron harboring it, as indicated by accumulation of unspliced transcript corresponding to intron1 and decreased levels of spliced product. A swapped pri-miR of the same length, also a substrate for MPC but devoid of such a tertiary structure, has no significant effect on splicing of its host transcript.

To determine the effect of TS-pri-miR on the kinetics of intron removal, I needed a reliable *in vitro* splicing assay. Such assays have been well established for mini genes and genes containing introns of less than 1 kb (Rooke and Underwood 2001; Mayeda and Krainer 2012), but the larger introns involved here required a modified assay. To this end, I optimized an *in vitro* splicing assay for these large RNAs. The kinetic analysis of these *in vitro* splicing reactions clearly suggested that the transcript containing the TS-pri-miR has a slower rate of splicing. This effect was more pronounced at the upstream intron which demonstrated ~3.2 times slower splicing compared to

the WT. *In cellulis* splicing studies on *C13orf25* host transcript indicated that the TS-pri-miR can regulate the proportion of different spliced forms that is a reflection of altered splice site choices of the host transcript.

Chakraborty *et al* showed that folding of pri-miR-17-92a into a tertiary structure makes Drosha binding sites and thus influences its processing into pre-miRNAs. I now propose that by folding into 3D-structure it could also mask or reveal accessibility to key splice sites and regulate splicing. Its removal by processing would also then be expected to lift this splicing inhibition **(Figure 3)**. Thus far, the role of the RNA structure in splicing has been limited to the formation of local secondary structures, which could influence splice site choice by masking or unmasking key regulatory sites required for binding to *trans*-regulatory proteins. The present study shows that the slow kinetics of intron removal associated with the TS-pri-miR may play a role in facilitating proper exon tethering and determining splicing choices of the host transcript. Understanding how molecular signatures on the intronic RNA can modulate its structural plasticity resulting in the host RNA auto-regulating its processing is an area of RNA-based gene regulation that is now ripe for investigation.
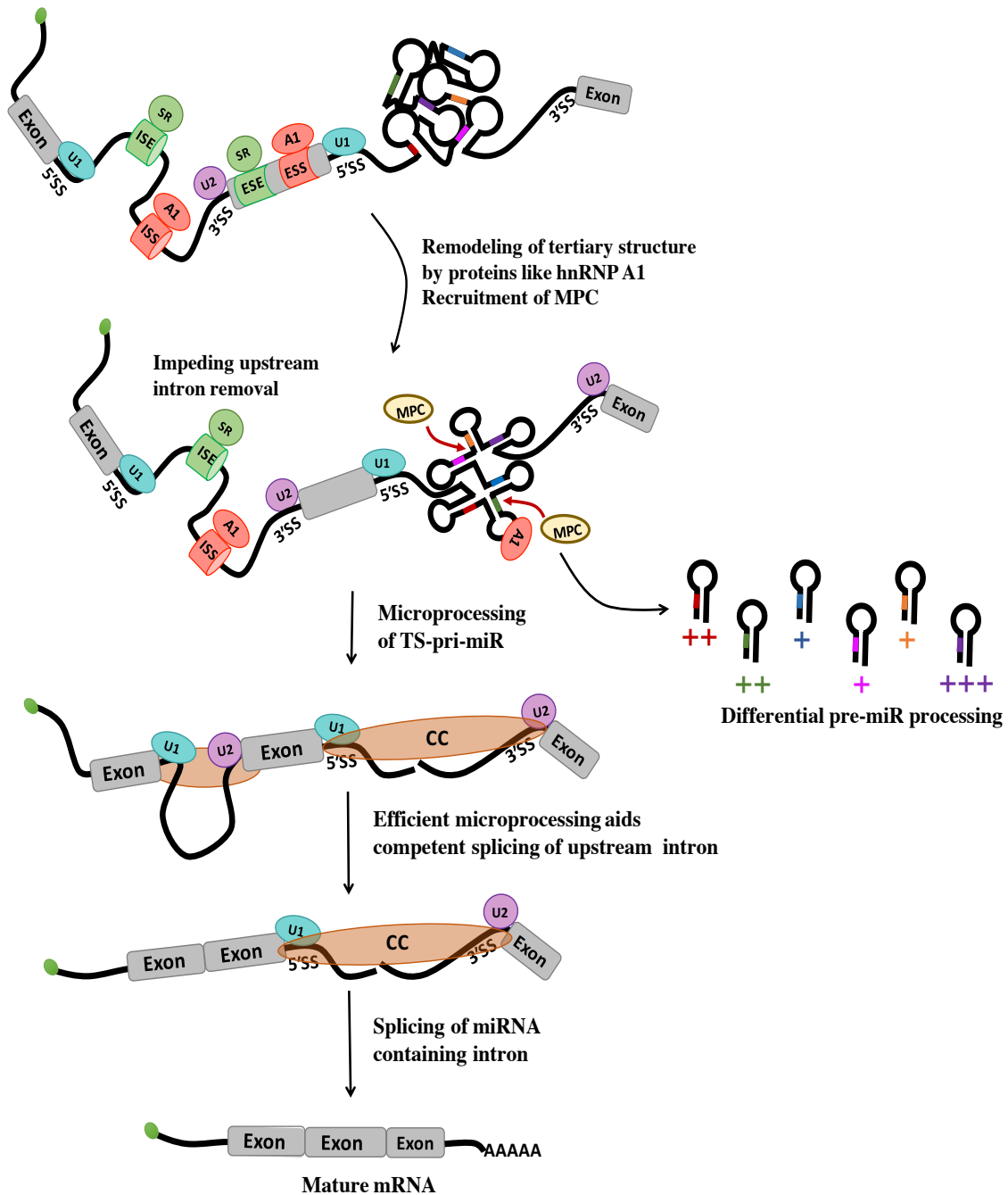
Figure 3: A model for splicing of tertiary structured intronic pri-miR transcript. Presence of tertiary structured pri-miR impedes splicing of its upstream intron possibly by masking the recognition sites for binding of *trans* regulatory proteins. Remodeling of the tertiary structure by proteins like hnRNP A1 result in efficient microprocessing of the intron harboring the pri-miR cluster. This then aids the splicing of upstream intron and because the exons flanking the intronic pri-miR have already been paired and tethered by the commitment complex (CC), splicing of this intron would still occur efficiently despite the discontinuity of the intron.

References:

Ambros V. 2004.The functions of animal microRNAs. *Nature* 431: 350-355.

Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136(2), pp. 215–233.

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* 72, pp. 291–336.

Bushati, N. and Cohen, S.M. 2007. microRNA functions. *Annual Review of Cell and Developmental Biology* 23, pp. 175–205.

Butcher, S.E. and Pyle, A.M. 2011. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of Chemical Research* 44(12), pp. 1302–1311.

Chakraborty, S., Mehtab, S., Patwardhan, A. and Krishnan, Y. 2012. Pri-miR-17-92a transcript folds into a tertiary structure and autoregulates its processing. *RNA (New York)* 18(5), pp. 1014–1028.

Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E.M., Veldhoen, R.A., Khanna, S.K., Goping, I.S., MacMillan, A.M., Mendell, J.T., Young, H.S., Fahlman, R.P. and Glover, J.N.M. 2011. Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biology* 8(6), pp. 1105–1114.

Chen, K. and Rajewsky, N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews*. *Genetics* 8(2), pp. 93–103.

Chen, M. and Manley, J.L. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews*. *Molecular Cell Biology* 10(11), pp. 741–754.

Davis, B.N. and Hata, A. 2009. Regulation of MicroRNA Biogenesis: A miRiad of mechanisms. *Cell Communication and Signaling* 7, p. 18.

De Conti, L., Baralle, M. and Buratti, E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley interdisciplinary reviews*. *RNA* 4(1), pp. 49–60.

Fong, N., Ohman, M. and Bentley, D.L. 2009. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. *Nature Structural & Molecular Biology* 16(9), pp. 916–922.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432(7014), pp. 235–240.

Green, M.R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annual review of cell biology* 7, pp. 559–599.

Guil, S. and Cáceres, J.F. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Structural & Molecular Biology* 14(7), pp. 591–596.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H. and Kim, V.N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development* 18(24), pp. 3016–3027.

Hendrix, D.K., Brenner, S.E. and Holbrook, S.R. 2005. RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics* 38(3), pp. 221–243.

Janas, M.M., Khaled, M., Schubert, S., Bernstein, J.G., Golan, D., Veguilla, R.A., Fisher, D.E., Shomron, N., Levy, C. and Novina, C.D. 2011. Feed-forward microprocessing and splicing activities at a microRNA-containing intron. *PLoS Genetics* 7(10), p. e1002330

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M. and Stamm, S. 2013. Function of alternative splicing. *Gene* 514(1), pp. 1–30.

Kim, E., Magen, A. and Ast, G. 2007. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research* 35(1), pp. 125–131.

Kim, Y.-K. and Kim, V.N. 2007. Processing of intronic microRNAs. *The EMBO Journal* 26(3), pp. 775–783.

Klaff, P., Riesner, D. and Steger, G. 1996. RNA structure and the regulation of gene expression. *Plant Molecular Biology* 32(1-2), pp. 89–106.

Krol, J., Loedige, I. and Filipowicz, W. 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews. Genetics* 11(9), pp. 597–610.

Liu, W., Zhou, Y., Hu, Z., Sun, T., Denise, A., Fu, X.-D. and Zhang, Y. 2010. Regulation of splicing enhancer activities by RNA secondary structures. *FEBS Letters* 584(21), pp. 4401–4407

Matlin, A.J., Clark, F. and Smith, C.W.J. 2005. Understanding alternative splicing: towards a cellular code. *Nature Reviews. Molecular Cell Biology* 6(5), pp. 386–398.

McManus, C.J. and Graveley, B.R. 2011. RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development* 21(4), pp. 373–379.

Mendell, J.T. 2008. miRiad roles for the miR-17-92 cluster in development and disease. *Cell* 133(2), pp. 217–222.

Megraw M., Sethupathy P., Corda B., Hatzigeorgiou A.G. 2007. miRGen: a database for the study of animal microRNA genomic organization and function. *Nucleic Acids Res.* 35: 149-155.

Mortimer, S.A., Kidwell, M.A. and Doudna, J.A. 2014. Insights into RNA structure and function from genome-wide studies. *Nature Reviews. Genetics* 15(7), pp. 469–479.

Onoa, B. and Tinoco, I. 2004. RNA folding and unfolding. *Current Opinion in Structural Biology* 14(3), pp. 374–379.

Ota, A., Tagawa, H., Karnan, S., Tsuzuki, S., Karpas, A., Kira, S., Yoshida, Y. and Seto, M. 2004. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Research* 64(9), pp. 3087–3095.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. and Bradley, A. 2004. Identification of mammalian microRNA host genes and transcription units. *Genome Research* 14(10A), pp. 1902–1910.

Schroeder, R., Grossberger, R., Pichler, A. and Waldsich, C. 2002. RNA folding in vivo. *Current*

Opinion in Structural Biology 12(3), pp. 296–300.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A. and Soreq, H. 2005. Function of alternative splicing. *Gene* 344, pp. 1–20.

Tang G.Q., Maxwell E.S. 2008. Xenopus microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post- transcriptional regulation. Genome Res. 18: 104-112.

Thomson J.M., Newman M., Parker J.S., Morin-Kensicki E.M., Wright T., Hammond S.M. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev*. 20: 2202-2207.

Wang, Z. and Burge, C.B. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA (New York)* 14(5), pp. 802–813.

Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E. and Chang, H.Y. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505(7485), pp. 706–709.

List of publications

1. Icosahedral DNA nanocapsules via modular assembly.

   Bhatia, D[#]., **Mehtab, S**[#]., Krishnan, R., Indi, S.S., Basu, A., Krishnan, Y**.*** *Angew. Chem. Int. Ed.*, 2009, *48*, 4134 – 4137. [#] co-first authors.

2. Pri-miR-17-92a transcript folds into a tertiary structure and autoregulates its processing.

   Chakraborty, S., **Mehtab, S**., Patwardhan, A.R., Krishnan, Y.* *RNA*, **2012**, *18*, 1014-1028.

3. A method to encapsulate molecular cargo within DNA icosahedra.

   Bhatia, D., Chakraborty, S., **Mehtab, S**., Krishnan, Y.* *Methods Mol. Biol.* 2013, *991*, 65-80.

4. The Predictive Power of Synthetic Nucleic Acid Technologies in RNA Biology.

   Saikat Chakraborty, **Shabana Mehtab** & Yamuna Krishnan*. *Acc. Chem. Res.* 2014, 47, 1710–1719.

5. Tertiary structure of intronic Pri-miR-17-92a regulates splicing.

   Shabana M Shaik, Joseph A Piccirilli* & Yamuna Krishnan* (Manuscript in preparation).

CHAPTER-1

INTRODUCTION

1. A: RNA Structure

The discovery of catalytic RNAs in the early 1980's opened new vistas on RNA function, whose importance formerly lay confined to protein production (Kruger et al. 1982). That's an important role, but only a sliver of its diverse set of functions. RNA's ability to act as both genetic template and biochemical catalyst led to the proposal of the 'RNA World' hypothesis for the origin of life. The RNA World hypothesis proposes that RNA preceded both DNA and proteins in evolution. The regulatory function of RNA was assumed to be based on its ability to base pair with DNA or RNA. Over the past three decades it has become clear that a variety of RNA molecules have important or essential biological functions in cells, beyond the well-established roles of ribosomal, transfer and messenger RNAs in protein biosynthesis. At the heart of the multifaceted regulatory potential of RNA lies its property to fold into intricate shapes (Tinoco and Bustamante 1999; Schroeder et al. 2002; Wan et al. 2014; Onoa and Tinoco 2004). Pairing of local nucleotides in RNA involving canonical Watson-Crick base pairing can result in secondary structures such as hairpin-loops, bulges and stem-junctions (Onoa and Tinoco 2004;Wan et al. 2014) **(Figure 1.1)**. Non-canonical hydrogen bonding interactions among distantly located sequences can form higher order tertiary structures (Hendrix et al. 2005; Laing et al. 2009; Butcher and Pyle 2011) **(Figure 1.2)**.
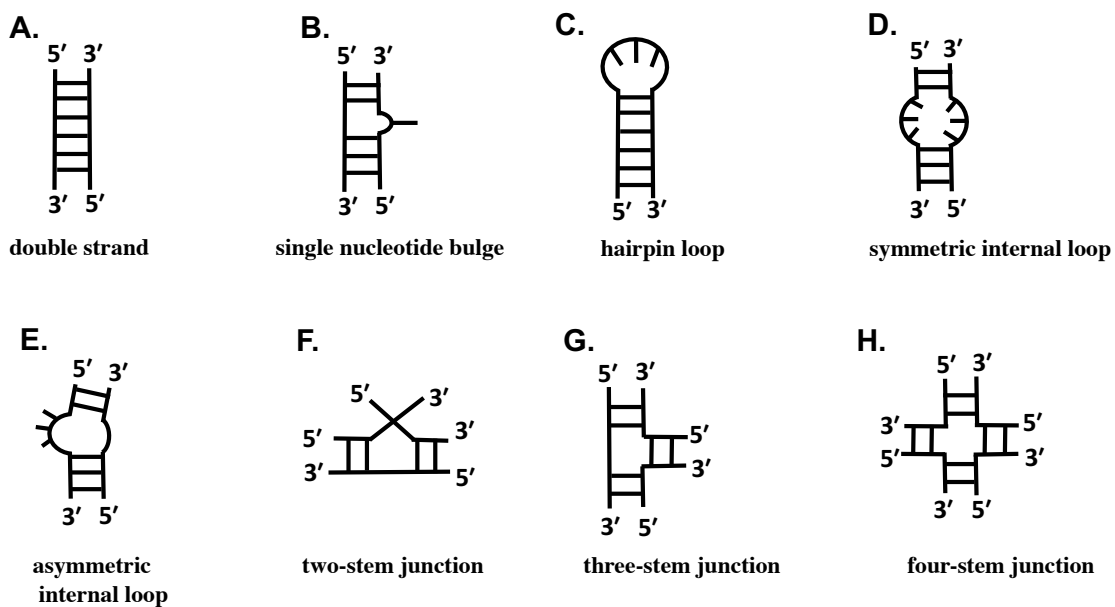
Figure 1.1: Schematic showing various secondary structures formed by RNA.
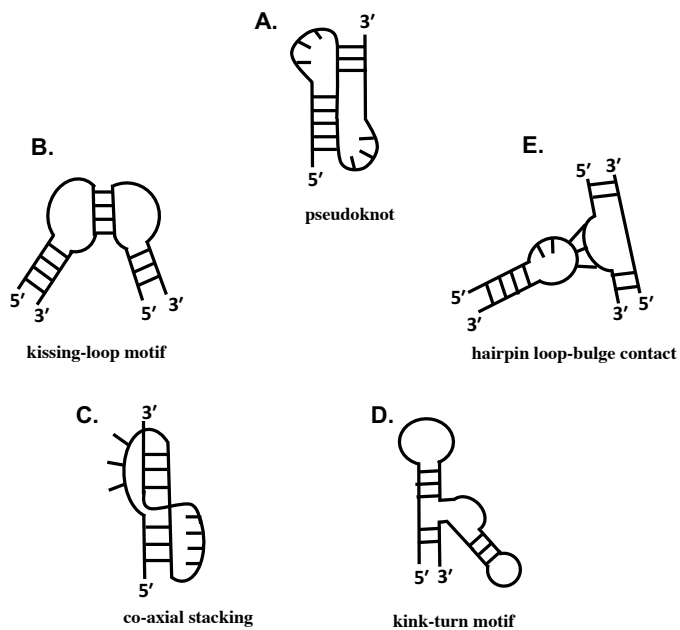
Figure 1.2: Schematic showing different tertiary interactions in RNA. **A**. pseudoknot, **B**. kissing-loop motif, **C**. co-axial stacking, **D**. kink-turn motif, **E**. hairpin loop-bulge contact.

RNA structures are shown to influence the transcription, splicing, cellular localization, translation and turnover of the RNA (Mortimer et al. 2014; Klaff et al. 1996; McManus and Graveley 2011; Jacobs et al. 2012). While the advent of high-throughput sequencing technologies has enabled the sequencing of whole genome and transcriptome annotation, coupling RNA structure probing to high-throughput sequencing reveals genome-wide RNA structural information, providing insights into the secondary structures of thousands of transcripts. Such RNA structures enable an RNA to interact with itself, other RNAs, with DNA, with ligands and with proteins (RNA binding proteins, RBPs). Deeper insight into the molecular understanding of RNA function has made it evident that RNA secondary and tertiary structures influence the function of almost all classes of RNAs that include both coding and non-coding RNAs.

1. B: RNA structure-mediated gene regulation

RNA sequences with enzymatic function, termed 'ribozymes' were first described in *Tetrahymena* where nuclear pre-mRNA of the 23S rRNA excised an intervening sequence from adjacent exons without protein assistance (Kruger et al. 1982; Berget et al. 1977). All identified, naturally occurring ribozymes may be functionally classified as either cleaving or splicing ribozymes. The general mechanism involves a nucleophilic attack of a polarized water molecule on an adjacent phosphate in the RNA backbone, resulting in well-defined cleavage products. However, unlike ribonucleases, ribozymes cleave at a unique location, determined by base-pairing and tertiary interactions mediated by divalent cations, particularly $Mg^{2+,}$ to form an 'active conformation' crucial for cleavage (Cate et al. 1996).

While the most abundant rRNAs that form the ribosome and tRNAs involved in translation have been extensively characterized to reveal the role of RNA structure, other classes of regulatory

RNA are also quite well studied. These include but not limited to riboswitches, catalytic RNAs or ribozymes, small nuclear RNAs (snRNAs) that compose the pre-mRNA splicing machinery, small nucleolar RNAs (snoRNAs) responsible for ribosomal RNA modification, guide RNAs involved in RNA editing, telomerase RNA required for chromosome end replication, signal recognition particle (SRP) RNA necessary for protein translocation, microRNAs required for gene silencing and long non-coding RNAs (lncRNAs) (Wan et al. 2011). MicroRNAs are short eukaryotic RNAs that modulate gene expression in normal development and disease pathogenesis (Tsunetsugu-Yokota and Yamamoto 2010; Ghelani et al. 2012).The interaction between miRNAs and 3′ UTRs of their target mRNAs can lead to mRNA degradation and translation inhibition. MiRNA based regulation is outlined in section 1.F (also **see 2.A**). Modulation of secondary structure of pre-mRNA in response to various protein factors is crucial to alternative splicing. This is important for generating complexity of genomes and this aspect of structural regulation is addressed in section 1.I.2 (also **see 3.A**). Multiple RNA structures can be formed from its primary sequence that are dynamic in response to various molecular inputs. Such a conformational flexibility results in changes in the gene expression, which adds another layer to the complexity of gene regulation. In the section below I outline two examples that demonstrate the specificity and dynamic character of RNA structures in mediating gene regulation.

1. B.1: Riboswitches

Riboswitch is one of the best demonstration of specificity and dynamics of RNA structures. These are regulatory domains on mRNAs found extensively in prokaryotes that comprise two domains: (i) an aptamer domain that recognizes its specific ligand and (ii) an expression domain that modulates gene expression either at the transcriptional or translational level (Breaker 2012)

(Figure 1.3). Thus, multiple classes of riboswitches that respond to wide range of cellular stimuli including amino acids, nucleotides, metal ions, small molecules and coenzymes have been reported (Breaker 2012).
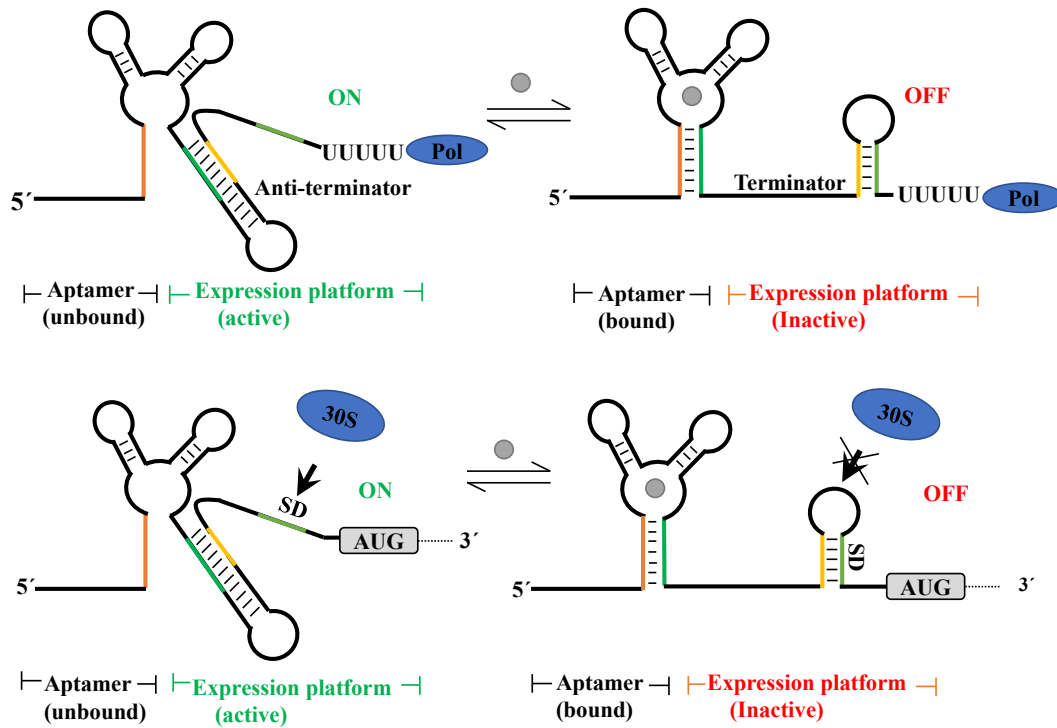
Figure 1.3: Common mechanisms of riboswitch mediated gene control. Transcription control involves metabolite binding and stabilization of a specific conformation of the aptamer domain that precludes formation of a competing anti-terminator stem. This allows formation of a terminator stem, which prevents transcription of the mRNA (top). Control of translation is achieved by metabolite-induced structural changes that sequester the Shine-Dalgarno sequence (SD), thereby preventing the 30S subunit of the ribosome from binding to the mRNA (bottom).

In eukaryotes such as *Neurospora crassa*, a TPP-specific riboswitch within the intron of NMT1 RNA was shown to regulate alternative splicing of the host transcript (Li and Breaker 2013). The aptamer domain of a riboswitch binds to a specific ligand through multiple interactions, such as hydrogen bonding and electrostatic interactions. For example, adenine riboswitch found in the 5′ UTR of the bacteria *ydhL* mRNA forms a secondary structure when bound to adenine

(Mandal and Breaker 2004). This prevents the formation of a transcription terminator loop and so transcription occurs (Lemay and Lafontaine 2007). However, a single base-pair change from U to C in the ligand binding site changes the affinity of the riboswitch from adenine to guanine. *S*-Adenosyl Methionine (SAM) riboswitch is another example where distinct classes of the SAM riboswitches can distinguish between SAM, a coenzyme for methylation reactions, and *S*-Adenosyl Cytosine (SAH), a by-product of the methylation reaction, though SAM and SAH are highly similar in structure. High resolution crystal structures of SAM riboswitches revealed multiple RNA structural conformation that are important to the recognition of specific substrates (Montange and Batey 2006; Gilbert et al. 2008; Lu et al. 2008).

The dynamics of RNA structure is also a recurring theme in mammalian RNAs. Modulation of RNA structure by binding of proteins which affect gene expression is an emerging field. For example, it is reported that human vascular endothelial growth factor A (*VEGFA*) mRNA contains a hypoxia-stability region in its 3′ UTR (Ray et al. 2009). The structure of this region changes depending on whether the cell is exposed to normoxic conditions or whether it is exposed to hypoxic conditions in the presence of interferon-γ (IFNγ). Under normal conditions the presence of the IFNγ-activated inhibitor of translation (GAIT) complex causes the *VEGFA* mRNA to form a structure that inhibits translation. However, during hypoxia, the binding of hnRNPAL results in a conformation switch to allow protein synthesis.

1. B.2: Long non-coding RNAs (LncRNAs)

Long noncoding RNAs (lncRNAs) are a class of RNAs that are larger than 200 nucleotides and that generally do not encode proteins. Recent studies have begun to uncover the potential of lncRNAs to regulate pathways by diverse mechanisms including chromatin modification,

transcriptional and post-transcriptional regulation, cell differentiation and subcellular trafficking (Tahira et al. 2011; Wang and Chang 2011; Qi et al. 2016). Their intrinsic nucleic acid nature confers on them the ability to function as ligands for trans-acting proteins and to mediate base-pairing interactions that guide lncRNA-containing complexes to specific RNA or DNA target sites. Though this activity is similar to that possessed by other small non-coding RNAs like microRNAs (miRNAs) and small nuclear RNAs (snRNAs), lncRNAs can fold into complex secondary and higher order structures to provide greater potential and versatility for both protein and target recognition. The flexible and modular scaffolding property of lncRNAs enables them to interact with proteins (RNA-protein interactions) in a co-operative manner compared to protein-protein interactions only. This then would result in combinatorial RNA-mediated tethering of proteins to enable a multiple steps of gene expression. High-throughput RNA sequencing (RNA-seq) experiments, have revealed that, lncRNAs are differentially expressed across various stages of differentiation, indicating that they may be novel 'fine-tuners' of cell fate.

Most of the lncRNAs identified so far, function by guiding chromatin modifiers to specific genomic loci. They have been shown to recruit DNA methyl transferase 3 (DNMT3) and histone modifiers, such as the Polycomb repressive complex (PRC2) and Histone H3 lysine 9 (H3K9) methyltransferases. The resultant DNA and histone modifications predominantly correlate with the formation of repressive heterochromatin and with transcriptional repression. Further, transcription of lncRNAs can regulate its own gene expression. Recruitment of chromatin-modifying complexes, such as the histone H3K4 methyltransferase mixed-lineage leukemia 1 (MLL1) complex, and by the activation of specific enhancer regions through changes to 3D-chromatin conformation. Depending on the target sites they are classified as (i) *cis*-acting lncRNAs, and (ii) *trans*-acting lncRNAs. *Cis-*

acting lncRNAs control the expression of their own genes which are positioned in the vicinity of their own transcription sites like Xist RNA, Kcnq1ot1, Airn, HOTTIP *etc* (**Figure 1.4A**). *Trans*-acting lncRNAs can either activate or repress the expression of genes positioned at distinct loci and examples include HOTAIR, lncRNA-ES1, Jpx 9 (**Figure 1.4B**).
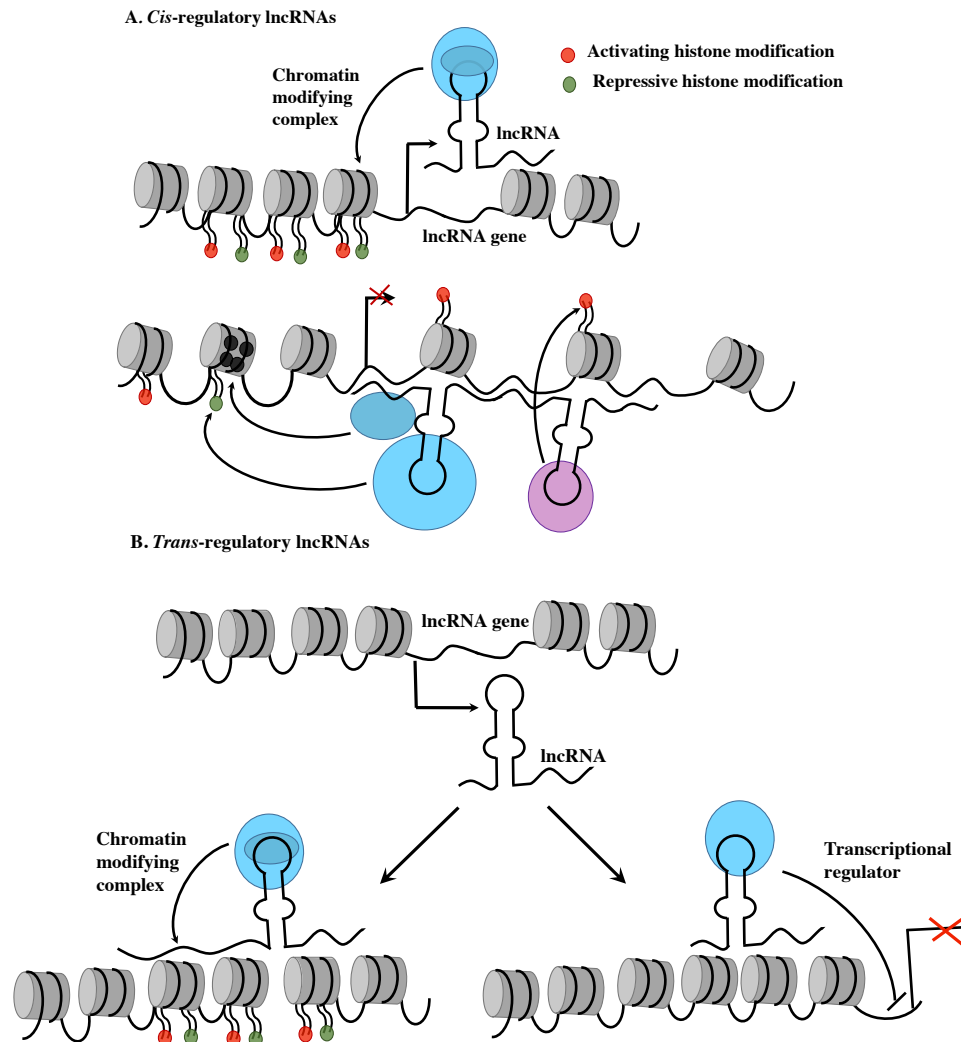
Figure 1.4: Schematic to show long non-coding RNAs (lncRNAs) mediated gene regulation. LncRNAs regulate transcription either in *cis* (A) or in *trans* (B) by recruiting specific transcriptional regulators onto specific chromosomal loci.

Nuclear lncRNAs have also been shown to mediate indirect regulatory effects on gene loci. For example, (i) they could act as decoys to sequester transcription factors, (ii) allosterically modulate the regulatory proteins, (iii) alter the nuclear domains and long-range three-dimensional chromosomal structures. The identification of X-inactive specific transcript (*XIST*) as a regulator of X chromosome inactivation in mammals provided one of the first examples of a *cis*-acting lncRNA that is directly involved in the formation of repressive chromatin (Plath et al. 2002). The lncRNA *Xist* is responsible for guiding X chromosome inactivation (XCI) which is a process that equalizes gene expression between mammalian males and females by inactivating one X in females (Tattermusch and Brockdorff 2011; Engreitz et al. 2013; Plath et al. 2002). During female development, Xist RNA is expressed from the active X and coats the X chromosome from which it is transcribed, leading to chromosome-wide repression of gene expression. An overlapping antisense lncRNA called Tsix represses Xist expression in *cis*, while the lncRNA Jpx, whose expression accumulates during XCI, activates Xist on the inactive. Hox transcript antisense RNA (*HOTAIR*) was one of the first trans-acting lncRNAs that was identified. *HOTAIR* is transcribed from HOXC gene cluster which represses the activity of HOXD cluster located on a different chromosome. It functions by recruiting two repressor complexes PRC2 and KDM1A-coREST-REST (lysine-specific histone demethylase 1A-REST corepressor 1-RE1-silencing transcription factor) that are histone-modifying complexes to the target genes (Rinn et al. 2007; Tsai et al. 2010).

The lncRNA *Air* involved in mouse imprinted gene silencing at the Igf2r locus , Braveheart (*Bvht*) was shown to be crucial for cardiac development in mice (Klattenhoff et al. 2013). Another lncRNA, *Fendrr*, controls chromatin modifications and, thus, developmental signaling in the rodent heart (Grote et al. 2013). A recent study identified cardiac hypertrophy–associated transcript (*Chast*),

that impedes expression of the autophagy regulator Pleckstrin homology domain–containing protein family M member 1 (Plekhm1), and this results in autophagic inhibition and cardiomyocyte hypertrophy (Viereck et al. 2016). PANDA is a lncRNA that is transcribed in response to DNA damage in a p53 dependent manner. It interacts with the transcription factor NF-YA to limit expression of pro-apoptotic genes and enables cell-cycle arrest (Puvvula et al. 2014). Genome-wide expression analysis projects predicted wide prevalence of independently transcribed intronic ncRNAs and recent studies revealed that introns in fact host a large number of lncRNAs (Louro et al. 2009; Wang and Chang 2011). The phenomenon of combinatorial transcriptional regulation by lncRNAs is also found in plants. Heo and Sung (2011) showed that a long intronic ncRNA termed cold-assisted intronic non-coding RNA (*COLDAIR*) is required for the vernalization-mediated epigenetic repression of Flowering locus (FLC) (Heo and Sung 2011). This is achieved via the establishment of stable repressive chromatins at *FLC* through its physical association with and recruitment of PRC2 to the locus.

It is interesting to note that the lncRNAs have been identified to have poorly conserved primary sequences. The most extreme example that shows lack of primary sequence conservation in long ncRNAs is Human Accelerated Region1 (*HAR1*), part of a novel RNA gene that is expressed specifically in Cajal–Retzius neurons in the developing human neocortex (Nishimura et al. 2002). However, sequence conservation might be relevant for ncRNAs that work in *trans*, when secondary structure is a requirement for these ncRNAs to bind at RNA-binding protein targets in order to exert their cellular functions. For example, the intergenic HOTAIR RNA that binds to a polycomb group factor to regulate the expression of homeobox genes has been shown to form intricate secondary structures that are crucial to its function (Somarowthu et al. 2015). In contrast, some long ncRNAs

exhibit an unexpected high level of nucleotide sequence conservation in mammalians, such as Metastasis-associated lung adenocarcinoma transcript1 (MALAT1) which is localized in nuclear speckles. It binds to and sequesters several serine/arginine (SR) splicing factors to nuclear speckle and thereby leading to altered pattern of alternative splicing (Tripathi et al. 2010).

RNA modifications have also emerged to be important modulators of the structure and function of cellular RNAs, of which N6-Methyladenosine (m^{6}A) is known to be the most abundant modification of mRNAs (Cao et al. 2016; Batista et al. 2014; Shi and He 2016). It is shown that m6A site in lncRNA called metastasis associated lung adenocarcinoma transcript1 (MALAT1) induces a local change in the structure that increases the accessibility of U rich sequence to hnRNPC. This m^{6}A dependent regulation of protein binding through conformational switch called m^{6}A-switch has been reported to affect mRNA expression as well as alternative splicing (Zhou et al. 2016).

Whole-genome structure analysis on HIV RNA genome using selective 2′ hydroxyl acylation analyzed by primer extension (SHAPE analysis) has revealed that both 5′ UTR and 3′ UTRs where highly structured (Wilkinson et al. 2008; Mortimer and Weeks 2007). A range of next-generation sequencing based methods in combination with parallel analysis of RNA structure (PARS) (Kertesz et al. 2010; Wan et al. 2013), fragmentation sequencing (FragSeq) (Underwood et al. 2010) are currently been applied to eukaryotic cells to gain comprehensive insights into RNA structure. However, understanding RNA tertiary structures and their mechanism of regulation lags behind secondary structure owing to the experimental challenges to decode long-range interactions. The examples discussed above clearly demonstrate that RNA structure-based changes result in meaningful functional outputs, and it is key to understand of how RNA structures have an impact on cellular function.

1. C: Intronic RNAs

1. C.1: Spliceosomal Introns

The presence of spliceosomal introns is one of the most defining features of eukaryotic genomes. The presence of introns in a genome was initially thought to impose substantial burden on the host and regarded as 'junk'. The excision of spliceosomal introns unlike self-splicing introns require the spliceosome which is a large ribonucleoprotein complex comprising five small nuclear RNAs (snRNAs) and more than 200 proteins. Intron-bearing genomes have to encode for all these proteins and snRNAs (Wahl et al. 2009) and dysregulation in any of these molecules that are necessary for proper splicing is detrimental to the cell. Further, replication and transcription of the introns is energetically expensive and time consuming. Though the energetic burden is probably tolerable (Lane and Martin 2010) an average RNA polymerase II (RNAP II) elongation rate of 60 bases per second (Singh and Padgett, 2009) indicates that some long introns may take few hours to get transcribed. The recognition of splice junctions by the spliceosome is directed by a host of *cis* regulatory elements which are prone to mutations. It is estimated that more than 50% of human genetic disorders are caused by disruption of the normal splicing pattern (López-Bigas et al. 2005; Wang and Cooper 2007). These potentially unfavorable nature of introns initiated a quest for functions that would overcome its deleterious effects. Based on these facts Walter Gilbert proposed the intron-early theory that suggested that introns were crucial in the formation of modern complex genes allowing for constant shuffling of primordial exons (Gilbert, 1987). Sequencing technologies on whole eukaryotic genomes have allowed for high resolution reconstruction of the evolutionary history of introns (Csuros et al. 2011; Carmel et al. 2007) Several studies now reveal that eukaryotic

genomes have gained several intron-related functions and introns are absolutely essential to their gene regulation through diverse mechanisms (Lynch 2007; Jo and Choi 2015).

1. C.2: Functions of introns

The life span of an intron can be broadly classified into five phases to address the functions: (i) the *genomic intron*, which is the DNA sequence of the intron, (ii) the *transcribed intron*, which is the phase in which the intron is under active transcription, (iii) the *spliced intron*, in which the spliceosome is assembled on the intron and actively excises it, (iv) the *excised intron*, which is the intronic RNA sequence released upon the completion of the splicing reaction, and (v) the exon-junction complex (EJC)-harboring transcript, which is the mature mRNA where the exon-exon junctions are marked by the EJC.

Introns are known to modify the expression levels of their host gene with very diverse mechanisms (Jo and Choi 2015). Intron-hosted DNA elements that include enhancers, silencers, and others are known to modulate the functions of the main promoter regions (Gaunitz et al. 2004; G. Zhang et al. 2011). Promoter-proximal introns that are characterized by unique sequence motifs are known to enhance the expression of genes in several species (Rose et al. 2008). Further, introns that do not harbor elements that modify the efficiency of the main promoter, but rather host an alternative promoter that when activated, gives an isoform with a different transcription start site are also reported (Davuluri et al. 2008). The lengths of the introns are also reported to be important in tissue specific genes as they host regulatory elements that could serve as scaffolds to allow correct assembly of nucleosomes (Vinogradov 2004). Depending on the length, introns require minutes to hours and sometimes even days to get transcribed indicating that they serve as tools to orchestrate time delays between the activation of a gene, and the appearance of its protein product. (Swinburne and Silver

2008). The ecdysone-inducible gene E74 that initiates metamorphosis of *D. melanogaster* consists of three transcripts of which the primary 60 kb long transcript gives 6kb mRNA after splicing. The protein product appears in the cytoplasm one hour post induction indicating that it is the intron transcription time that contributes to this delay (Thummel et al. 1990).

RNA splicing is co-transcriptional and these two cellular processes are strongly coupled through the carboxy terminal domain (CTD) of RNA Pol II (Beyer and Osheim 1988; LeMaire and Thummel 1990; Akhtar et al. 2009; Moore and Proudfoot 2009). The rate of formation of the first phosphodiester bond by RNAP II is known to be stimulated by the association of U1 snRNA with TFIIH, a general transcription initiation factor (Kwek et al. 2002). Other transcription initiation factors like TFIID and TFIIB, are also reported to be preferentially associated with donor splice junctions. This indicates that 5′-most introns stimulate transcription initiation at the upstream promoter through U1 snRNA-mediated preinitiation complex assembly at the donor splice site and that this function is independent of splicing (Damgaard et al. 2008; Spiluttini et al. 2010). U2 snRNP promotes transcription elongation by interacting with transcription elongation factors TAT-SF1 and p-TEFb (Fong and Zhou 2001). Depletion of SC35 has been reported to attenuate transcription via its interactions with p-TEFb, which was rescued by addition of recombinant SC35 (Lin et al. 2008). The final step of transcription wherein the mRNAs undergo 3′-end processing, involving endonucleolytic cleavage and the addition of a poly(A) tail, is also coupled with the 3′-most intron (Millevoi and Vagner 2010; Proudfoot 2011). The splicing regulatory elements (SRE) are harbored in both exons and introns. The SRE found in introns are intronic splicing silencers (ISSs) and intronic splicing enhancers (ISE)s (Havlioglu et al. 2007; Culler et al. 2010) that are key players in regulating splicing and thereby gene expression. Nova-1 a neuron-specific RNA binding protein found mainly in the

brain (Ule et al. 2005), regulates alternative splicing by binding to intronic motifs – such as YCAY – and enhances splicing of the downstream splice site. Fox-1 is another splicing factor that induces exon skipping in heart and skeletal muscles by binding to the intronic motif GCAUG (Jin et al. 2003). Autoregulation of ADAR2 gene that is important in A to I RNA editing is an interesting example wherein AA dinucleotide 47 bases upstream to one of the acceptor splice sites with AG dinucleotide is edited to AI. This is recognized at AG and preference of this new splice site over the canonical acceptor splice site leads to the production of inactive form of ADAR2, following a decrease in its levels (Rueter et al.1999).

Introns that are excised after splicing are part of post-splicing complexes that lead to efficient debranching and degradation. However, when an RNA gene is embedded within the intron, it is expressed upon the removal of intron and outlives its intronic host. Several studies have explored many families of non-coding RNAs like small nucleolar RNAs (snoRNAs), microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), small interfering RNAs (siRNAs), and various long non-coding RNAs (lncRNAs) that are preferentially associated with introns showing that genes autoregulate their expression by hosting relevant ncRNAs (Rearick et al. 2011). snoRNAs are a large family of small RNAs mainly known for their role in posttranscriptional methylation and pseudouridylation of various RNA genes like rRNAs, tRNAs, and snRNAs. They are known to reside in intergenic regions and have their own transcriptional promoter, or dwell in introns and rely on splicing for their maturation (Dieci et al. 2009) In fact, snoRNAs are found to be abundant in introns of vertebrates where they are processed by the exonucleolytic digestion of debranched introns after their excision from the pre-mRNA (Filipowicz and Pogacić 2002). Further, genes that function only to harbor

snoRNAs in their introns which do not have protein coding potential are also reported (Makarova and Kramerov 2009).

MicroRNA are small ncRNAs that bind to target sites along mRNAs, usually within their 3′ UTRs, and direct them for degradation or translation repression (Bartel 2009). They can lie in intergenic regions with their own transcriptional promoter or in intronic regions and co-expressed with their host genes (Baskerville and Bartel 2005). The processing of intronic miRNAs have been coupled to splicing wherein they are cropped from an unexcised intron (Kim and Kim 2007). Recent studies have found a large number of potential hairpin endogenous siRNAs within introns in human (Rearick et al. 2011).

In metazoans, the splicing reaction leaves traces in the form of a protein complex deposited 20–24 nucleotides upstream of the exon–exon junction, known as the EJC (Le Hir et al. 2003). The fact that even non-degraded spliced introns can be selectively exported to the cytoplasm (Wei et al. 2016) and be involved in global gene expression regulation makes the function of independently transcribed intronic sequences even more intriguing. The most crucial intronic function in contemporary metazoans is the increase in protein abundance of intron-bearing genes which was first reported in simian vacuolating virus 40 (SV40) constructs whose protein product was undetectable upon the elimination of their introns (Hamer et al. 1979). This phenomenon is associated with numerous introns suggesting that this intronic function is wide spread across many eukaryotic species (Le Hir et al. 2003). In fact, this function of intronic sequences is exploited to engineer hybrid introns from adenovirus 5′ splice site and an immunoglobulin G 3′ splice site to boost the expression of various genes both for *in vitro* and *in vivo* models (Choi et al. 1991).

1. D: MicroRNAs biogenesis

MicroRNAs (miRNAs) belong to a class of RNAs 20-22nt long that are regulators of eukaryotic gene expression. The miRNA genes are intergenic, or intronic and are transcribed by RNA Polymerase II to yield primary microRNA transcripts (pri-miRNA) which are hundred bp to several kb long (Lee et al. 2004). Pri-miRNAs are often 5′ capped and 3′ polyadenylated (Cai et al. 2004). The longest human miRNA cluster called chromosome 19 miRNA cluster (C19MC) that encodes ~50 miRNAs is exclusively transcribed by RNA polymerase III indicating that a small subset of microRNAs are also transcribed by RNA polymerase III (Borchert et al. 2006). Pri-miRNAs contain one or several local stem-loop structures in which one or both of the arms could form a mature miRNA. Transcripts that give rise to a single microRNA are called monocistronic and those that give two or more microRNAs are called polycistronic pri-miRNAs.

1. D.1: Canonical pathway

In a canonical miRNA biogenesis pathway, the pri-miRNAs are first cleaved by a nuclear RNase III enzyme, Drosha which forms a large multi-protein complex called microprocessor complex (MPC) with DiGeorge Critical Region 8 (DGCR8) and other proteins (**Figure 1.5**). This complex is ~650 kDa in human which crops the pri-miRNA releasing 70 to 80-nt long stem-loop structured precursor-miRNAs (pre-miRNAs). The two essential components of the minimal microprocessor complex are RNase III enzyme Drosha (~160 kDa) and the double-stranded RNA binding protein DGCR8, (known as Pasha in D. melanogaster and C. elegans) (Denli et al. 2004). Drosha contains two tandem RNase III domains and a double-stranded RNA-binding domain (dsRBD) that are important for catalysis. DGCR8 is ~120 kDa double stranded RNA binding protein that contains

two dsRBDs, and a putative WW domain, which is an interaction module for specific proline-rich

sequences.
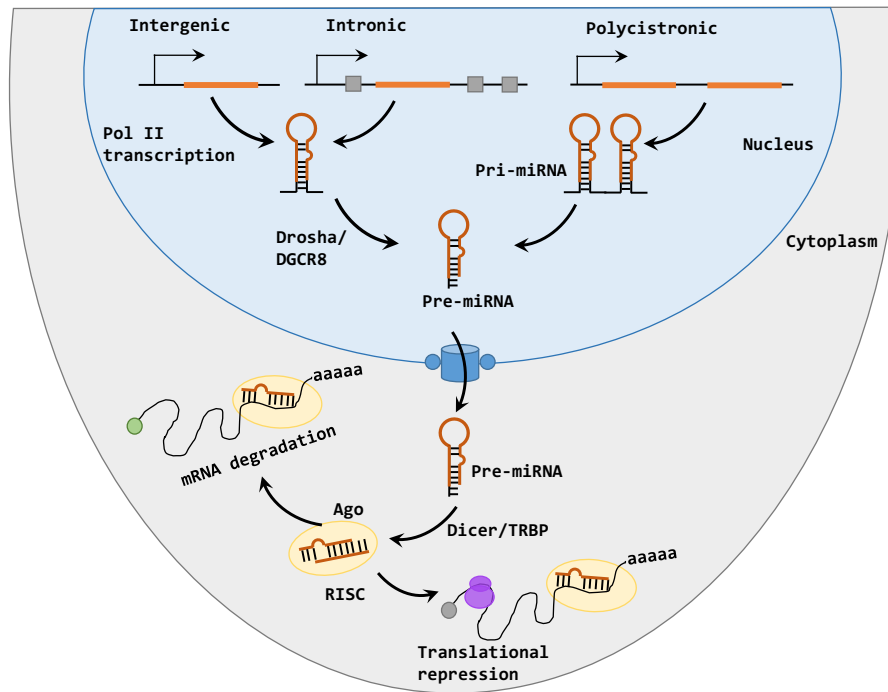
Figure 1.5: Schematic showing canonical miRNA biogenesis pathway. Canonical microRNA (miRNA) genes are transcribed by RNA polymerase II (Pol II) to generate the primary transcripts (pri-miRNAs). The first nuclear step (cropping) is mediated by the Drosha–DGCR8 complex (also known as the Microprocessor) that generates ~65-80 nucleotide (nt) pre-miRNAs with ~2-nt 3′ overhang. Post export into the cytoplasm by Exportin 5 the cytoplasmic RNase III Dicer catalyzes the second processing (dicing) step to produce miRNA duplexes. Dicer-TRBP complex mediates the processing of pre-miRNA and the assembly of the RISC. One strand of the duplex targets a specific mRNA and results either in its degradation or translational inhibition.

The larger microprocessor complex contains several accessory proteins like heterogeneous

nuclear ribonucleoproteins (hnRNPs), DEAD- box helicases p68, and p72 and Ewing's sarcoma

proteins (Gregory et al. 2004). During the first step of microprocessing, DGCR8 first recognizes

the single strand double strand junction (ss-ds) of pri-miRNA and binds to the pri-miRNA via its

RBDs and then recruits Drosha to crop the pri-miRNA. The binding of DGCR8 that is one helical

turn from the ss-ds junction serves as a molecular ruler for the cleavage site (Han et al. 2004). The

cropping of pri-miRNA by Drosha results in a 2 nt overhang at the 3′ end of the pre-miRNA that functions as a signature for recognition by Exportin-5. Further, Exportin-5 complexes with Ran-GTP and exports the pre-miRNA from nucleus to the cytoplasm (Yi et al. 2003). Following their export into the cytoplasm, pre-miRNAs are further processed by a cytoplasmic RNase III enzyme Dicer. Human Dicer is a 1922 amino acid residue protein that contains a helicase domain, a domain of unknown function (DUF283), a PAZ domain on the N-terminal side of the RNase III domains, and a dsRNA binding domain (dsRBD) on the C-terminal side *D. melanogaster*, has two homologues, Dicer-1 required for pre-miRNA cleavage, and Dicer-2 for siRNA generation. The PAZ domain is also found in a group of highly conserved proteins known as Argonaute (Ago) proteins and has been implicated in binding the protruding 3′-end of small RNAs. In human Dicer forms a multi-protein complex refered as miRNA RISC loading complex (miRLC) with several proteins such as TAR RNA binding protein (TRBP) (Chendrimada et al. 2005), protein activator of PKR (PACT, Lee et al. 2006), Argonaute-2 (Ago2, Gregory et al. 2005) and Loquacious (Loqs) in Drosophila. This complex cleaves the terminal loop by cutting both strands of the pre-miRNAs, leaving a transient, ~ 22-nt long, miRNA/miRNA* duplex with 2-nt overhangs at both the 3′-ends. The overall hairpin length and loop size influences the efficiency of Dicer processing, and the imperfect nature of the miRNA/miRNA* pairing also affects its cleavage (Park et al. 2011).

Following cleavage, Dicer and interacting proteins TRBP or PACT dissociate from the miRLC initiating the transition of the miRLC into the active RNA induced silencing complex (miRISC). During miRISC formation, the miRNA/miRNA* duplex is unwound and one of the strands called guide strand targets a specific mRNA. The complementary miRNA* strand or the passenger strand is degraded but sometimes the passenger stand can function as a guide strand and

target another complementary mRNA. The thermodynamic stability of the 5′ end of miRNA duplex determines the guide strand that is incorporated into miRISC. The 5′ end of the miRNA/miRNA* duplex with low base pair stability is incorporated into RISC and the unwinding of the duplex is known to be mediated by different RNA helicases such as p68, p72, RNA helicase A (RHA), TNRC6B, Mov10 (Gregory et al. 2004; Welker et al. 2010). The processing of Drosha and DCGR8 on some precursors is known to produce miRNA variants with heterogenic ends with different 5′ end stabilities which would consequently, alter selection of the guide strand (Carthew and Sontheimer 2009).

1. D.2: Non-canonical pathway

In addition to the well-defined canonical miRNA biogenesis pathway wherein the miRNAs are processed in two steps to form mature miRNAs, non-canonical pathways for miRNAs are also reported. Such pathways generate miRNAs independent of either Drosha or Dicer **(Figure 1.6)**. Some introns harbor miRNAs that are processed independent of Drosha and are called 'mirtrons'. The mirtrons are first subjected to splicing to excise the intron and then the lariat intron carrying the pri-miRNA undergoes debranching to generate pre-miRNA hairpins thereby bypassing the initial Drosha cropping to release pre-miRNA (Okamura et al. 2007; Berezikov et al. 2007). Unlike the canonically processed intronic miRNAs, mirtrons are located within very short introns with their ends containing the splice sites for splicing. Following RNA refolding, mirtrons are transported to the cytoplasm by Exportin-5-Ran-GTP complex where they are further processed by Dicer (Okamura et al. 2007; Ruby et al. 2007) to form mature miRNAs. Another non-canonical miRNA biogenesis pathway that is independent of Dicer has been shown for miR-451 (Liu et al. 2013). Though its pri-miRNA is processed by Drosha (also known as Rnasen), its maturation does not require Dicer. The pre-miR-

451 which has a shorter stem compared to other pre-miRs is directly loaded on Ago2 containing an RNase H-like endonuclease activity. This results in the formation of an intermediate 3′ end which is exonucleolytically trimmed by cellular RNAses to form mature miR-451.
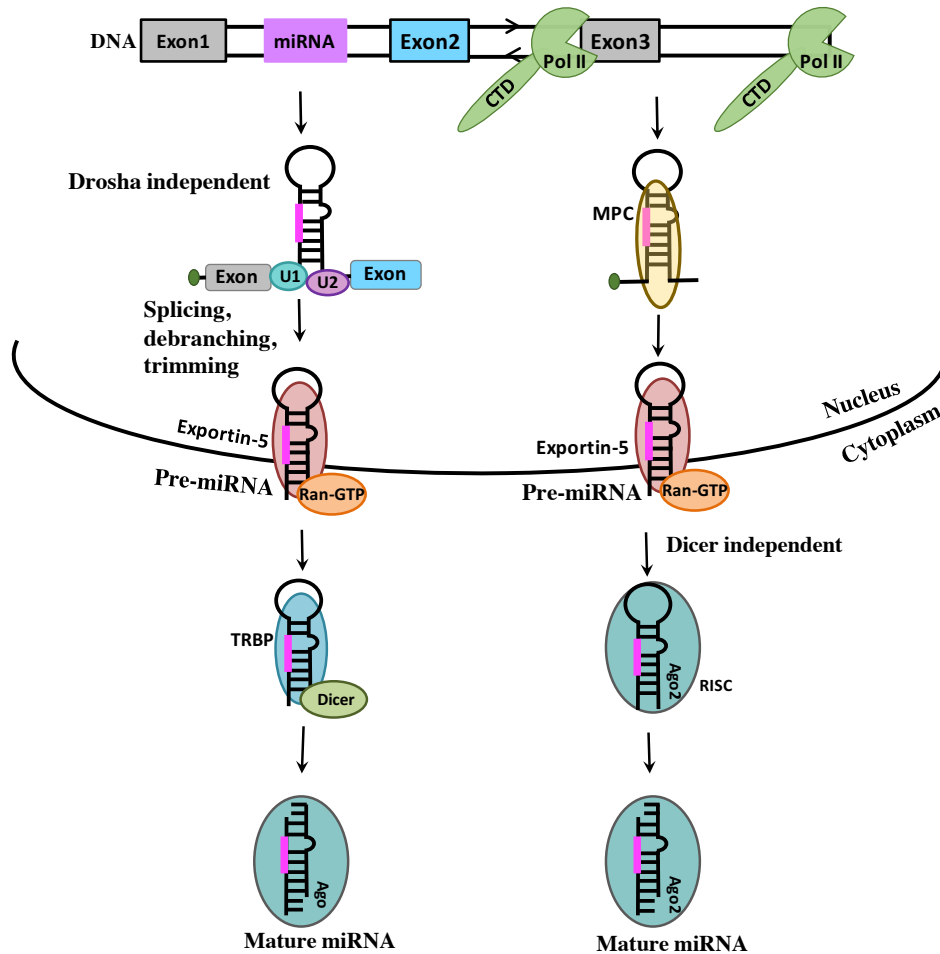
Figure 1.6: Non-canonical miRNA biogenesis pathways involve primary microRNAs (pri-miRs) that bypass Drosha or Dicer processing steps to give mature miRNAs. Small intronic RNAs called mirtrons bypass the Drosha are produced from spliced introns which are debranched to form pre-miRNAs. Pre-miRNA-451 formed from Drosha processing of pri-miR-451 can bypass Dicer and get loaded into RISC complex. Both result in functional mature miRNA that target specific mRNAs.

1. E: Mechanisms of miRNA-mediated regulation

The past decades have explored the regulatory functions of miRNAs and the key players in miRNA mediated gene silencing are Argonaute (Ago) proteins which form miRISC. Humans have four Ago proteins (hAgo1-4) that are 100 kDa in size and ubiquitously expressed. All four hAgo proteins are capable of loading miRNAs and guiding them to their target mRNAs. hAgo2, possesses cleavage activity and cleave the fully complementary sequences guided by short interfering RNAs (siRNAs) or miRNAs (Liu et al. 2004; Meister et al. 2004). In the majority of cases, miRISC targets the 3′ untranslated region (3′-UTR) of the mRNAs by forming a partially complementary RNA-RNA duplex. The target specificity is achieved by seed sequence complementary to the target site in 3′-UTR spanning positions 2-8 at the 5′ end of the miRNA (Lewis et al. 2005; Rajewsky 2006) miRNA-mediated gene silencing can occur at pre-translational, co-translational, and post-translational steps. Interestingly a process called RNA-induced transcriptional silencing (RITS) in mammals is known to be mediated by miRNAs (Bühler et al. 2006). In this, miRNA in complex with nuclear Ago proteins results in transcriptional gene silencing through chromatin remodeling (Shimada et al. 2016). Argonaute proteins are known to interact with several translational initiation factors that include eukaryotic translation initiation factor 4E (eIF4E) that directs ribosomes to bind to the 5′ cap. Further, polyA binding protein (PABP) and EIF4G, an RNA helicase that is involved in ATP-dependent unwinding of 5′-terminal secondary structure and recruitment of mRNA to the ribosome are associated with eIF4E. miRNAs are also reported to bind to 5′ UTR or coding regions of certain mRNAs to inhibit their translation (Lytle et al. 2007). Eukaryotic initiation factor 6 (eIF6) is reported to be recruited by Argonaute, which prevents the large ribosomal subunit from joining to miRNA-targeted mRNA (Chendrimada et al. 2007). Further, miRISC can interfere with elongation factors

resulting in dissociation or premature termination in translation and subsequent degradation of the premature polypeptide (Pillai et al. 2007).

MiRNAs were also demonstrated to regulate protein synthesis by altering the stability of their mRNA targets. MiRISCs bound to 3′-UTR of target mRNA is capable of recruiting deadenylase, CCR4:NOT1 to nibble the poly(A) tail followed by removal of the m7GpppG cap by the decapping enzyme (Behm-Ansmant et al. 2006; Wu et al. 2006). The deadenylated and decapped transcripts are subsequently degraded by exonucleolytic enzyme, Xrn1. mRNAs without a poly(A) are also subjected to cytoplasmic exonucleases or to sequence specific cleavage by polysomal ribonuclease 1 (PMR1). The mRNAs targeted by miRNAs are found to be localized in processing bodies (P-bodies) or GW bodies (Liu et al. 2005; Sen and Blau 2005). The decapping complex containing DCP1, and DCP2 are required for the mRNA decay inside P-bodies, are crucial for miRNA-mediated gene silencing (Rehwinkel et al. 2005). GW182 proteins are known to be major components of P-body's integrity and stability. The physical interaction between GW182 and Ago proteins suggest a functional link between the cytoplasmic P-bodies and miRNAs to translationally suppress the target mRNAs or to sequester them for storage till they are required for translation (Liu et al. 2005; Behm-Ansmant et al. 2006).

While the majority of miRNAs are modulate gene expression solely by negative regulation of target mRNA, recent observations by Vasudevan *et al* indicate that miRNAs oscillate between repression and stimulation in response to specific cellular conditions, sequences, and cofactors (Vasudevan et al. 2007). miR-145 is reported to upregulate Myocardin levels during muscle differentiation (Cordes et al. 2009). Krueppel-like factor 4 (KLF-4) a key transcription factor during embryonic development is shown to be upregulated by miR-206 (Lin et al. 2011).

1. F: Regulation of miRNA expression

Studies to understand miRNA-based gene regulation clearly state that proper functioning of an organism depends on their tight and controlled expression. Though the basic mechanisms of miRNA biogenesis and function have been unraveled in recent years, the regulation of miRNAs themselves are still not clear. Thus, the spatiotemporal expression of a microRNA remains an active area of research (Krol et al. 2010).

1. F.1: Transcriptional regulation of miRNAs

MiRNA genes are located within various genomic contexts and can be transcriptionally regulated in a manner similar to that of protein-coding genes. Often, several miRNA loci are in close proximity to each other and constitute a polycistronic transcription unit. MiRNA genes that reside in the introns of protein-coding genes usually share the promoter of the host gene. However, it is known that some miRNA genes have multiple transcription start sites and the promoters of such intronic miRNAs are distinct from the promoters of their host genes (Zhou et al. 2006). miRNA genes are usually transcribed by RNA Pol II and are controlled by RNA Pol II associated transcription factors and epigenetic regulator (Lee et al. 2004; Cheng and Sharp 2003). A small subset of miRNAs that includes C19MC miRNAs are transcribed by RNA Pol III (Borchert et al. 2006). Transcription factors like p53, c-Myc, Zinc Finger E-box Binding Homeobox-1 (ZEB1) and ZEB2 are reported to regulate the miRNA expression. Epigenetic factors like DNA methylation and histone modifications also contribute to the miRNA regulation (Ozsolak et al. 2007).

MiRNAs also uniquely participate in regulatory loops involving specific transcription factors that modulate their own expression. The transcription factor PITX3 and miR-133b form a negative autoregulatory loop that controls dopaminergic neuron differentiation (J. Kim et al. 2007). PITX3

stimulates transcription of miR-133b, which in turn suppresses PITX3 expression. miR-145 was identified to induce a pro-apoptotic effect, dependent on TP53 activation, and TP53 activation could stimulate miR-145 expression, thus establishing a death- promoting loop between miR-145 and TP53 (Cordes et al. 2009; Spizzo et al. 2010). Beyond regulation of miRNAs through such feedback loops, it is also reported that certain environmental conditions to influence miRNA expression. Cells undergoing hypoxic stress have demonstrated the up-regulation of miR-210 by interaction with Hypoxia-inducible factor 1A (HIF1A). HIF1A antagonizes c-Myc activity and induces cell-cycle arrest and the induction of miR-210 promotes cell survival under hypoxia (Huang et al. 2009).

1. F.2: Post-transcriptional regulation of miRNAs

Several post-transcriptional mechanisms have been reported to control the miRNA expression through various intracellular cues that impact the microprocessing step involving trans-acting protein factors that interface with components of the microprocessor complex. I outline a few mechanisms in this section **(Figure 1.7)**.
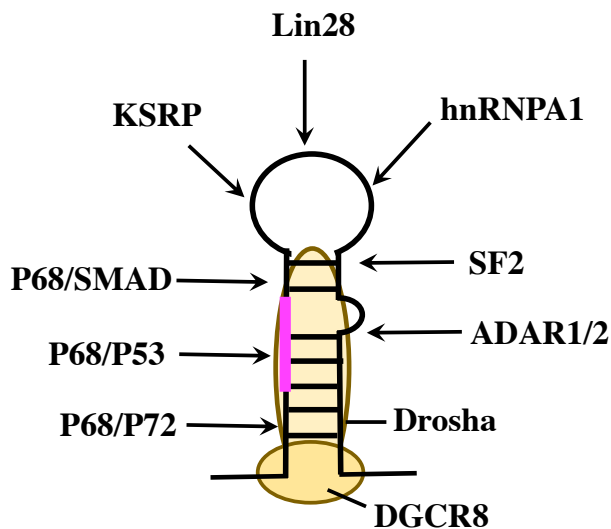
Figure 1.7: Post-transcriptional regulation of miRNA biogenesis. Trans-acting proteins can either facilitate (hnRNPA1, KSRP, P68/72, P68/SMAD, P53/P68, SF2) or inhibit (lin28, ADAR1, ADAR2) processing of primary microRNAs.

p53 is a tumor suppressor protein that responds to DNA damage, oncogene activation and other stress signals through the induction of cell cycle arrest, senescence and apoptosis (Sengupta and Harris 2005; Junttila and Evan 2009). It interacts with the Drosha processing complex through the association with the DEAD-box RNA helicase p68 (also known as DDX5) and facilitates the processing of pri-miRNAs to pre-miRNAs (Suzuki et al. 2009). One of the best studied negative regulator of miRNA biogenesis is Lin-28, which can act at different levels to control the biogenesis of the let-7 family of miRNAs. It is shown to bind to the conserved terminal loop sequences of pri-let-7 and interfere with Drosha cleavage. Further, it can also bind to pre-let-7 and induce the 3′-terminal polyuridylation of pre-let-7 by recruiting the TuT4 thereby preventing Dicer processing and targeting pre-let-7 for degradation (Heo et al. 2008; Newman et al. 2008).

HnRNP A1, a protein involved in splicing regulation was shown to bind to the terminal loop of pri-miR-18a present in pri-miR-17-92a cluster and cause the structural remodeling of the stem to generate a more favorable Drosha cleavage site (Guil and Cáceres 2007). Another splicing regulatory protein called KH-type splicing regulatory proteins (KSRP) is reported to bind several miRNAs and enhance the processing by Drosha-DGCR8 complex post-transcriptionally in response to DNA damage (Trabucchi et al. 2009; X. Zhang et al. 2011). RNA editing by Adenosine deaminases that act on RNA (ADARs) catalyze deamination of adenosine and converts it into inosine in dsRNA segments. This results in altered base-pairing and structural properties of transcripts that can affect both Drosha and Dicer-mediated cleavage. RNA editing of human pri-miR-142 is shown to change its structure and inhibit Drosha cleavage activity and the edited pri-miR-142 is degraded by Tudor-SN (Scadden 2005). Editing of pre-miR-151 prevents Dicer processing, resulting in accumulation of pre-miR-151 (Kawahara et al. 2007). Pre-miRNAs, when translocated to the cytoplasm, are cleaved

near the terminal loop by Dicer to generate a 22-nt double stranded mature miRNA. Similar, to Drosha, several Dicer-associated proteins have been identified, including TAR RNA-binding protein (TRBP) and protein kinase R-activating protein (PACT) (Chendrimada et al. 2005). The association of these proteins is shown to enhance the stability and processing activity of Dicer. Deletion mutations of TRBP resulted in decreased TRBP thereby destabilization of Dicer that allowed inefficient processing (Melo et al., 2009). In mammals, Argonaute (Ago) proteins show different potencies to repress translation and hence, differences in the relative abundance of individual Ago proteins may result in altered miRNA repression. Steady state concentrations of miRNA reflect the balance between the miRNA biogenesis and turnover. Like miRNA biogenesis, the degradation of individual miRNA species is also probably subjected to intense regulation. In *C.elegans* it is shown that the turnover of mature miRNA is mediated by the 5′-3′ exoribonuclease XRN-2 (Chatterjee and Grosshans 2009). The stability of mature miR-29b, but not the co-transcribed miR-29a, is reported to be modulated in a cell cycle-dependent manner (Hwang et al. 2007). However, the identification of factors that regulate both general and specific miRNA turnover are yet to be discovered.

1. F.3: Impact of pri-miRNA structure on processing

While a number of trans-binding proteins that are reported act on the microprocessor (MPC) to positively or negatively affect pre-miRNA formation from pri-miRNA substrate, the effect of the underlying RNA scaffold on the microprocessing has not been considered. RNA has the intrinsic ability to modulate or regulate the binding of RNA binding proteins during any cellular processes. For example, hnRNP A1 is shown to bind pri-miR-18a as well as pri-let-7a though they differ in their sequence and structure (Guil and Cáceres 2007; Michlewski and Cáceres 2010). The binding of hnRNP A1 facilitates processing of miR-18a while inhibits processing in let-7a suggesting that RNA

structure is crucial to determining such processing fates (**Figure 1.8**).
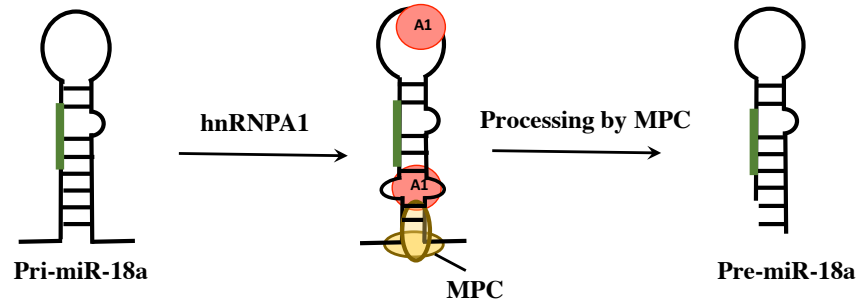
Figure 1.8: Trans-binding protein mediates structural changes in the pri-miRNA. Binding of hnRNP A1 to the pri-miR-18a conserved loop recruits another to the stem which then remodels the stem region and facilitate Drosha cleavage**.**

It is hence reasonable to hypothesize that pri-miRNAs are transcribed and folded into inactive substrates for microprocessing rather than being an authentic cleavage substrates. Subtle structural rearrangements result in activation of the pri-miRNA resulting in the formation of the authentic sites. Such structural remodeling can be brought about by trans-acting proteins such as RNA helicases or factors that can recruit them. However, the molecular details of these mechanisms still need to be elucidated. Footprint analysis of pri-miR-18a with hnRNP A1 revealed two hnRNP A1 binding regions, a primary site corresponding to the terminal loop of pri- miR-18a, and a secondary site that is internal loop in the stem. Binding of hnRNP A1 to the internal loop pri-miR-18a confers protection to specific nucleotides, and also results in the relaxation of a few nucleotides that are involved in canonical Watson-Crick pairing in the unbound pri-miR-18a (Michlewski and Cáceres 2010). This is first study that outlined the structural features on a pri-miRNA that can block Drosha processing and indicates the potential of other auxiliary factors to mediate such a regulation. For example, Lin28 has been reported to inhibit processing of pri-let-7 (Nam et al. 2011). Lin28 contains two nucleic acid interaction domains, a cold shock domain (CSD) and two tandem Cys-Cys-His-Cys (CCHC)-type zinc-binding motifs. The crystal structure

of Lin28-let-7 complex reveals that the CSD domain of Lin28 recognizes the terminal loop and CCHC motifs bind to a conserved GGAG motif and adopts a conformation that is not recognized by Dicer thereby inhibiting its processing. It is evident from the above examples that RNA structure and its remodeling can greatly affect its processing.

1. G: pre-mRNA splicing

One of the most fascinating findings in molecular biology was the discovery that eukaryotic genes are discontinuous, with protein-coding segments or exons disrupted by noncoding segments or introns (Berget et al. 1977). With advances in genome sequencing, it is now clear that precursor messenger RNA (pre-mRNA) splicing can occur to a great extent that scales with organismal complexity (E. Kim et al. 2007; Kim et al. 2004). The precise excision of introns and ligation of exons from pre-messenger RNA is performed by the spliceosome, a dynamic macromolecular machine containing five small nuclear RNAs and numerous proteins via a two-step trans-esterification reaction: (i) attack of the 2′ hydroxyl of a branch point adenosine at the 5′ splice site (SS), resulting in formation of an intermediate 2′-5′ branched intron-3′exon, (ii) attack of the 3′ hydroxyl of the 5′ exon at the 3′ SS producing the ligated exons and liberating the 2′-5′ branched lariat intron. The spliceosome assembles *de novo* on each intron from its canonical subunits U1, U2, U4, U5 and U6 snRNPs and various other protein factors. Yeast and human spliceosomes have sedimentation values of 40 to 60S and masses of ~4.8 MDa (Will and Lührmann 2011).

The spliceosome undergoes a cascade of assembly events and conformational rearrangement before forming an active complex on the pre-mRNA (Hoskins et al. 2011; Ilagan et al. 2013; Rino and Carmo-Fonseca 2009) **(Figure 1.9)**. The first step in splicing is typically the ATP-independent recognition of the 5′ SS by the U1 snRNA and the association of the U1 snRNP

with this region, which results in the formation of the early commitment (E) complex. The 3′ SS of the pre-mRNA is recognized by the U2 snRNP and associated factors, splicing factor 1 (SF1) and U2 auxiliary factors (U2AFs) also components of E complex. In a subsequent ATP-dependent process catalyzed by the DExD/H helicases pre-mRNA-processing 5 (Prp5) and Sub2, U2 snRNA recognizes sequences around the branch point (BP) adenosine and interacts with U1 snRNP to form the complex A. Metazoan genes contain relatively short exons (50–250 nucleotides) that are separated by larger introns (few hundred bp up to few hundred kb), splice sites are predominantly recognized in pairs across exons through the interaction of U1 and U2 snRNPs. This process is called exon definition, and the U1, U2 snRNP complex that forms across exons is known as the exon definition complex (Sterner et al. 1996). In a subsequent transition step, U1 and U2 snRNPs undergo structural rearrangements, forming an intron-spanning interaction known as the intron definition complex that brings the 5′ SS, branch point and 3′ SS into close proximity (De Conti et al. 2013). Thus, the metazoan intron definition complex is generally considered to be the equivalent of complex A in yeast, whereas the metazoan exon definition complex is similar to complex E. After the assembly of complex A, the U4, U6 and U5 snRNPs are recruited as a preassembled U4/U6. U5 to form complex B, in a reaction catalyzed by the DExD/H helicase Prp28. The resulting complex B goes through a series of compositional and conformational rearrangements to form a catalytically active complex B* (Bessonov et al. 2010).

Brr2, 114 kDa U5 small nuclear ribonucleoprotein component, and Snu114, two RNA helicases and Prp2 are required for the activation of complex B, resulting in rearrangements that lead to the formation of the U2–U6 snRNA complex. This U2–U6 base-paired complex forms the active site of the spliceosome, where the catalytic transesterification reactions of intron excision

and exon joining occurs. This structure bears remarkable similarity to the domain V region of self-splicing group II introns (Fica et al. 2014; Marcia et al. 2013), which also use a lariat 2′–5′ mechanism for group II intron removal. On the basis of the similarity between the U2–U6 snRNA base pairingand the group II domain V structure and mechanism, it is now clear that the spliceosome used RNA-mediated catalysis, much like the ribosome.
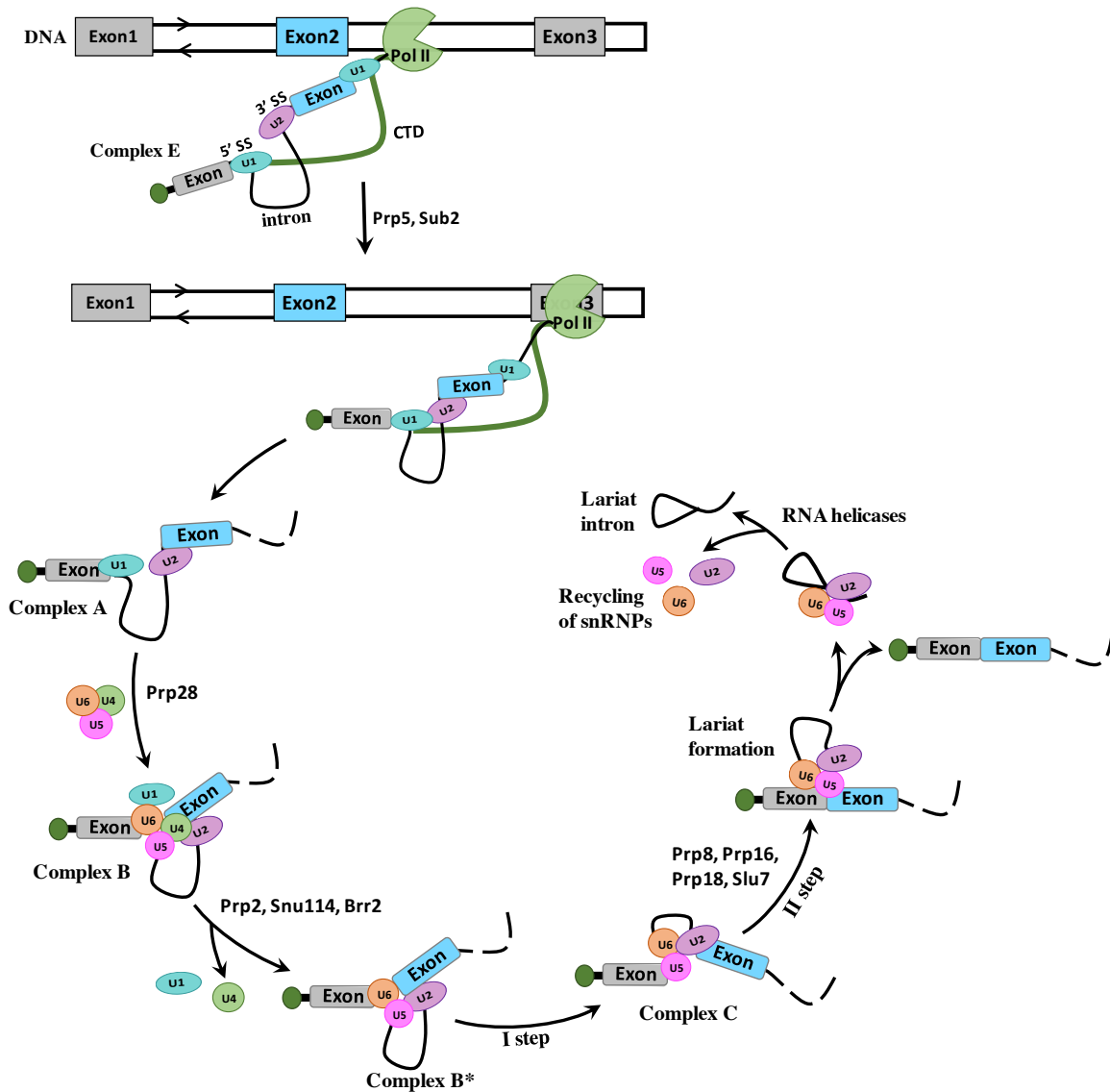
Figure 1.9: Schematic showing step-wise assembly of the spliceosome and the catalytic steps of splicing. Spliceosome assembly takes place at sites of transcription. The U1 and U2 snRNPs (Continued on next page)

31

assemble on the intron in a co-transcriptional manner through recognition of the 5′ splice site (5′SS) and 3′ SS, which is mediated by the carboxy-terminal domain (CTD) of polymerase II (Pol II). This is called the early complex (Complex E). The U1 and U2 snRNPs interact with each other to form the Complex A. This process is dependent on DExD/H helicases pre-mRNA-processing 5 (Prp5) and Sub2. In a subsequent reaction catalyzed by Prp28, the preassembled tri-snRNP U4–U6•U5 is recruited to form Complex B. The resulting Complex B undergoes a series of rearrangements to form a catalytically active complex B (Complex B*), which requires Brr2, Snu114 and Prp2 and results in the release of U4 and U1 snRNPs. Complex B* then carries out the first catalytic step of splicing, generating Complex C, which contains free exon1 and the intron–exon2 lariat intermediate. Complex C undergoes additional rearrangements and then carries out the second catalytic step, resulting in a post-spliceosomal complex that contains the lariat intron and spliced exons. Finally, the U2, U5 and U6 snRNPs are released from the mRNP particle catalyzed by the DExD/H helicase Prp22 and recycled for additional rounds of splicing.

Galej et al showed 3.8 Å cryo-electron microscopy structure of the spliceosome immediately after lariat formation wherein the 5′ SS that is cleaved was in close proximity to the catalytic in U2/U6 snRNA and the 5′ phosphate of the intron nucleotide G (+1) is linked to 2′ OH of the Adenosine at the branch point. The 5′ exon is held between amino-terminal and linker domains of Prp8 and base-pairs with U5 snRNA. Non-Watson–Crick interactions between the 5′ SS and branch helix dock the BP adenosine into the active site, stabilized by Isy I and factors Yju2, and Cwc25, while intron nucleotides +3 to +6 base-pair with the U6 snRNA <u>ACAG</u>AGA sequence. Isy1 and factors Yju2 and Cwc25 stabilize docking of the branch helix (Galej et al. 2016). The activation of complex B also opens the U4 and U6 snRNAs, releasing U4 and U1 from the complex, which is thought to unmask the 5′ end of U6 snRNA (Raghunathan and Guthrie 1998). Complex B* then completes the first catalytic step of splicing, generating complex C, which contains the free exon 1 and the intron–exon 2 lariat intermediate. Complex C undergoes additional ATP- dependent rearrangements to form complex C* before carrying out the second catalytic step of splicing, which is dependent on Prp8, Prp16 and synthetic lethal with U5 snRNA 7 (Slu7). This

results in a post-spliceosomal complex that contains the lariat intron and spliced exons (Wahl et al. 2009). A very recent cryo-EM structure of C* complex from yeast at 4A° resolution identified the components of second step of spliceosome, providing a basis for rationalization of the interactions required for this step (Yan et al. 2015; Yan et al. 2017). Finally, the U2, U5 and U6 snRNPs are released and recycled for additional rounds of splicing. The spliced product is then released from the spliceosome which is catalyzed by the DExD/H helicase Prp22 (Ilagan et al. 2013). Disassembly of the post-catalytic spliceosome is also driven by several RNA helicases like Brr2, Snu114, Prp22 and Prp43 in an ATP-dependent manner (Fourmann et al. 2013) (**Figure 1.9**).

1. H: Regulation of splicing

1. H.1: Core splicing signals

The primary signals that determine where the process of splicing occurs, are included within the pre-mRNA transcript. These are the donor site or the 5′ splice site (5′ SS), the acceptor site or the 3′ splice site (3′ SS). Also, the branch point (BP) contains Adenosine residue, and the polypyrimidine tract (PPT) between the BP nucleotide and the 3′ SS (Hall et al. 1988; Gao et al. 2008; Aebi et al. 1987) (**Figure 1.10**). The 5′ SS motif in eukaryotes contains partially conserved nucleotides, MAG/GURAGU (M indicates A or C, R indicates purines and the slash the exon-intron boundary) at the exon-intron junction. However, the GU dinucleotide of the 5′ SS consensus sequence is universally conserved and mutations in one of these two nucleotides is known to completely abolish the splicing (Langford et al. 1984). The 5′ SS is recognized by the U1 snRNA during early assembly of the spliceosome (Zhuang et al. 1989; Horowitz and Krainer 1994). It is shown that the interaction between the U1 snRNP and the 5′ SS is not highly complementary and it is possible that U1 can bind

33

to a nonspecific 5′ SS (Hicks et al. 2010). During the assembly of the spliceosome other factors like

U6 snRNP also mediate the recognition of 5′ SS. The interaction of U6 snRNP with 5′ SS is enhanced

in the presence of U1 snRNP which is displaced by U6 that binds the 5′ SS and stimulates the first

transesterification reaction.
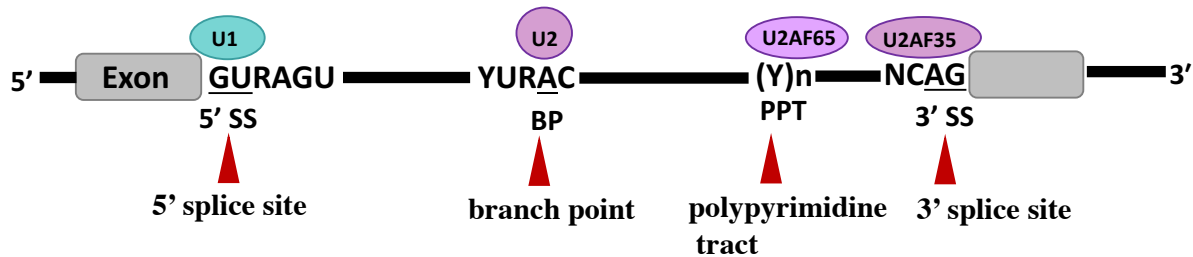
Figure 1.10: Schematic showing the core signals of splicing present on the introns. The 5′ SS is a conserved GU recognized by U1snRNP and a conserved AG at the 3′ SS that binds to U2AF35. The branch point contains an Adenosine followed by a polypyrimidine tract.

In yeast the conserved branch point (BP) sequence is UACUA**A**C, while in the metazoans this

sequence conservation is limited to the conserved adenosine. Mammalian BP is specified primarily

by its proximity to the intron/exon junction and the general consensus YNYUR**A**C motif (R-purine,

Y-pyrimidine). Most branch points are within 18-40 nt of the 3′ SS and mutation of the BP is known

to strongly reduce the splicing of downstream exon (Reed and Maniatis 1988; Reed 2000). It is

recognized by the splicing factor SF1 during assembly of Complex E and subsequently involves

binding of U2 snRNP through SF3a and SF3b to form Complex A (Gao et al. 2008). The

polypyrimidine tract (PPT) is a stretch of pyrimidines located between the branch point and the 3′ SS.

Several proteins such as auxiliary factor U2AF65 and polypyrimidine tract binding protein (PTB)

bind to this region (Wagner and Garcia-Blanco 2001). Deletion of polypyrimidine tract is known to

abolish the formation of the lariat during the second transesterification reaction. The terminal AG

dinucleotide defines the 3′ SS of the intron which has a short Y**AG**/G sequence (Hall et al. 1988) The

small subunit U2AF35 is involved in the recognition of the essential AG dinucleotide of the 3′ SS

during the earliest stage of spliceosome assembly (Sander et al. 2006). Both U2AF subunits (U2AF35

and U2AF65) bind to the 3′ SS and are subsequently replaced by the U5 snRNP during the formation

of the Complex B.

1. H.2: *Cis*-regulatory elements

Besides the core signals in splice site selection additional cis-regulatory elements that serve

as either splicing enhancers or silencers are important to maintain the fidelity of splicing. If they are

present on exons they are classified as exonic splicing enhancers (ESEs) or silencers (ESSs) and they

function to promote or inhibit inclusion of the exon they reside in. If present on introns they are

refered as intronic splicing enhancers (ISEs) or silencers (ISSs) and they enhance or inhibit usage of

adjacent splice sites or exons from an intronic location **(Figure 1.11)**.
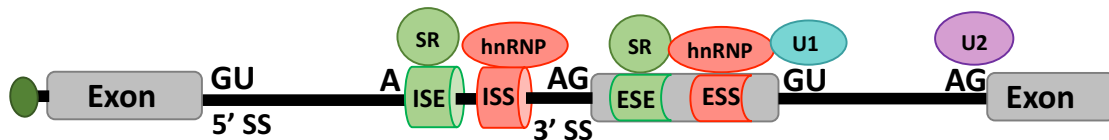
Figure 1.11: Schematic showing auxiliary regulatory elements on pre-mRNA (i) *cis*-regulatory elements: intronic splicing enhancers (ISE), intronic splicing silencers (ISS) and exonic splicing enhancers (ESE), exonic splicing silencers (ESS) and (ii) *trans*-regulatory elements: SR proteins and hnRNPs.

In general, these splicing regulatory elements (SREs) function by recruiting *trans*-acting

splicing factors that activate or suppress splice site recognition or spliceosome assembly by various

mechanisms (Wang and Burge 2008). ESEs include a diverse range of sequences, are usually located

downstream of suboptimal 3′ SS and regulate splicing by binding to SR proteins. This is achieved by

the interaction of C-terminal RS domains of SR proteins with ESE and mediate protein–protein interactions that facilitate spliceosome assembly (Graveley et al. 1998). ESSs are often bound by splicing repressors of the hnRNP class, a diverse group of proteins containing one or more RNA-binding domains and sometimes splicing inhibitory domains such as glycine-rich motifs (Cartegni et al. 2002). A silencer acts in antagonistic manner to the nearby ESE or by recruiting factors that interfere with the splicing machinery by steric hindrance or through exon looping out of the pre-mRNA (Gaunitz et al. 2004). The G triplet (GGG) is a well characterized ISE that often occur in clusters and can enhance recognition of adjacent 5′ SS or 3′ SS. Intronic CA repeats are known to enhance splicing of upstream exons, probably through binding of hnRNP L (Hui et al. 2005). ISSs are shown to bind splicing repressors like polypyrimidine tract binding protein (PTB) and heterogeneous ribonucleoprotein A1 (hnRNP A1) (Matlin et al. 2005).

1. H.3: *Trans*-regulatory factors.

In addition to the snRNPs and splicing factors that are important in the assembly of spliceosome two families of RNA binding proteins: (i) the serine-arginine rich proteins (SR proteins) and (ii) the heterogeneous nuclear ribonucleoproteins (hnRNP factors), are the main components of distinct regulatory complexes required for the process of splicing (Chen and Manley 2009). Serine-arginine (SR) proteins, also called SR splicing factors (SRSFs) are families of structurally related RNA binding proteins, highly conserved in metazoan cells (Mayeda et al. 1999). They contain one or two N-terminal RNA-recognition motifs (RRM) and a specific C-terminal domain rich in repeating arginines and serines, the 'RS' domain (Zhou and Fu 2013). The RRMs mediate sequence-specific binding to the RNA, and determine substrate specificity whereas the RS domain is involved mainly in protein-protein interactions (Long and Caceres 2009). SR proteins function by their ability to bind

ESEs and through their RS domain to recruit and stabilize U1 snRNP and U2AF binding to the 5′ and 3′ splice site respectively and, thus, facilitating the recruitment of the spliceosome (Corrionero et al. 2011). SR protein activity is regulated through phosphorylation and dephosphorylation of the RS domains, during the spliceosome maturation by several protein families such as: Serine/Arginine protein kinase (SRPKs), the CDC2-like kinase family (CLKs) and the AKT family (Zhou and Fu 2013). This post-translational modification is a crucial step for organization of splicing inside the nucleus by affecting the RNA-binding activity and sub-nuclear localization of the SR proteins (Misteli and Spector 1997).

Heterogeneous nuclear ribonucleoproteins (hnRNP) proteins are pre-mRNA/mRNA-binding proteins that associate with these transcripts and profoundly influence their function and fate (Dreyfuss et al. 2002). They contain one or more of a small number of RNA-binding motifs. The most common of these are the RNP motifs (RBD, also called RNA-recognition motif; RRM), KH domains and RGG (Arg–Gly–Gly) boxes. hnRNP proteins frequently mediate splicing repression, particularly through binding to splicing silencer elements or by sterical interference with other protein interaction (Jean-Philippe et al. 2013) **(Figure 1.11).**

1. I: Alternative splicing

Alternative splicing is an important mechanism to significantly expand the form and function of the human proteome (Kelemen et al. 2013). There are different ways in which alternative splicing can occur: (i) exon skipping, (ii) alternative 5′ SS, (iii) alternative 3′ SS, (iv) intron retention, and (v) mutually exclusive exons (Cartegni et al. 2002).
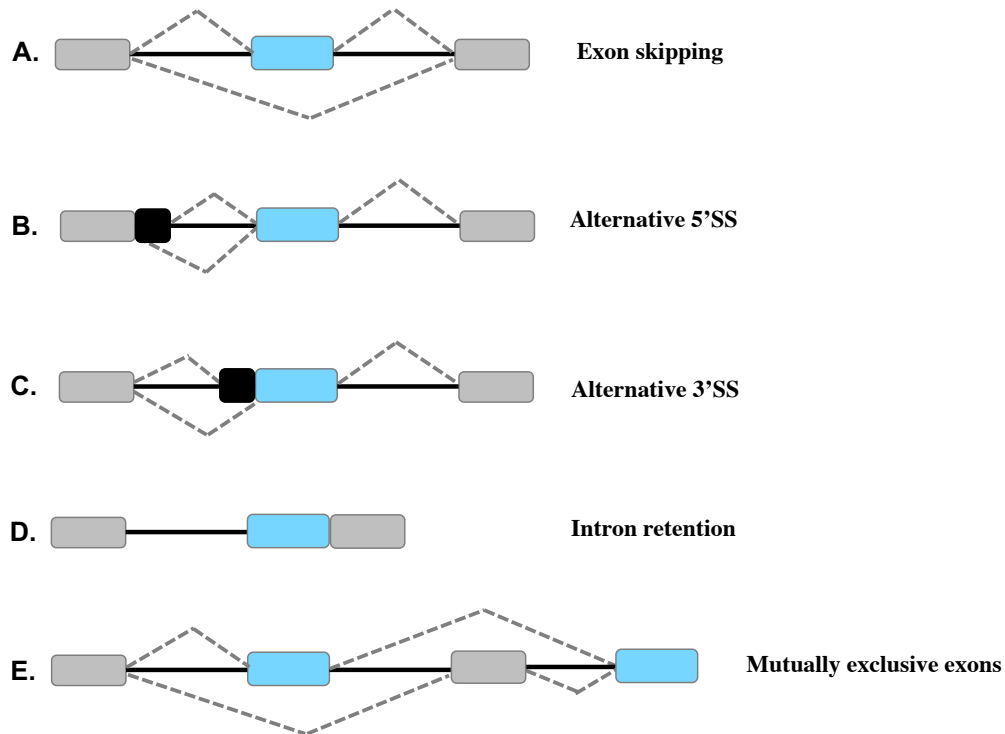
Figure 1.12: Schematic showing different modes of alternative splicing. **A.** exon skipping **B.** alternative 5′ SS **C.** alternative 3′ SS **D.** intron retention **E.** mutually exclusive exons.

1. I.1: Regulation of alternative splicing:

The current ensemble annotations data indicate that, for multiexon protein-coding genes, *Caenorhabditis elegans* has 25% that undergo alternative splicing, *Drosophila* has 45%, mice have 63%, and humans have 88%. The most current estimates, based on RNA-seq data, indicate that >95% of human genes generate at least two alternative pre-mRNA isoforms (Wang et al. 2008; Barberan-Soler et al. 2009; Lewis et al. 2003; Ben-Dov et al. 2008; Pan et al. 2008). Regulated splicing choices are mostly determined by non-splice-site sequences, located in alternative exons or neighboring introns, which recruit special regulatory proteins (Wang and Burge 2008). Non-spliceosomal proteins like hnRNPs, SR proteins (Rooke et al. 2003) and tissue specific RNA binding proteins such as Nova, neuronal PTB/hnRNPI, Rbfox family and muscleblind/CELF family proteins are known to play

important roles in splice site selection and mechanism of alternative splicing (Ule et al. 2005; Jin et al. 2003; Paradis et al. 2007; Wang et al. 2015).

Alternative splicing is not only regulated by splicing factors but also by other processes that involve the transcription machinery. In addition to the splicing process there are the capping and polyadenylation processes that together with the splicing process modify the mRNA respectively at the 5′ ends and the 3′ ends. Changes in alternative splicing can modulate transcript expression levels subjecting mRNAs to nonsense-mediated decay (NMD) by creating a stop codon within the coding sequence or by altering the structure of the gene product that could possible arise due to insertion, or deletion of corresponding sequences (McGlincy and Smith 2008; Boutz et al. 2015; Lee and Rio 2015).

1. I.2: Role of RNA secondary structure in alternative splicing

The first evidence of involvement of secondary structural elements from both exonic and intronic sequences in regulating alternative splicing came from the analysis of the chicken-*Tropomyosin* gene (Libri et al. 1991) and the *Dystrophin* gene (Matsuo et al. 1992). RNA secondary structures which affect the recognition of conserved splice site consensus sequences resulted in the generation of human growth hormone isoforms (A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms). Mutations in the 5′ SS of exon10 of the *Tau* gene affect the formation of stem-loop structure that regulates alternative splicing results in aberrant splicing (Grover et al. 1999). The EDA exon in fibronectin gene contains and ESS that stabilizes the secondary structure of an ESE that is 13nt upstream to allow binding of SR protein to the ESE and regulates its alternative splicing (Muro et al. 1999). Binding of several positive regulators like B52, SRp55, and Nova-1 and negative regulators like hnRNPA1

have been shown to depend on specific nucleotide sequences as well as RNA secondary structures (Buckanovich and Darnell 1997; Shi et al. 1997). Binding of specific proteins to different $(CNG)_n$ trinucleotide repeats studied in the transcripts of 16 genes associated with Triplet Repeat Expansion Diseases (TREDs) resulted in the formation of characteristic hairpin structures (Sobczak et al. 2003). Disruption of secondary structure of a conserved 24 nt stem-loop of ISE element in intron 7 of survival motor of neuron (SMN) pre-mRNA abolishes the binding of the trans factor thereby disrupting the splicing (Miyaso et al. 2003).

In addition to the local RNA secondary structures, long-range structures have been reported to affect alternative splicing and such structures may play a larger role in regulating alternative splicing than currently appreciated (Pervouchine et al. 2012). The most striking example of RNA structure in alternative splicing comes from the drosophila *Dscam* gene that encodes a cell adhesion protein important for neuronal wiring and immune responses in development of fly (Graveley 2005; Pervouchine et al. 2012). Through mutually exclusive splicing of 4 cassette exon clusters (exon 4, 6,9 and 17) the Dscam forms >38,000 isoforms which is approximately three times higher than the number of genes in the fruit fly genome. In support of this, Raker et al. found evidence for a large number of long range structures in Drosophila introns within the exon6 cluster which contains highly conserved (Raker et al. 2009). The exon6 cluster contains 48 alternative exons that are used in mutually exclusive manner. A conserved "docking" site is complementary to a conserved "selector" site upstream from each of the 48 alternative exons and this RNA-RNA base pairing dictates which of 48 alternative exons are used to make the mature Dscam mRNA.
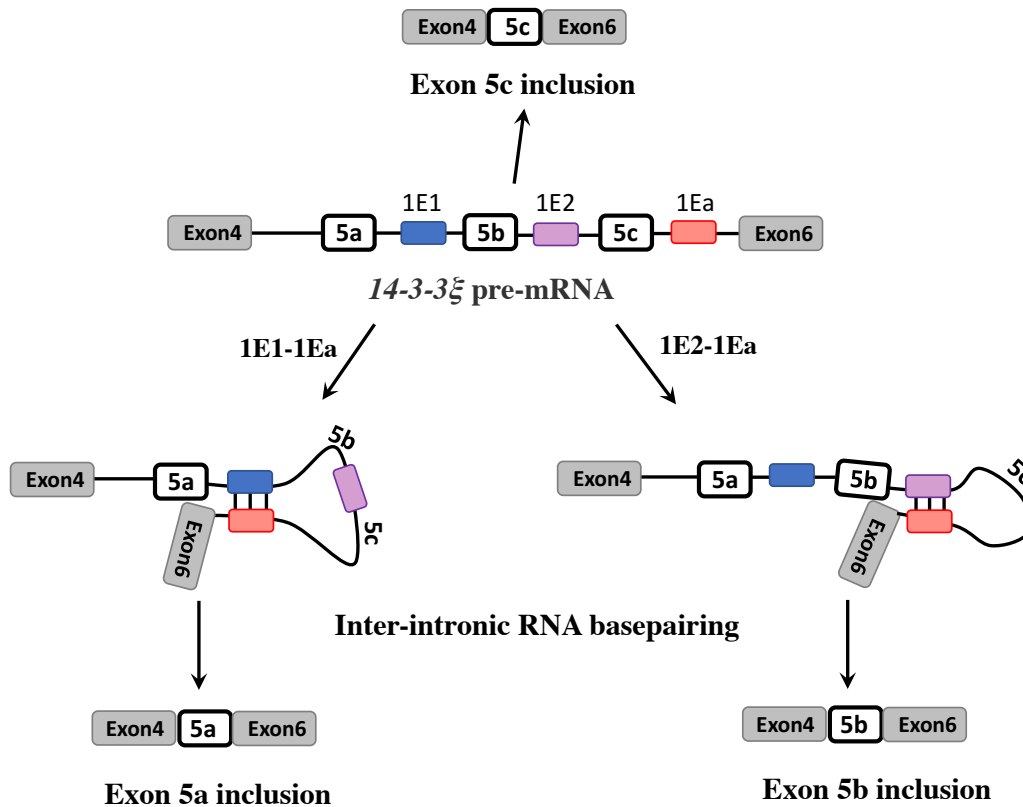
Figure 1.13: Schematic showing regulation of mutually exclusive splicing. *14-3-3ε* pre-mRNA contains intronic element (IE1, IE2) that can compete with the docking site called IEa to form a splicing-activating complex. This inter-intronic RNA basepairing influences the exon5 variant (5a or 5b) that would be included. Conversely, exon 5c would be included if the docking site assumed a linear conformation without specific RNA pairing interactions, whereas exon 5a and exon 5b are not included.

Studies on DSCAM exons4 and 9 and 14-3-3ε pre-mRNA also revealed that inter-intronic RNA pairings that result in RNA loops to ensure the selection of only one of the mutually exclusive exons (Graveley 2005). This is achieved through activation of the proximal variable exon outside the loop and simultaneous repression of the exon within the loop by formation of competitive RNA pairing **(Figure 1.13)**. These findings suggest a broadly applicable mechanism to ensure mutually exclusive splicing (Yang et al. 2011; May et al. 2011; Yue et al. 2013).

A very recent study reported a class of introns that depend upon secondary structure of its RNA for correct splicing. These introns that do not require U2AF2 component of spliceosome, contain simple repeat expansions of complementary AC and GT dimers that occur at the boundary of the intron to form a bridging structure that ensures correct splice site pairing (Lin et al. 2016). Thus, alternative pre-mRNA splicing plays key roles in gene expression and in the diversification of both the transcriptome and the encoded proteome.

1. J: References

Aebi, M., Hornig, H. and Weissmann, C. 1987. 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* 50(2), pp. 237–246.

Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D. and Ansari, A.Z. 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular Cell* 34(3), pp. 387–393.

Barberan-Soler, S., Lambert, N.J. and Zahler, A.M. 2009. Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in C. elegans. *RNA (New York)* 15(9), pp. 1652–1660.

Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* 136(2), pp. 215–233.

Baskerville, S. and Bartel, D.P. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA (New York)* 11(3), pp. 241–247.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., Carter, A.C., Flynn, R.A., Zhou, C., Lim, K.-S., Dedon, P., Wernig, M., Mullen, A.C., Xing, Y., Giallourakis, C.C. and Chang, H.Y. 2014. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* 15(6), pp. 707–719.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P. and Izaurralde, E. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & Development* 20(14), pp. 1885–1898.

Ben-Dov, C., Hartmann, B., Lundgren, J. and Valcárcel, J. 2008. Genome-wide analysis of alternative pre-mRNA splicing. *The Journal of Biological Chemistry* 283(3), pp. 1229–1233.

Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E. and Lai, E.C. 2007. Mammalian mirtron genes. *Molecular Cell* 28(2), pp. 328–336.

Berget, S.M., Moore, C. and Sharp, P.A. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 74(8), pp. 3171–3175.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H. and Lührmann, R. 2010. Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA (New York)* 16(12), pp. 2384–2403.

Beyer, A.L. and Osheim, Y.N. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & Development* 2(6), pp. 754–765.

Borchert, G.M., Lanier, W. and Davidson, B.L. 2006. RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology* 13(12), pp. 1097–1101.

Boutz, P.L., Bhutkar, A. and Sharp, P.A. 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes & Development* 29(1), pp. 63–80.

Breaker, R.R. 2012. Riboswitches and the RNA world. *Cold Spring Harbor Perspectives in Biology* 4(2).

Buckanovich, R.J. and Darnell, R.B. 1997. The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Molecular and Cellular Biology* 17(6), pp. 3194–3201.

Bühler, M., Verdel, A. and Moazed, D. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* 125(5), pp. 873–886.

Butcher, S.E. and Pyle, A.M. 2011. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of Chemical Research* 44(12), pp. 1302–1311.

Cai, X., Hagedorn, C.H. and Cullen, B.R. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA (New York)* 10(12), pp. 1957–1966.

Cao, G., Li, H.-B., Yin, Z. and Flavell, R.A. 2016. Recent advances in dynamic m6A RNA modification. *Open biology* 6(4), p. 160003.

Carmel, L., Rogozin, I.B., Wolf, Y.I. and Koonin, E.V. 2007. Patterns of intron gain and conservation in eukaryotic genes. *BMC Evolutionary Biology* 7, p. 192.

Cartegni, L., Chew, S.L. and Krainer, A.R. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews. Genetics* 3(4), pp. 285–298.

Carthew, R.W. and Sontheimer, E.J. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4), pp. 642–655.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273(5282), pp. 1678–1685.

Chatterjee, S. and Grosshans, H. 2009. Active turnover modulates mature microRNA activity in Caenorhabditis elegans. *Nature* 461(7263), pp. 546–549.

Chendrimada, T.P., Finn, K.J., Ji, X., Baillat, D., Gregory, R.I., Liebhaber, S.A., Pasquinelli, A.E. and Shiekhattar, R. 2007. MicroRNA silencing through RISC recruitment of eIF6. *Nature* 447(7146), pp. 823–828.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K. and Shiekhattar, R. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436(7051), pp. 740–744.

Cheng, C. and Sharp, P.A. 2003. RNA polymerase II accumulation in the promoter-proximal region of the dihydrofolate reductase and gamma-actin genes. *Molecular and Cellular Biology* 23(6), pp. 1961–1967.

Chen, M. and Manley, J.L. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews. Molecular Cell Biology* 10(11), pp. 741–754.

Choi, T., Huang, M., Gorman, C. and Jaenisch, R. 1991. A generic intron increases gene expression in transgenic mice. *Molecular and Cellular Biology* 11(6), pp. 3070–3074.

De Conti, L., Baralle, M. and Buratti, E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley interdisciplinary reviews. RNA* 4(1), pp. 49–60.

Cordes, K.R., Sheehy, N.T., White, M.P., Berry, E.C., Morton, S.U., Muth, A.N., Lee, T.-H., Miano, J.M., Ivey, K.N. and Srivastava, D. 2009. miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature* 460(7256), pp. 705–710.

Corrionero, A., Raker, V.A., Izquierdo, J.M. and Valcárcel, J. 2011. Strict 3' splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease. *RNA (New York)* 17(3), pp. 401–411.

Csuros, M., Rogozin, I.B. and Koonin, E.V. 2011. A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Computational Biology* 7(9), p. e1002150.

Culler, S.J., Hoff, K.G., Voelker, R.B., Berglund, J.A. and Smolke, C.D. 2010. Functional selection and systematic analysis of intronic splicing elements identify active sequence motifs and associated splicing factors. *Nucleic Acids Research* 38(15), pp. 5152–5165.

Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H. and Kjems, J. 2008.

A 5' splice site enhances the recruitment of basal transcription initiation factors in vivo. *Molecular Cell* 29(2), pp. 271–278.

Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C. and Huang, T.H.-M. 2008. The functional consequences of alternative promoter use in mammalian genomes. *Trends in Genetics* 24(4), pp. 167–177.

Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F. and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* 432(7014), pp. 231–235.

Dieci, G., Preti, M. and Montanini, B. 2009. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 94(2), pp. 83–88.

Dreyfuss, G., Kim, V.N. and Kataoka, N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nature Reviews. Molecular Cell Biology* 3(3), pp. 195–205.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., Plath, K. and Guttman, M. 2013. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341(6147), p. 1237973.

Fica, S.M., Mefford, M.A., Piccirilli, J.A. and Staley, J.P. 2014. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nature Structural & Molecular Biology* 21(5), pp. 464–471.

Filipowicz, W. and Pogačić, V. 2002. Biogenesis of small nucleolar ribonucleoproteins. *Current Opinion in Cell Biology* 14(3), pp. 319–327.

Fong, Y.W. and Zhou, Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414(6866), pp. 929–933.

Fourmann, J.-B., Schmitzová, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.-L., Fabrizio, P. and Lührmann, R. 2013. Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes & Development* 27(4), pp. 413–428.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J. and Nagai, K. 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature* 537(7619), pp. 197–201.

Gao, K., Masuda, A., Matsuura, T. and Ohno, K. 2008. Human branch point consensus sequence is yUnAy. *Nucleic Acids Research* 36(7), pp. 2257–2267.

Gaunitz, F., Heise, K. and Gebhardt, R. 2004. A silencer element in the first intron of the glutamine synthetase gene represses induction by glucocorticoids. *Molecular Endocrinology* 18(1), pp. 63–69.

Ghelani, H.S., Rachchh, M.A. and Gokani, R.H. 2012. MicroRNAs as newer therapeutic targets: A big hope from a tiny player. *Journal of pharmacology & pharmacotherapeutics* 3(3), pp. 217–227.

Gilbert, S.D., Rambo, R.P., Van Tyne, D. and Batey, R.T. 2008. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature Structural & Molecular Biology* 15(2), pp. 177–182.

Graveley, B.R. 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123(1), pp. 65–73.

Graveley, B.R., Hertel, K.J. and Maniatis, T. 1998. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *The EMBO Journal* 17(22), pp. 6747–6756.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432(7014), pp. 235–240.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M. and Herrmann, B.G. 2013. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental Cell* 24(2), pp. 206–214.

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K. and Hutton, M. 1999. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *The Journal of Biological Chemistry* 274(21), pp. 15134–15143.

Guil, S. and Cáceres, J.F. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Structural & Molecular Biology* 14(7), pp. 591–596.

Hall, K.B., Green, M.R. and Redfield, A.G. 1988. Structure of a pre-mRNA branch point/3' splice site region. *Proceedings of the National Academy of Sciences of the United States of America* 85(3), pp. 704–708.

Hamer, D.H., Smith, K.D., Boyer, S.H. and Leder, P. 1979. SV40 recombinants carrying rabbit beta-globin gene coding sequences. *Cell* 17(3), pp. 725–735.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H. and Kim, V.N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development* 18(24), pp. 3016–3027.

Havlioglu, N., Wang, J., Fushimi, K., Vibranovski, M.D., Kan, Z., Gish, W., Fedorov, A., Long, M. and Wu, J.Y. 2007. An intronic signal for alternative splicing in the human genome. *Plos One* 2(11), p. e1246.

Hendrix, D.K., Brenner, S.E. and Holbrook, S.R. 2005. RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics* 38(3), pp. 221–243.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J. and Kim, V.N. 2008. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Molecular Cell* 32(2), pp. 276–284.

Heo, J.B. and Sung, S. 2011. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331(6013), pp. 76–79.

Hicks, M.J., Mueller, W.F., Shepard, P.J. and Hertel, K.J. 2010. Competing upstream 5' splice sites enhance the rate of proximal splicing. *Molecular and Cellular Biology* 30(8), pp. 1878–1886.

Le Hir, H., Nott, A. and Moore, M.J. 2003. How introns influence and enhance eukaryotic gene expression. *Trends in Biochemical Sciences* 28(4), pp. 215–220.

Horowitz, D.S. and Krainer, A.R. 1994. Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends in Genetics* 10(3), pp. 100–106.

Hoskins, A.A., Friedman, L.J., Gallagher, S.S., Crawford, D.J., Anderson, E.G., Wombacher, R., Ramirez, N., Cornish, V.W., Gelles, J. and Moore, M.J. 2011. Ordered and dynamic assembly of single spliceosomes. *Science* 331(6022), pp. 1289–1295.

Huang, X., Ding, L., Bennewith, K.L., Tong, R.T., Welford, S.M., Ang, K.K., Story, M., Le, Q.-T. and Giaccia, A.J. 2009. Hypoxia-inducible mir-210 regulates normoxic gene expression involved in tumor initiation. *Molecular Cell* 35(6), pp. 856–867.

Hui, J., Hung, L.-H., Heiner, M., Schreiner, S., Neumüller, N., Reither, G., Haas, S.A. and Bindereif, A. 2005. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *The EMBO Journal* 24(11), pp. 1988–1998.

Hwang, H.-W., Wentzel, E.A. and Mendell, J.T. 2007. A hexanucleotide element directs microRNA nuclear import. *Science* 315(5808), pp. 97–100.

Ilagan, J.O., Chalkley, R.J., Burlingame, A.L. and Jurica, M.S. 2013. Rearrangements within human spliceosomes captured after exon ligation. *RNA (New York)* 19(3), pp. 400–412.

Jacobs, E., Mills, J.D. and Janitz, M. 2012. The role of RNA structure in posttranscriptional regulation of gene expression. *Journal of Genetics and Genomics = Yi Chuan Xue Bao* 39(10), pp. 535–543.

Jean-Philippe, J., Paz, S. and Caputi, M. 2013. hnRNP A1: the Swiss army knife of gene expression. *International Journal of Molecular Sciences* 14(9), pp. 18999–19024.

Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K. and Inoue, K. 2003. A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *The EMBO Journal* 22(4), pp. 905–912.

Jo, B.-S. and Choi, S.S. 2015. Introns: the functional benefits of introns in genomes. *Genomics & informatics* 13(4), pp. 112–118.

Junttila, M.R. and Evan, G.I. 2009. p53--a Jack of all trades but master of none. *Nature Reviews. Cancer* 9(11), pp. 821–829.

Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhattar, R. and Nishikura, K. 2007. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Reports* 8(8), pp. 763–769.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M. and Stamm, S. 2013. Function of alternative splicing. *Gene* 514(1), pp. 1–30.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y. and Segal, E. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467(7311), pp. 103–107.

Kim, E., Magen, A. and Ast, G. 2007. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research* 35(1), pp. 125–131.

Kim, H., Klein, R., Majewski, J. and Ott, J. 2004. Estimating rates of alternative splicing in mammals and invertebrates. *Nature Genetics* 36(9), pp. 915–6; author reply 916.

Kim, J., Inoue, K., Ishii, J., Vanti, W.B., Voronov, S.V., Murchison, E., Hannon, G. and Abeliovich, A. 2007. A MicroRNA feedback circuit in midbrain dopamine neurons. *Science* 317(5842), pp. 1220–1224.

Kim, Y.-K. and Kim, V.N. 2007. Processing of intronic microRNAs. *The EMBO Journal* 26(3), pp. 775–783.

Klaff, P., Riesner, D. and Steger, G. 1996. RNA structure and the regulation of gene expression. *Plant Molecular Biology* 32(1-2), pp. 89–106.

Klattenhoff, C.A., Scheuermann, J.C., Surface, L.E., Bradley, R.K., Fields, P.A., Steinhauser, M.L., Ding, H., Butty, V.L., Torrey, L., Haas, S., Abo, R., Tabebordbar, M., Lee, R.T., Burge, C.B. and Boyer, L.A. 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152(3), pp. 570–583.

Krol, J., Loedige, I. and Filipowicz, W. 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews. Genetics* 11(9), pp. 597–610.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* 31(1), pp. 147–157.

Kwek, K.Y., Murphy, S., Furger, A., Thomas, B., O'Gorman, W., Kimura, H., Proudfoot, N.J. and Akoulitchev, A. 2002. U1 snRNA associates with TFIIH and regulates transcriptional initiation.

Nature Structural Biology 9(11), pp. 800–805.

Laing, C., Jung, S., Iqbal, A. and Schlick, T. 2009. Tertiary motifs revealed in analyses of higher-order RNA junctions. *Journal of Molecular Biology* 393(1), pp. 67–82.

Lane, N. and Martin, W. 2010. The energetics of genome complexity. *Nature* 467(7318), pp. 929–934.

Langford, C.J., Klinz, F.J., Donath, C. and Gallwitz, D. 1984. Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell* 36(3), pp. 645–653.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H. and Kim, V.N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal* 23(20), pp. 4051–4060.

Lee, Y. and Rio, D.C. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual Review of Biochemistry* 84, pp. 291–323.

LeMaire, M.F. and Thummel, C.S. 1990. Splicing precedes polyadenylation during Drosophila E74A transcription. *Molecular and Cellular Biology* 10(11), pp. 6059–6063.

Lemay, J.-F. and Lafontaine, D.A. 2007. Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA (New York)* 13(3), pp. 339–350.

Lewis, B.P., Burge, C.B. and Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120(1), pp. 15–20.

Lewis, B.P., Green, R.E. and Brenner, S.E. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America* 100(1), pp. 189–192.

Libri, D., Piseri, A. and Fiszman, M.Y. 1991. Tissue-specific splicing in vivo of the beta-tropomyosin gene: dependence on an RNA secondary structure. *Science* 252(5014), pp. 1842–1845.

Lin, C.-C., Liu, L.-Z., Addison, J.B., Wonderlin, W.F., Ivanov, A.V. and Ruppert, J.M. 2011. A KLF4-miRNA-206 autoregulatory feedback loop can promote or inhibit protein translation depending upon cell context. *Molecular and Cellular Biology* 31(12), pp. 2513–2527.

Lin, C.-L., Taggart, A.J., Lim, K.H., Cygan, K.J., Ferraris, L., Creton, R., Huang, Y.-T. and Fairbrother, W.G. 2016. RNA structure replaces the need for U2AF2 in splicing. *Genome Research* 26(1), pp. 12–23.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S. and Fu, X.-D. 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nature Structural & Molecular Biology*

15(8), pp. 819–826.

Li, S. and Breaker, R.R. 2013. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Research* 41(5), pp. 3022–3031.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J. and Parker, R. 2005. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nature Cell Biology* 7(7), pp. 719–723.

Liu, Y.P., Schopman, N.C.T. and Berkhout, B. 2013. Dicer-independent processing of short hairpin RNAs. *Nucleic Acids Research* 41(6), pp. 3723–3733.

Long, J.C. and Caceres, J.F. 2009. The SR protein family of splicing factors: master regulators of gene expression. *The Biochemical Journal* 417(1), pp. 15–27.

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. and Guigó, R. 2005. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters* 579(9), pp. 1900–1903.

Louro, R., Smirnova, A.S. and Verjovski-Almeida, S. 2009. Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 93(4), pp. 291–298.

Lu, C., Smith, A.M., Fuchs, R.T., Ding, F., Rajashankar, K., Henkin, T.M. and Ke, A. 2008. Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nature Structural & Molecular Biology* 15(10), pp. 1076–1083.

Lynch, K.W. 2007. Regulation of alternative splicing by signal transduction pathways. *Advances in Experimental Medicine and Biology* 623, pp. 161–174.

Lytle, J.R., Yario, T.A. and Steitz, J.A. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proceedings of the National Academy of Sciences of the United States of America* 104(23), pp. 9667–9672.

Makarova, J.A. and Kramerov, D.A. 2009. Analysis of C/D box snoRNA genes in vertebrates: The number of copies decreases in placental mammals. *Genomics* 94(1), pp. 11–19.

Mandal, M. and Breaker, R.R. 2004. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Structural & Molecular Biology* 11(1), pp. 29–35.

Marcia, M., Somarowthu, S. and Pyle, A.M. 2013. Now on display: a gallery of group II intron structures at different stages of catalysis. *Mobile DNA* 4(1), p. 14.

Matlin, A.J., Clark, F. and Smith, C.W.J. 2005. Understanding alternative splicing: towards a cellular code. *Nature Reviews. Molecular Cell Biology* 6(5), pp. 386–398.

Matsuo, M., Nishio, H., Kitoh, Y., Francke, U. and Nakamura, H. 1992. Partial deletion of a dystrophin gene leads to exon skipping and to loss of an intra-exon hairpin structure from the predicted mRNA precursor. *Biochemical and Biophysical Research Communications* 182(2), pp. 495–500.

Mayeda, A., Screaton, G.R., Chandler, S.D., Fu, X.D. and Krainer, A.R. 1999. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Molecular and Cellular Biology* 19(3), pp. 1853–1863.

May, G.E., Olson, S., McManus, C.J. and Graveley, B.R. 2011. Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA (New York)* 17(2), pp. 222–229.

McGlincy, N.J. and Smith, C.W.J. 2008. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in Biochemical Sciences* 33(8), pp. 385–393.

McManus, C.J. and Graveley, B.R. 2011. RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development* 21(4), pp. 373–379.

Michlewski, G. and Cáceres, J.F. 2010. Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nature Structural & Molecular Biology* 17(8), pp. 1011–1018.

Millevoi, S. and Vagner, S. 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research* 38(9), pp. 2757–2774.

Misteli, T. and Spector, D.L. 1997. Protein phosphorylation and the nuclear organization of pre-mRNA splicing. *Trends in Cell Biology* 7(4), pp. 135–138.

Miyaso, H., Okumura, M., Kondo, S., Higashide, S., Miyajima, H. and Imaizumi, K. 2003. An intronic splicing enhancer element in survival motor neuron (SMN) pre-mRNA. *The Journal of Biological Chemistry* 278(18), pp. 15825–15831.

Montange, R.K. and Batey, R.T. 2006. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* 441(7097), pp. 1172–1175.

Moore, M.J. and Proudfoot, N.J. 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136(4), pp. 688–700.

Mortimer, S.A., Kidwell, M.A. and Doudna, J.A. 2014. Insights into RNA structure and function from genome-wide studies. *Nature Reviews. Genetics* 15(7), pp. 469–479.

Mortimer, S.A. and Weeks, K.M. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *Journal of the American Chemical Society* 129(14), pp. 4144–4145.

Muro, A.F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E. and Baralle, F.E. 1999. Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Molecular and Cellular Biology* 19(4), pp. 2657–2671.

Nam, Y., Chen, C., Gregory, R.I., Chou, J.J. and Sliz, P. 2011. Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* 147(5), pp. 1080–1091.

Newman, M.A., Thomson, J.M. and Hammond, S.M. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA (New York)* 14(8), pp. 1539–1549.

Nishimura, R., Hayashi, M., Wu, G.-J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M., Nagasawa, M., Harada, K. and Kawaguchi, M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. *Nature* 420(6914), pp. 426–429.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M. and Lai, E.C. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* 130(1), pp. 89–100.

Onoa, B. and Tinoco, I. 2004. RNA folding and unfolding. *Current Opinion in Structural Biology* 14(3), pp. 374–379.

Ozsolak, F., Song, J.S., Liu, X.S. and Fisher, D.E. 2007. High-throughput mapping of the chromatin structure of human promoters. *Nature Biotechnology* 25(2), pp. 244–248.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* 40(12), pp. 1413–1415.

Paradis, C., Cloutier, P., Shkreta, L., Toutant, J., Klarskov, K. and Chabot, B. 2007. hnRNP I/PTB can antagonize the splicing repressor activity of SRp30c. *RNA (New York)* 13(8), pp. 1287–1300.

Park, J.-E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J. and Kim, V.N. 2011. Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* 475(7355), pp. 201–205.

Pervouchine, D.D., Khrameeva, E.E., Pichugina, M.Y., Nikolaienko, O.V., Gelfand, M.S., Rubtsov, P.M. and Mironov, A.A. 2012. Evidence for widespread association of mammalian splicing and conserved long-range RNA structures. *RNA (New York)* 18(1), pp. 1–15.

Pillai, R.S., Bhattacharyya, S.N. and Filipowicz, W. 2007. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends in Cell Biology* 17(3), pp. 118–126.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A. and Panning, B. 2002. Xist RNA and the mechanism of X chromosome inactivation. *Annual Review of Genetics* 36, pp. 233–278.

Proudfoot, N.J. 2011. Ending the message: poly(A) signals then and now. *Genes & Development* 25(17), pp. 1770–1782.

Puvvula, P.K., Desetty, R.D., Pineau, P., Marchio, A., Moon, A., Dejean, A. and Bischof, O. 2014. Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nature Communications* 5, p. 5323.

Qi, P., Zhou, X.-Y. and Du, X. 2016. Circulating long non-coding RNAs in cancer: current status and future perspectives. *Molecular Cancer* 15(1), p. 39.

Raghunathan, P.L. and Guthrie, C. 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Current Biology* 8(15), pp. 847–855.

Rajewsky, N. 2006. microRNA target predictions in animals. *Nature Genetics* 38 Suppl, pp. S8–13.

Raker, V.A., Mironov, A.A., Gelfand, M.S. and Pervouchine, D.D. 2009. Modulation of alternative splicing by long-range RNA structures in Drosophila. *Nucleic Acids Research* 37(14), pp. 4533–4544.

Ray, P.S., Jia, J., Yao, P., Majumder, M., Hatzoglou, M. and Fox, P.L. 2009. A stress-responsive RNA switch regulates VEGFA expression. *Nature* 457(7231), pp. 915–919.

Rearick, D., Prakash, A., McSweeny, A., Shepard, S.S., Fedorova, L. and Fedorov, A. 2011. Critical association of ncRNA with introns. *Nucleic Acids Research* 39(6), pp. 2357–2366.

Reed, R. 2000. Mechanisms of fidelity in pre-mRNA splicing. *Current Opinion in Cell Biology* 12(3), pp. 340–345.

Reed, R. and Maniatis, T. 1988. The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes & Development* 2(10), pp. 1268–1276.

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D. and Izaurralde, E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA (New York)* 11(11), pp. 1640–1647.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. and Chang, H.Y. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129(7), pp. 1311–1323.

Rino, J. and Carmo-Fonseca, M. 2009. The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends in Cell Biology* 19(8), pp. 375–384.

Rooke, N., Markovtsov, V., Cagavi, E. and Black, D.L. 2003. Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. *Molecular and Cellular Biology* 23(6), pp. 1874–1884.

Rose, A.B., Elfersi, T., Parra, G. and Korf, I. 2008. Promoter-proximal introns in Arabidopsis thaliana are enriched in dispersed signals that elevate gene expression. *The Plant Cell* 20(3), pp. 543–551.

Rueter, S.M., Dawson, T.R. and Emeson, R.B. 1999. Regulation of alternative splicing by RNA editing. *Nature* 399, pp.75-80

Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Lührmann, R. and Stark, H.

2006. Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Molecular Cell* 24(2), pp. 267–278.

Scadden, A.D.J. 2005. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. *Nature Structural & Molecular Biology* 12(6), pp. 489–496.

Schroeder, R., Grossberger, R., Pichler, A. and Waldsich, C. 2002. RNA folding in vivo. *Current Opinion in Structural Biology* 12(3), pp. 296–300.

Sen, G.L. and Blau, H.M. 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature Cell Biology* 7(6), pp. 633–636.

Sengupta, S. and Harris, C.C. 2005. p53: traffic cop at the crossroads of DNA repair and recombination. *Nature Reviews. Molecular Cell Biology* 6(1), pp. 44–55.

Shi, H. and He, C. 2016. A glance at N(6)-methyladenosine in transcript isoforms. *Nature Methods* 13(8), pp. 624–625.

Shi, H., Hoffman, B.E. and Lis, J.T. 1997. A specific RNA hairpin loop structure binds the RNA recognition motifs of the Drosophila SR protein B52. *Molecular and Cellular Biology* 17(5), pp. 2649–2657.

Shimada, Y., Mohn, F. and Bühler, M. 2016. The RNA-induced transcriptional silencing complex targets chromatin exclusively via interacting with nascent transcripts. *Genes & Development* 30(23), pp. 2571–2580.

Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. and Krzyzosiak, W.J. 2003. RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Research* 31(19), pp. 5469–5482.

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F. and Pyle, A.M. 2015. HOTAIR forms an intricate and modular secondary structure. *Molecular Cell* 58(2), pp. 353–361.

Spiluttini, B., Gu, B., Belagal, P., Smirnova, A.S., Nguyen, V.T., Hébert, C., Schmidt, U., Bertrand, E., Darzacq, X. and Bensaude, O. 2010. Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *Journal of Cell Science* 123(Pt 12), pp. 2085–2093.

Spizzo, R., Nicoloso, M.S., Lupini, L., Lu, Y., Fogarty, J., Rossi, S., Zagatti, B., Fabbri, M., Veronese, A., Liu, X., Davuluri, R., Croce, C.M., Mills, G., Negrini, M. and Calin, G.A. 2010. miR-145 participates with TP53 in a death-promoting regulatory loop and targets estrogen receptor-alpha in human breast cancer cells. *Cell Death and Differentiation* 17(2), pp. 246–254.

Sterner, D.A., Carlo, T. and Berget, S.M. 1996. Architectural limits on split genes. *Proceedings of the National Academy of Sciences of the United States of America* 93(26), pp. 15081–15085.

Suzuki, H.I., Yamagata, K., Sugimoto, K., Iwamoto, T., Kato, S. and Miyazono, K. 2009. Modulation of microRNA processing by p53. *Nature* 460(7254), pp. 529–533.

Swinburne, I.A. and Silver, P.A. 2008. Intron delays and transcriptional timing during development. *Developmental Cell* 14(3), pp. 324–330.

Tahira, A.C., Kubrusly, M.S., Faria, M.F., Dazzani, B., Fonseca, R.S., Maracaja-Coutinho, V., Verjovski-Almeida, S., Machado, M.C.C. and Reis, E.M. 2011. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. *Molecular Cancer* 10, p. 141.

Tattermusch, A. and Brockdorff, N. 2011. A scaffold for X chromosome inactivation. *Human Genetics* 130(2), pp. 247–253.

Thummel, C.S., Burtis, K.C. and Hogness, D.S. 1990. Spatial and temporal patterns of E74 transcription during Drosophila development. *Cell* 61(1), pp. 101–111.

Tinoco, I. and Bustamante, C. 1999. How RNA folds. *Journal of Molecular Biology* 293(2), pp. 271–281.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R. and Rosenfeld, M.G. 2009. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459(7249), pp. 1010–1014.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., Blencowe, B.J., Prasanth, S.G. and Prasanth, K.V. 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell* 39(6), pp. 925–938.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E. and Chang, H.Y. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329(5992), pp. 689–693.

Tsunetsugu-Yokota, Y. and Yamamoto, T. 2010. Mammalian MicroRNAs: Post-Transcriptional Gene Regulation in RNA Virus Infection and Therapeutic Applications. *Frontiers in microbiology* 1, p. 108.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.-S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., Zeeberg, B.R., Kane, D., Weinstein, J.N., Blume, J. and Darnell, R.B. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nature Genetics* 37(8), pp. 844–852.

Underwood, J.G., Uzilov, A.V., Katzman, S., Onodera, C.S., Mainzer, J.E., Mathews, D.H., Lowe, T.M., Salama, S.R. and Haussler, D. 2010. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nature Methods* 7(12), pp. 995–1001.

Vasudevan, S., Tong, Y. and Steitz, J.A. 2007. Switching from repression to activation:

microRNAs can up-regulate translation. *Science* 318(5858), pp. 1931–1934.

Viereck, J., Kumarswamy, R., Foinquinos, A., Xiao, K., Avramopoulos, P., Kunz, M., Dittrich, M., Maetzig, T., Zimmer, K., Remke, J., Just, A., Fendrich, J., Scherf, K., Bolesani, E., Schambach, A., Weidemann, F., Zweigerdt, R., de Windt, L.J., Engelhardt, S., Dandekar, T., Batkai, S. and Thum, T. 2016. Long noncoding RNA Chast promotes cardiac remodeling. *Science Translational Medicine* 8(326), p. 326ra22.

Vinogradov, A.E. 2004. Compactness of human housekeeping genes: selection for economy or genomic design? *Trends in Genetics* 20(5), pp. 248–253.

Wagner, E.J. and Garcia-Blanco, M.A. 2001. Polypyrimidine tract binding protein antagonizes exon definition. *Molecular and Cellular Biology* 21(10), pp. 3281–3288.

Wahl, M.C., Will, C.L. and Lührmann, R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* 136(4), pp. 701–718.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456(7221), pp. 470–476.

Wang, G.-S. and Cooper, T.A. 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews*. *Genetics* 8(10), pp. 749–761.

Wang, K.C. and Chang, H.Y. 2011. Molecular mechanisms of long noncoding RNAs. *Molecular Cell* 43(6), pp. 904–914.

Wang, Y., Liu, J., Huang, B.O., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M. and Wang, X.-Z. 2015. Mechanism of alternative splicing and its regulation. *Biomedical reports* 3(2), pp. 152–158.

Wang, Z. and Burge, C.B. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA (New York)* 14(5), pp. 802–813.

Wan, Y., Qu, K., Ouyang, Z. and Chang, H.Y. 2013. Genome-wide mapping of RNA structure using nuclease digestion and high-throughput sequencing. *Nature Protocols* 8(5), pp. 849–869.

Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E. and Chang, H.Y. 2014. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505(7485), pp. 706–709.

Wei, M., Zhao, X., Liu, M., Niu, M., Seif, E. and Kleiman, L. 2016. Export of Precursor tRNAIle from the Nucleus to the Cytoplasm in Human Cells. *Plos One* 11(4), p. e0154044.

Welker, N.C., Pavelec, D.M., Nix, D.A., Duchaine, T.F., Kennedy, S. and Bass, B.L. 2010. Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs.

RNA (New York) 16(5), pp. 893–903.

Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C. and Weeks, K.M. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology* 6(4), p. e96.

Will, C.L. and Lührmann, R. 2011. Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology* 3(7).

Wu, L., Fan, J. and Belasco, J.G. 2006. MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America* 103(11), pp. 4034–4039.

Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C.L. and Shi, Y. 2015. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* 349(6253), pp. 1182–1191.

Yan, C., Wan, R., Bai, R., Huang, G. and Shi, Y. 2017. Structure of a yeast step II catalytically activated spliceosome. *Science* 355(6321), pp. 149–155.

Yang, Y., Zhan, L., Zhang, W., Sun, F., Wang, W., Tian, N., Bi, J., Wang, H., Shi, D., Jiang, Y., Zhang, Y. and Jin, Y. 2011. RNA secondary structure in mutually exclusive splicing. *Nature Structural & Molecular Biology* 18(2), pp. 159–168.

Yi, R., Qin, Y., Macara, I.G. and Cullen, B.R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development* 17(24), pp. 3011–3016.

Yue, Y., Li, G., Yang, Y., Zhang, W., Pan, H., Chen, R., Shi, F. and Jin, Y. 2013. Regulation of Dscam exon 17 alternative splicing by steric hindrance in combination with RNA secondary structures. *RNA Biology* 10(12), pp. 1822–1833.

Zhang, G., Li, X., Cao, H., Zhao, H. and Geller, A.I. 2011. The vesicular glutamate transporter-1 upstream promoter and first intron each support glutamatergic-specific expression in rat postrhinal cortex. *Brain Research* 1377, pp. 1–12.

Zhang, X., Wan, G., Berger, F.G., He, X. and Lu, X. 2011. The ATM kinase induces microRNA biogenesis in the DNA damage response. *Molecular Cell* 41(4), pp. 371–383.

Zhou, K.I., Parisien, M., Dai, Q., Liu, N., Diatchenko, L., Sachleben, J.R. and Pan, T. 2016. N(6)-Methyladenosine Modification in a Long Noncoding RNA Hairpin Predisposes Its Conformation to Protein Binding. *Journal of Molecular Biology* 428(5 Pt A), pp. 822–833.

Zhou, X., Ruan, J., Wang, G. and Zhang, W. 2006. Computational Characterization and Identification of Core Promoters of MicroRNA Genes in C. elegans, H. sapiens and A. thaliana. In: Eskin, E., Ideker, T., Raphael, B., and Workman, C. eds. *Systems biology and regulatory genomics*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 235–248.

Zhou, Z. and Fu, X.-D. 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 122(3), pp. 191–207.

Zhuang, Y.A., Goldstein, A.M. and Weiner, A.M. 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America* 86(8), pp. 2752–2756.

CHAPTER 2

TERTIARY STRUCTURE OF PRI-MIR-17-92A AUTOREGULATES ITS PROCESSING

2. A: Introduction

MicroRNAs (miRNAs) are short 20–22 nucleotide RNA molecules that function as regulators of gene expression in eukaryotes at the post transcriptional level (Filipowicz et al. 2008). RNA-mediated gene silencing pathways have essential roles in development, cell differentiation, cell proliferation, cell apoptosis, chromosome structure, immunity and metabolism (Hammond and Wood 2011; Bushati and Cohen 2007; Schratt 2009). Mature miRNAs target specific mRNA by binding to the complementary sequences on the 3′ UTR, and regulate gene expression via diverse mechanisms ranging from mRNA cleavage, to translational repression and heterochromatin formation (Valinezhad Orang et al. 2014; Breving and Esquela-Kerscher 2010; Abbott et al. 2005; Chen and Rajewsky 2007) They show unique spatio-temporal expression patterns owing to their extensive control over the transcriptome.

Misregulation of several miRNAs have been implicated in several diseases including cancers, cardiovascular and neurodegenerative diseases (Hammond 2006). Given that miRNAs are crucial across wide range of biological pathways, their expression requires stringent control. A large number of studies have been directed at understanding the processing of mature miRNAs and the mechanism of target recognition (Krol et al. 2010; Valinezhad Orang et al. 2014). Such investigations have primarily focused on monocistronic transcripts containing a single pre-miRNA hairpin, which are processed in two steps into pre-miRNA and mature miRNA (Davis and Hata

2009; Michlewski et al. 2008). This essentially requires a few trans-acting proteins in addition to the Microprocessor complex (MPC) for efficient primary transcript processing (Davis et. al. 2008; Michlewski et al. 2008; Viswanathan et. al. 2010).

It is known that primary miRNA gene loci are intergenic, intronic or polycistronic and, in vertebrates, 40% of miRNAs are present as a part of polycistronic cluster, i.e. two or more miRNAs are present in tandem on the primary transcript (Megraw et al. 2007; Griffiths-Jones et al. 2008). The clustering patterns suggest that miRNAs in the same cluster are transcribed from a single polycistronic unit (Saini et al. 2008; Yang et al. 2006; Ryazansky et al. 2011), similar to the operon regulation systems in prokaryotes (Lawrence 1999; Price et al. 2005). As genes located in the same operon often have related functions (Jacob and Monod 1961), miRNAs in the same cluster were hypothesized to regulate overlapping or complementary sets of targets to elicit a biological response (Ventura et al. 2008; Kim et al. 2009). However, there is mounting evidence that although the miRNAs from a cluster are processed from a common polycistronic RNA, their biogenesis can be differentially regulated and give rise to changes in relative levels of the mature miRNAs in a tissue specific manner (Tang and Maxwell 2008; Thomson et al. 2006). Such differences in the abundances of the component miRNAs from a cluster have been ascribed to tissue specific differences in transcription, differential pri-miRNA processing or differential stabilities of mature miRs (Bail et al. 2010; Tahira et al. 2011). But the mechanisms to explain how this differential processing is achieved are not yet clear. Ribosomal RNAs in the ribosome and snRNAs in the spliceosome are well established examples of RNA scaffolds that undergoes structural remodeling to enable multiprotein recognition in the course of their assembly (Nissen et al. 2001; Filipowicz and Pogacić 2002; Talkington et al. 2005). We considered a possibility that polycistronic miRNA

transcripts might self-orchestrate the binding of various proteins required for the processing of the microRNAs that result in their differential levels.

We therefore chose a human polycistronic pri-miR-17-92a cluster which resides in intron3 of a ~7 kb primary transcript known as *C13orf25* (Ota et al. 2004) present on the genomic locus of chromosome 13 (13q31.3) to address such a differential processing (**Figure 2.1**).
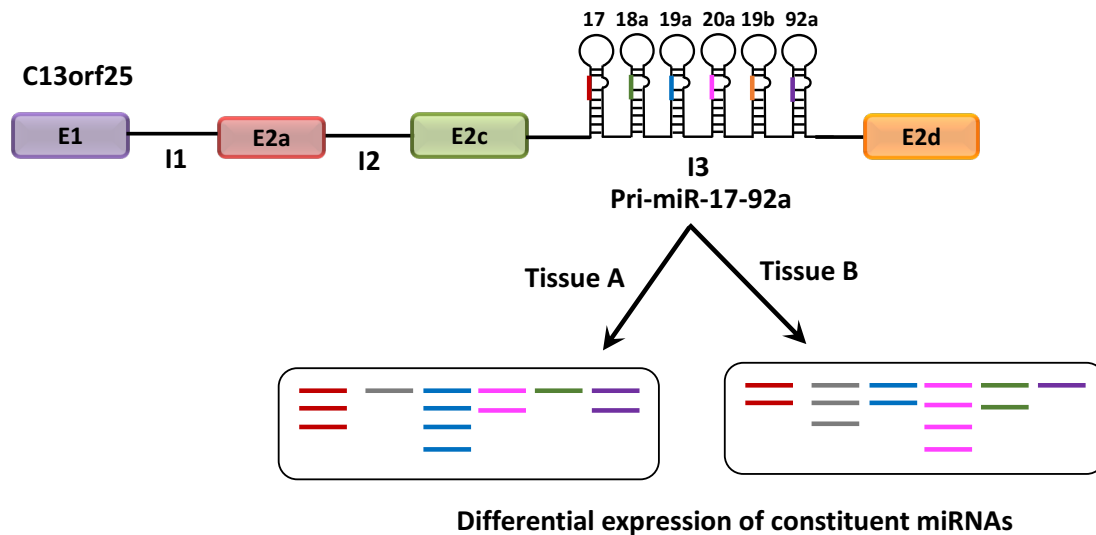
Figure 2.1: Schematic representation of genomic locus of pri-miR-17-92a cluster. The cluster resides in intron 3 (I3) of the gene C13orf25. It contains six stem loop structures which gives six individual microRNAs (17, 18a, 19a, 20a, 19b, 92a). It undergoes differential processing to yield component miRNAs in different amounts in different tissues (A & B).

The 0.8 kb pri-miR-17-92a undergoes processing to give six individual mature miRs i.e. miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1, and miR-92a (Mendell 2008). The organization of this cluster and the sequences of the mature miRs are highly conserved in all vertebrates, indicating the evolutionary pressure to preserve them in clusters. Despite the high degree of conservation of the miRNA sequences, the exonic sequences of C13orf25 are not conserved between species, suggesting that the sole function of this transcript is to produce these miRNAs. MicroRNAs

(miRNAs) encoded by the pri-miR-17-92a cluster and its paralogs are known to act as oncogenes. Expression of miRNAs from this cluster are known to promote cell proliferation and suppress apoptosis of cancer cells, tumor angiogenesis and metastasis (Ota et al. 2004; Zhang et al. 2007) Further, these miRs are also thought to be important for the development of cardiovascular system, hematopoiesis and the immune system (Li et al. 2012). Since it encodes miRNAs of such diverse functions it is very important to understand how the pri-miR-17-92a cluster is regulated. It has been reported that the expression of miRs from this cluster is regulated transcriptionally by c-Myc, however a few studies have indicated that post transcriptional regulation also plays a role in achieving this differential expression (Tang and Maxwell 2008; Thomson et al. 2006; O'Donnell et al. 2005)

Pri-miR-17-92a is represented as a series of six stem-loops indicating the presence of a secondary structure. Structural studies indicate Drosha-DGCR8 (Microprocessor, MPC) binding requires a single strand-double strand junction (ss-ds) for recognition and cropping (Han et al. 2004; Gregory et al. 2004; Denli et al. 2004). If this was the case then all the pre-miRs would be equally processed from the cluster due to equal accessibility of the Drosha-DGCR8 complex. However, if the pri-miR-17-92a cluster adopted a higher order structure it would cause suboptimal display of recognition sites for MPC complex. The formation of tertiary structure from the helices of the transcript could mask key recognition and binding junctions that would lead to the sequestration of MPC recognition sites (**Figure 2.2**).

Our hypothesis is supported by the study that show the recruitment of hnRNPA1 protein to enable the remodeling of the local stem loop of pre-miR-18a and facilitate its processing (Guil and Cáceres 2007). Further, association of RNA helicase activity with the MPC due to the

presenceof p68/72 suggests that remodeling of the RNA scaffold could contribute to such differential expression patterns of the miRs from the cluster. We were interested to determine if the pri-miR-17-92a can adopt a well-defined higher order structure and if that tertiary structure could regulate its differential miRNA processing.
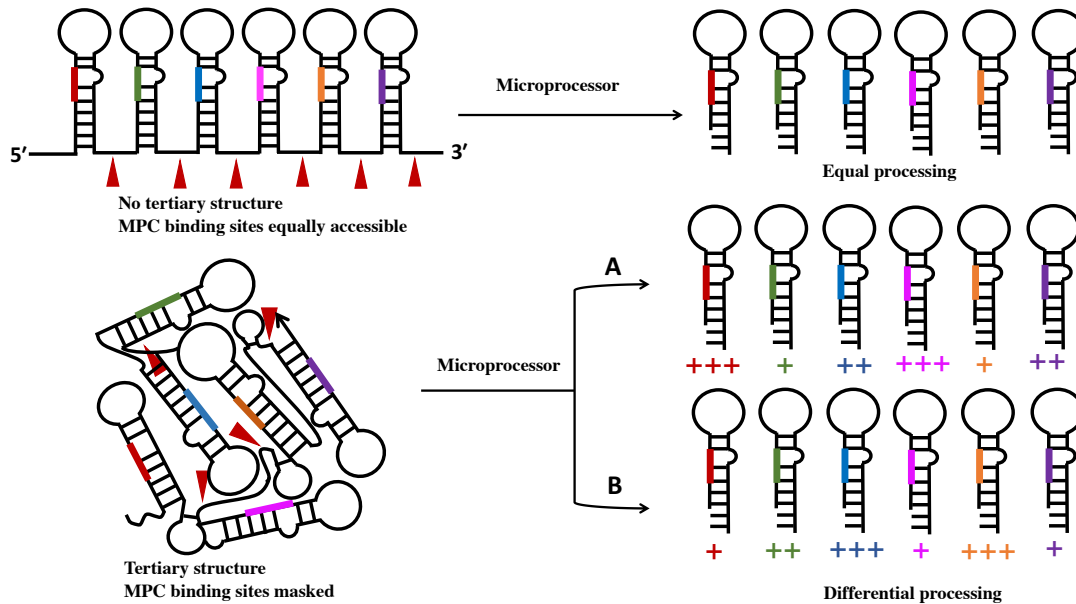
Figure 2.2: Schematic showing the pri-miR-17-92a cluster. In the absence of a tertiary structure the pri-miR would undergo equal processing to give equimolar amounts of individual pre-miRs (colors represent the six individual miRs) due to equal accessibility of the microprocessor (shown by arrowheads, red). Folding of the pri-miRNA into a distinct tertiary structure would allow differential processing in different tissues (A, B) by masking the recognition sites of the microprocessor to give different levels of its component pre-miRs.

In this chapter, I address the effect of tertiary structured pri-miR-17-92a on its pre-miR processing. To do this, the 0.8kb primiR-17-92a was cloned for overexpression in mammalian cells along with a swapped mutant construct. This construct was made by shuffling the hairpins (defined as shuffled pri-miR) from their native order that abrogates folding and hence lacks a distinct tertiary structure. Using *in cellulis* processing assays I observed that the levels of pre-miRs

from the native pri-miR and shuffled pri-miRs were very different. The relative abundances of pre-miRs were higher in the swapped construct and the pri-miR levels were undetectable, indicating that their processing times were different. Finally, using *in vitro* processing assays, I have shown that the shuffled cluster undergoes processing at a ~4-fold increased rate compared to the native cluster indicating that the tertiary structure present in the native cluster acts as a barrier to its pre-miRNA processing.

2. B: Materials and methods

2. B.1: Cloning of pri-miR 17-92a gene

The genomic region corresponding to the pri-miR-17-92a (~0.8 kb), its related miniclusters, pri-miR-17-19a (Mini Cluster1) and pri-miR-20a-92a (Mini Cluster2), were PCR amplified using specific DNA primers from HeLa genomic DNA (see **Table 2.1** for primer sequences). All sequences were then cloned into TOPO-pCR 2.1 TA cloning vector (Invitrogen) and linearized with SalI and EcoRI before *in vitro* transcription to generate the corresponding RNA transcripts. The shuffled or a swapped pri-miR-20a-19a cluster was derived from overlap PCR of 17-19a and 20a-92a. Primers were designed against miR-92a and miR-17 regions with 20nt overhangs complementary to each other (see Table 2.1 for primer sequences). After the first round of separate PCR cycles of 17-19a and 20a-92a, the two PCR amplicons were mixed in a 1:1 ratio. This was followed by an overlap extension PCR for 10 cycles. The extended product was used as a template, and further amplified using terminal primers. The forward primer used for making each construct for transcription contains T7 promoter sequence and the reverse primer contains a SalI site. For overexpression of pri-miR-17-92a and pri-miR-20a-19a in mammalian cell lines, the PCR

amplified DNA was cloned into pcDNA3.1/V5-His-TOPO TA vector (Invitrogen). The identity of all the constructs were confirmed by sequencing.

2. B.2: Cell culture and transfection

HeLa and HEK 293T cells were maintained in DMEM medium supplemented with 50 IU ml^{-1} of penicillin, 50 mg ml^{-1} streptomycin and 10% (v/v) FBS. Cells were cultured overnight at 37 °C in 5% CO_2 to a confluency of 70%, and then transiently transfected with empty vector (EV), pcDNA3-pri-miR-17-92a (native), pcDNA3-pri-miR-20a-19a (shuffled), and pcDNA3 Mini cluster 1(MC1) using HeLa Monster transfection reagent (Mirus Bio) following the manufacturer's instructions. Typically, 20 µg of plasmid DNA was used for cells grown in T75 cell culture flask and 0.3 volumes of transfection polymer was used. The transfected medium was replaced with fresh medium after 4 h of incubation and the cells were collected 24 h post transfection and used for RNA isolation.

2. B.3: Isolation of RNA

Total RNA from the transfected HeLa cells was isolated using Trizol reagent (Invitrogen) according to manufacturer's instructions with the following additional steps. Samples were extracted with an equal volume of Tris-Cl buffered phenol-chloroform pH 4.5 (Sigma) after the standard chloroform extraction and prior to precipitation with an equal volume of isopropanol. The nucleic acid pellets were resuspended in nuclease-free water (Ambion) and treated with 1U of TurboDNase (Ambion) per 20 µg of RNA at 37° C for 30 min. The RNA was then extracted with an equal volume of Tris-Cl-buffered phenol-chloroform pH 4.5 (Sigma) and precipitated with 2 volumes of absolute ethanol and one-tenth volume of 3M sodium acetate pH 5.6. The RNA was

resuspended in nuclease-free water. Total small RNAs from HeLa and HEK 293 cells transfected with native and shuffled pri-miR expression plasmids were isolated using miRVana miRNA isolation kit (Ambion) according to the manufacturer's protocol.

2. B.4: *In cellulis* processing of native and shuffled pri-miRs

2. B.4.1: RT-PCR analysis of pri-miR

RT-PCR analysis was used to assess the steady state levels of pri-miR in HeLa cells and HEK 293T cells transfected with native or shuffled pri-miR expression constructs. 3 μg of total RNA was reverse-transcribed with gene-specific primers using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. Briefly the RNAs were denatured at 80° C for 2 min and transferred to an ice-bath for 5 min. Reactions were prepared by addition of first-strand mix containing 0.5 mM dNTP mix, 0.25 mM RT primer and 50 U of SSIII reverse transcriptase and incubated at 37° C for 1 h followed by 42° C for 30 min. 4 μL of the cDNA was used for a standard PCR using Taq DNA Polymerase (New England Biolabs) for 30 cycles using the primers specific for native pri-miR-17-92a or shuffled pri-miR as indicated in Table 2.1. As a control, 18S rRNA was reverse transcribed and amplified. The RT-PCR products were analyzed on 1% agarose gel and visualized by ethidium bromide staining.

2. B.4.2: Northern analysis of pri-miRNAs

Northern blotting was employed to assess the steady state levels of pri-miR post transfection. Total RNA (10 μg) isolated from HeLa cells transfected with native pri-miR and shuffled pri-miR constructs were resolved on a 6% denaturing polyacrylamide gel in 1x TBE and then transferred to Amersham Hybond XL Nylon membrane (GE Healthcare). The membranes were UV cross

linked at 1200 mJ in Strategene UV Stratalinker, baked at 80° C for 30 min and pre-hybridized for 1 h at 40° C in hybridization buffer (1x SSC, 1% SDS and 200 µg/ml of Salmon sperm SS DNA (Sigma). Probes for pri-miR transcripts and U6 RNA (as a loading control) end-labeled using γ-P^{32} ATP and T4 Polynucleotide Kinase (New England Biolabs) were left to hybridize overnight in hybridization buffer at 37° C. Following hybridization, the membranes were washed twice in wash buffer I (1x SSC and 0.2% (w/v) SDS) for 30 min each and once in wash buffer II (0.2x SSC and 1% SDS) at 42° C for 15 min. The blots were then analyzed using phosphor imager (Typhoon Trio, GE Healthcare).

2. B.5: RT-PCR analysis of pre-miRNAs

RT-PCR analysis was used to assess the steady state levels of pre-miR transcripts from HeLa cells and HEK 293T cells transfected with native or shuffled pri-miR expression constructs. 1 µg of total small RNA was reverse-transcribed with primers specific for each pre-miR using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. Briefly, the RNAs were denatured at 80° C for 2 min and transferred to an ice-bath for 5 min. First-strand mix containing 0.5 mM dNTP mix, 0.25 mM specific RT primer and 50 U of SSIII reverse transcriptase was added and the reactions were incubated at 37° C for 1 h. 3 µL of the cDNA was used for a standard PCR using Taq DNA Polymerase (New England Biolabs) for 30 cycles using the primers specific for each pre-miR as indicated in **Table 2.1**. As a control, 5S rRNA was reverse transcribed and amplified. The RT-PCR products were analyzed on a 2% agarose gel and visualized by ethidium bromide staining.

2. B.6: qPCR analysis

Total RNAs from cultured HeLa cells were isolated using Trizol reagent (Invitrogen). RNA samples were reverse-transcribed into cDNA with Superscript III reverse transcriptase (Invitrogen). Real-time PCR was performed with gene-specific primers and Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA) using the 7500 Fast Real-Time PCR System (Applied Biosystems). For qPCR of pre-miRNAs, total small RNAs were isolated using miRVana miRNA isolation kit (Ambion) according to the manufacturer's protocol. Relative quantities of pri-miRNAs and pre-miRNAs were determined using the comparative C_T Method ($\Delta\Delta C_T$ method) as outlined by the manufacturer. Briefly, relative quantities of pri-miRNAs were determined by this method normalized to an endogenous reference (28S rRNA) and relative to the calibrator (empty vector, EV). Similarly the pre-miRNA levels of the shuffled transcript and mini cluster1 were normalized to an endogenous reference (5.8S rRNA) and relative to the native transcript.

2. B.7: *In vitro* transcription

To generate RNAs for *in vitro* experiments, TOPO-TA plasmids (20 μg each) containing DNA sequences for the required RNAs were digested with fast digest SalI first followed by EcoRI (3 units each, Fermentas) using manufacturer's protocol. Digested fragments of required sizes were gel purified from 0.8% agarose gel using Quiquick gel extraction kit (Qiagen). Gel extracted template DNAs were then precipitated with 1/10 volume of 3 M sodium acetate pH 5 and 3 volumes absolute ethanol followed by chilled 70% ethanol wash. Pri-miRNA substrates for all the studies were prepared by standard *in vitro* transcription using the purified template using T7 MegaScript kit (Ambion) according to the manufacturer's protocol. RNA samples were folded by heating them in nuclease free water (Ambion) at 90° C for 3 min followed by flash cooling on ice

for 1 min and incubating at 37° C for 1 h in folding buffer (50mM HEPES pH 7.8, 100mM KCl and 3mM $MgCl_2$).

2. B.8: *In vitro* pri-miRNA processing assays

Total HeLa extracts were prepared from ~3 x 10^6 control resuspended in 1 mL of buffer D (20 mM HEPES-KOH (pH 7.9), 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 5% (w/v) glycerol). The suspension was sonicated and centrifuged for 5 min at 10,000 g and the supernatant was used for *in vitro* assays. Pri-miRNA substrates (native, shuffled, MC1, and MC2) were prepared by standard *in vitro* transcription with T7 RNA polymerase (Thermo Scientific) in the presence of [α-32P] UTP. Before transcription, template DNAs were linearized with SalI. Assays were done in 30 μL reaction mixtures containing 50% (v/v) total HeLa extract, 0.5 mM ATP, 20 mM creatine phosphate, 3.2 mM $MgCl_2$ and 40,000 c.p.m. (~10 fmol) of each pri-miRNA. Reactions were incubated at 30° C for 90 min, then subjected to phenol-chloroform extraction, precipitation and then resolved on 8% (w/v) denaturing gel electrophoresis. The gels were dried using a gel drier and then subjected to phosphor imaging.

2. B.9: *In vitro* processing analysis by RT-PCR

10 μM of unlabeled *in vitro* transcribed RNAs corresponding to native and shuffled pri-miRs were incubated with total HeLa extracts as described in a time dependent manner after which the reactions were quenched by putting the tubes in dry-ice bath. The RNAs were subjected to ethanol precipitation after phenol-chloroform extraction and resuspended in nuclease-free water. The RNA was used for reverse transcription using specific primers to generate the corresponding cDNAs which were then used for PCR using primer set specific for native or shuffled pri-miR. The RT-PCR products were resolved on 1% agarose gel and visualized using ethidium bromide staining.

2. B.10: Northern blot and kinetic assays

Northern blotting of the *in vitro* processing reactions assesses the kinetics of the native and shuffled pri-miR transcripts. 5 μg of each *in vitro* transcribed pri-miR was incubated under processing conditions in the presence of HeLa cell extract in a 50 μL reaction at 37° C for the indicated times followed by phenol-chloroform extraction, ethanol precipitation sequentially. The RNAs were mixed with gel loading buffer and resolved on a 6% denaturing polyacrylamide gel in 1x TBE and then transferred to Hybond XL Nylon membrane (GE Healthcare). The membranes were UV cross-linked at 1200 mJ in Strategene UV Strata linker, baked at 80° C for 30 min and pre-hybridized for 1 h at 40° C in hybridization buffer (1x SSC, 1% SDS and 200 μg/ml of Salmon sperm SS DNA (Sigma). Probes for pri-miR transcripts (Nat-miR-probe, Shu-miR probe) and U6 snRNA RNA (as a control) end-labeled using a αP^{32}-ATP and T4 Polynucleotide Kinase (New England Biolabs) were left to hybridize overnight in hybridization buffer at 37° C. Following hybridization, the membranes were washed twice in wash buffer I (1x SSC and 0.2% (w/v) SDS) for 30 min each and once in wash buffer II (0.2x SSC and 1% SDS) at 42° C. The blots were then analyzed using a phosphor imager (Typhoon Trio, GE Healthcare).

For kinetics of processing, the Northern blots of the *in vitro* processing experiment were quantified. The amount of substrate remaining after time t, was quantified from the blots (Image J, NIH) and normalized to the substrate at t=0. For determining the initial rates, the traces from three independent processing experiments were constructed for both the native and shuffled transcripts and expressed as mean ± standard deviation (s.d.)

Table 2.1: List of primers used in the study

Primer	Sequence (5'-3')
Pri-miR-17-92a FP	TAATACGACTCACTATAGGGAAAACTGAAGATTGTGACC
Pri-miR-17-92a RP	GTCGACTCTTCTGGTCACAATCCCCACCAA
Pri-miR-17-19a FP	TAATACGACTCACTATAGGGAAAACTGAAGATTGTG
Pri-miR-17-19a RP	GTCGACCATTCATTTGAAGGAAATAGCAG
Pri-miR-20a-92a FP	TAATACGACTCACTATAGGGATTGTGTCGATGTAGAATCT
Pri-miR-20a-92a RP	GTCGACTCTTCTGGTCACAATCCCCACCAA
Pri-miR-20a-19a FP	TAATACGACTCACTATAGGGATTGTGTCGATGTAGAATCT
Pri-miR-20a-19a RP	GTCGACCATTCATTTGAAGGAAATAGCAG
Pri-miR overlap PCR FP	AAAAGAGAACATCACCTTGTAAAACTGAAGATTGTGACC
Pri-miR overlap PCR RP	ACAAGGTGATGTTCTCTTTTTCTTCTGGTCACAATCCCC
Pri-miR-17-92a pcDNA FP	AAAACTGAAGATTGTGACCAGTCAGA
Pri-miR-17-92a pcDNA RP	CATTCATTTGAAGGAAATAGCAG
Pri-miR-20a-19a pcDNA FP	ATTGTGTCGATGTAGAATCTGCCT
Pri-miR-20a-19a pcDNA RP	CATTCATTTGAAGGAAATAGCAGGC
Nat-miR-Probe	AAAACTGAAGATTGTGACCAGTCAGA
U6 snRNA probe	ATATGGAACGCTTCACGATT
Nat qPCR FP	GGGAAACTCAAACCCCTTTCTAC
Nat qPCR RP	CAACAGGCCGGGACAAGT
Shu qPCR FP	GCCCAATCAAACTGTCCTGT

71

Shu qPCR RP	ACAATCCCCACCAAACTCAA
MC1 qPCR FP	TGCCCTAAGTGCTCCTTCT
MC1 qPCR RP	AAATAGCAGGCCACCATCAG
28S rRNA FP	CAGGGGAATCCGACTGTTTA
28S rRNA RP	ATGACGAGGCATTTGGCTAC
pre-miR-17 FP	CAAAGTGCTTACAGTGCAG
pre-miR-17 RP	CTACAAGTGCCTTCACTG
pre-miR-18a FP	TAAGGTGCATCTAGTGCA
pre-miR-18a RP	CCAGAAGGAGCACTTAGG
pre-miR-19a FP	AGTTTTGCATAGTTGCACTAC
pre-miR-19a RP	CAGTTTTGCATAGATTTGCAC
pre-miR-20a FP	CTTTAAGTGCTCATAATGCAG
pre-miR-20a RP	CTTTAAGTGCTCATAATGCAG
pre-miR-19b FP	AGTTTTGCAGGTTTGCATCC
pre-miR-19b RP	CAGTTTTGCATGGATTTGCAC

pre-miR-92a FP	AAGGTTGGGATCGGTTGC
pre-miR-92a RP	TTACAGGCCGGGACAAGT
5.8S rRNA FP	GTGCGTCGATGAAGAAC
5.8S rRNA RP	TCAATGTGTCCTGCAATT

2. C: Results and discussion

2. C.1: *In cellulis* processing of native pri-miR-17-92a transcript

The pri-miR-17-92a transcript is a polycistronic cluster encoding six precursor miRs on chromosome 13 (miR-17, miR-18a, miR-19a, miR-20a, mir-19b and mir-92a). It is schematically depicted as six tandem stem loops possessing defined secondary structural elements (**Figure 2.3A**). Though previous studies have indicated plausible hairpin interactions in this cluster whether or not the pri-miR cluster could form a tertiary structure was unknown. In order to test this hypothesis, we cloned pri-miR-17-92a into a pcDNA3 mammalian expression vector after amplifying the 0.8 kb sequence corresponding to pri-miR-17-92a (Native pri-miR, N) from HeLa genomic DNA shown in **Figure 2.3C**. Additionally, a swapped or shuffled pri-miR-20a-19a (Shuffled, S) construct was made swapping the two half domains of the native transcript (**Figure 2.3D**) and a mini cluster was made containing the first three hairpins refered as pri-miR-17-19a or

MC1 in pcDNA3 mammalian expression vector. **Figure 2.3B** shows the schematic representation

of the order of hairpins in these constructs.
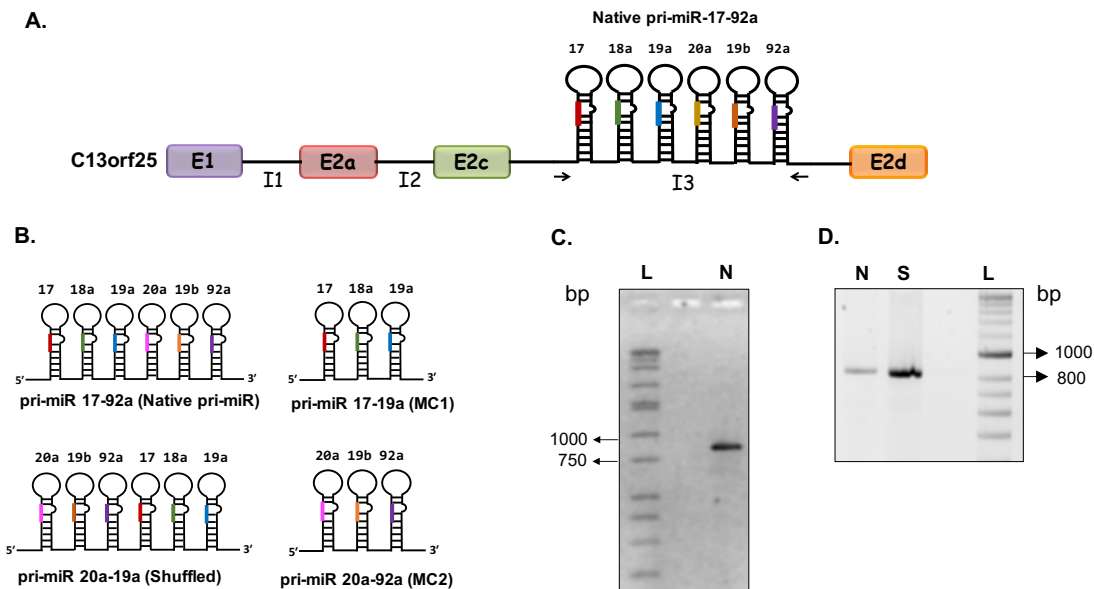
Figure 2.3: **A**. Genomic structure of the c13orf25 gene, harboring the intronic pri-miR- 17-92a cluster **B**. Diagrammatic representation of the pri-miR-17-92a and its related transcripts. **C.** 1% agarose gel to show the 0.8kb amplicon corresponding to pri-miR 17-92a from HeLa genomic DNA. **D.** 1% agarose gel to show the formation of 0.8kb fragment by shuffling the two half domains using overlap extension PCR (lane N-native pri-miR, lane S-shuffled pri-miR).

RT-PCR was performed on total RNA isolated from HeLa cells transfected with pri-miR-

17-92a to analyze its expression of the pri-miR transcript. To do this, pri-miR-17-92a RP was used

to generate cDNA which was subjected to PCR using specific set of primers. **Figure 2.4A** shows

a single amplicon of 0.8 kb corresponding to the pri-miR 17-92a. *In cellulis* processing of pri-miR-

17-92a was then analyzed by measuring the levels of individual pre-miRs from this cluster using

qPCR and Northern blot. The qRT-PCR analysis revealed that the pre-miRNAs from the pri-miR-

17-92a transcript had differential levels of expression **Figures 2.4B & C**. Previous investigation

of processing kinetics, which measured pre-miR processing *in vitro* using purified Drosha-DGCR8 also indicated differential processing for the pre-miRs (Chaulk et al. 2011).
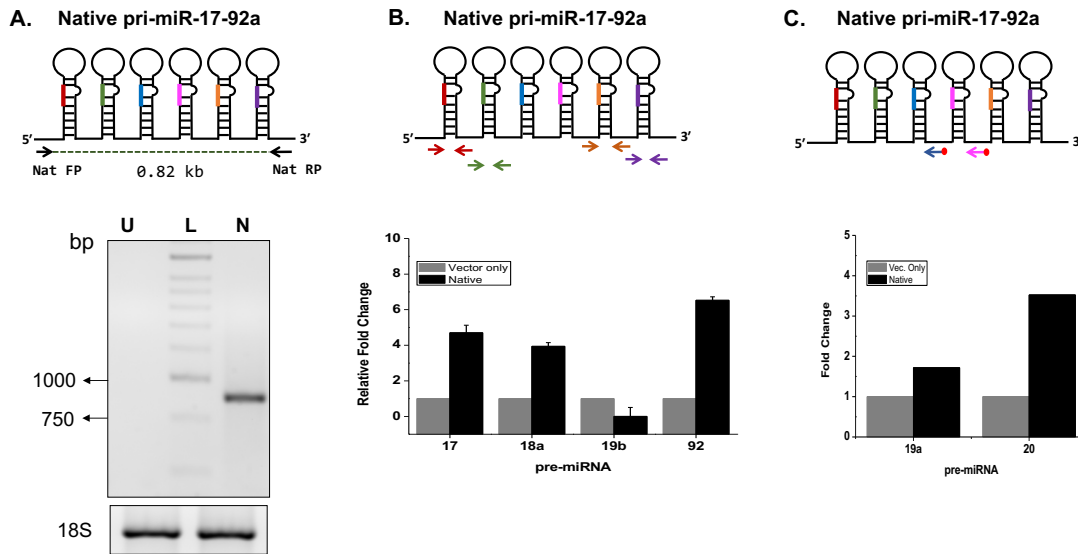
Figure 2.4: **A**. RT-PCR to address the steady state levels of pri-miR-17-92a from HeLa cells transfected with pcDNA3-pri-miR-17-92a. 1% agarose gel showing the levels of pri-miR 17-92a, U represents total RNA from untransfected HeLa cells. 18S rRNA was amplified as control. **B**. Small RNAs isolated from HeLa cells transfected with pri-miR-17-92a in pcDNA3 vector (black) relative to those transfected with the vector only (grey) are shown. **C.** Northern blot to show relative levels of pre-miR-19a and 20a from total small RNAs isolated from HeLa transfected with pri-miR-17-92a and resolved on 15% denaturing PAGE gel and hybridized using radiolabeled probes post transfer to a membrane

2. C.2: Steady state levels of native and shuffled pri-miRs

Northern blot analysis was performed to assess steady-state levels of pri-miRa after overexpression. Total RNA from HeLa cells and HEK293T cells transfected with native pri-miR, or shuffled pri-miR was isolated and resolved on 8% denaturing agarose gel. Northern blot was performed using 5′ end-P^{32}-labeled probes (Nat-miR probe, see **Table 2.1** for sequences) that hybridized to the 3′ end of the pri-miRs. As show in **Figure 2.5A** lane N shows a single band

corresponding to native pri-miR but lane S lacks a detectable band. The same probes were used to detect *in vitro* transcribed native and shuffled pri-miRs indicating that the probes are insensitive to cluster organization.

Next RT-PCR was employed to analyze the steady-state expression levels of pri-miRs from total RNA isolated from HeLa and HEK 293T cells transiently transfected with native pri-miR, and shuffled pri-miR. cDNA was synthesized using RT primers specific to each transcript and then PCR amplified using pri-miR-17-92a FP and RP for native pri-miR, and pri-miR-20a-19a FP and RP for shuffled pri-miR. **Figure 2.5B** shows the 0.8 kb band corresponding to native pri-miR-17-92a in both HeLa and HEK 293T cells. Lane S did not show any detectable levels of shuffled pri-miR from HeLa and a faint band from HEK 293T. This was further confirmed by a qRT-PCR analysis of the native, shuffled, and MC1 transcripts. This decrease in the levels of shuffled pri-miR could possibly be due to low expression or degradation of the shuffled transcript.
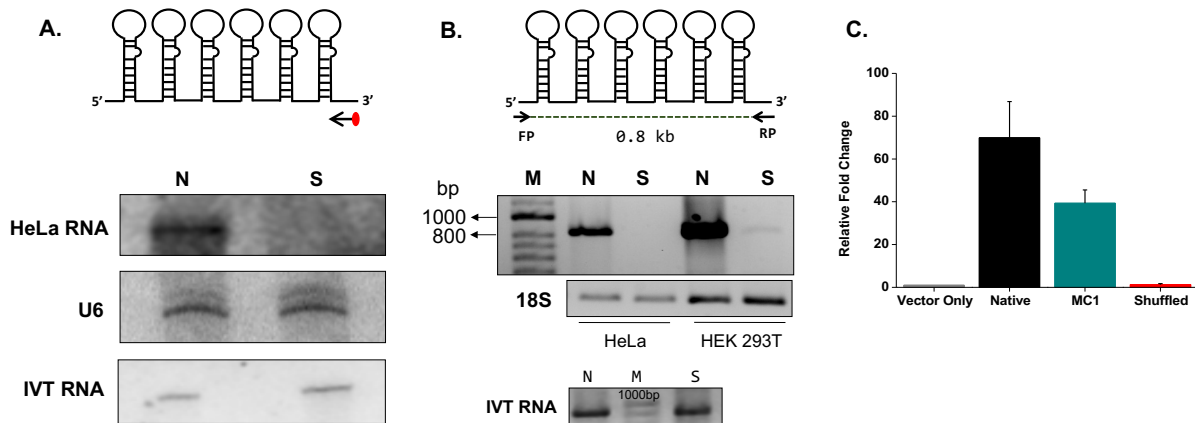
Figure 2.5: The relative equilibrium levels of the pri-miR-17-92a (N) and pri-miR-20a-19a transcripts (S) *in cellulis* are different. **A**. Northern Blot of total RNA fractionated on 8% Urea-PAGE and probed using a radiolabeled complementary oligo. **B**. Total RNA isolated from HeLa and HEK 293T cells 24 h post transfection of the plasmids expressing the native and the shuffled transcripts were used for RT-PCR analysis. Products were analyzed on a 1% agarose gel. RT-PCR of 18S rRNA was used as a control. *In vitro* transcribed (IVT) native (N) and shuffled (S) RNA were used as control to show the efficiency of probes and primers used.**C.** Quantitative RT–PCR was used to measure levels of the pri-miR from total RNA isolated from HeLa cells transfected

either with native (N), shuffled (S) or Mini cluster 1 (MC1). The relative expression levels as determined by $\Delta\Delta C_T$ analyses are shown.

2. C.3: *In cellulis* processing of native and shuffled pre-miRs

To test if the low abundance of the shuffled pri-miR was due its weak expression or due to spontaneous degradation, we sought to investigate the abundance of pre-miRs for both native and shuffled transcripts *in cellulis*. The goal of these investigations was to determine if higher order structure in the native pri-miR-17-92a influenced its pre-miR processing. Low abundance of constituent pre-miRs within the shuffled would indicate weak expression of the shuffled. To study *in cellulis* processing, total small RNAs isolated from HeLa and HEK293T cells were reverse transcribed with RT primer specific to each pre-miR of the cluster. The resultant cDNAs were used for PCR with primers specific for each pre-miR as shown in the schematic **(Figure 2.6A).** As shown in **Figure 2.6B** we observed bands of same length corresponding to each of the six pre-miR of the native and shuffled clusters in both the cell lines indicating that both shuffled pri-miR undergoes processing to give its constituent pre-miRs. In the absence of any tertiary interaction in pri-miR-17-92a, the shuffled transcript which contains all six pre-miRNAs but in a different order, should be processed similarly in cells due to availability of all the relevant binding sites to the microprocessor.

To confirm this, quantitative real time PCR (qRT-PCR) analysis was performed to measure the relative levels of the processed pre-miRNAs from HeLa cells over expressing either native or shuffled transcript (for details, see Chakraborty et al. 2012). Total small RNAs isolated were reverse transcribed with RT primer specific for each pre-miR and the resultant cDNAs were used for qRT-PCR using specific forward and reverse primers for each pre-miR **(See Table 2.1** for sequences**).** The levels of the pre-miRs from native, shuffled and MC1 pri-miR were determined

by quantitative PCR using $\Delta\Delta C_T$ method. **Figure 2.6C** shows that compared to native pri-miR-17-92a the individual pre-miRNAs from shuffled pri-miR-20a-19a as well as the mini cluster pri-miR-17-19a show much higher levels.
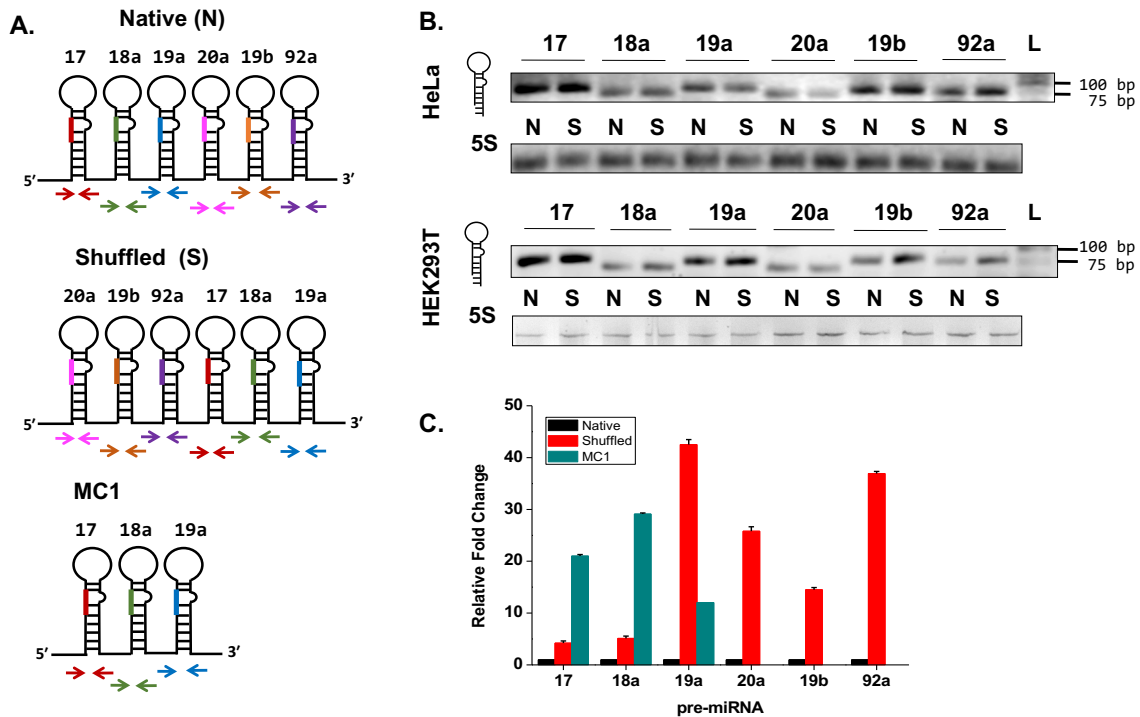
Figure 2.6: A. Schematic representation of the pre-miRs on native, shuffled and MC1 and the primer binding sites on the individual pre-miRs. **B**. Relative levels of pre-miRs from native and shuffled transcripts. Small RNAs isolated 24 h post transfection from HeLa and HEK 293T cells transfected with pri-miR-17-92a-pcDNA3 and pri-miR-20a-19a-pcDNA3 were used for RT-PCR using primers specific for each pre-miR indicated. 5S RNA used for RT-PCR serves as a loading control. **C**. The levels of the pre-miRs from native, shuffled and MC1 pri-miR were determined by quantitative PCR using $\Delta\Delta C_T$ method.

Further, this increase is not uniform for each pre-miR, especially for pre-miR-20a, pre-miR-19a and pre-miR-92a from pri-miR-20a-19a, indicating a positional significance in the context of the entire transcript. Thus, a mere shuffling of discrete, pre-miR containing, hairpin domains are sufficient to alter the relative abundance of the processed pre-miRNAs. Thus far, *in cellulis* processing studies indicate structural differences between the native transcript and the

78

shuffled transcript that impacts their processing, which is beyond a simplistic secondary structure model.

2. C.4: *In vitro* processing of native and shuffled pri-miRs

Biochemical and biophysical investigations of pri-miR-17-92a revealed that the transcript adopts a well-defined tertiary structure *in vitro* (Chakraborty et al. 2012). Next, we sought to determine how this tertiary structure influences microRNA processing *in vitro*. To do this native, shuffled, MC1, and MC2 RNA sequences were *in vitro* transcribed (**Figure 2.7A)**. The integrity of *in vitro* transcribed RNAs was analyzed on 8% denaturing PAGE gel (**Figure 2.7B**). To address the processing of the pri-miRs internally radiolabeled RNAs using αP^{32}-UTP and incubated with total HeLa extracts.
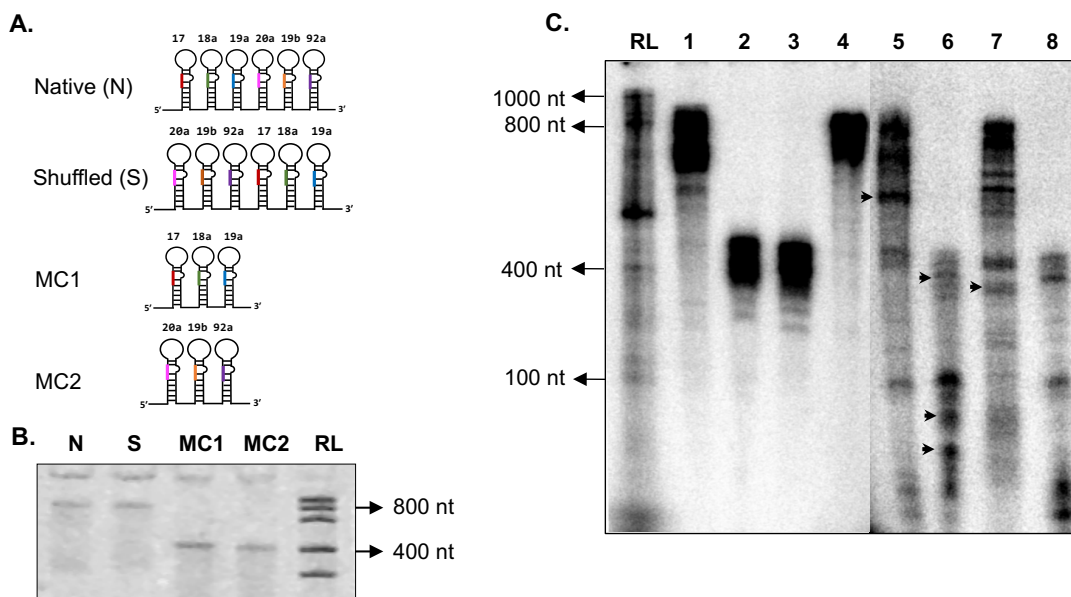
Figure 2.7: In vitro processing (IVP) of pri-miR transcripts. **A.** Schematic representation of native, shuffled, MC1, and MC1 transcripts. **B.** 8% denaturing PAGE showing the *in vitro* transcribed products corresponding to native (N), shuffled (S), MC1, and MC2. RL indicated Ribo ruler, Low range. **C.** 8% denaturing PAGE showing the *in vitro* processed products. Lane-1: native pri-miR-17-92a unprocessed (devoid of HeLa extracts), lane-2: MC1 unprocessed, lane-3: MC2

unprocessed, lane-4: shuffled pri-miR unprocessed, lane-5: IVP native, lane-6: IVP MC1+ MC2, lane-7: IVP shuffled, lane-8: IVP MC1.

Figure 2.7C shows the products of *in vitro* processing of native, shuffled, MC1, and MC2. The relative abundances of the in vitro processed products from native (lane 5) are different from shuffled (lane 8) as indicated with arrowheads, recapitulating the *in cellulis* processing. Further, the fact that the native cluster behaves differently from a simple mini cluster suggests interaction between pri-miR-17-19a (MC1) and pri-miR-20a-92a (MC2) when they are fused to form the native transcript (lane 6).

2. C.5: Kinetic analysis of *in vitro* processing of native and shuffled pri-miRs

The data provided above demonstrates that the native pri-miR-17-92a with a well-defined tertiary structure and a shuffled pri-miR-20a-19a transcript containing the same sequence information but lacking a distinct higher order structure were processed inside HeLa cells into their respective pre-miRs. However, when the relative levels of the full-length transcripts pri-miR-17-92a and pri-miR-20a-19a were analyzed with their relevant primers 24 hr post transfection, it was found that although pri-miR-17-92a was present at reasonable levels, pri-miR-20a-19a level were below detection limit. This indicated that the equilibrium levels of pri-miR-17-92a and pri-miR-20a-19a *in cellulis* were quite different. The rate of production of the transcripts *in cellulis* can be considered comparable since they are under same CMV promoter and have same *in vitro* lifetimes (Chakraborty et al 2012). This implies that their kinetics of processing *in cellulis* might bedifferent. Hence, investigation of the kinetics of processing of native and shuffled pri-miRs was required. To this end, *in vitro* transcribed and folded pri-miR-17-92a or pri-miR-20a-19a transcripts were incubated in HeLa cell extract in a time dependent manner. The amount of full length transcript

was first detected by RT-PCR using specific primers **(Figure 2.8).** The semi-quantitative RT-PCR

gels indicate that the amount of substrate pri-miR left after processing into its constituent pre-miRs

was higher for native compared to the shuffle.
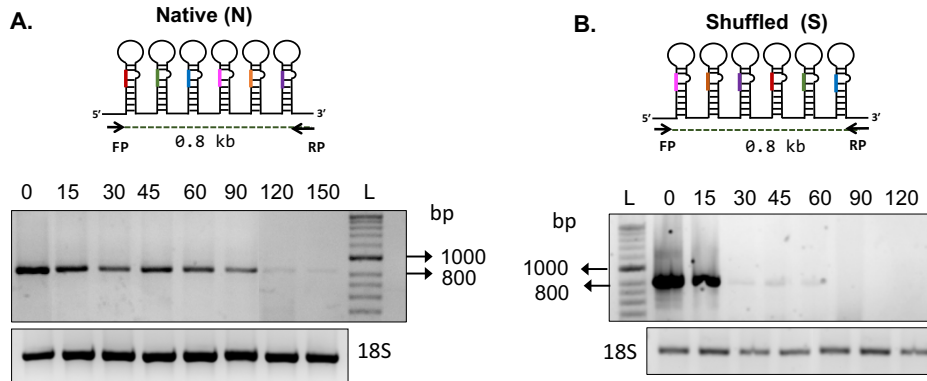
Figure 2.8: *In vitro* processing of pri-miRs. Native and shuffled pri-miR were incubated with total HeLa extracts in timed reactions. The purified RNAs were reverse transcribed using specific RT primer and the resultant cDNA was PCR amplified using primers specific for native and shuffled pri-miRs. 1% agarose gel showing the amount of substrate pri-miR left corresponding to native pri-miR **(A)** and shuffled pri-miR **(B).**

To quantitate the amount of the substrate remaining during the course of its processing,

either the native or shuffled pri-miR transcripts were incubated with HeLa cell extract in a time

dependent manner. The full-length transcript was detected using the radiolabeled probe. Northern

blot analysis showed that processing is indeed faster in the case of the shuffled transcript pri-miR-

20a-19a than in the native, folded transcript pri-miR-17-92a **(Figure 2.9A & B).** The time taken

for 50% of the transcript to be processed in case of native pri-miR-17-92a and shuffled pri-miR-

20a-19a was about ~120 minutes and 40 minutes, respectively. The rate of processing for pri-miR-

17-92a and pri-miR-20a-19a was quantified by the initial rate method, given that the order of the

reaction is not known and this showed that the rate of processing of pri-miR-20a-19a is about ~4

fold higher than the rate of pri-miR-17-92a processing **(Figure 2.9C)**. This shows that the tertiary

81

structure adopted by the primary microRNA transcript has considerable influence on its processing *in vitro* and *in cellulis*.
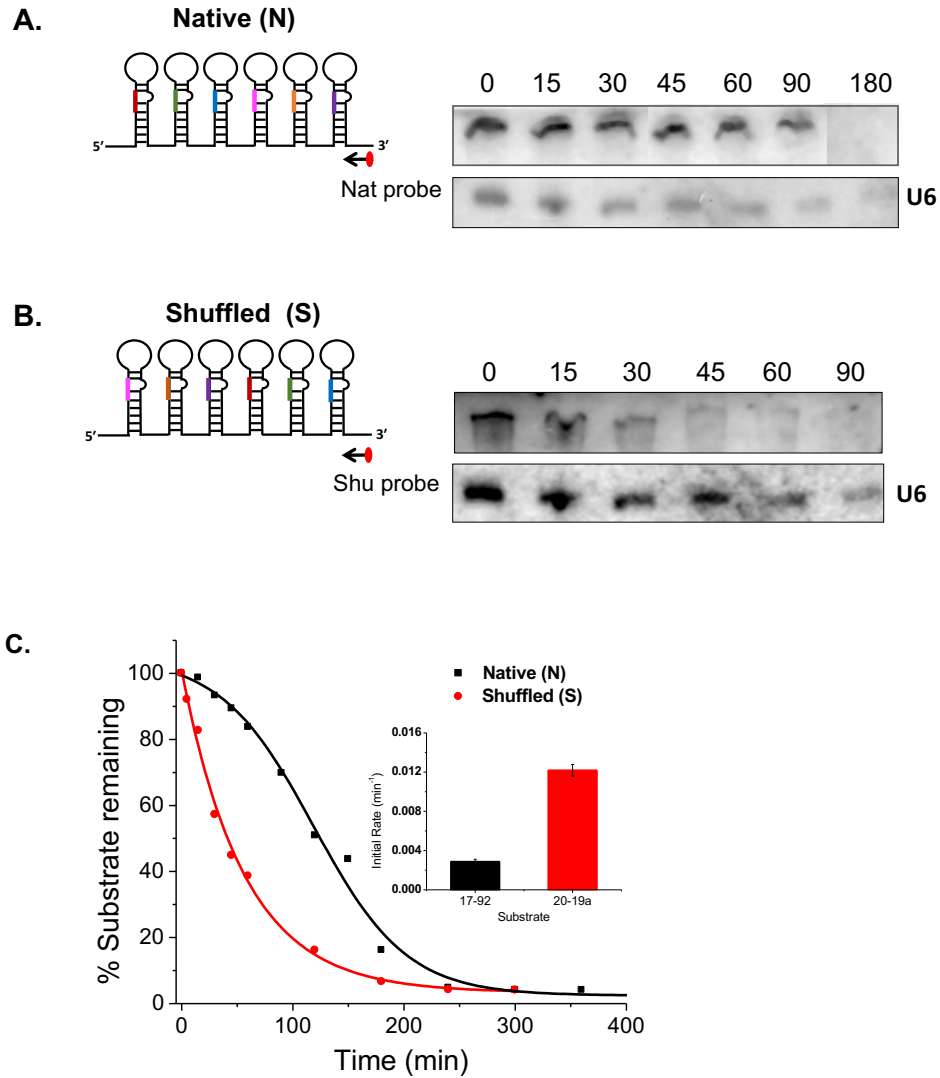
Figure 2.9: Tertiary structure of the pri-miR-17-92a transcript acts as a kinetic barrier for its processing. Northern blot to probe full length native (N) and shuffled transcripts (S) remaining after *in vitro* processing with total HeLa extracts for the indicated time points. U6 is shown as an internal control. **C.** Quantification of Northern blots to probe full length native and shuffled transcripts remaining after *in vitro* processing show the kinetics of processing of native transcript is much slower. Initial rates were determined from the kinetic traces expressed as mean ± s.d from three independent experiments, shown in inset.

2. D. Conclusions:

Transcription of a polycistronic pri-miRNA permits coordinated expression of individual mature miRNAs which can share targets and functions or target genes with opposing functions. Unlinking their expression by post-transcriptional regulation in response to different cues has the potential to change the expression of multiple miRNA targets to facilitate a biological response. The most studied polycistronic pri-miRNA transcript pri-miR-17-92a cluster, encodes miRNAs that target both positive and negative cell cycle regulators, and pro-apoptotic and anti-apoptotic proteins (O'Donnell et al. 2005; Li et al. 2012). Using extensive phylogenetic sequence analysis, biochemical and spectroscopic studies, our group has shown that pri-miRNA-17-92a cluster folds into a compact tertiary structure (Chakraborty et al. 2012). The folding of this RNA into its tertiary structure might be facilitated by the conserved terminal loops of the pre-miR domains, the inter-pre-miR regions or both.

Here, I have shown that tertiary structure formation by pri-miR-17-92a influences its processing into its constituent pre-miRNAs. The disruption of such structured inter pre-miRNA elements, such as the region between pre-miR-19a and pre-miR-20a, yields a shuffled transcript pri-miR-20a-19a with impaired tertiary structure. Overexpression of native pri-miR and shuffled pri-miR showed that although native pri-miR-17-92a was present at reasonable levels, shuffled pri-miR-20a-19a level was below detection limit, indicating that *in cellulis* equilibrium levels of native and shuffled transcript were quite different. *In cellulis* processing of the native and shuffled transcripts to its constituent pre-miRs revealed that both the transcripts had altered processing efficiencies with shuffled pri-miR showing elevated levels of pre-miRs compared to native. This was further confirmed by *in vitro* processing studies wherein the tertiary structured native pri-miR-

17-92a undergoes slow (~4-fold) microprocessing compared to the shuffled transcript. Thus, a mere shuffling of discrete, pre-miR containing, hairpin domains is sufficient to alter the relative abundance of the processed pre-miRNAs. Using *in cellulis* and *in vitro* processing studies we have shown that structural differences between the native transcript and the shuffled transcript impacts their processing, which is beyond a simplistic secondary structure model. This suggests a model where tertiary structure formation by a primary miRNA transcript imposes a kinetic barrier to its processing where the conformation adopted is transparent to the MPC leading to an inhibition imposed at its earliest stage of its processing into pre-miRs.

A parallel study by (Chaulk et al. 2011) also showed complementary evidence of tertiary structure formation by pri-miR-17-92a. Cryo EM at 2 nm resolution showed a compact globular structure where miRNAs 18a, 19b, and 92a internalized within the core of the folded structure are processed less efficiently than miRNAs on the surface of the structure. Deletion of the 5′ region of the cluster exposed these miRs to ribonuclease digestion indicating that the 5′ region protects the 3′ core domain region within the tertiary structure of the cluster. Expression of a mutant cluster that disrupted the structure of the pri-miR-17-92a and exposed the 3′ core resulted in a ~3-fold decrease in integrin subunit alpha5 (ITGA5) mRNA level, a validated miR-92a target compared to expression of the full-length miR-17~92 cluster. This suggests that there is an apparent correlation between surface accessibility of the individual miRNA-containing hairpins and microprocessor processing efficiency. Further, using a site-specific photo-cross-linker and mutagenesis experiments Chaulk *et al* have identified a tertiary contact between a non-miRNA stem-loop (NMSL) and pre-miR-19b hairpin. Based on their studied they proposed that this tertiary

contact within the 3′ core of the pri-miR serves as a molecular scaffold to mask the expression of anti-angiogenic miR-92a (Chaulk et al. 2014).

Interestingly miR-92a was shown to be constitutively expressed in mouse embryonic stem cells by Du *et al* suggesting that there could be a structural reorganization to regulate the release of specific miRs in a cell type specific manner (Du et al. 2015). This study revealed two complementary cis-regulatory repression domains within pri-miR-17-92a, one at the 5′ end defined as repression domain (RD) and another between pri-miR-19b and pri-miR-92a defined as repression domain* (3′-RD*), which base pair to form an auto inhibitory RNA conformation that blocks the all the constituent miRNA processing from the cluster except miR-92a. Cleavage of pri-miR-17-92a to remove this inhibitory 5′ fragment results in an intermediate termed progenitor microRNA (pro-miRNA) which is dynamically regulated by an endonuclease component the cleavage and polyadenylation specificity factor complex, CPSF3 (also known as CPSF73 or CPSF-73) and a splicing factor ISY1. This developmentally regulated generation of pro-miRNA explains the post-transcriptional control of pri-miR-17-92a expression in mouse ESCs.

It is thus becoming increasingly evident that post-transcriptional mechanisms play an important role controlling miRNA expression. Such mechanisms that control individual miRNA from clustered pri-miRNAs need a tight and coordinated regulation to release a specific miRNA in a spatio-temporal manner. Gene amplification and increased expression of miRNAs from pri-mR-17-92a cluster is observed in several cancers (Li et al. 2012; Fuziwara and Kimura 2015; Concepcion et al. 2012). While overexpression of this OncomiR-1 has been known to promote B cell lymphoma, and T cell leukemia (Mu et al. 2009; Mihailovich et al. 2015), individual miRNAs from this cluster promote cell proliferation, inhibit apoptosis and differentiation and promote

angiogenesis to drive tumor formation (Zhang et al. 2007; Li et al. 2012; Chamorro-Jorganes et al. 2016; Zhang et al. 2014; Chang et al. 2013). Overexpression of the entire pri-miR-17-92a cluster in mice resulted in normal myeloid and lymphoid lineage differentiation but overexpression of individual miR-19a or miR-92a from this locus resulted in B-cell hyperplasia and erythroleukemia respectively. Co-expression of miR-17 with miR-92a abolished the miR-92a induced effect (Li et al. 2012). Inactivation of P53 contributed to upregulation of miR-19a and oncogenic miR-92a, and down regulation of suppressive miR-17. These studies reveal that imbalanced expression of miRs from this cluster result in hematopoietic malignancies (Li et al. 2012). Expression of miR-19 promotes lymphoma which is suppressed by co-expression of miR-92 and in c-Myc induced mouse lymphoma the ratio of miR-19: miR-92 expression regulates the tumor progression (Olive et al. 2009). The pri-miR-17-92a cluster is also highly expressed in human endothelial cells and miR-92a, controls angiogenesis (Bonauer et al. 2009) by targeting mRNAs corresponding to several pro-angiogenic proteins including integrin subunit alpha5 (*ITGA5*), a suppressor of angiogenesis and endothelial nitric oxide-synthase (*eNOS*) which controls vascular tone and is crucial for postnatal neovascularization. Overexpression of miR-92a in endothelial cells in mice blocked angiogenesis *in vitro* and *in vivo*. Systemic administration of an antagomir designed to inhibit miR-92a in mouse models of limb ischemia and myocardial infarction (Bonauer et al. 2009)(Bonauer et al. 2009)(Bonauer et al. 2009)(Bonauer et al. 2009)(Bonauer et al. 2009) resulted in enhanced blood vessel growth and functional recovery of damaged tissues.

The present study shows that the folding of the transcript into a tertiary structure might function as the first step of post transcriptional regulation by precluding equal accessibility of the constituent pre-miR domains to their relevant protein partners. Absence of a tertiary structure in a

transcript such as shuffled pri-miR-20a-19a leads to a 4-fold faster cleavage rate, due to easy accessibility of binding sites for the MPC. The proposed model suggests auto regulation imposed early on by the RNA scaffold adopting a tertiary structure. Thus far, work done by our group and others has shown the potential of pri-miR-17-92a to form folded three-dimensional structure that enables different presentation of the pre-miR hairpins for processing. Distinct ways to remodel this tertiary structure influenced by various proteins in specific cellular environments would result in differential processing of the individual miRNAs. The impact of this layer of regulation on normal or perturbed developmental processes, as well as potential for therapeutic intervention awaits further investigation.

2. E: References

Abbott, A.L., Alvarez-Saavedra, E., Miska, E.A., Lau, N.C., Bartel, D.P., Horvitz, H.R. and Ambros, V. 2005. The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans. *Developmental Cell* 9(3), pp. 403–414.

Bail, S., Swerdel, M., Liu, H., Jiao, X., Goff, L.A., Hart, R.P. and Kiledjian, M. 2010. Differential regulation of microRNA stability. *RNA (New York)* 16(5), pp. 1032–1039.

Bonauer, A., Carmona, G., Iwasaki, M., Mione, M., Koyanagi, M., Fischer, A., Burchfield, J., Fox, H., Doebele, C., Ohtani, K., Chavakis, E., Potente, M., Tjwa, M., Urbich, C., Zeiher, A.M. and Dimmeler, S. 2009. MicroRNA-92a controls angiogenesis and functional recovery of ischemic tissues in mice. *Science* 324(5935), pp. 1710–1713.

Breving, K. and Esquela-Kerscher, A. 2010. The complexities of microRNA regulation: mirandering around the rules. *The International Journal of Biochemistry & Cell Biology* 42(8), pp. 1316–1329.

Bushati, N. and Cohen, S.M. 2007. microRNA functions. *Annual Review of Cell and Developmental Biology* 23, pp. 175–205.

Chakraborty, S., Mehtab, S., Patwardhan, A. and Krishnan, Y. 2012. Pri-miR-17-92a transcript folds into a tertiary structure and autoregulates its processing. *RNA (New York)* 18(5), pp. 1014–1028.

Chamorro-Jorganes, A., Lee, M.Y., Araldi, E., Landskroner-Eiger, S., Fernández-Fuertes, M., Sahraei, M., Quiles Del Rey, M., van Solingen, C., Yu, J., Fernández-Hernando, C., Sessa, W.C.

and Suárez, Y. 2016. VEGF-Induced Expression of miR-17-92 Cluster in Endothelial Cells Is Mediated by ERK/ELK1 Activation and Regulates Angiogenesis. *Circulation Research* 118(1), pp. 38–47.

Chang, C.-C., Yang, Y.-J., Li, Y.-J., Chen, S.-T., Lin, B.-R., Wu, T.-S., Lin, S.-K., Kuo, M.Y.-P. and Tan, C.-T. 2013. MicroRNA-17/20a functions to inhibit cell migration and can be used a prognostic marker in oral squamous cell carcinoma. *Oral Oncology* 49(9), pp. 923–931.

Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E.M., Veldhoen, R.A., Khanna, S.K., Goping, I.S., MacMillan, A.M., Mendell, J.T., Young, H.S., Fahlman, R.P. and Glover, J.N.M. 2011. Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biology* 8(6), pp. 1105–1114.

Chaulk, S.G., Xu, Z., Glover, M.J.N. and Fahlman, R.P. 2014. MicroRNA miR-92a-1 biogenesis and mRNA targeting is modulated by a tertiary contact within the miR-17~92 microRNA cluster. *Nucleic Acids Research* 42(8), pp. 5234–5244.

Chen, K. and Rajewsky, N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews. Genetics* 8(2), pp. 93–103.

Concepcion, C.P., Bonetti, C. and Ventura, A. 2012. The microRNA-17-92 family of microRNA clusters in development and disease. *Cancer Journal* 18(3), pp. 262–267.

Davis, B.N. and Hata, A. 2009. Regulation of MicroRNA Biogenesis: A miRiad of mechanisms. *Cell Communication and Signaling* 7, p. 18.

Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F. and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* 432(7014), pp. 231–235.

Du, P., Wang, L., Sliz, P. and Gregory, R.I. 2015. A Biogenesis Step Upstream of Microprocessor Controls miR-17~92 Expression. *Cell* 162(4), pp. 885–899.

Filipowicz, W., Bhattacharyya, S.N. and Sonenberg, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews. Genetics* 9(2), pp. 102–114.

Filipowicz, W. and Pogacić, V. 2002. Biogenesis of small nucleolar ribonucleoproteins. *Current Opinion in Cell Biology* 14(3), pp. 319–327.

Fuziwara, C.S. and Kimura, E.T. 2015. Insights into Regulation of the miR-17-92 Cluster of miRNAs in Cancer. *Frontiers in medicine* 2, p. 64.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432(7014), pp. 235–240.

Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research* 36(Database issue), pp. D154–8.

Guil, S. and Cáceres, J.F. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Structural & Molecular Biology* 14(7), pp. 591–596.

Hammond, S.M. and Wood, M.J.A. 2011. Genetic therapies for RNA mis-splicing diseases. *Trends in Genetics* 27(5), pp. 196–205.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H. and Kim, V.N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development* 18(24), pp. 3016–3027.

Jacob, F. and Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3(3), pp. 318–356.

Kim, V.N., Han, J. and Siomi, M.C. 2009. Biogenesis of small RNAs in animals. *Nature Reviews. Molecular Cell Biology* 10(2), pp. 126–139.

Krol, J., Loedige, I. and Filipowicz, W. 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews. Genetics* 11(9), pp. 597–610.

Li, Y., Vecchiarelli-Federico, L.M., Li, Y.-J., Egan, S.E., Spaner, D., Hough, M.R. and Ben-David, Y. 2012. The miR-17-92 cluster expands multipotent hematopoietic progenitors whereas imbalanced expression of its individual oncogenic miRNAs promotes leukemia in mice. *Blood* 119(19), pp. 4486–4498.

Megraw, M., Sethupathy, P., Corda, B. and Hatzigeorgiou, A.G. 2007. miRGen: a database for the study of animal microRNA genomic organization and function. *Nucleic Acids Research* 35(Database issue), pp. D149–55.

Mendell, J.T. 2008. miRiad roles for the miR-17-92 cluster in development and disease. *Cell* 133(2), pp. 217–222.

Michlewski, G., Guil, S., Semple, C.A. and Cáceres, J.F. 2008. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Molecular Cell* 32(3), pp. 383–393.

Mihailovich, M., Bremang, M., Spadotto, V., Musiani, D., Vitale, E., Varano, G., Zambelli, F., Mancuso, F.M., Cairns, D.A., Pavesi, G., Casola, S. and Bonaldi, T. 2015. miR-17-92 fine-tunes MYC expression and function to ensure optimal B cell lymphoma growth. *Nature Communications* 6, p. 8725.

Mu, P., Han, Y.-C., Betel, D., Yao, E., Squatrito, M., Ogrodowski, P., de Stanchina, E., Andrea, A. D', Sander, C. and Ventura, A. 2009. Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Genes & Development* 23(24), pp. 2806–2811.

Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. and Steitz, T.A. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proceedings of the National Academy of Sciences of the United States of America* 98(9), pp. 4899–4903.

O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T. 2005. c-Myc-regulated

microRNAs modulate E2F1 expression. *Nature* 435(7043), pp. 839–843.

Olive, V., Bennett, M.J., Walker, J.C., Ma, C., Jiang, I., Cordon-Cardo, C., Li, Q.-J., Lowe, S.W., Hannon, G.J. and He, L. 2009. miR-19 is a key oncogenic component of mir-17-92. *Genes & Development* 23(24), pp. 2839–2849.

Ota, A., Tagawa, H., Karnan, S., Tsuzuki, S., Karpas, A., Kira, S., Yoshida, Y. and Seto, M. 2004. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Research* 64(9), pp. 3087–3095.

Ryazansky, S.S., Gvozdev, V.A. and Berezikov, E. 2011. Evidence for post-transcriptional regulation of clustered microRNAs in Drosophila. *BMC Genomics* 12, p. 371.

Saini, H.K., Enright, A.J. and Griffiths-Jones, S. 2008. Annotation of mammalian primary microRNAs. *BMC Genomics* 9, p. 564.

Schratt, G. 2009. microRNAs at the synapse. *Nature Reviews. Neuroscience* 10(12), pp. 842–849.

Tahira, A.C., Kubrusly, M.S., Faria, M.F., Dazzani, B., Fonseca, R.S., Maracaja-Coutinho, V., Verjovski-Almeida, S., Machado, M.C.C. and Reis, E.M. 2011. Long noncoding intronic RNAs are differentially expressed in primary and metastatic pancreatic cancer. *Molecular Cancer* 10, p. 141.

Talkington, M.W.T., Siuzdak, G. and Williamson, J.R. 2005. An assembly landscape for the 30S ribosomal subunit. *Nature* 438(7068), pp. 628–632.

Tang, G.-Q. and Maxwell, E.S. 2008. Xenopus microRNA genes are predominantly located within introns and are differentially expressed in adult frog tissues via post-transcriptional regulation. *Genome Research* 18(1), pp. 104–112.

Thomson, J.M., Newman, M., Parker, J.S., Morin-Kensicki, E.M., Wright, T. and Hammond, S.M. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes & Development* 20(16), pp. 2202–2207.

Valinezhad Orang, A., Safaralizadeh, R. and Kazemzadeh-Bavili, M. 2014. Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. *International journal of genomics* 2014, p. 970607.

Ventura, A., Young, A.G., Winslow, M.M., Lintault, L., Meissner, A., Erkeland, S.J., Newman, J., Bronson, R.T., Crowley, D., Stone, J.R., Jaenisch, R., Sharp, P.A. and Jacks, T. 2008. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* 132(5), pp. 875–886.

Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R. and Nishikura, K. 2006. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nature Structural & Molecular Biology* 13(1), pp. 13–21.

Zhang, B., Pan, X., Cobb, G.P. and Anderson, T.A. 2007. microRNAs as oncogenes and tumor suppressors. *Developmental Biology* 302(1), pp. 1–12.

Zhang, L., Zhou, M., Qin, G., Weintraub, N.L. and Tang, Y. 2014. MiR-92a regulates viability and angiogenesis of endothelial cells under oxidative stress. *Biochemical and Biophysical Research Communications* 446(4), pp. 952–958.

CHAPTER 3

TERTIARY STRUCTURE OF INTRONIC PRI-MIR-17-92A REGULATES SPLICING

3. A: Introduction

RNA is a versatile informational molecule. Besides carrying information in the form of its linear primary sequence, RNA can adopt secondary structures by canonical base pairing of complementary nucleotides. These further fold into a compact tertiary structure by invoking non-canonical interactions between sequences that are far apart in one dimension. Secondary and tertiary structured RNA offer a higher level of structural information that influences various steps of gene regulation (Wan et al. 2011; De Conti et al. 2013).

The presence of introns is a defining feature of eukaryotic genes that undergo splicing within the nucleus to remove introns and join exons in a process catalyzed by a multi-mega Dalton ribonucleoprotein complex called the Spliceosome (Will and Lührmann 2011; Rino and Carmo-Fonseca 2009; Valadkhan 2007; Hoskins and Moore 2012). Pre-mRNA splicing can be either constitutive or alternative (Green 1991). During constitutive exon splicing, a dynamic process involving the spliceosome joins the 5′ and 3′ splice sites (SS) that define the exon-intron boundary through the interplay of small nuclear riboproteins (snRNPs) and other associated proteins with the pre-mRNA (Han et al. 2011; Hang et al. 2015; De Conti et al. 2013). The accuracy of splicing is defined by consensus splice sites and regulatory elements such as exon and intron splicing enhancers (ESE, ISE) and silencers (ESS, ISS) (Liu et al. 2010; Wang and Burge 2008). These regulatory elements are recognized by members of the SR and hnRNP family of proteins (Zhou and Fu 2013; Simard and Chabot 2000) **(Figure 3.1)**.

Alternative splicing is an important mechanism for regulating gene expression. It expands the coding capacity of a single gene to produce different proteins with distinct functions (Stamm et al. 2005; Matlin et al. 2005; Chen and Manley 2009; Kelemen et al. 2013). It is now established that close to 90% of human genes undergo alternative splicing and encode for at least two isoforms (Wang et al. 2008; Pan et al. 2008). The divergence observed in gene expression due to alternative splicing may be tissue-specific (Wang et al. 2008; Ellis et al. 2012; Xu et al. 2002) or developmentally regulated (Revil et al. 2010; Wang et al. 2016; Traunmüller et al. 2016). Alternative splicing can occur through various mechanisms such as exon skipping, intron retention, alternative 3′ SS usage, alternative 5′ SS usage, or alternative polyadenylation site usage (Black 2003; Lee and Rio 2015; Green 1991). The information present in the canonical splicing signals that define exon-intron boundaries is not sufficient for correct assembly of the spliceosome for pre-mRNA splicing.
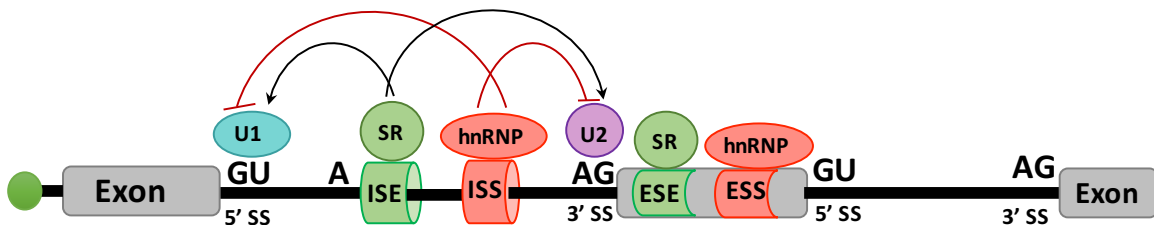
Figure 3.1: Schematic showing *cis* and *trans* regulatory elements in pre-mRNA splicing. Core signals like 5′ SS and 3′ SS are recognized by spliceosomal proteins (U1, U2 are shown), trans-factors like SR proteins and hnRNPA1 proteins bind to enhancers (ESE, ISE). Binding of SR proteins has positive effect on exon recognition by recruiting splicing factors and silencers (ESS, ISS) while hnRNPs inhibit splicing by sterically interfering with other splicing factors.

Additional signals exist in the pre-mRNA as auxiliary *cis*-elements that recruit trans-acting factors to promote splicing. ESE and ISE often bind the serine-arginine rich nuclear factors (SR proteins) to promote the choice of splice sites in the pre-mRNA. Members of the heterogeneous

nuclear ribonucleoprotein (hnRNP) family and polypyrimidine tract binding protein (PTB) bind the ESS and ISS and function as splicing repressors (Ghetti et al. 1992; Paradis et al. 2007; Rahman et al. 2013). Thus, the regulation of alternative splicing is often the result of dynamic orchestrated interplay of *trans*-acting factors binding such *cis*-elements. Interaction between these proteins and spliceosomal components establishes a '*splicing code*' that defines the nature and level of a particular spliced isoform (Matlin et al. 2005; Blencowe 2006; (Barash et al. 2010).

Further, specific cellular stimuli such as cellular metabolic changes, DNA damage, neuronal depolarization and immune signaling can favor the binding of certain trans-factors over others, thereby modifying the splicing pattern (Tarn 2007; Heyd and Lynch 2011; Lynch 2007). As most of the pre-mRNA splicing is co-transcriptional, an additional level of splicing regulation occurs through coupling transcription to pre-mRNA processing. The C-terminal domain of RNA Polymerase II (Pol II) large subunit acts as a landing pad and provides a platform for recruitment of RNA processing factors to act on the emerging pre-mRNA transcript (Maniatis and Reed 2002; Kornblihtt et al. 2004; (Naftelberg et al. 2015). This co-transcriptional regulation of splicing has been studied extensively (Goldstrohm et al. 2001; Shukla and Oberdoerffer 2012; Aitken et al. 2011; Pandya-Jones and Black 2009; Lai et al. 2013).

Secondary structures within the pre-mRNAs are also known to regulate splicing especially those transcripts that can undergo alternative splicing (Warf and Berglund 2010; McManus and Graveley 2011; May et al. 2011). However, there are only few reports of RNA secondary structure influencing pre-mRNA splicing **(Figure 3.2)**. There are two major functional mechanisms by which RNA secondary structure influences splicing. The first mechanism is by occlusion or exposure of primary *cis*-acting elements that results in modulation of their accessibility to splicing

94

factors. The classical examples are that of human *Tau* gene where formation of a hairpin structure interferes with the recognition of the donor splice site (Grover et al. 1999; Jiang et al. 2000; Kar et al. 2011). **(See Figure 3.2A)** The drosophila *Adh* gene is another example wherein formation of a stem structure in the intron organizes the branch point into a single stranded conformation, promoting its use (Chen and Stephan 2003). The second mechanism is one where structural changes in the spatial positioning of *cis*-acting elements with respect to each other **(Figure 3.2B)**.

A

splice site suppression

B

masking/unmasking of *cis*-elements

C

cis-elements
in proximity

steric
hindrance

D

RNA looping-out

E

competitive RNA structures
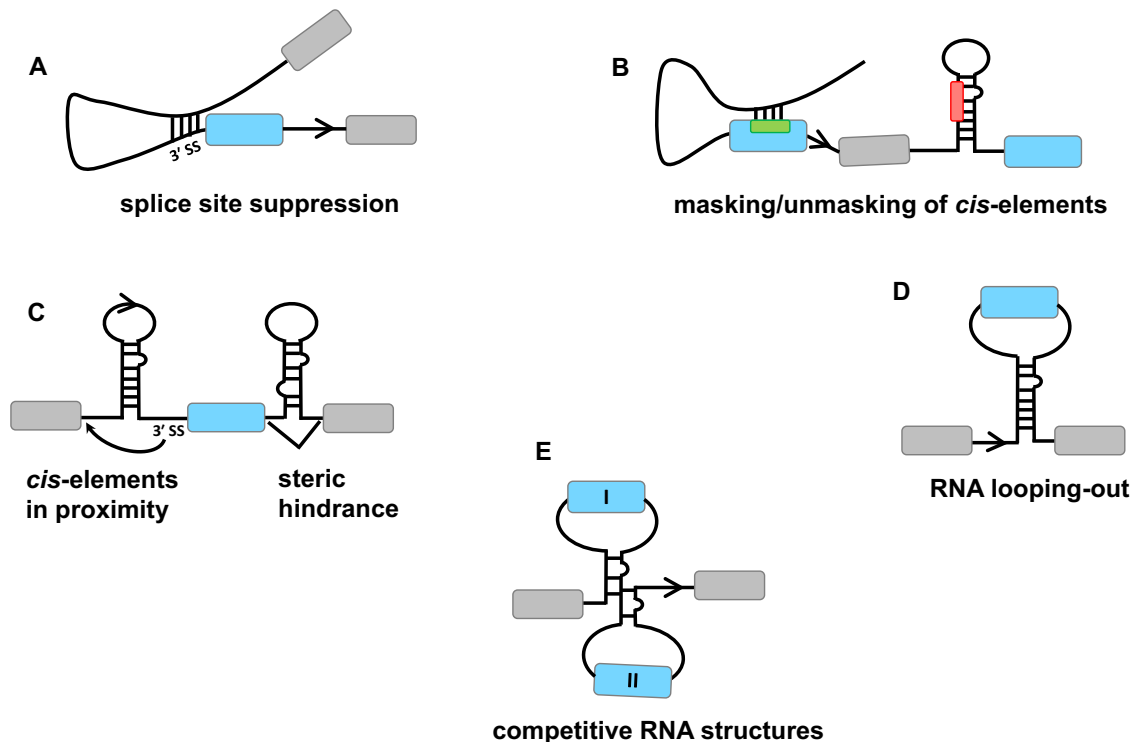
Figure 3.2: Schematic to show the regulatory role of RNA secondary structure in splicing. The following functional mechanisms are represented **A**. splice site suppression **B.** masking or unmasking *cis*-elements, ESS, ESE, ISS, ISE **C.** *cis* elements can be brought into close proximity or result in steric hindrance **D.** RNA looping-out mechanism. **E.** Formation of competitive RNA structures.

Tropomyosin gene from chicken and *Dystrophin* gene from human are shown to form RNA loop structures that triggers the spliceosome to remove the intron (Matsuo et al. 1992; Sirand-Pugnet et al. 1995). Formation of such loop structures results in spatial reorganization of distant

cis elements and bringing the splice sites enclosed in the loop closer (Nasim et al. 2002) **(Figure 3.2C)**. The most well studied gene that is spliced by looping out mechanism is the *Drosophila melanogaster* Down Syndrome Cell Adhesion Molecule (*Dscam*) gene in Drosophila **(Figure 3.2D)**. *Dscam* contains 95 alternatively spliced exons that encode ~38,000 distinct isoforms where competing RNA structures regulate alternative splicing of ~48 mutually exclusive exons (May et al. 2011) **(Figure 3.2E)** Computational analysis reveals ~200 highly conserved RNA secondary structures in introns of Drosophila genes (Raker et al. 2009) indicating that RNA secondary structure based modulation of splicing could actually be more widespread than currently assumed.

Dysregulated splicing arising from altered splice-site choice results in several human diseases. Splicing mutations described as early as the discovery of splicing was, in Hemoglobin sub-unit beta (HBB) which encodes for β-globin where a point mutation in the intron generates an alternative 3′ SS that results in a condition β-thalassemia that is characterized by reduced β- globin protein causing severe anemia (Spritz et al 1981). The other recent examples include: splice site mutations in dystrophin (DMD), which result in loss of dystrophin function and causes Duchenne muscular dystrophy (Fletcher et al. 2013; Wein et al. 2014), polymorphic UG and U tracts near the 3′ SS of cystic fibrosis transmembrane conductance regulator (CFTR) exon 9, resulting in cystic fibrosis (Chu et al. 1993) and ESE, ESS, and 5′ SS mutations in microtubule-associated protein tau (MAPT) exon 10, which causes frontotemporal dementia with parkinsonism linked to chromosome 17 (Grover et al. 1999; Niblock and Gallo 2012) and Lamin A/C (LMNA) gene resulting in Hutchinson-Gilford Progeria Syndrome, a rare genetic disease (Lopez-Mejia et al. 2014; Chatzispyrou and Houtkooper 2014).

There is a wealth of information on regulatory signatures within introns that dictate splicing in terms of *cis*-elements such as primary sequence, and *trans* elements such as proteins. Yet it is not known whether the next level of information coded by RNA, i.e., its tertiary structure, can regulate splicing. Hence, I wanted to investigate if the tertiary structure of intronic polycistronic microRNA cluster, pri-miR-17-92a can influence the splicing of its host transcript. I reasoned that microRNAs would be particularly relevant as potential regulators of splicing in that 80% of the human microRNA loci genes have been reported to be located within the intronic regions of coding or non-coding transcription units and are transcribed along with their host genes (Rodriguez et al. 2004; Kim and Kim 2007).
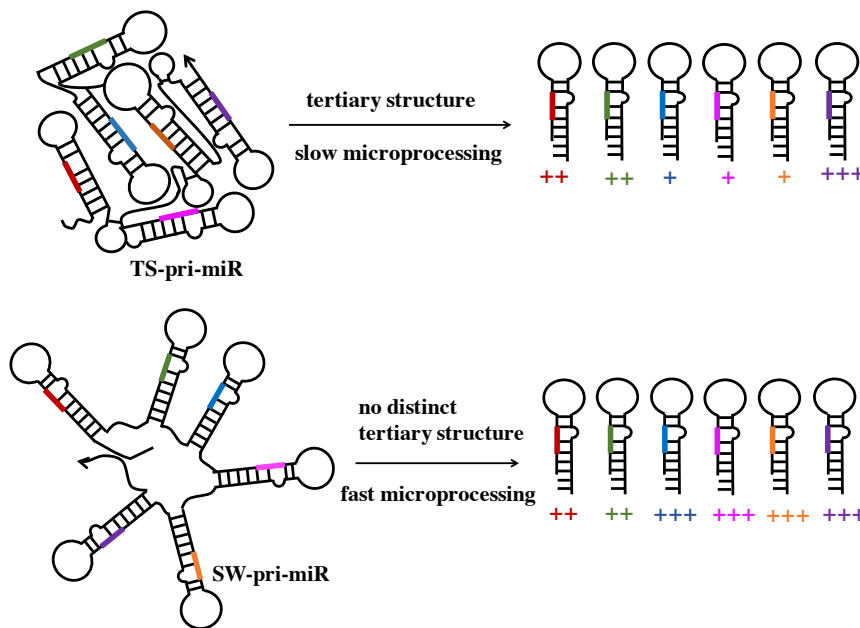
Figure 3.3: Schematic showing that in TS-pri-miR the tertiary structure of pri-miR-17-92a kinetically regulates slow microprocessing to allow differential release of pre-miRNAs, while SW-pri-miR which is devoid of such tertiary structure is processed much faster to give high abundances of its pre-miRNAs.

In the chapter 2 of my thesis I have shown that, the intronic pri-miR-17-92a cluster, harboring six miRs in quick succession, folds into a distinct tertiary structure (TS) and that this 'TS' regulates

its processing into its constituent mature miRNAs by slowing down the processing kinetics by 4-fold compared to a swap pri-miR (SW) which lacks such as distinct tertiary structure **(Figure 3.3)**.

Bioinformatic analysis of 10 different human microRNA clusters have previously shown that not only the regions of individual pre-miRNAs are conserved but the 'non-functional' inter pre-miRNA regions are also conserved (Chakraborty et al. 2012). These regions were shown to be conserved as they were crucial for 3D-structure of the pri-miRNA. Given that 40% of human pri-miRs are clustered (Yuan et al. 2009; Griffiths-Jones et al. 2008), we hypothesized that the structure of clustered pri-miRNAs could potentially could function to regulate splicing of the host gene **(Figure 3.4)**.
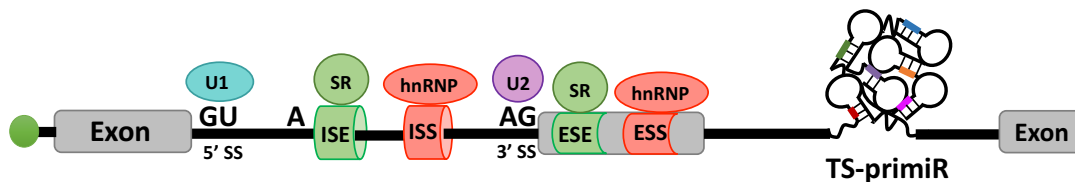
Figure 3.4: Schematic representation of the *cis*-splicing and auxiliary regulatory elements that facilitate exon recognition by *trans*-binding factors in the first intron and a tertiary structured pri-miR (TS-pri-miR) in the second intron.

Studies on processing of intronic miRNAs have shown that splicing is not a prerequisite for miRNA processing and that unspliced introns can also serve as the substrate for the microprocessor complex (MPC). This was consistent with continuous introns being unnecessary for successful pre-mRNA splicing. Co-transcriptional cleavage of the intron by insertion of a cleavage element (CoTC) or ribozyme into the intron of a RNA transcript resulted in efficient splicing of the flanking exons suggesting that the preceding exon may be tethered to the elongating Pol II complex until the next exon emerges (Dye et al. 2006) **(Figure 3.5)**.
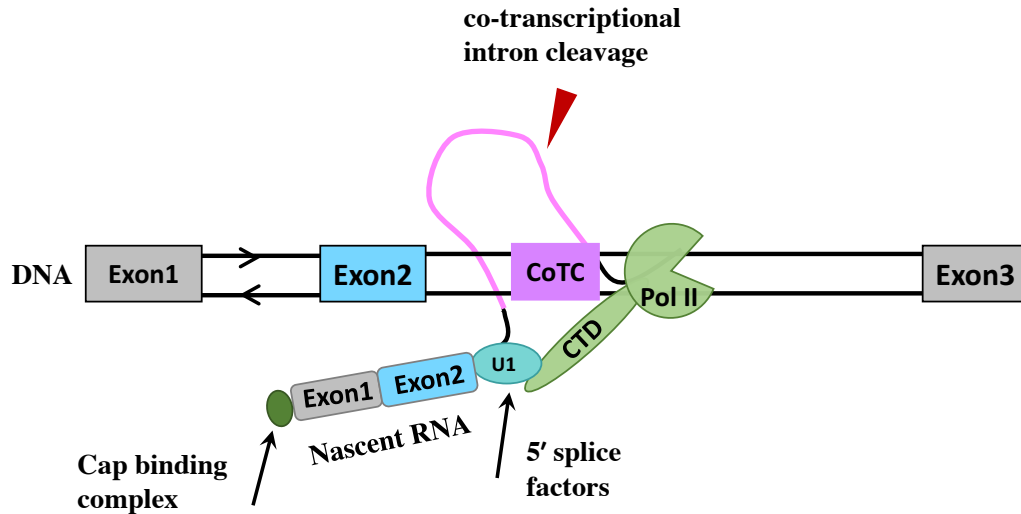
Figure 3.5: Schematic to show exon tethering model. Co-transcriptional cleavage of intron by cleavage element (CoTC) or ribozyme results in efficient splicing due to tethering of flanking exon to C-terminal domain (CTD) of Pol II.

RNA splice site choice is likely to be regulated kinetically. Recognition of splice sites could be masked by proteins or presented in a different order, all of which can guide the splicing outcome. This is revealed by the influence of kinetics of cleavage on splicing. Insertion of fast or slow self-cleaving ribozymes within the intron of β-globin transcript show that fast cleavage leads to impaired splicing due to inhibition of co-transcriptional assembly of the spliceosome, while slow cleavage allows for efficient assembly, leading to proper exon tethering that results in effective pre-mRNA splicing (Fong et al. 2009) **(Figure 3.6)**.
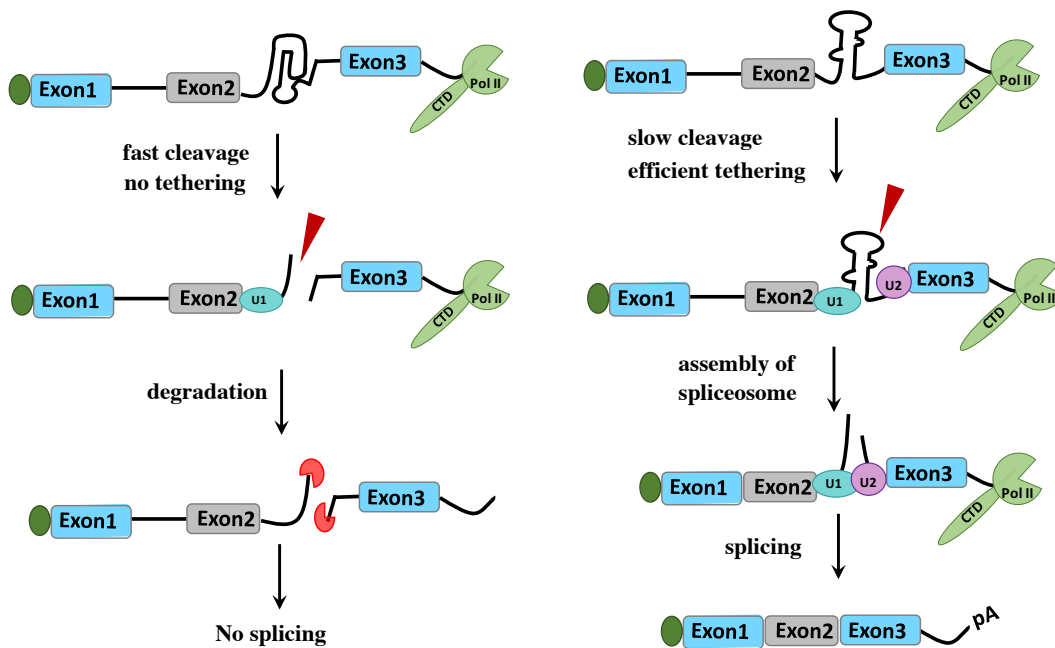
Figure 3.6: Schematic showing the influence of kinetics on intron cleavage. Fast cleavage by ribozyme inhibits exon tethering of RNA transcript to Pol II and subsequent degradation by exonucleases and thus abolishes RNA processing. Slow intron cleavage allows the assembled spliceosomal components to remain tethered to Pol II via the exons. As the spliced site contacts are already paired, this results in efficient splicing of the reporter RNA.

Kim *et al* showed that the adjacent introns are spliced more rapidly than introns encoding miRNA. A construct with mutated splice sites showed reduced levels of pre-miR suggesting that splicing may facilitate the first step of microRNA processing (Kim and Kim 2007). Based on their studies they proposed a model for the processing of intronic miRNAs where the exons flanking the miRNA are held to each other by a splicing commitment complex (CC). The MPC cleaves the intron to release the pre-miRNA and since exons are tethered, the spliceosome can bring about *trans* splicing (bimolecular ligation) ensuring the miRNA biogenesis and formation of mRNA (**Figure 3.7**).
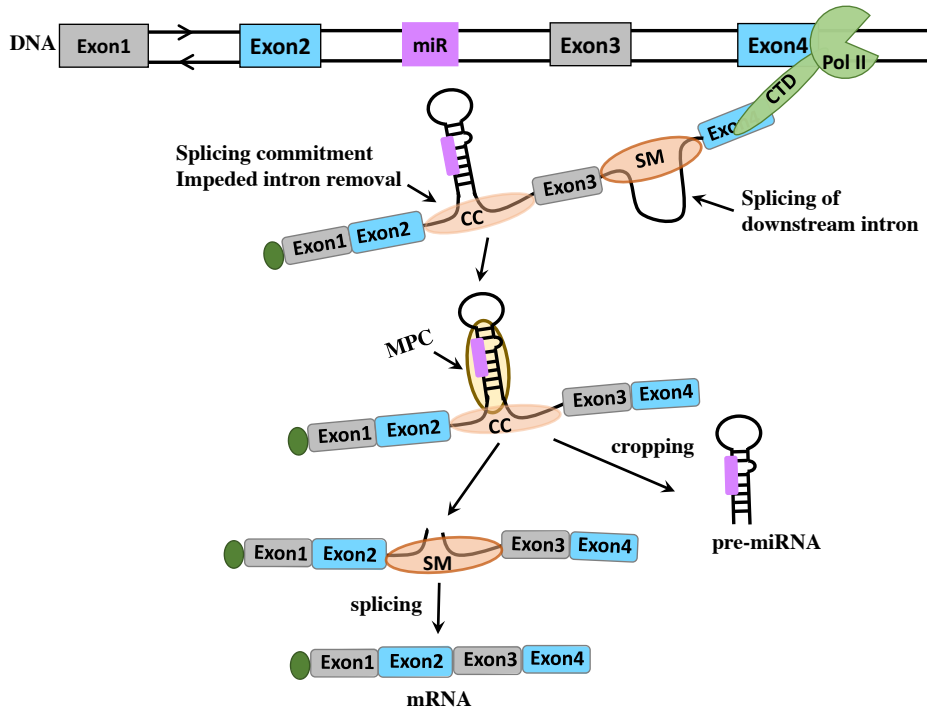
Figure 3.7: Schematic representation of processing of intronic microRNAs. The exons flanking the microRNA are bound to the splicing commitment complex (CC). Splicing of miRNA harboring intron is impeded while downstream introns are rapidly spliced. Microprocessor (MPC) cleaves the intron to release its pre-miRNA. As the exons are paired and tethered, splicing of the intron would occur efficiently.

In vitro studies with nuclear extracts where pri-miR processing and mRNA splicing were simultaneous detected revealed that cropping of the pri-miRNA into pre-miRNA occurred prior to mRNA splicing and impeding splicing of the miR-containing intron facilitated pre-miR production (Kataoka et al. 2009).

Immunoprecipitation with an anti-Drosha antibody precipitated the RNA components: pre-miR, a Drosha-cleaved product, a Y-shaped intron and a Y-shaped splicing intermediate, indicative of *trans* splicing and the proteins U2, U5, and U6 but not U1 or U4 snRNPs indicating the cropping happens at the step of formation of B complex in the assembly of the spliceosome. The model

proposed suggested that the MPC associates with the spliceosome, allowing production of two

different functional RNAs from one pre-mRNA molecule (Kataoka et al. 2009) **(Figure 3.8).**
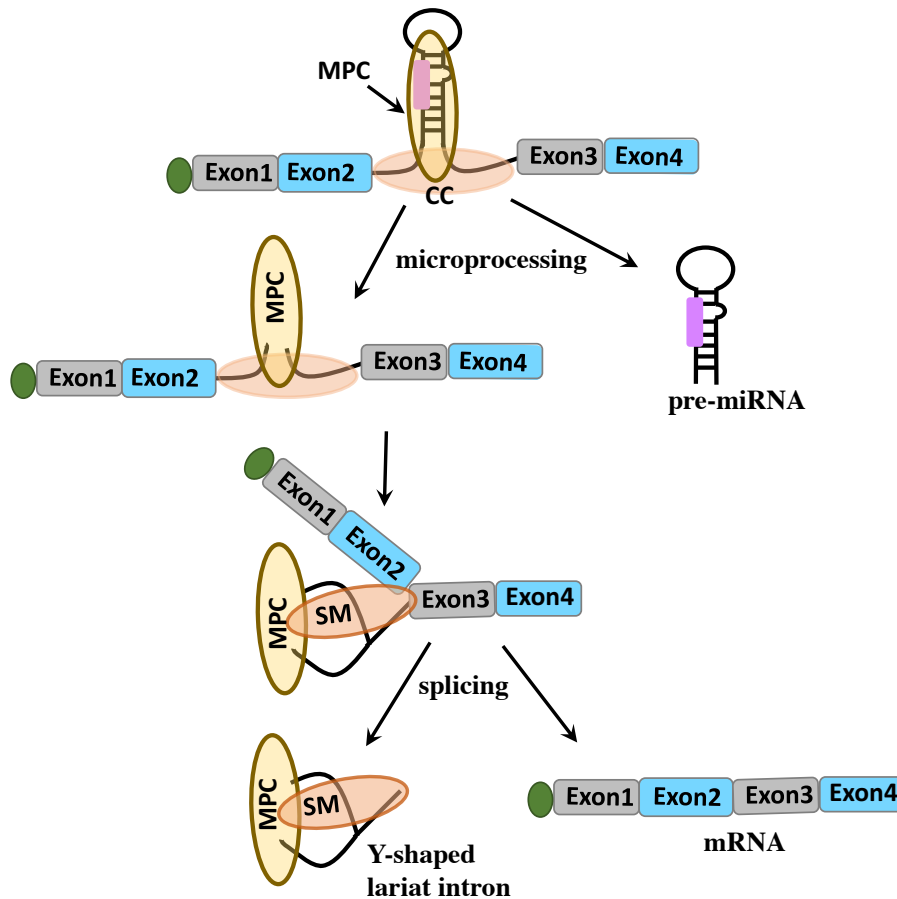
Figure 3.8: Model showing functional association of microprocessor (MPC) and spliceosome (SM). MPC containing Drosha associates with SM on introns harboring microRNAs and results in the removal of pre-miRNA followed by splicing of the pre-mRNA. Pre-miRNA, cropped pre-mRNA fragment and Y-shaped lariat intron containing products are associated with Drosha.

Co-immunoprecipitation identified that several RNA binding proteins and RNA helicases

that were components of the MPC were involved in pre-mRNA splicing (Gregory et al. 2004; Lee

et al. 2006). Micro processing of miR-211 located in intron6 of melastatin gene promoted the

splicing of exon6-exon7 by a mechanism that required cleavage by Drosha (Janas et al. 2011).

Mutations in the 5′ SS but not in 3′ SS, BP or polypyrimidine tract of intron6, reduced miR-211

biogenesis indicating that 5′ SS recognition by the spliceosome promotes cropping of its miR. Knockdown of U1 spliceosomal component reduced expression of intronic miRNAs but not intergenic miRNAs. Based on these studies a feed-forward regulation between miRNA processing and splicing was proposed, whereby 5′ SS recognition by the U1 complex promotes miRNA processing of intronic miR-211 by Drosha and this cropping of miR-211 promotes splicing at its host melastatin intron 6 (Janas et al. 2011) **(Figure 3.9)**.
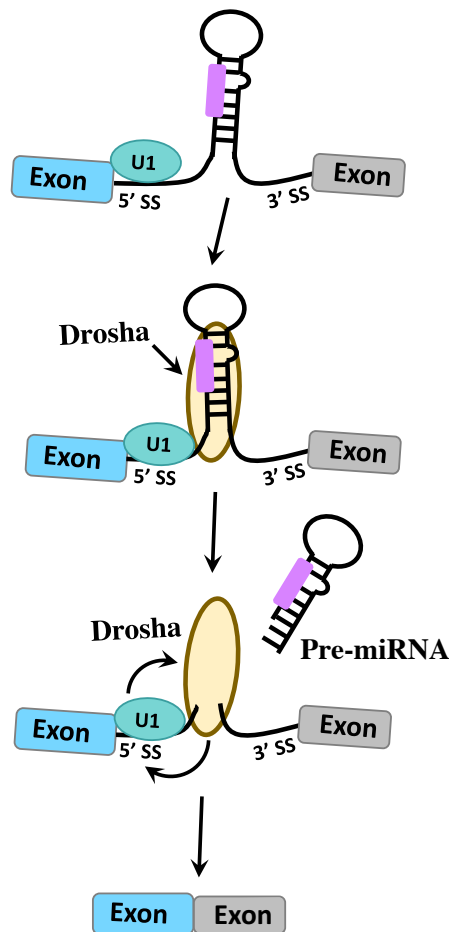
Figure 3.9: Schematic showing feed-forward regulation of microprocessor and spliceosome. At intronic miRNA loci, the U1 snRNP of the spliceosome binds to 5′ SS and promotes the recruitment of Drosha, a microprocessor component and leads to increased microprocessing of the pre-miRNA hairpin. This microprocessed intron further aids in promoting the splicing reaction. Mutually co-operative activities of microprocessor and spliceosome results in increase exon-exon junction at intronic microRNA loci.

These studies clearly revealed the crosstalk between the microprocessor and spliceosome that resulted in the production of two different functional RNAs: (i) regulatory miRNAs and (ii) spliced mRNA from a single pre-mRNA transcript. However, the contribution of tertiary structure to influence the splicing needed investigation. I therefore decided to investigate the effect of this tertiary structured pri-miR 17-92a (TS-pri-miR) in regulating the splicing of the host gene.

To address this hypothesis, I first chose a β-globin reporter gene for two important reasons: (i) The regulatory elements of splicing are quite well established (Marotta et al. 1974; Green et al. 1983; Leach et al. 2003) (ii) It undergoes efficient *in vitro* splicing (Krainer et al. 1984) and the mini gene system for the β-globin (intron1 flanking with exon1 and exon2) has been the gold standard for splicing. Mutations in the splice site of β-globin have been implicated in hematological disorders like β-Thalassemia and have significant medical relevance. Further, it was interesting that when I began my thesis *in vitro* studies with large RNAs were not well established due to the experimental challenges concerning the stability of such RNAs. Having tested my hypothesis with β-globin, I then addressed the effect of TS-pri-miR on its native host gene *C13orf25* which is alternately spliced. To do this, I constructed C13orf25 reporter having its endogenous pri-miR-17-92a or a swapped pri-miR lacking a distinct tertiary structure. In this chapter I outline the details of this study to show that the tertiary structured pri-miR-17-92a impedes the splicing of its host transcript. Using overexpression studies, I have addressed the effect of this tertiary structure on removal of both the introns of a β-globin reporter transcript and optimizing an *in vitro* splicing assay on these transcripts that contain large introns I have addressed the kinetics of splicing. The presence of tertiary structured pri-miR in a constitutively spliced β-globin showed that it impedes splicing suggesting that effective miRNA processing is required for

splicing. This could be a general mechanism by which intronic pri-miR clusters could regulate splicing. In its endogenous context of C13orf25, this tertiary structured pri-miR is important in regulating the abundance of various spliced forms that is a reflection of altered splice site choices of the host transcript.

Design of the system

The tertiary structured pri-miR-17-92a (TS-pri-miR), a polycistronic miRNA cluster is part of a non-coding RNA called C13orf25 (*Chromosome 13 open reading frame*) also referred as *MIR17HG* (MIR17 host gene) (Ota et al. 2004) located on chromosome 13 (13q31). This 6.7 kb transcript is known to undergo alternative splicing to give two variants: spliced variant1 (Sp1) of 0.9 kb and spliced variant 2 (Sp2) of ~5.6 kb. Sp1 has all three introns removed to ligate all four exonic sequences whereas Sp1 has only intron1 spliced and retained intron2 and intron3 **(Figure 3.10)**. However, no functional protein has been reported from Sp1 and the only function of this gene has been to give rise to the six component miRNAs (miR-17, miR-18a, miR-19a, miR-20a, miR-19b, miR-92a) processed from the primary microRNA transcript, pri-miR-17-92a located in its third intron (Tagawa and Seto 2005).
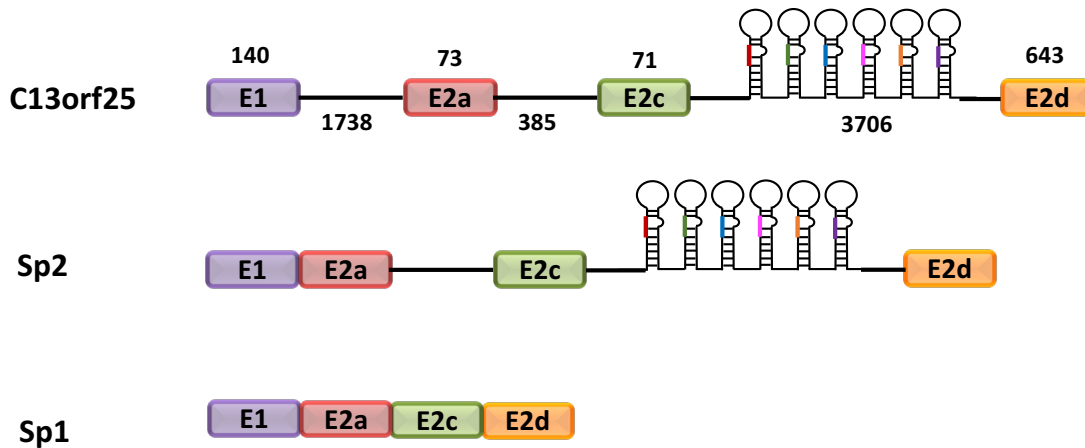
Figure 3.10: Genomic structure of *C13orf25* gene harboring the intronic pri-miR 17-92a cluster. C13orf25 is a non-coding RNA undergoes alternative splicing to form two variants Sp1 and Sp2.

Given that 40% of pri-miRNAs are clustered we hypothesized that splicing of many host transcripts could be regulated by their intronic tertiary structured miRNAs. Fong *et al* have previously reported the influence of kinetics on splicing of β-globin gene engineered with self-cleaving ribozymes in its intron. Hence I chose the β-globin reporter gene and inserted the pri-miR-17-92a (TS-pri-miR) in its intron2 to address the influence of tertiary structure on its splicing and this system is described below (See **Figure 3.11).**

β-globin pre-mRNA (~1.6 kb) is a three exon and two intron reporter system that undergoes constitutive splicing in two steps to remove intron1 and intron2 to give 0.6 kb β-globin mRNA (**Figure 3.11**). The sizes of exonic and intronic regions are as follows: exon1 (E1)-142 nt, intron1 (I1)-130 nt, exon2 (E2)-223 nt, intron2 (I2)-850 nt and exon3 (E3)-261 nt (as confirmed from the sequencing analysis of β-globin plasmid).
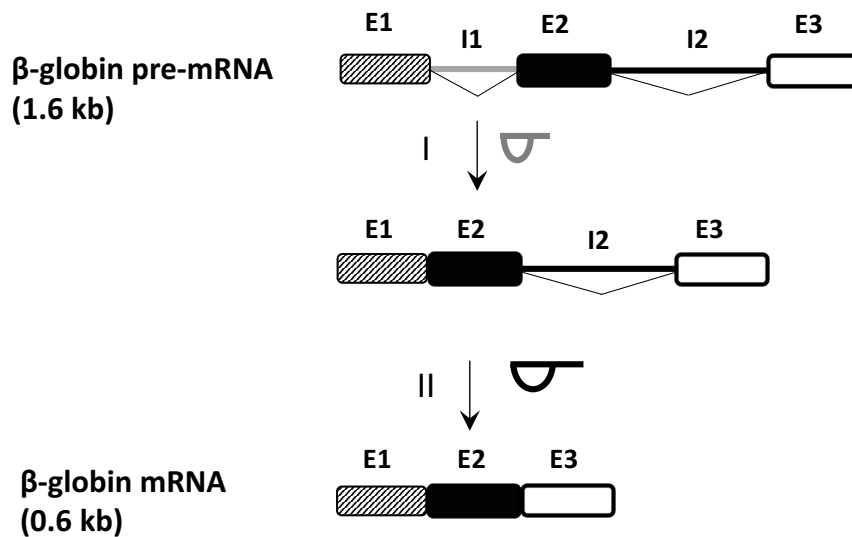
Figure 3.11: Schematic of the β-globin reporter gene. Exons are shown in rectangles; grey and black lines represent introns. β-globin pre-mRNA undergoes constitutive splicing in two steps to remove the introns and form its mRNA.

3.B: Materials and Methods

3. B.1: Plasmids

3. B.1.1: Cloning of β-globin reporter constructs

The β-globin WT containing human β-globin fragment cloned into pCI-Neo expression vector (Bird et al. 2004) was a kind gift from Prof. David Bentley's Lab (University of Colorado, Colorado, US). This plasmid contains a CMV promoter for expression in mammalian cell lines and in my thesis, this construct would be referred to as **pCMV-β-globin WT** in this thesis. The pri-miR-17-92a was cloned into TOPO-pCR2.1 as described in 2.B.1 Briefly, the genomic DNA corresponding to pri-miR-17-92a was PCR amplified by Pfu DNA Polymerase using specific primers. Swap-pri-miR sequence was generated by overlap extension PCR and was cloned

similarly. A 0.8 Kb DNA fragment of pri-miR-17-92a was PCR amplified with a forward and reverse primer containing Mfe I site, digested and inserted into the Mfe I site present in the intron2 of β-Globin WT DNA to construct **pCMV-β-globin TS-pri-miR** plasmid. **pCMV-β-globin Swap-pri-miR** plasmid was constructed by PCR amplifying the Swap-pri-miR sequence with specific primers having Mfe I sites and inserted into intron2 of β-globin WT. The single hairpin domain region of 100 bp corresponding to pri-miR-17 sequence of the pri-miR-17-92a cluster was similarly inserted into the Mfe I site to construct **pCMV-β-globin HP**-miR plasmid. The identity of all the constructs was confirmed by sequencing. The list of the primers used in making the plasmids is given in Table 3.1.

3. B.1.2: Cloning of C13orf25 gene

A 4.5 kb C13orf25 DNA fragment containing the entire exon2a, intron2, exon2c, intron 3 and 5′ end of the exon2d was amplified from the BAC clone RP11-94P6 (Children's Hospital Oakland Research Institute) using C13orf25 Full FP and RP with Long Amplification Taq DNA Polymerase (New England Biolabs). The sizes of the C13orf25 gene with pri-miR 17-92a cluster are as follows: E2a-73 bp, In2-386 bp, E2c-71 bp, In3-3701 bp, E2d-293 bp (full size-4528bp). The fragment was inserted into pcDNA3.1-Topo-His/V5 vector (Invitrogen) to construct **pcDNA3-C13orf25-TS-pri-miR**. The pri-miR-17-92a cluster in the intron 3 was replaced by swapped miRNA cluster described in 2.B.1 by Long Multiple fusion (LMF) (Shevchuk et al. 2004; Heckman and Pease 2007) using Phusion Hot start Taq Polymerase (Thermo Scientific) to construct **pcDNA3-C13orf25-Swap-pri-miR. C13orf25-inv-pri-miR** was made by replacing the pri-miR-17-92a sequence with its reverse and complementary sequence. A deletion construct, **pCDNA3-**

C13orf25-Δpri-miR, was made by Splicing by Overlap Extension (SOE) (Senanayake and Brian 1995; Heckman and Pease 2007) to delete the 0.8 kb pri-miR region using Long Amplification Taq DNA Polymerase (New England Biolabs). The list of primers used to make C13orf25 constructs is given in Table 3.2.

3. B.2: Transformation and plasmid DNA isolation

Chemically competent DH5α cells were made using Inoue Method (Molecular Cloning, Sambrook and Maniatis). Competent cells were thawed on ice, then incubated with 20 ng of cloned plasmid DNA, mixed gently and incubated on ice for 30 min. The cells were then subjected to heat-shock for 20 secs at 42° C and placed on ice for 2 min. The transformation volume was then made up to 1ml with pre-warmed LB medium and shaken for 1 hr at 37° C. The cells were spun at 3000 rpm for 5 min, supernatant was discarded and cells were resuspended in 200 μL of media and were spread on LB agar plates containing a selective antibiotic (100 μg/mL) and incubated overnight at 37° C. Additionally, negative (no plasmid DNA) and positive plasmid control transformations were carried out. Individual colonies were picked for further analysis.

Generating large quantities of a plasmid generally involved the inoculation of an individual colony of *E. coli* DH5α into 5 mL of LB media cultured at 37° C for 2 h and this starter culture was used to inoculate 100 mL of antibiotic selective LB media overnight 37° C. Transformed bacterial cultures were centrifuged to yield a pellet of bacteria (6000 g for 5 min at 4° C) from which the plasmid DNA was extracted. Plasmid Maxi kit (Qiagen) was typically used for the extraction of plasmid DNA from 100 mL bacterial cultures, according to the manufacturer's

recommended instructions. The plasmid DNA was resuspended either in nuclease free water for transfections or in TE buffer pH 7.6 for long term storage.

3. B.3: Cell culture and Transfection

HeLa cells were maintained in DMEM medium supplemented with 50 IU ml^{-1} of penicillin, 50 mg ml^{-1} streptomycin and 10% (v/v) FBS. Cells were cultured overnight at 37° C in 5% CO_2 to a confluency of 70% and then transiently transfected with β-globin and its related pri-miR expression plasmid using Xfect Polymer transfection reagent (Clontech) following the manufacturer's instructions. Typically, 20 μg of plasmid DNA was used for cells grown in 10 cm cell culture dish and 0.3 volumes of transfection polymer was used. The transfected medium was replaced after 4 h of incubation and replaced with fresh medium and the cells were collected 24 h post transfection and used for RNA isolation.

3. B.4: Isolation of RNA from HeLa cells

3. B.4.1: Isolation of total RNA

Total RNA from the transfected HeLa cells was isolated using Trizol reagent (Invitrogen) according to manufacturer's instructions with the following additional steps. Samples were extracted with an equal volume of Tris-Cl buffered phenol-chloroform pH 4.5 (Sigma) after the standard chloroform extraction and prior to precipitation with an equal volume of isopropanol. The nucleic acid pellets were resuspended in nuclease free water (Ambion) and treated with 1U of TurboDNase (Ambion) per 20 μg of RNA at 37° C for 30 min. The RNA was then extracted with an equal volume of Tris-Cl-buffered phenol-chloroform pH 4.5 (Sigma) and precipitated with 2

volumes of absolute ethanol and one-tenth volume of 3M sodium acetate pH 5.6. The RNA was resuspended in nuclease free water.

3. B.4.2: Isolation of nuclear RNA from HeLa cells

For isolation of nuclear RNA nuclear-cytoplasmic fractionation of HeLa cells was carried out using a standard procedure. Briefly, post-transfection the cells were collected and washed with 1x PBS pH 7.4. 500 μL of hypotonic buffer (10mM HEPES pH 7.9) was added to the dish and aspirated out. 400 μL of ice cold hypotonic buffer was added onto the dish of cells and allowed stand for 10 min. The buffer was removed and 300 μL of sucrose-EDTA solution was added and the cells were gently scraped into an eppendorf tube. A 1 mL disposable syringe with a 25-gauge x 5/8-inch needle is filled with Hypotonic buffer and emptied again by ejection to remove the air from the syringe barrel and the needle. The cell suspension is drawn slowly into the syringe and then ejected back into the tube with a rapid push on the plunger (10 strokes). The crude nuclei are collected by centrifugation at 3300 g for 10 min. The supernatant contained the cytoplasm and the pelleted nuclei were used for proceed with isolation of nuclear RNA using Trizol. The RNAs isolated were quantified by UV spectroscopy.

3. B.5: *In cellulis* splicing of β-globin and its related pri-miR transcripts

3. B.5.1: RT-PCR analysis

RT-PCR analysis was used to assess the steady state levels of spliced β -globin mRNA from HeLa cells transfected with WT reporter gene or reporter harboring relevant pri-miRs. 2.5 μg of total RNA or 1 μg of nuclear RNA isolated from transfected HeLa cells was reverse transcribed using E3R primer which is complementary to Exon3 of the globin transcript and SuperScript III reverse transcriptase (Invitrogen) in a 20 μL reaction according to the manufacturer's instructions. Briefly

the RNAs were denatured at 80° C for 2 min and transferred to an ice-bath for 5 min. First-strand mix containing 0.5 mM dNTP mix, 0.25 mM RT primer and 50 U of SSIII reverse transcriptase was added and the reactions were incubated at 37°C for 1 h followed by 42° C for 30 min. Control reactions (-RT) reactions were performed without the reverse transcriptase. 1 μL of the reaction mix was amplified by PCR using Taq DNA Polymerase (New England Biolabs) for 30 cycles using the primers specific for spliced or unspliced forms as indicated in **Table 3.3.** 18S rRNA was amplified as control and Neo (Neomycin phosphotransferase gene) amplification serves as control for transfection. PCR products were resolved on 1% agarose gels unless otherwise specified and visualized by ethidium bromide staining. PCR using a primer set specific for TS-pri-miR present in the intron2 of β-globin TS transcript and SW-pri-miR in the β-globin SW-pri-miR construct was used to study their relative abundances.

3. B.5.2: Northern Blot

For quantification of spliced β-globin product using Northern blot analysis, 10 μg of RNA was electrophoresed through 1.2% agarose-formaldehyde gels in 1x MOPS buffer pH 8.0 (20 mM MOPS pH 7.0, 0.2 mM sodium acetate, 1 mM EDTA pH 8.0). The long RNAs were efficiently transferred overnight onto a positively charged Hybond-XL membrane (GE Healthcare Life Sciences) by conventional capillary transfer using 5x SSC (20x SSC buffer: 3M Sodium Chloride, 0.3M Sodium Citrate). After UV cross linking at 254 nm, 1200mJ using Strata linker (Strata gene, USA) the membranes were hybridized with 5′ P^{32}-labeled probes overnight at 42° C in 3 mL of Oligo Hybridization Buffer (Ambion). The membranes were washed 1x 10 min with 5 mL of wash buffer I (1x SSC, 0.1% SDS) at 37° C and 2x 10 min with wash buffer II (0.1x SSC, 0.1%SDS).

The membranes were covered with saran wrap and subjected to phosphor imaging. The images were recorded using a FLA-5000 system (Fuji) with Image Quant software.

For making radiolabeled probes, the DNA probes complementary to two consecutive sites on exon3 (E3-1, E3-2) and on intron2 (I2-1, I2-2) were designed and subjected to a kinase reaction using γ^{32}P-ATP (Perkin Elmer) and T4 Polynucleotide Kinase (New England Biolabs) at 37° C for 1 h. The labeled probes were purified using Biospin P6 Gel column (Bio-Rad) according to the manufacturer's protocol to remove the unincorporated radionucleotides.

3. B.6: *In vitro* transcription

To generate RNAs for *in vitro* experiments, template plasmids containing DNA sequences for the required RNAs were amplified using Phusion Hot start II High Fidelity DNA Polymerase (Thermo Scientific). The forward primer contained T7 promoter sequence and reverse primer incorporated $(A)_{10}$ to increase the stability of the transcripts. The PCR-generated DNA was purified with a PCR cleanup kit (Promega) and eluted in nuclease free water to be used as template. *In vitro* transcription was carried out using 5 μg/mL template DNA and T7 RNA Polymerase (New England Biolabs) using manufacturer's instructions. Reactions were supplemented with 1 U/ml of Thermostable inorganic pyrophosphatase (New England Biolabs), 400 U/ml of RNase inhibitor (Roche) and incubated for 3 h at 37° C. RNA Cap Structure Analog (New England Biolabs) was added to the reaction to generate 5′ capped RNAs. Following transcription, the mixture was incubated with 20 U/ml TurboDNase (Ambion) for 30 min at 37° C. A small aliquot of the reaction was precipitated and resuspended in nuclease free water and the integrity of RNA was checked on

denaturing agarose gel as indicated in 3. B.8. A small scale of capped RNAs were also generated using mMessage mMachine T7 kit (Ambion) as control to check for size of the RNAs.

3. B.7: Native Purification of RNAs

To retain co-transcriptional folding of the long RNAs, established native purification methods were followed (Chillón et al. 2015). Briefly, post transcription reactions were diluted with Filtration buffer (1x Filtration buffer: 10 mM Tris-Cl pH 7.4, 5 mM $MgCl_2$, 100 mM KCl) and purified by subjecting it to Amicon Ultra-0.5-50 kDa MWCO centrifugal filters (Millipore) and centrifuged at 4000 g for 15 min at room temperature. The column filtered products were purified by HiLoad Superdex 200PG AKTA Chromatography (GE Life Sciences) using 1x filtration buffer as running buffer. The columns were washed with 5 column volumes (CV) of Millipore water and equilibrated with 10 CV of filtration buffer before each run. The large RNAs were eluted at 40-45 mL while the unincorporated NTPs eluted after 100 min on the column. The peak corresponding to the RNA transcripts were collected and were concentrated using 3 kDa amicon. The concentration of the RNA was estimated by measuring the absorbance at 260 nm.

3. B.8: Gel electrophoresis of RNAs

The *in vitro* transcribed RNAs purified under native conditions were electrophoresed in gels containing 1% agarose. The running buffer was 0.5x TBE (90 mM Tris, 90 mM boric acid and 2 mM EDTA) with 5mM $MgCl_2$. Prior to loading, 6x loading dye (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/w) glycerol in H_2O) was added to the samples to a final concentration of 1x. The RNAs were electrophoresed for 1 h at 35 V and 10 mA. The gels were

visualized by staining with ethidium bromide. Riboruler (RL), High Range (Thermo Scientific) and 1kb DNA ladder (L) were loaded as molecular markers.

To check the integrity and the lengths of the *in vitro* transcripts, the RNAs were resolved on 1% agarose-formaldehyde gel using standard protocol. Briefly, 1 g agarose was mixed with 72 ml of sterile water and boiled. After cooling to 60° C, 10 mL of 10X MOPS buffer (0. 2M MOPS (3-(N-Morpholino) propanesulfonic acid pH 7.0), 5 mM sodium acetate pH 6.0, 10 mM EDTA pH 8.0) and 18 mL of 37% (12.3M) formaldehyde and 5 μL of ethidium bromide solution (10 mg/mL) was added. Before loading the RNA was mixed with Gel Loading buffer II (Ambion) and denatured by heating at 80° C for 5 min and snap cooling on ice. The running buffer was 1x MOPS buffer and the gel was electrophoresed for 2 h at 100 V and 15 mA.

3. B.9: Preparation of splicing competent nuclear extracts

Nuclear extracts active in pre-mRNA splicing were prepared by few changes of the procedure originally reported (Dignam et al. 1983). The following protocols reported were considered to make efficient extracts (Zerivitz and Akusjärvi 1989; Kataoka and Dreyfuss 2008; Webb and Hertel 2014; Reichert and Moore 2000). HeLa cells were harvested from ten 10 cm plates at 90% confluency after aspirating the media, washing the cells with 1x PBS and scraping the cells using a cell lifter. The cells were centrifuged at 300 g for 5 min at 4° C to pellet them and PBS was carefully aspirated to estimate the packed cell volume (PCV). The cells were resuspended in 5x PCV of Buffer A (Buffer A or hypotonic buffer: 10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT and 0.2 mM PMSF (freshly added) and kept on ice for 10 min to swell the cells and which were then pelleted by centrifuging at 100 g for 5 min. 2x PCV of fresh hypotonic buffer

was added to resuspend the cells which were incubated on ice for 10 min. The cells were homogenized using a dounce pestle with 15-18 strokes and the cell lysis was checked on the microscope with every two dounces after 10 strokes to prevent breaking open nuclei. The cells were assayed for lysis using Trypan blue (Sigma), which would stain the nuclei blue.

The lysate was transferred to a pre-chilled eppendorf tube and centrifuged at 1500 g for 10 min at 4° C to pellet the nuclei and estimate packed nuclear volume (PNV). The nuclei were first resuspended in 1x PNV of Cell Extraction Buffer (Thermo Scientific) and 1x PNV of Buffer C (Buffer C or High Salt Buffer: 20 mM HEPES pH 7.9, 25% glycerol, 1.2 M KCl, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 0.5 mM DTT and 0.2 mM PMSF) was added drop by drop while stirring on ice to avoid the lysis of nuclei and also to prevent the protein precipitation due to quick changes in the salt concentration. The samples were incubated at 4° C for 30 min on a tilting board to extract the soluble proteins. The samples were centrifuged at 18000 g for 15 min at 4° C to collect the supernatant containing salt nuclear extract (HS-NE). The HS-NE was transferred into pre-chilled Amicon 3 kDa (Millipore) and spun for 30 min at 14000 g at 4° C. The HS-NE was recovered by inverting the Amicon into a fresh eppendorf tube and spinning for 5 min at 1500 g. The HS-NE was transferred into Slide A-Lyzer (Thermo Scientific) and placed in 500 ml of chilled Buffer D (Buffer D or Dialysis Buffer: 20 mM HEPES pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT and 0.2 mM PMSF (freshly added) and dialysed for 1 h at 4° C. The nuclear extracts were aliquoted, flash frozen in liquid nitrogen and stored at -80° C.

All bright field images taken during nuclei isolation from HeLa cells were acquired using Olympus IX83 (Olympus, Japan), an inverted microscope illuminated with mercury halide lamp (Olympus, Japan). All the images were acquired using a 512×512, EM CCD camera (Photometrix

116

evolve). Images of cells were acquired using a 100X, 1.4 NA, oil immersion objective (UPlanFLN, Olympus, Japan).

3. B.10: Bradford assay

Total protein concentrations were measured from the nuclear extracts prepared in section 3. B.8. The Bradford protein quantification assay was performed using a standard protocol. A standard curve was constructed using bovine serum albumin BSA (Sigma) between concentrations of 0 to 1 mg/ml mixed with Bradford reagent (0.01% w/v Coomassie Brilliant Blue G-250, 4.8% w/v Ethanol and 8.5% Phosphoric acid). Absorbance readings were taken at 595 nm on xMark™ Microplate Absorbance Spectrophotometer (BioRad). All measurements were performed in triplicates. The absorbance readings from the nuclear extracts were plotted against the corresponding absorbance on the standard curve to determine the total protein in the extracts.

3. B.11: *In vitro* splicing assay

In vitro splicing reactions of the capped RNA transcripts generated were performed according to the procedure previously described with slight modifications (Zerivitz and Akusjärvi 1989). Splicing reactions were carried out in a 25 μL volume containing 30% HeLa nuclear extracts. As the nuclear extracts were in Buffer D which is 20 mM HEPES pH 7.9, 20% (v/v) glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF and 0.5 mM DTT, the final concentrations of the various components in the *in vitro* splicing buffer were 2 mM ATP, 20 mM creatine phosphate, 3.6 mM MgCl$_2$, 80 mM potassium acetate, 20 mM HEPES–KOH pH 7.9, 1 mM DTT, 0.2 U/μL RNasin and 3% PVA. Splicing reactions are incubated at 30° C for a different time points indicated in the figures (0-210 min). Following incubation, the reactions were stopped by addition of 200 μL

splicing stop solution (0.3 M sodium acetate pH 5.2 and 0.1% SDS). To this 0.2 ml Tris-saturated Phenol was added, mixed and centrifuged for 5 min at 10,000 g. The supernatant was transferred into fresh eppendorf tube and the RNA fragments were precipitated using 2 volumes absolute ethanol supplemented with 2 μL of 5mg/ml glycogen. The precipitated RNAs were resuspended in nuclease free water and used for further analysis.

For control experiments, the stability of the RNAs generated by *in vitro* transcription under the reaction conditions was analyzed by incubating 2 μL of each RNA transcript in splicing reaction buffer (devoid of nuclear extracts) for indicated time points. The RNAs were then precipitated and checked on 0.8% agarose gel, visualized by ethidium bromide staining.

3. B.12: *In vitro* RNA splicing analysis by RT-PCR

RT-PCR analysis was used to understand the relative abundances of spliced and unspliced RNAs from β-globin and its relevant pri-miR transcripts *in vitro*. The precipitated RNAs were reverse transcribed using Superscript IV (Invitrogen) and E3R primer, which hybridizes to the exon3 of the transcript. The RNA-primer mix was heated at 80° C for 2 min and transferred to an ice-bath for 5 min. The first strand synthesis components were added to the tubes and incubated at 52° C for 15 min to generate cDNA. 2 μL of the cDNA was used for PCR with Taq DNA Polymerase (New England Biolabs). To quantitate the levels of unspliced and spliced RNAs 5′ P^{32}-labeled forward primer was spiked in the PCR reaction mix. A first reaction mix to amplify the unspliced and spliced RNA corresponding to intron1 contained a primer set of unlabeled E1F complementary to exon1, unlabeled E2R complementary to exon2 and 5′ P^{32}-labeled E1F. Another set of this PCR reaction mix was used as a control for PCR using cDNA formed from in vitro transcribed RNA

(unspliced reactions that were devoid of extracts). A second reaction mix to amplify the unspliced and spliced RNA corresponding to intron2 contained a primer set of unlabeled E2F complementary to exon2, unlabeled E2R complementary to exon2 and 5′ P^{32}-labeled E2F. The RT-PCR products from splicing of the intron1 were analyzed on 8% native PAGE gel and the products from splicing of intron2 were analyzed on 1.2% native agarose gel. The running buffer for the gels was 0.5X TBE and the electrophoresis was done at 100 V, 15mA for 1 h. The gels were dried on Model 583 Gel Dryer (BioRad) at 80° C for 1 h and exposed to phosphor plates.

Control RT-PCR experiments were performed on RNAs precipitated post incubation with splicing buffer (devoid of extracts). These RNAs were used for cDNA synthesis using E3R primer as mentioned above. PCR was done using 2 μL of cDNA with E1F, E3R primer set spiked with 5′ P^{32}-labeled E1F. The products were resolved on 0.8% native agarose gel and visualized both by ethidium bromide staining and by phosphor imaging.

3. B.13: Kinetic analysis of *in vitro* RNA splicing

The kinetic analysis was done according to the methods previously described (Mueller and Hertel 2014; Hicks et al. 2005). The RNA bands were analyzed and quantitated with BAS-2500 (Fuji Film, Japan) and Image J (NIH). The averages were calculated from two independent experiments, and statistical analysis (Student's *t* test) was performed using Origin (Origin Labs). Significance is denoted by *. * indicate p value <0.05, ** p<0.01 and *** indicate p <0.001.

3. B.14: *In cellulis* splicing of C13orf25 and its related transcripts

3. B.14.1: Transfection of C13orf25 plasmids

HeLa cells were maintained as mentioned in 3. B.3. For transient transfection of C13orf25 related genes, cells were cultured in $75cm^2$ cell culture flask (Nalgene) overnight at 37° C in 5% CO_2 to a confluency of 70% and then transfected with C13orf25 harboring either the TS-pri-miR, SW-pri-miR, or the pri-miR deletion construct using HeLa Monster transfection reagent (Mirus Bio) following the manufacturer's instructions. Typically, a 1:3 ratio of DNA to transfection reagent was used. The transfected medium was replaced after 4 h of incubation and replaced with fresh medium and the cells were collected 24 h post transfection and used for RNA isolation using Trizol (Invitrogen).

3. B.14.2: RT-PCR

2.5 μg of total RNA isolated from HeLa cells transfected with C13orf25 related plasmids was reverse transcribed using an RT primer complementary to the last exon (E2d) and SuperScript III Reverse Transcriptase (Invitrogen). Briefly the RNAs were denatured at 80° C for 5 min and transferred to an ice-bath for 5 min. First-strand mix containing 0.5 mM dNTP mix, 0.25 mM RT primer and 100 U of SSIII reverse transcriptase was added and the reactions were incubated at 37° C for 1 h followed by 52° C for 30 min. Resultant cDNAs were PCR amplified with a specific forward primer which hybridizes to exon2a (E2a) and reverse primer to exon2d (E2d) using Long Amplification Taq DNA Polymerase (New England Biolabs). The PCR products were analyzed on 0.8% agarose gels and visualized by ethidium bromide staining. 18S rRNA was amplified as housekeeping RNA control and Neo (Neomycin phosphotransferase gene) was amplified as control for transfection.

Table 3.1: Primers and probes used for β-globin related studies

PRIMER	SEQUENCE (5′-3′)
TS-pri-miR-Mfe I FP	CCGGAATTCCGGCAATTGAAAACTGAAGATTGTGAC
TS-pri-miR-Mfe I RP	CCGGAATTCCGGCAATTGTCTTCTGGTCACAA
SW-pri-miR-Mfe I FP	CCGGAATTCCGGCAATTGATTGTGTCGATGTAG
SW-pri-miR-Mfe I RP	CCGGAATTCCGGCAATTGCATTCATTTGAAGG
E1F	CTGAGGAGAAGTCTGCCGTTAC
E2F	CTGGACAACCTCAAGGGCACCT
I2R	CAGAATAATCCAGCCTTATCCC
E2R	GGTGAGCCAGGCCATCACTAAA
E3R	AATCCAGATGCTCAAGGCCC
18S FP	CAGCCACCCGAGATTGAGCA
18S RP	TAGTAGCGACGGGCGGTGTG
Neo FP	TATCCATCATGGCTGATGCAATGC
Neo RP	CAGCAATATCACGGGTAGCCAAC
Globin for	ACATTTGCTTCTGACACAACTGTGTTCA
Globin rev	GCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGA
Globin for-T7	GCCCTTTAATACGACTCACTATAGGGACA
Globin rev-A10	TTTTTTTTTTTTTTTTTTTTGCAATGAAAATAAATGT
TS-pri-miR FP	AAAACTGAAGATTGTGACCAGTCAGA
TS-pri-miR RP	TCTTCTGGTCACATCCCCACCAA

("Table 3.1, continued")

Swap-miR FP	ATTGTGTCGATGTAGAATCTGCCT
Swap-miR RP	CATTCATTTGAAGGAAATAGCAGGC
miR probe	AGCATTGCAACCGATCCCAACCT
U6 RNA	ATATGGAACGCTTCACGAATT
E3-1 probe	AGTGATGGGCCAGCACACAGACCAGCACGTTGCGGAG
E3-2 probe	TGATAGGCAGCCTGCACTGGTGGGATGAATTCTTTGCCA
I2-1 probe	AAGAAGGGGAAAGAAAACATCAAGCGTCCCATAGACTCAC
I1-1 probe	ATTGGTCTCCTTAAACCTGTCTTGTAACCTTGATACCAA C

Table 3.2: Primers and probes used for C13orf25 related studies

PRIMER	SEQUENCE (5′-3′)
C13orf25 Full FP	GTGGGGCTTGTCCGTATTTACGTTGAGGCG
C13orf25 Full RP	ACCTTGAAGTTTTTATTTCAATATTCTCGTTCTGGACAATTTCTTA
E2a FP	GTCATACACGTGGACCTAAC
E2d RP	CTGAAGTCTCAAGTGGGCAT
C13-S-for1	GAACCTCTCCTCGCGGGG
C13-S- OE-rev1	GTCCACGTGTTATGACTGGAATAGGCAAAATAAGGAAAAAAGATAAA
C13-S-OE-for3	GTGGCCTGCTATTTCCTTCAAATGAATGTTTTGAAAATTAAATATTACTGAAGATTTCGACTTCCACTGTTAAATGT
C13-S-rev3	CTGAAGTCTCAAGTGGGCATGATGA
C13-Δ OE-rev	TTCAGTAATATTTAATTTTCAAAAACAAGGTGATGTTCTCTTTTTTCCTGC
C13-Δ OE-for	GAAAAAAGAGAACATCACCTTGTTTTTGAAAATTAAATATTACTGAAGATTTCGACTTCCACTGT
Neo for	AGAGGCTATTCGGCTATGACTG
Neo rev	TTCGTCCAGATCATCCTGATC

3. C: Results and discussion

3. C.1: *In cellulis* splicing of β-globin WT and β-globin pri-miR transcripts

To analyze the effect of tertiary structure of intronic pri-miR-17-92a on the splicing of its host transcript, I made three mammalian expression constructs, shown in Figure 3.12.
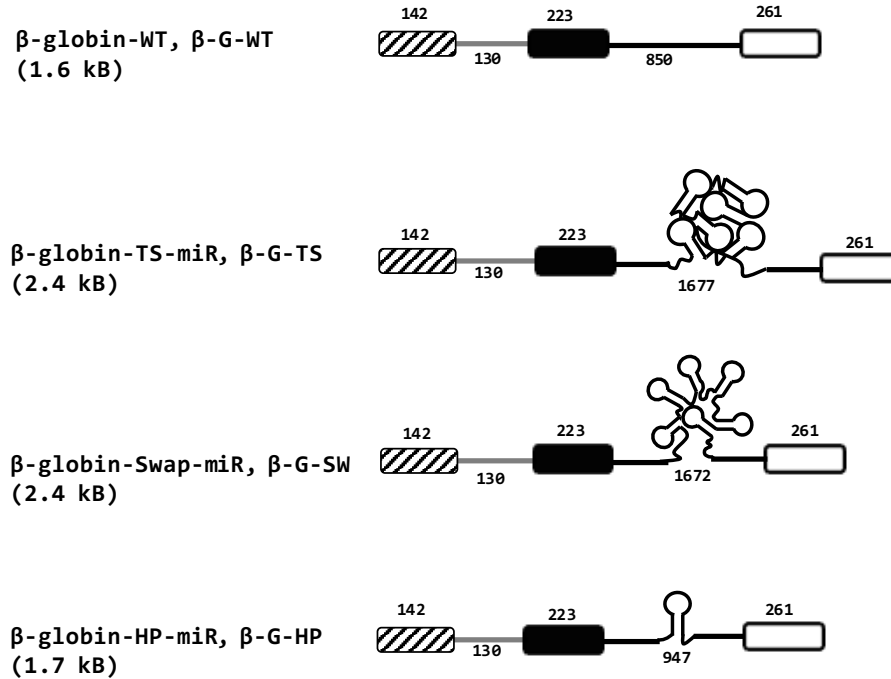
Figure 3.12: Details of the β-globin splicing reporter constructs used in the study. Rectangles represent exons and grey and black thick lines represent intron1, and intron2, respectively β-globin TS-pri-miR (β-G-TS) was generated by inserting the tertiary structured pri-miR-17-92a in intron2 of β-globin gene. β-globin Swap-pri-miR (β-globin-SW) contains the swapped-pri-miR which is devoid of tertiary structure and β-globin HP-miR (β-G-HP) contains a single miRNA hairpin domain inserted in intron2.

 (i) **pCMV-β-globin TS-pri-miR**: This corresponds to 0.8 kb sequence TS-pri-miR-17-92a was engineered in the MfeI site of intron2 of the β-globin WT plasmid, (ii) **pCMV-β-globin Swap-pri-miR**: Here in the place of TS-pri-miR a "swap-pri-miR" is inserted. "swap pri-miR" is a construct where the 5′ and 3′ halves of TS-pri-miR are swapped to yield a pri-miR of identical

length yet devoid of a distinct tertiary structure, and is also processed faster that TS-pri-miR (See 2.C) (iii) **pCMV-β-globin HP-miR**: Here, in the place of TS-pri-miR a single microRNA hairpin domain (HP) corresponding to miR-17, the first microRNA of the pri-miR-17-92a cluster was inserted. These constructs would be refered to as β-G-WT, β-G-TS, β-G-SW, β-G-HP.

I first analyzed the effect of inserting a 0.8 kb pri-miRNA sequence in intron2 of the β-globin reporter gene on its splicing by RT-PCR analysis of total RNA from HeLa cells transfected with β-globin WT (**β-G-WT**), β-globin TS-pri-miR (**β-G-TS**), β-globin Swap pri-miR (**β-G-SW**), or β-globin HP-miR (**β-G-HP**) expression plasmids. cDNA from each transcript was synthesized by using E3R primer, which is complementary to the sequence in exon3 and then PCR amplified using E3R and E1F, which is complementary to sequence in exon1. As a control cDNA was synthesized using dT18 oligo and PCR was performed using primers specific for 18S rRNA transcript as well as Neomycin phosphotransferase encoded in the transfected vector (Neo). Thus, 18S RNA serves as loading control while Neo serves as a control for transfection efficiency.

Figure 3.13A shows RT-PCR performed on transfected constructs encoding for each of the splicing substrates described above. The wild type globin, in lane W showed a single 500 bp band corresponding to the fully spliced β-globin mRNA (E1-E2-E3) from the β-G-WT reporter. β-G-HP (lane H) gave a single amplicon corresponding to fully spliced mRNA similar to WT. β-G-TS and β-G-SW reporter transcripts (lane T and lane S respectively) also resulted in a single fully spliced RNA amplicon. This indicates that insertion of a 0.8 kb long pri-miR did not affect the identity of the spliced products. While the β-G-SW transcript mirrored the β-G-TS RNA in every way, the intensity of the amplicon of the β-G-TS was lower. To explore whether this

125

decrease was significant, a quantitative analysis of the relative levels of spliced form was done

using Northern blot. Total RNA from HeLa cells transfected with either β-G-WT, β-G-TS, β-G-

SW, or β-G-HP was isolated and resolved on 1.2% denaturing agarose gel. **Figure 3.13B** shows

ethidium bromide staining of RNAs (upper panel) and methylene blue (lower panel) to check for

complete transfer.
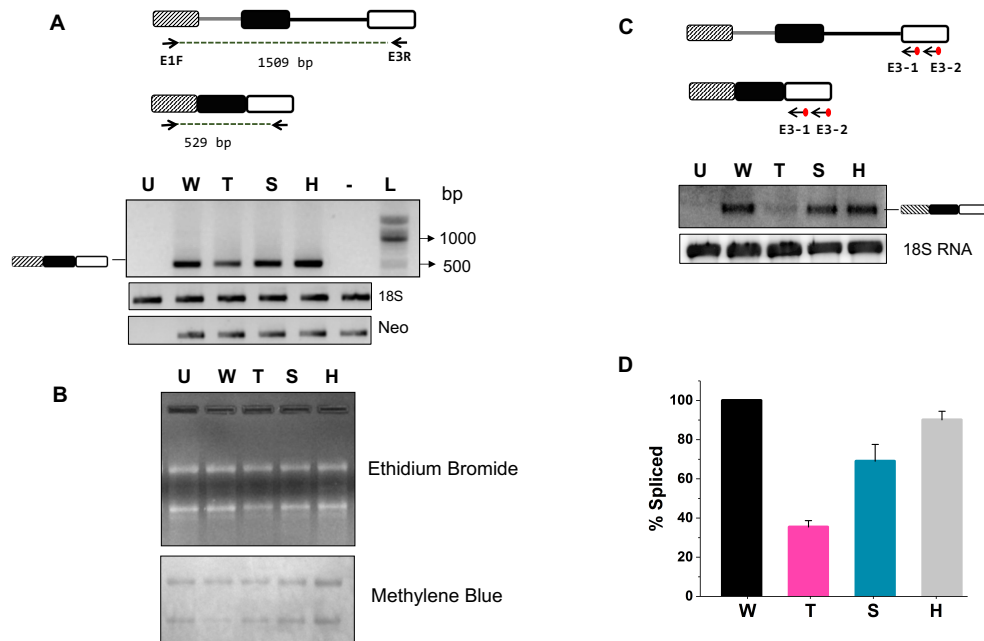
Figure 3.13: Effect of the insertion of pri-miR on β-globin splicing. **A.** RT-PCR to address the steady levels of the spliced form of β-globin using total RNA isolated from HeLa cells transfected with β-G-WT (W), β-G-TS (T), β-G-SW (S), and β-G-HP (H). U indicates RNA from untransfected cells and - indicates no RT. 2.5 μg of total RNA was used for cDNA synthesis using E3R primer and E1F and E3R primers for PCR. **B.** 10 μg of total RNA was resolved on 1.2% agarose-formaldehyde gel and transferred to Hybond membrane. Gel stained with ethidium bromide shows rRNA abundances and integrity and methylene blue staining indicates complete transfer to membrane. **C.** Northern blot using radiolabeled E3 probes showing relative levels of spliced globin from each transcript. U6 snRNA serves as a control. **D.** Quantification of spliced β-globin measured by Northern blot from three independent experiments, error bars indicate standard error of mean (SEM).

Northern blot was performed with probes designed to bind only exon3 such that they only hybridized to fully spliced intact mRNAs, and so that RNAs that are not fully transcribed or spliced, or truncated transcripts, would be non-responsive. The blot showed a significant decrease in the levels of the spliced form detected in the total RNAs of β-G-TS reporter construct (**Figure 3.13C**). The relative abundances of the spliced β-globin mRNA from each reporter transcript normalized to 18S RNA levels clearly indicated that the host gene containing TS-pri-miR showed a 2.5-fold decrease in the steady state levels of spliced RNA compared to the WT (**Figure 3.13D**).

3. C.2: Analysis of nuclear RNA for splicing of the reporter gene

The reduced levels of the spliced form from β-G-TS could be either the result of splicing or due to its altered abundances in the nucleus and cytoplasmic, possibly arising from differential transport or stabilities. To address this, I performed RT-PCR analysis on nuclear RNA isolated from HeLa cells transiently transfected with either β-G-WT or β-G-TS or β-G-SW or β-G-HP constructs. cDNA was synthesized using E3R primer, which is complementary to exon3 (**Table 3.1**) and then PCR amplified using E3R and primer E1F, which is complementary to exon1. The RT-PCR results revealed that the equilibrium levels of the fully spliced mRNA (E1-E2-E3) was lowered in the case of β-G-TS compared to β-G-WT, and this was true in both nuclear and cytoplasmic fractions (**Figure 3.14A**).

Next northern analysis (**Figure 3.14B**) was performed on the nuclear RNA isolated from HeLa cells transfected with β-G-WT or β-G-TS or β-G-SW constructs using radiolabeled probes E3-1, E3-2 and I2-1, I2-2 that bound to exon3 and intron2 regions, respectively. (**Table 3.1** for sequences and **Figure 3.14B**, schematic for location on the transcript). Importantly, because full

length transcripts could not be detected with exonic probes (E3-1, E3-2) which hybridized to exon3 in earlier northern blots **(Figure 3.13C),** a mix of radiolabeled probes that hybridized to exon3 (E3-1 probe, E3-2 probe) and to intron2 (I2-1 probe, I2-2 probe) were used in this experiment.
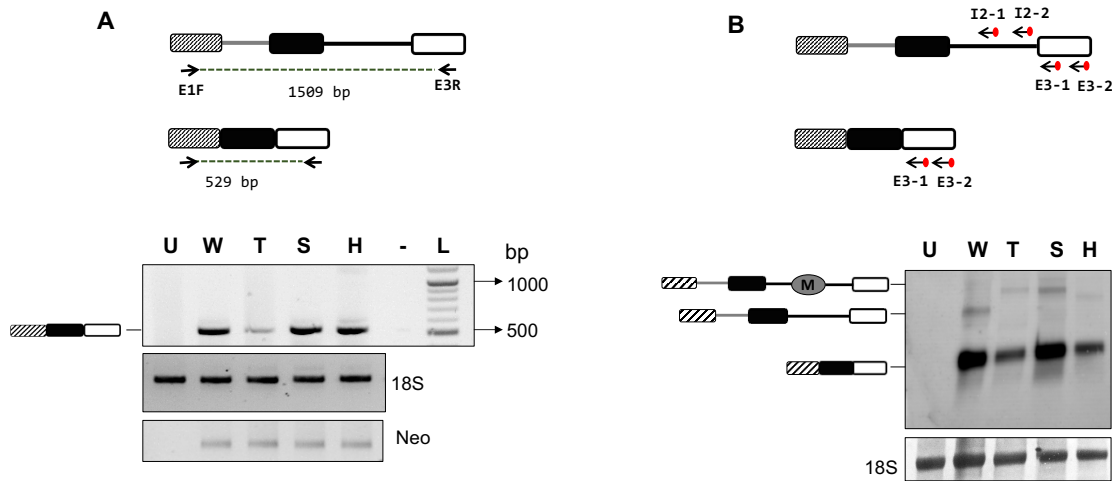
Figure 3.14: Effect of TS-pri-miR on the formation of β-globin mRNA. **A.** RT-PCR to adddress the steady levels of the spliced form of β-globin using nuclear RNA isolated from HeLa cells transfected with β-G-WT (W) or β-G-TS (T) or β-G-SW (S) or β-G-HP (H). 2.5 μg of total RNA was used for cDNA synthesis using E3R primer, and E1F and E3R primers were used for PCR. U indicates RNA from untransfected HeLa cells and - indicates no RT. **B.** 10 μg of nuclear RNA was resolved on 1.2% agarose-formaldehyde gel and transferred to Hybond membrane. Northern blot using radiolabeled E3 probes showing the relative levels of spliced globin from each transcript. In the schematic of unspliced globin 'M' represents the pri-miR (TS-pri-miR or SW-pri-miR) in the intron. 18S rRNA serves as a control.

Here, I observed a consistent decrease in the fully spliced globin mRNA for β-G-TS, and also able to detect the lower mobility band corresponding to the full length, unspliced RNA for all transfected constructs (**Figure 3.14B**). This revealed that host transcripts containing a polycistronic pri-miR cluster, can undergo efficient splicing despite the inclusion of multiple cleavage sites introduced into its intron in the form of pri-miR-17-92a or its swapped form.

3. C.3: Effect of tertiary structure on splicing of intron1 and intron2

Given that 'β-G-TS' showed decreased levels of fully spliced β-globin mRNA (E1-E2-E3), I sought to decouple the effect of the TS-pri-miR on each of the constituent introns i.e., the intron which harbored the pri-miR (intron2) as well as its upstream intron (intron1). To do this, nuclear RNA from HeLa cells transiently transfected with β-G-WT, β-G-TS, β-G-SW, β-G-HP was reverse transcribed using primer E3R which is complementary to exon3. The resulting cDNAs were then amplified by using two different combinations of PCR primers that each address the splicing of intron1 and intron2. The E1F primer that hybridizes to exon1 was used in combination with E2R primer that is complementary to exon2 to indicate the relative levels of unspliced and spliced intron1 (**Figure 3.15A**).
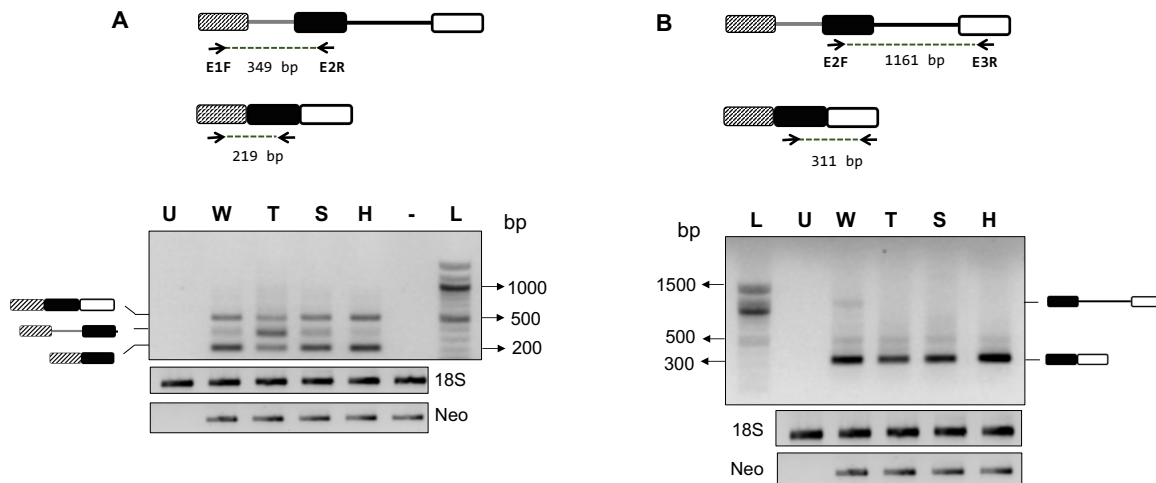
Figure 3.15: Effect of TS-pri-miR on the splicing β-globin intron1 and intron2. Total nuclear RNA isolated from HeLa transfected with β-G-WT (W) or β-G-TS (T), β-G-SW (S), β-G-HP (H) was used for cDNA synthesis using E3R primer. **A**. E1F and E2R primers used for PCR to measure the levels of splicing of intron1. **B**. Primers E2F and E3R were used to measure the levels of intron2. U indicates RNA from untransfected HeLa cells and - indicates no RT.

Using the same cDNA, another PCR was setup using E1F and E3R primers that amplify fully spliced RNA. The amplicons thus generated by PCR using E1F, E2R primer set and E1F,

E3R primer set were loaded in the same lane. 18S rRNA was used as loading control. As seen in **Figure 3.15A,** I observed a ~ 0.35 kb amplicon corresponding to unspliced intron1 (E1-I1-E2) and a 0.2 kb band corresponding to spliced intron1 (E1-E2). The lower mobility band of 0.5 kb corresponds to the fully spliced globin mRNA (E1-E2-E3). Lane T showing the RT-PCR products from 'TS' indicated that the levels of the 0.35 kb band corresponding to unspliced intron1 (E1-I1-E2) is higher while the 0.2 kb band corresponding to spliced intron1 (E1-E2) is lower compared to those observed for β-G-WT, β-G-SW, β-G-HP. This accumulation of the unspliced intron1 and consistent decrease in the spliced form seen in β-G-TS but not in β-G-SW, this suggests the exciting possibility that tertiary structure of pri-miR-17-92a influences the splicing of the upstream intron (intron1).

The cDNAs generated using E3 primer were also used to probe splicing efficiency of intron2 by PCR. For this, E3R primer in combination with an E2F primer complementary to exon2 was used to measure the relative levels of unspliced and spliced intron2. **Figure 3.15B** showed a major 0.3 kb amplicon corresponding to spliced intron2 (E2-E3) for β-G-WT, β-G-TS, β-G-SW, and β-G-HP. Interestingly, β-G-TS (Lane T) showed a slight decrease in the levels of E2-E3 compared to β-G-SW (lane S), β-G-WT (Lane W) or β-G-HP (Lane H). Even though β-G-WT showed a faint amplicon of ~ 1 kb corresponding to unspliced intron2 (E1-I1-E2), this species could not be detected for β-G-TS and β-G-SW with this primer set under these PCR conditions. The identity of the bands was confirmed by sequencing. Taken together these results suggest that presence of TS-pri-miR in the intron2 of the β-globin impedes splicing of its host transcript and that this effect is more evident on its upstream intron compared to the intron in which it resides.

3. C.4: RT-PCR analysis of unspliced intron2 transcripts

Next, I decided to investigate the levels of unspliced intron2 in β-G-WT, β-G-TS, β-G-SW to check the contribution of TS and SW which have different kinetics of microprocessing on the abundances of unspliced intron2. To measure the levels of intron2 RT-PCR was performed with primers that hybridize to intron2. Nuclear RNA from HeLa cells transiently transfected with β-G-WT, β-G-TS, and β-G-SW constructs were reverse transcribed using primer E3R, which is complementary to exon3 and the resulting cDNAs were then amplified by using two different combinations of PCR primers which could report on the unspliced intron2.

In the first set, E1F primer in combination with I2R was used to detect any unspliced transcripts. I2R primer hybridizes to a region in intron2 (771-793 nt of intron2 in WT; 1604-1626 nt in TS; 1599-1612 in SW) which is 55 nt upstream to the 3′ splice site AG. Thus, WT is expected to yield a 1.2 kb amplicon corresponding to unspliced intron1 (E1-I1-E2-I2) and a 1 kb amplicon for spliced intron 1 (E1-E2-I2) with this primer set. Accordingly, β-G-WT (lane W) showed two amplicons corresponding to transcript with unspliced and spliced intron1 (**Figure 3.16A**). For β-G-TS and β-G-SW this is expected to result in analogous amplicons of 2.0 kb and 1.9 kb for E1-I1-E2-I2 and E1-E2-I2, respectively. β-G-TS (Lanes T) and β-G-SW (lane S) both showed amplicons of ~2 kb. Note that, as the unspliced and spliced intron1 transcripts of β-G-TS and β-G-SW differ by 130 nt, the two bands are not easily distinguishable under the electrophoresis conditions.

Next, a second set of primers, E2F and I2R complementary to exon2 and intron2 respectively were used probe the above cDNAs. E2F and I2R primer are expected to yield amplicon sizes of 0.87 kb for β-G-WT and 1.7 kb for β-G-TS and β-G-SW corresponding to

unspliced intron2 (E2-I2). Accordingly, β-G-WT showed a single amplicon of ~0.9 kb (lane W,

Figure 3.16B) and ~ 2 kb major amplicons for β-G-TS and β-G-SW (lane T and S respectively).
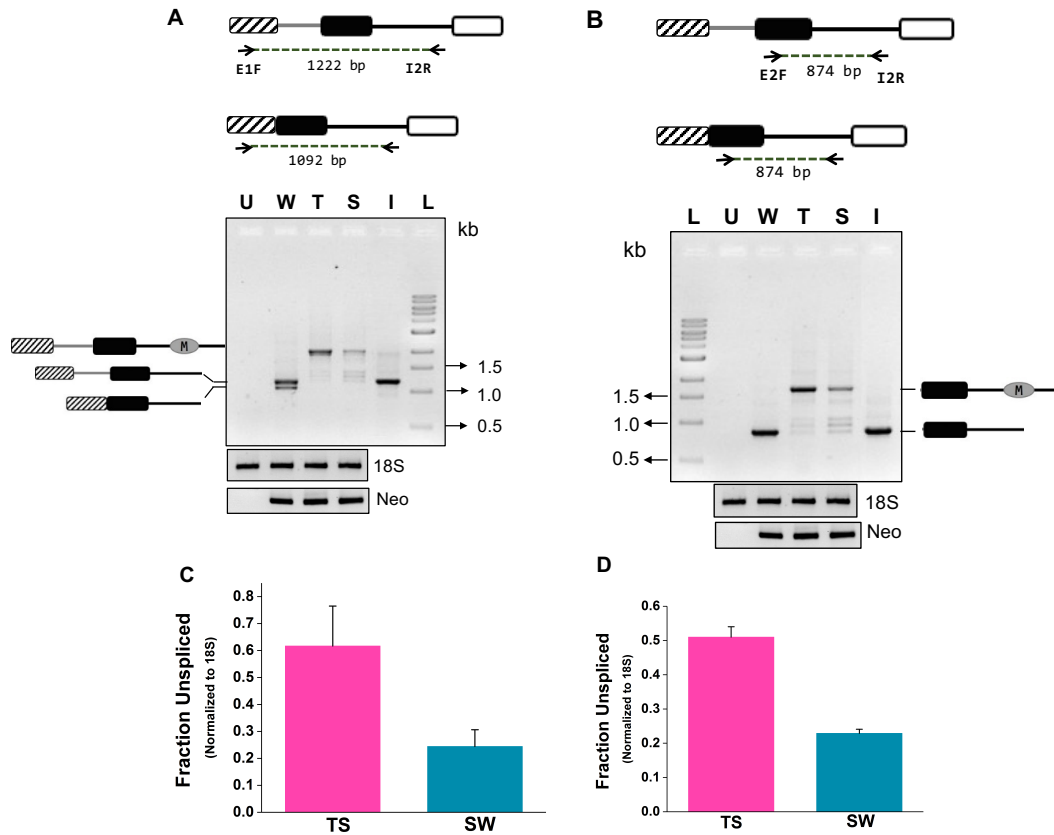
Figure 3.16: RT-PCR to measure the levels of β-globin unspliced intron using the indicated primers. 1μg of nuclear RNA isolated from HeLa transfected with β-G-WT (W), β-G-TS (T), β-G-SW (S), or β-G-HP (H) was used for cDNA synthesis using E3R primer. I indicate *in vitro* transcribed β-G-WT RNA (unspliced) used as control for RT-PCR **A**. E1F and I2R primer set **B**. E2F and I2R for intron2 were used for PCR to detect the unspliced intron. In the schematic 'M' represents the pri-miR (TS-pri-miR or SW-pri-miR) in the intron. **C.** Quantification of unspliced band in lanes T and S of the gel in A, normalized to 18S rRNA control. **D.** Quantification of unspliced band in lanes T and S of gel in B, normalized to 18S rRNA control.

In vitro transcribed β-G-WT transcript (unspliced) was used as a control for reverse

transcription using E3R primer and the cDNA generated was used for PCR using E1F and I2R

primers (lane I). Quantification of amplicons from two independent experiments showed a 2.5-

fold greater abundance of unspliced RNA in the case of β-G-TS than β-G-SW (**Figure 3.16C &**

D). The identity of all the amplicons represented schematically on either side of the gels have been

confirmed by sequencing of the RT-PCR products (data not shown).

3. C.5: Relative levels of pri-miRs harbored in β-globin intron

In cellulis analysis of spliced and unspliced β-G-WT, β-G-TS, and β-G-SW transcripts was done

to study the effect of tertiary structured pri-miR on its host transcript. To address the levels of pri-

miR present in the host pre-mRNA nuclear RNA isolated from HeLa cells transfected with β-G-

WT, β-G-TS, β-G-SW and β-G-HP was reverse transcribed using E3R primer. The cDNAs were

then used for PCR with primers specific for TS-pri-miR, SW-pri-miR, or HP-miR. The resulting

PCR products were resolved on 0.8% agarose gel shown in **Figure 3.17A**. β-G-WT (lane W),

showed no amplification product, indicating that no detectable endogenous pri-miR-17-92a in

HeLa cells. However, β-G-TS and β-G-SW (lanes T and S respectively) showed a single 0.8 kb

amplicon corresponding to TS-pri-miR or SW-pri-miR. Lane H showed a faint amplification

product corresponding to 100 nt single hairpin domain. Lanes P_{ivt} and W_{ivt} correspond to PCR

product from cDNAs of *in vitro* transcribed 0.8 kb TS-pri-miR and β-G-WT respectively as

controls. The same cDNA was used for PCR using primers E1F and E3R, to amplify the spliced

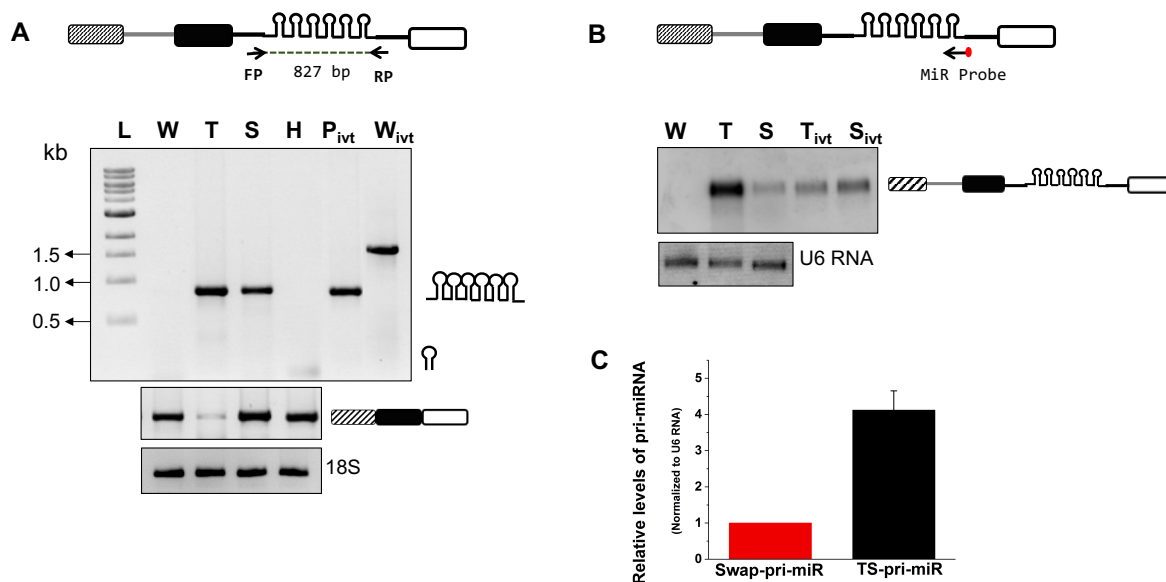mRNA. 18S rRNA was used as a loading control.

Figure 3.17: Steady state levels of pri-miRs harbored in the β-globin reporter transcripts. **A.** RT-PCR analysis to show the steady levels of pri-miR from nuclear RNA isolated from HeLa cells transfected with β-G-WT (W) or β-G-TS (T), β-G-SW (S), or β-G-HP (H). P_{ivt} and W_{ivt} indicates *in vitro* transcribed 0.8 kb TS-pri-miR and *in vitro* transcribed β-globin-WT (W), respectively. **B.** Northern analysis using a radiolabeled miR probe to indicate the levels of unspliced β-globin from nuclear RNA isolated from HeLa cells transfected with β-G-WT (W), β-G-TS (T), or β-G-SW (S). T_{ivt} and S_{ivt} indicate *in vitro* transcribed β-G-TS and β-G-SW respectively. U6 snRNA serves as control. **C.** Quantification of bands in lanes T and S normalized to U6 RNA.

We confirmed the above findings by Northern blot analysis of nuclear RNA isolated from HeLa cells transfected with β-G-WT, β-G-TS, or β-G-SW using specific radiolabeled probes that hybridize to TS-pri-miR or SW-pri-miR. β-G-WT (lane W, **Figure 3.17B)** showed no detectable levels of endogenous pri-miR with selected probes. β-G-TS and β-G-SW (lanes T and S respectively) showed a greater level of unprocessed pri-miRs present in β-G-TS and β-G-SW. The probe for U6 snRNA was used as a control. T_{ivt} and S_{ivt} correspond to *in vitro* transcribed β-G-TS and β-G-SW respectively. **Figure 3.17C** is a quantification of three independent replicates, where **Figure 3.17B** shows a typical example. This indicates that the levels of unprocessed pri-miR from

β-G-TS is ~4.2 times higher compared to β-G-SW. This is consistent with my previous study on the kinetics of pri-miR processing, which showed that TS-pri-miR undergoes 4-fold slower miRNA processing compared to the SW-pri-miR, thus leading to higher levels of unprocessed pri-miR substrate (**See 2.C**).

3. C.6: *In vitro* splicing studies

Using semi-quantitative RT-PCR and quantitative Northern blots I have shown that the presence of TS- pri-miR in the intron of a β-globin reporter gene impedes the splicing of its host introns *in cellulis* and that this effect is more pronounced on the upstream intron. To address the contribution of a tertiary structured RNA element present in the transcript on the kinetics of splicing of the host pre-mRNA, I chose to perform the *in vitro* splicing assay on a ~2 kb full length three exon- two intron system. This *in vitro* splicing assay offers a reductionist system to study the influence of specific RNA elements, and *trans*-acting protein factors on splicing.

The two biggest challenges in performing *in vitro* splicing studies on these large RNAs were: (i) to generate large amounts of homogenous RNA transcript and (ii) to produce splicing competent nuclear extracts, and I outline below the two procedures that I developed to achieve each of these in detail. Most of the *in vitro* studies done so far have focused on divalent ion initiated refolding of fully synthesized transcripts. However co-transcriptional RNA folding studies better mimic how RNA folds in the cellular environment (Schroeder et al. 2002; Pan and Sosnick 2006), and native purification methods for co-transcriptionally folded RNAs have been established (Wong and Pan 2009; Pereira et al. 2010; Chillón et al. 2015) and are currently being explored for structural studies of large RNAs (Somarowthu et al. 2015). Thus, the present study explores the

effect of the insertion of a tertiary structured primary microRNA on the *in vitro* splicing of host RNAs undergoing co-transcriptional folding.

3. C.6.1: *In vitro* transcription to generate the RNA for splicing assays.

To make DNA templates for *in vitro* transcription the plasmids pCMV-β-globin WT, pCMV-β-globin TS-pri-miR and pCMV-β-globin SW-pri-miR were used to amplify the globin region containing E1-I1-E2-I2-E3 using specific primers and a high-fidelity Phusion DNA polymerase. Using this amplicon as template, a subsequent PCR was performed with a forward primer (T7-FP) containing the T7 promoter sequence and a reverse primer containing $(A)_{10}$. The DNA templates thus prepared contain a T7 promoter at the 5′ end and $(A)_{10}$ at the 3′ end, schematically depicted in **Figure 3.18A** (upper panel). This was done as 5′ cap has been reported to increase the stability and efficiency of *in vitro* splicing (Konarska et al. 1984; Edery and Sonenberg 1985). The sizes of the amplicons analyzed on agarose gel correspond to 1.7 kb β-G-WT (lane 1) and 2.5 kb β-G-TS and β-G-SW (lane 2, lane 3) are shown in the lower panel. The 5′ capped and $(A)_{10}$ tailed RNAs were generated by *in vitro* transcription from these DNA templates and purified under native conditions as described (3.B.6).

A small-scale reaction of capped RNAs was also generated using mMessage mMachine T7 kit to check for size of the RNAs. The lengths of the RNA transcripts formed by *in vitro* transcription were analyzed on 1.2% agarose-formaldehyde gel (**Figure 3.18B).** β-G-WT (lane W) showed a transcript length of ~2.0 kbs while that of β-G-TS and β-G-SW (lane T and lane S) showed a length of ~3.0 kbs The RNAs generated on ~10 mg scale, purified under native

conditions were electrophoresed on 1% native agarose gel with 0.5x TBE containing 5mM $MgCl_2$ (**Figure 3.18C**).
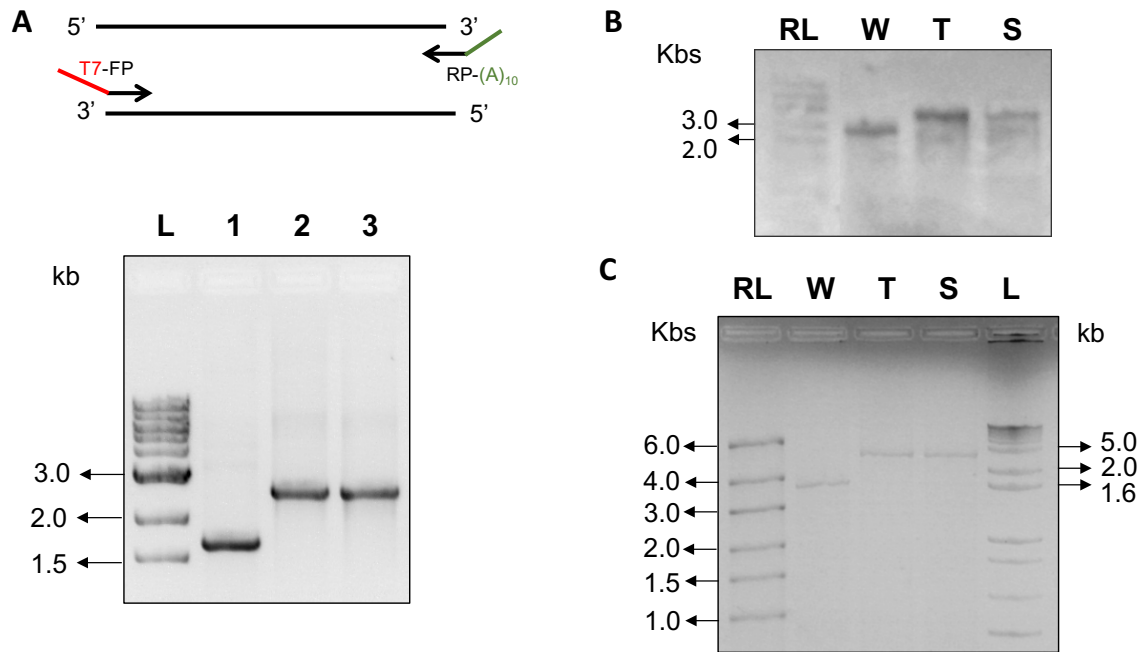
Figure 3.18: Generation of RNA transcripts by IVT. **A**. The DNA template for IVT was generated by PCR using a forward primer that incorporated a T7 promoter sequence and a reverse primer for (A) $_{10}$ to enhance the stability. 0.8% agarose gel showing the PCR products corresponding to the templates lane-1: β-G-WT, lane-2: β-G-TS, lane-3: β-G-SW. **B.** 1.2% agarose-formaldehyde gel showing the IVT products: β-G-WT (W), β-G-TS (T) and β-G-SW (S). **C.** 1% native agarose to show the RNA transcripts purified under native conditions. RL: RiboRuler High range, L: DNA Ladder

3. C.6.2: Splicing competent HeLa nuclear extracts

Preparation of HeLa nuclear extracts competent for pre-mRNA splicing were prepared as I have outlined in the materials and methods section 3. B.9. Each step was monitored under an optical microscope (**Figure 3.19**). The extracts were kept in Buffer D, aliquoted, flash frozen and stored at -80° C until use. The total protein concentration of the extract was estimated by Bradford assay to be about 8-10 mg/ml.

137
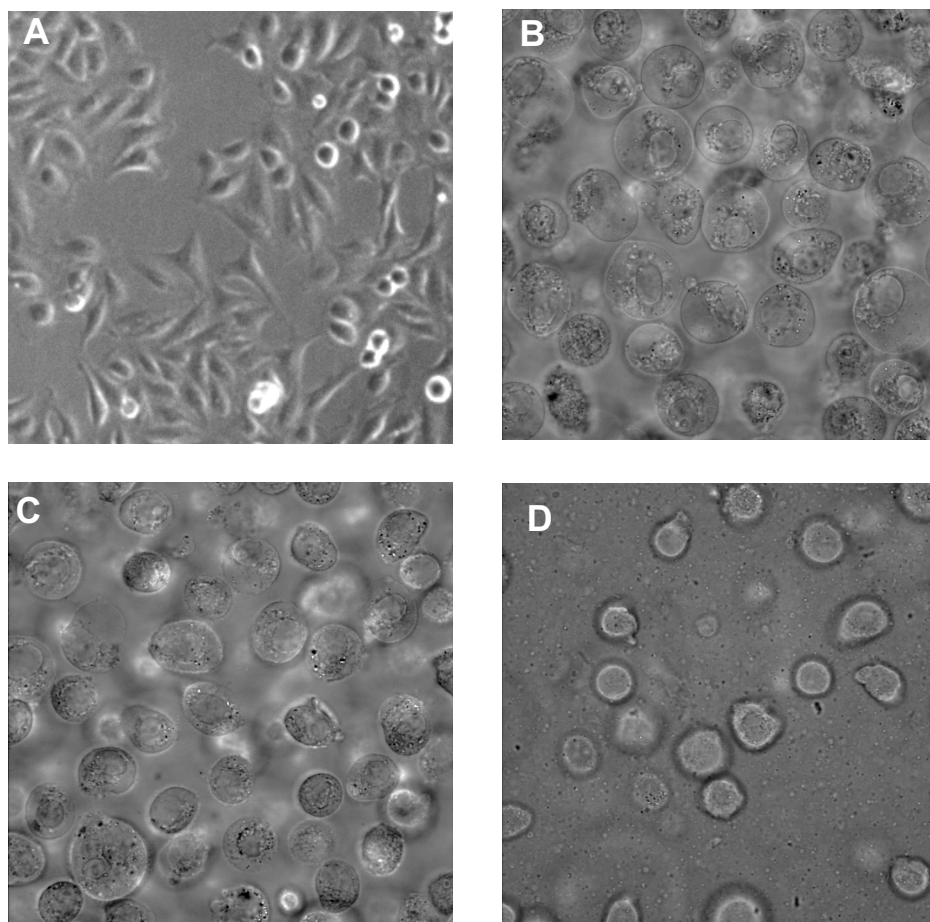
Figure 3.19: Microscopic images taken during fractionation of HeLa cells. HeLa cells after washing with PBS **(A)**, after soaking them in hypotonic buffer **(B)** douncing (10 strokes) **(C)** and nuclei after fractionation used for extraction of nuclear proteins (**D**).

3. C.7: Stability of *in vitro* transcripts.

First, the stabilities of the *in vitro* transcribed 5′ capped and poly-A tailed RNAs were studied under the assay conditions that were to be used for splicing. To do this, 2 µM of each RNA transcript (β-G-WT, β-G-TS and β-G-SW) was incubated in the splicing buffer (devoid of nuclear extracts) at 30° C for the indicated times. The RNA was then precipitated, resuspended in nuclease free water, loaded on 0.8% native agarose gel electrophoresis and analyzed by staining with

ethidium bromide. **Figure 3.20A** shows that the *in vitro* generated RNAs are stable for 3 h with insignificant hydrolysis in splicing buffer.
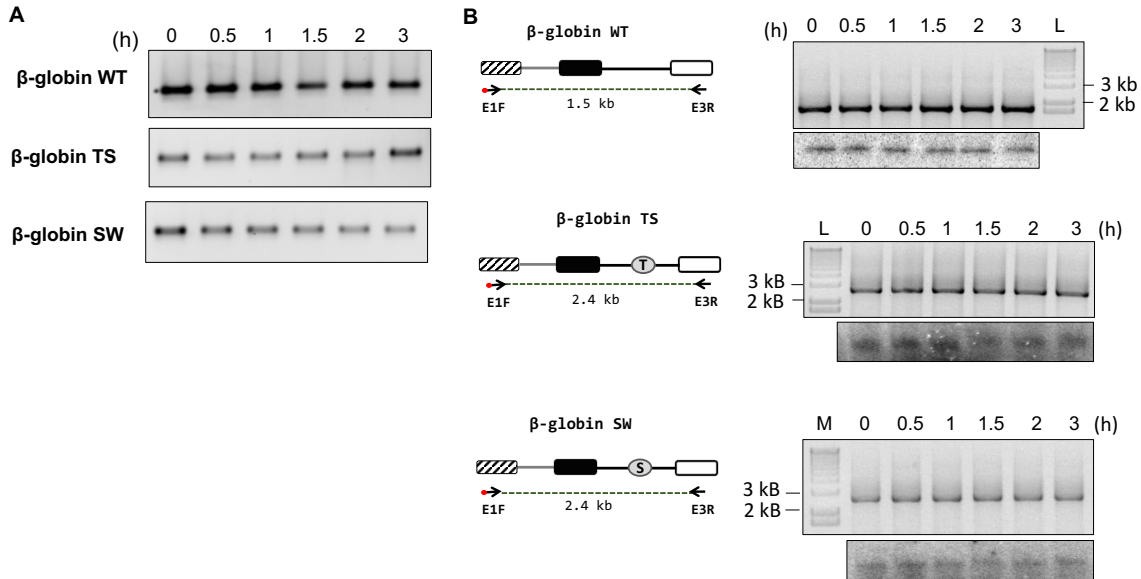
Figure 3.20: Stability of RNA transcripts in splicing buffer: **A.** 1% native agarose gel showing β-G-WT, β-G-TS, and β-G-SW RNA transcripts incubated in splicing reaction buffer devoid of nuclear extracts for indicated time points. **B.** The RNA transcripts incubated in the splicing buffer devoid of extracts were purified and used for RT-PCR analysis using E1F and E3R primers. PCR products resolved on 0.8% agarose gel and stained with ethidium bromide (top panel). 5′ P^{32}-labeled E1F primer was spiked into the PCR and analyzed by phosphor imaging (bottom panel).

As we chose a quantitative RT-PCR based assay (see 3.C.8) to address the kinetics of *in vitro* splicing, the stabilities of the *in vitro* transcribed RNAs were also studied using RT-PCR. To do this, the RNAs that were incubated in splicing buffer (devoid of extracts) at specific times of incubation were reverse transcribed using E3R primer that binds to exon3. The cDNA thus generated was used for PCR using E1F and E3R primers, analyzed on 0.8% agarose gel after staining with ethidium bromide. The same samples of cDNA were also amplified by PCR using unlabeled E1F, and E3R primers with 5′end labeled E1F spiked into the PCR reaction mix and analyzed by phosphor imaging (**Figure 3.20B**). This revealed that the full length unspliced

transcripts were stable under assay conditions and that this method could be used to detect unspliced and spliced transcripts obtained by *in vitro* splicing.

3. C.8: *In vitro* splicing assay and RNA analysis

To understand the effect of tertiary structured pri-miR on splicing of its host pre-mRNA, I sought to perform *in vitro* splicing of these RNAs. Biochemical methods to understand the removal of introns greater than 0.5 kb *in vitro* has been technically challenging. The methods previously established to study splicing for mini genes or small RNA routinely use internally radiolabeled RNA or a 5′ end labeled primer to convert the RNA fragments to their respective cDNAs by primer extension, which are then resolved on denaturing gels to address the unspliced transcripts and spliced product (Mayeda and Krainer 2012; Movassat et al. 2014; Erster et al. 1988). RNase protection assays have also been extensively used for small RNAs to probe for exonic and intronic regions of a given transcript (Fong et al. 2009). As I wanted to understand the effect of TS-pri-miR on the kinetics of each intron from the full β-globin gene (containing both the introns), I set out to standardize the *in vitro* splicing assays on these large RNAs and used radiolabeled primers in an optimized RT-PCR based assay to understand the kinetics of splicing.

Standard splicing reactions (25 μL) were set up containing 20 pmoles of RNA, and 30% HeLa nuclear extracts with splicing buffer as described. Importantly, the splicing buffer contained potassium acetate instead of potassium chloride, as acetate ions provide better splicing conditions (Reichert and Moore 2000). Reactions were incubated at 30° C in timed manner (typically 15 min to 180 min). The timepoint t= 0 corresponds to the time immediately after the addition of nuclear extracts that is flash frozen to prevent the splicing reaction from starting. This timepoint was used

to adjust for background intensity associated with unspliced product for all the reaction analyses. Post-splicing, the RNAs are precipitated using standard precipitation procedures and reverse transcribed using E3R primer. The cDNAs thus synthesized were used for RT-PCR using two sets of primers which would report on the removal of intron1 and intron2 in each transcript.
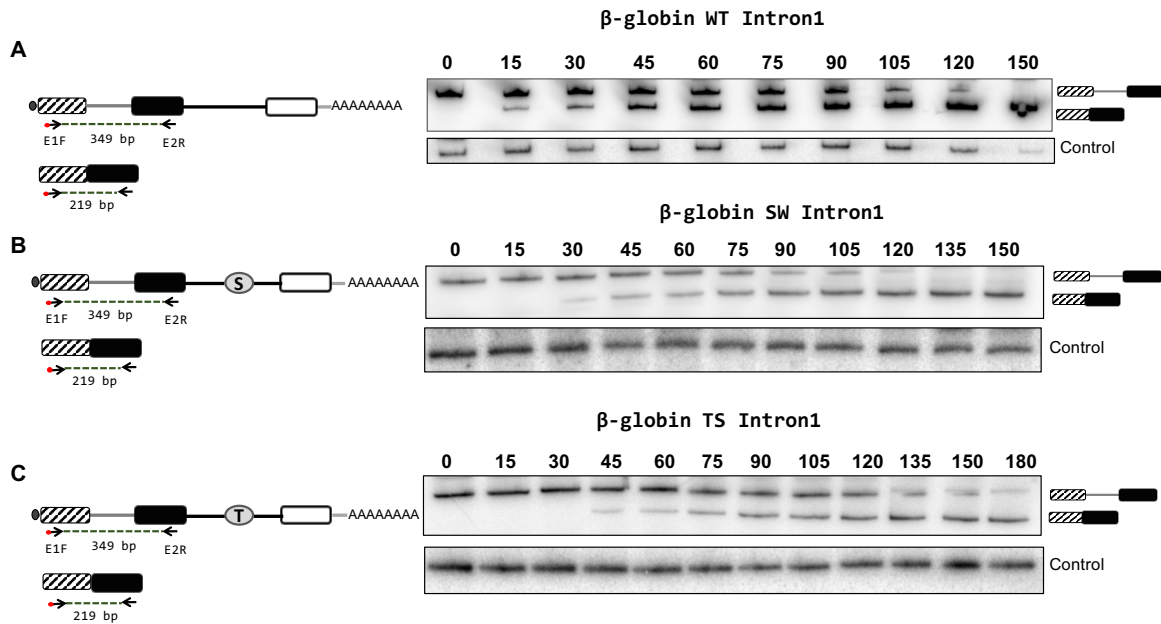
Figure 3.21: *In vitro* splicing assay to address the kinetics of intron1 removal. *In vitro* transcribed β-G-WT, β-G-TS and β-G-SW transcripts were incubated in splicing buffer containing HeLa nuclear extracts in timed reactions. The RNAs were purified and used as template for cDNA synthesis using E3R. E1F and E2R primers were used for PCR in a reaction spiked with P^{32}-labeled E1F. The products resolved on 8% native PAGE gels. Representative autoradiogram of the *in vitro* splicing profile of intron1 of β-G-WT (**A**), β-G-SW (**B**), and β-G-TS (**C**).

A first set of primers for PCR used unlabeled E1F and E2R primers with 5′end P^{32}-labeled E1F primer was spiked in to analyze the removal of intron1 by phosphor imaging. Products were analyzed by 8% native PAGE to resolve 0.3 kb unspliced and 0.2 kb spliced transcripts derived from intron1 of β-G-WT (**A**), β-G-TS (**B**) and β-G-SW (**C**) (**Figure 3.21**). A second set of primers used unlabeled E2F and E3R primers with spiked in 5′end P^{32}-labeled E2F to analyze the removal

of intron2. E2F and E3R primer would result in amplicons of 1.2 and 0.3 kb for WT and 1.9 and

0.3 kb for TS and SW corresponding to unspliced and spliced intron2, respectively. **Figure 3.22**

shows a 1.2% native agarose gel used to separate the RT-PCR products of intron2. The use of P^{32}-

labeled primer in RT-PCR enabled us to perform kinetic analysis of the spliced reactions.
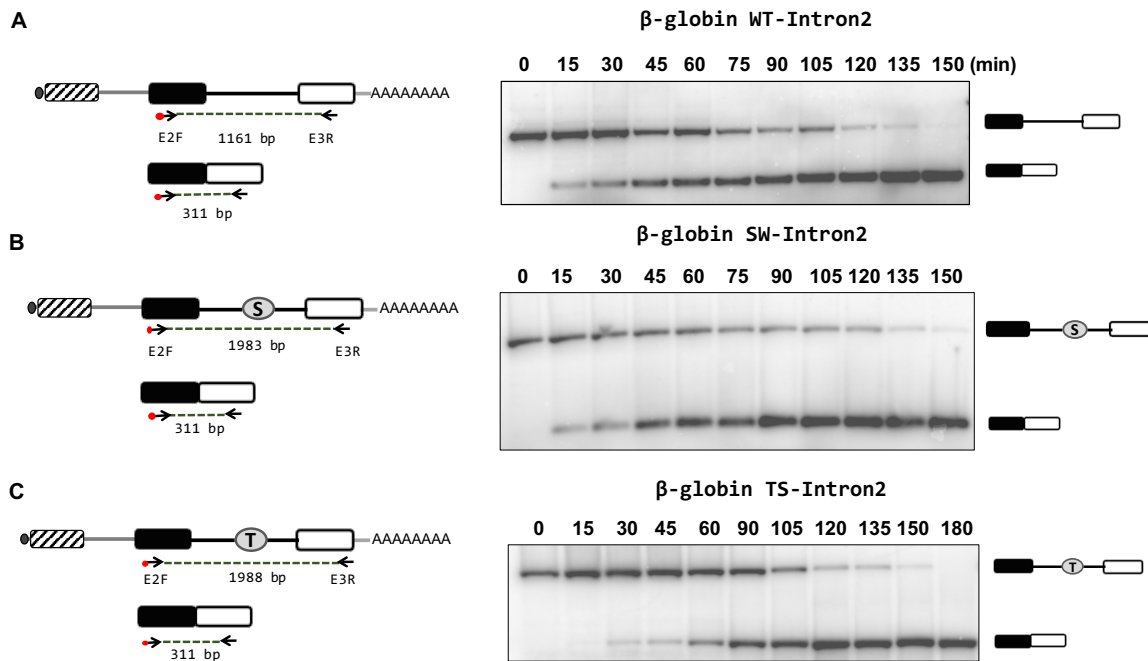
Figure 3.22: *In vitro* splicing assay to address the kinetics of intron2 removal. β-G-WT, β-G-TS and β-G-SW were incubated in splicing buffer containing HeLa nuclear extracts in timed reactions. The RNAs were purified and used as template for cDNA synthesis using E3R. E2F and E3R primers were used for PCR in a reaction spiked with P^{32}-labeled E2F. The products resolved on 1.2 % native agarose gels. Representative autoradiogram of the *in vitro* splicing profile of intron2 of β-G-WT (**A**), β-G-SW (**B**) and β-G-TS (**C**).

3. C.9: Kinetic analysis of *in vitro* splicing of β-globin and its related pri-miR transcripts.

The kinetic analysis of *in vitro* splicing reaction relies on the presence of a small amount of pre-

mRNA substrate and an excess of splicing factors contained within the nuclear extracts throughout

the reaction. Quantification of spliced products (after background subtraction and normalization

to time 0 reaction) is performed using Image J (NIH).

The fraction spliced for each intron is given by equation below:

Fraction Spliced = $I_{spliced}/I_{spliced + }I_{unspliced}$

Where,

$I_{spliced}$ = signal intensity from spliced amplicon in each lane

$I_{unspliced}$ = signal intensity from unspliced amplicon in each lane

Similarly, the substrate left was plotted for each intron and given by equation below:

Substrate left = $I_{unspliced}/I_{spliced + }I_{unspliced}$

The raw data for the quantification of fraction spliced and substrate left as a function of time are shown in **Figure 3.23**. We observed three distinct phases for each reaction: (i) a short lag phase before the appearance of product (ii) a linear phase where most of the product is formed, and (iii) the endpoint of the reaction. The quantification of the spliced amplicon for intron1 from β-G-WT, β-G-SW, and β-G-TS is shown in **Figure 3.23A** and for intron2 in **Figure 3.23B**. The amount of substrate left with time is quantified and plotted for intron1 (**Figure 3.23C**) and intron 2 (**Figure 3.23D**) of each transcript. It is evident from the kinetic profiles that measurable amounts of spliced product appear only after 15 min of incubation for WT, consistent with previous studies that in a time course of splicing reaction there is a slight delay in the appearance of the spliced product. Such a delay has been observed in other splicing assays and has been suggested to correspond to the time taken for the assembly of the spliceosome on the intron of the transcripts *in vitro* (Hicks et al. 2005).
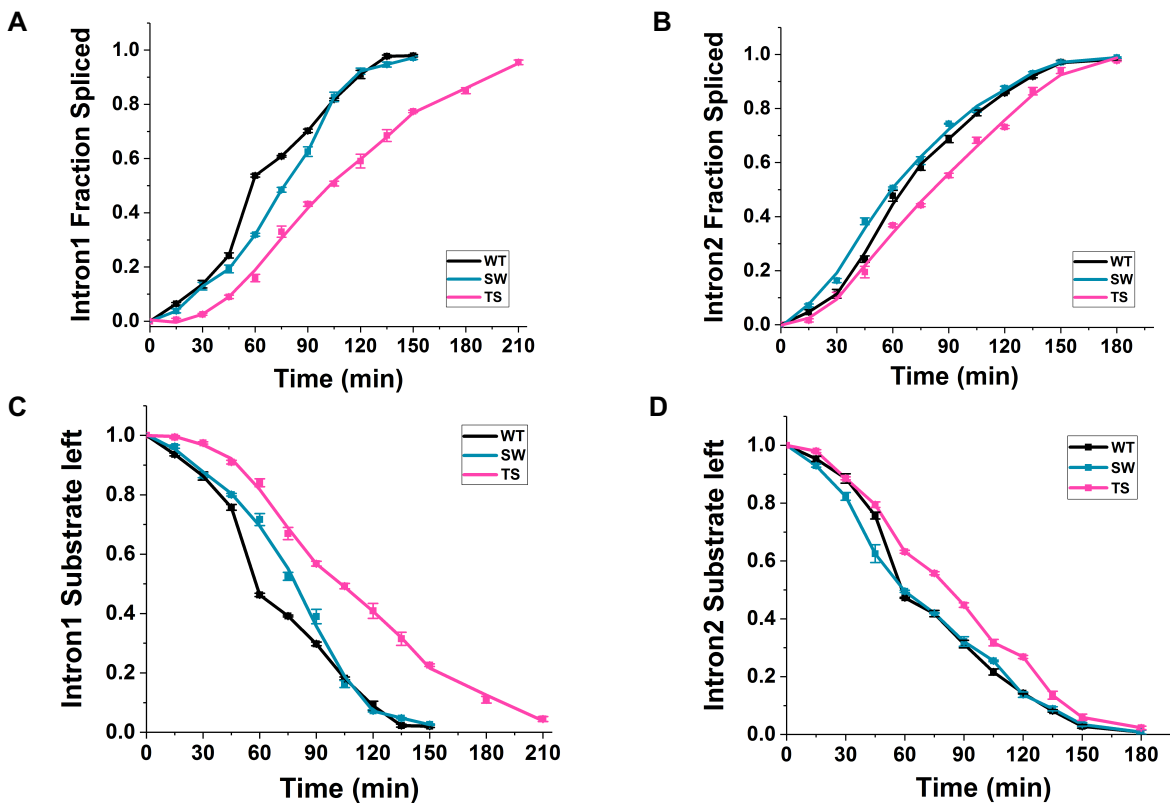
Figure 3.23: Kinetic analysis of the *in vitro* splicing reaction. *In vitro* splicing profile of intron1 and intron2 of the transcripts as a function of product appearance from β-G-WT (WT), β-G-SW (SW), and β-G-TS (TS) quantitated from three experiments. Fraction spliced is defined as intensity of spliced/(intensity of unspliced + intensity of spliced). The kinetic data was plotted and fit with a smooth curve to show the phases of the reaction for intron1 **(A)** and intron2 **(B)**. Substrate left is calculated from the intensity of unspliced / (intensity of unspliced + intensity of spliced) and plotted for intron1 **(C)** and intron2 **(D)**.

Interestingly, the delay observed in the appearance of the spliced product from each intron was significantly higher for β-G-TS compared to β-G-WT or β-G-SW. The time for appearance of spliced product corresponding to intron1 was 15 min for WT, 30 min for SW and 45 min for TS. Intron2 spliced product appearance was 15 min for WT & SW and 30 min for TS. To more rigorously analyze this, I defined the time taken for 20% formation of spliced product as Lag Time (T_{lag}). **Figure 3.24** shows the T_{lag} for intron1 and intron2 of β-G-WT, β-G-SW and β-G-SW.
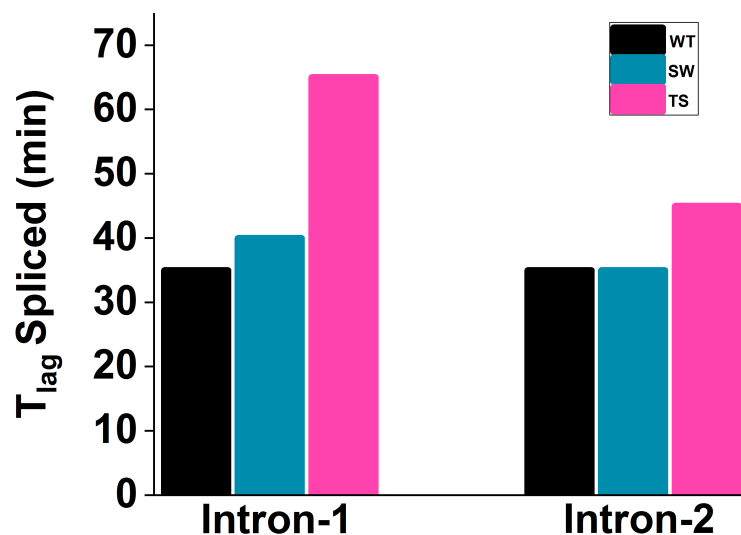
Figure 3.24: The time taken for ~20% of spliced product to be formed is represented by T_{lag}. The T_{lag} for splicing of intron1 and intron2 for β-G-WT (WT, black), β-G-SW (SW, cyan) and β-G-TS (TS, magenta) are shown.

Table 3.3 shows the summary of percentage completion of the splicing reaction and associated times for each of the transcripts β-G-WT, β-G-SW and β-G-TS.

Substrate	Time taken for 50% completion of splicing reaction (T_{50}) (min)	Time taken for 90% completion of splicing reaction (T_{90}) (min)
β-G-WT intron1	60	120
β-G-SW intron1	75	120
β-G-TS intron1	105	175
β-G-WT intron2	60	135
β-G-SW intron2	45	130
β-G-TS intron2	75	145

After the lag period the product formation follows the profile characteristic of a first-order reaction. Hence, I chose to consider the time taken to reach 90% of the reaction without the lag and I refer to this time window as $T_{reaction}$ (T_{rx}) in which to make splicing reaction rate measurements. **Figure 3.25** shows exponential fitting of reaction profiles of intron1 and intron2 of WT (black), SW (cyan), and TS (magenta) after the T_{rx} dataset was normalized to T_{lag}.
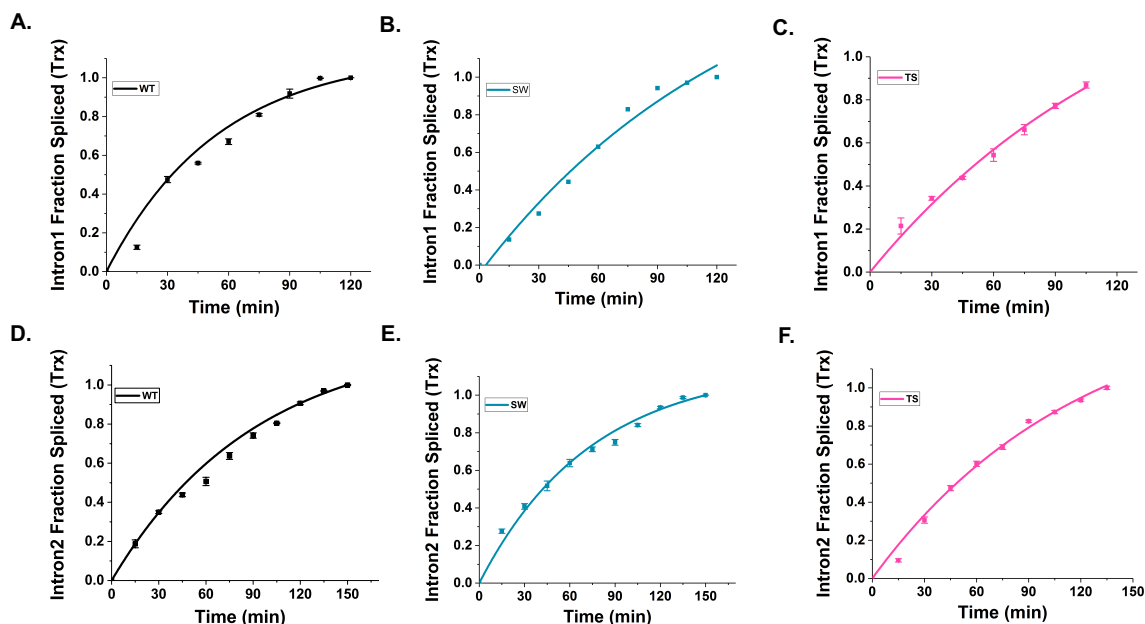
Figure 3.25: The exponential fit of the normalized data for product appearance of Trx for intron1 and intron2 of β-G-WT (WT, black), β-G-SW (SW, cyan), and β-G-TS (TS, magenta).

The rate of splicing for intron1 and intron2 of β-G-WT, β-G-SW and β-G-TS were quantified by the initial rate method from the slope of the plots in 3.20. Plots from **Figure 3.26** show that the initial rate of splicing of intron1 of β-G-TS was ~3.2-fold and intron2 was ~1.4 slower compared to β-G-WT. The β-G-SW showed 1.2-fold slower splicing of intron1 and 0.8-

fold faster splicing of intron2 compared to the WT. These results clearly indicate that the tertiary structured pri-miR impeded the splicing of its host.
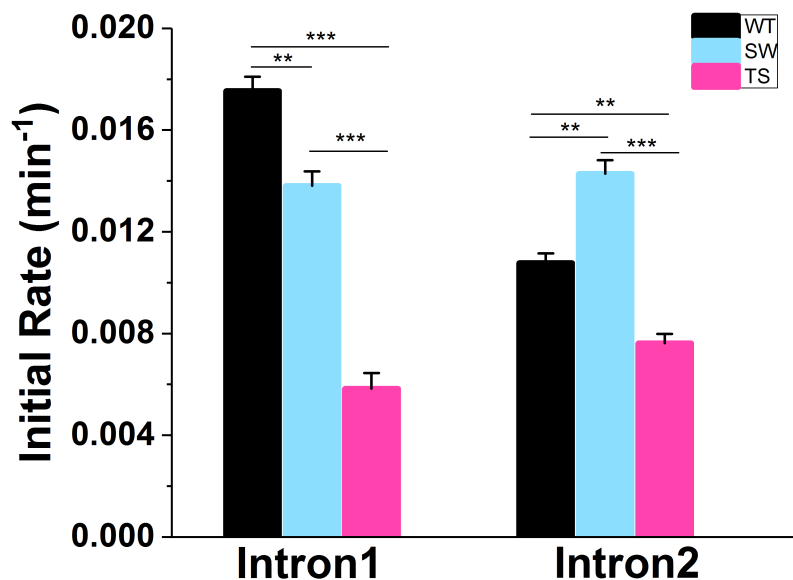
Figure 3.26: Initial rates for each intron were determined from the exponential fit of kinetic data after normalization to the lag (time taken for formation of approximate 20% product) and expressed as mean ± SEM from two independent experiments. Statistical analysis (Student's *t* test) was performed using Origin (Origin Labs). Significance is denoted by *. * indicate p value <0.05, ** p <0.01 and *** indicate p <0.001.

Next, the normalized data set was fitted to a first order description of product appearance given by the equation below,

$A = C (1-e^{-kt})$,

Where,

A = fraction spliced

C = fraction spliced at the endpoint of the reaction,

k = apparent rate constant

t = time.

Table 3.4 summarizes the observed rate constants for each intron. The apparent rate constant for *in vitro* splicing of intron1 from a full β-globin WT transcript is 0.021 min^{-1} and is consistent with the rate constant estimated for intron1 from the β-globin mini gene (two exon one intron system, E1-I1-E2) estimated at 0.029 min^{-1} (Mueller and Hertel 2014). Such measurement for intron2 removal was not done previously and our studies indicate that the rate constant for intron2 of β-globin WT is 0.014 min^{-1}.
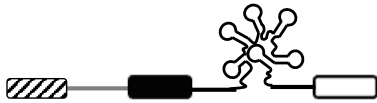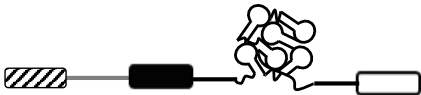
	Rate constant (min^{-1}) intron1 splicing	Rate constant (min^{-1}) intron2 splicing
β-G-WT	0.021± 0.002	0.014± 0.005
β-G-SW	0.018± 0.002	0.016± 0.003
β-G-TS	0.015± 0.001	0.010± 0.002

Table 3.4: Rate constants calculated by fitting the T$_{rx}$ data to a first order rate description for product appearance, A = C (1- e^{-kt}), where A is the fraction spliced, C is the fraction spliced at the endpoint of the reaction, k is the apparent rate constant, and t is the time. Each value was determined from at least two independent splicing reactions and expressed as mean± standard error of mean (SEM).

3. C.10: *In cellulis* splicing of C13orf25, the endogenous pri-miR 17-92a host gene.

Next, I wanted to understand the effect of the tertiary structured pri-miR-17-92a on its endogenous host gene, C13orf25, which undergoes alternative splicing. C13orf25 is a non-coding RNA

transcript which contains 4 exons (E1, E2a, E2c, E2d) and three introns (I1, I2, I3), and the 3 kb intron3 harbors the pri-miR-17-92a. Three Bacterial Artificial Chromosome (BAC) clones RP11-383J16, RP11-282D2, and RP11-94P6 containing human Chr13 DNA sequence were procured from Children's Hospital Oakland Research Institute (CHORI) and first checked for the presence of the pri-miR 17-92a sequence using PCR (**Figure 3.27B**).
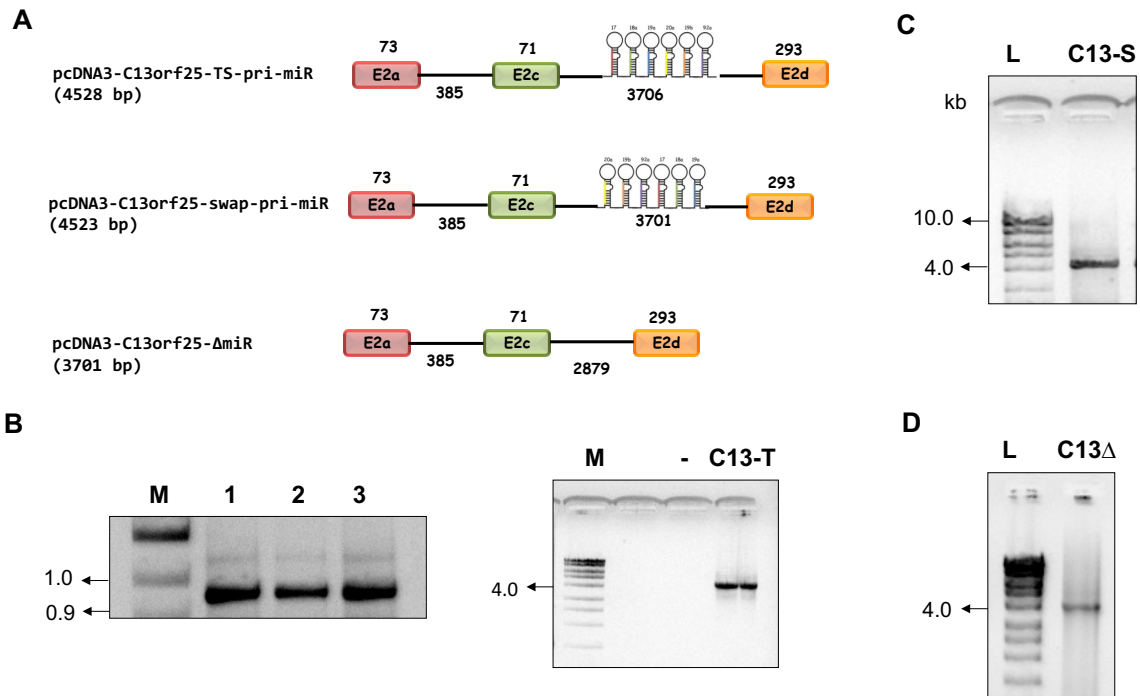
Figure 3.27: **A**. Schematic representation of C13orf25 and its related constructs. The C13orf25 gene containing sequences containing exon2 (E2a), intron2, exon3 (E2c), intron3 and exon4 (E2d) was cloned into pcDNA3-TOPO vector. Pri-miR-17-92a is located in intron3 of C13orf25 and this construct is refered as pcDNA3-C13orf25-TS. The 0.8 kb DNA sequence corresponding to pri-miR-17-92a was replaced by a swap-pri-miR to generate pcDNA3-C13orf25-SW and pcDNA3-ΔmiR was generated by deleting the pri-miR-17-92a sequence. **B**. 0.8% agarose gel showing the PCR amplification of endogenous 0.8kb pri-miR-17-92a from three BAC clones obtained from CHORI as DNA templates. Lane-1: PCR from BAC clone RP11-383J16, lane-2: PCR from RP11-282D2, and lane-3: PCR from RP11-94P6. C13T indicates amplification of 4.5 kb C13orf25-TS. **C.** 0.8% agarose gel to show amplification of 0.45kb C13orf25-SW made by Long Multiple Fusion method. **D.** 0.8% agarose gel showing the amplification of 0.37 kb C13orf25-ΔmiR constructed by SOE.

A mini gene was constructed with the 4.5 kb region of C13orf25 containing E2a-I2-E2c-I3-E2d from a BAC clone containing Chr13 sequence, and cloned it into the pcDNA3 mammalian expression vector, C13orf25-TS-pri-miR. Two mutant constructs: pcDNA3-C13orf25-Swap-pri-miR (TS-pri-miR replaced by swap-pri-miR), pcDNA3-C13orf25-inv-miR (TS-pri-miR sequence replaced by its reverse complementary sequence), and a deletion construct, pcDNA3-ΔmiR were also constructed. The 4.5 kb amplicon corresponding to C13orf25 with swap-pri-miR (C13-S) and 3.7 kb amplicon of C13orf25 where pri-miR is deleted (C13-ΔmiR) are shown in **Figure 3.27 C & D**, respectively. The schematic of the three constructs is shown in **Figure 3.27A** and the sequences of primers used for cloning are given in **Table 3.2**. These will be referred to as C13orf25-TS, C13orf25-SW, C13orf25-inv and C13orf25-ΔmiR respectively.

We used RT-PCR to measure levels of the spliced variants of C13orf25. Total RNA from HeLa cells transiently transfected with C13orf25-TS, C13orf25-SW, C13orf25-inv or C13orf25-ΔmiR expression plasmids was reverse transcribed by using E2d RP, which is complementary to exon2d. The resulting cDNAs were then amplified by using E2a FP, which hybridizes to exon2a, and E2d RP primers (**Table 3.2** for sequences). –RT indicates reactions that omitted reverse transcriptase enzyme to check for DNA contamination. The resulting PCR products were electrophoresed on 0.8% agarose gel shown in **Figure 3.28**. C13orf25-TS (lane TS-miR) shows three PCR amplicons of C13orf25 transcripts corresponding to a 5-kb band of E2a-I2-E2c-I3-E2d, a 0.4 kb band of E2a-E2c-E2d, and a ~0.2 kb band of E2a-E2d. C13orf25-SW (lane SW-miR) shows a ~ 4.5 kb amplicon of high intensity corresponding to E2a-I2-E2c-I3-E2d and fainter amplicons of 0.4 kb corresponding to the spliced E2a-Ec-E2d form.
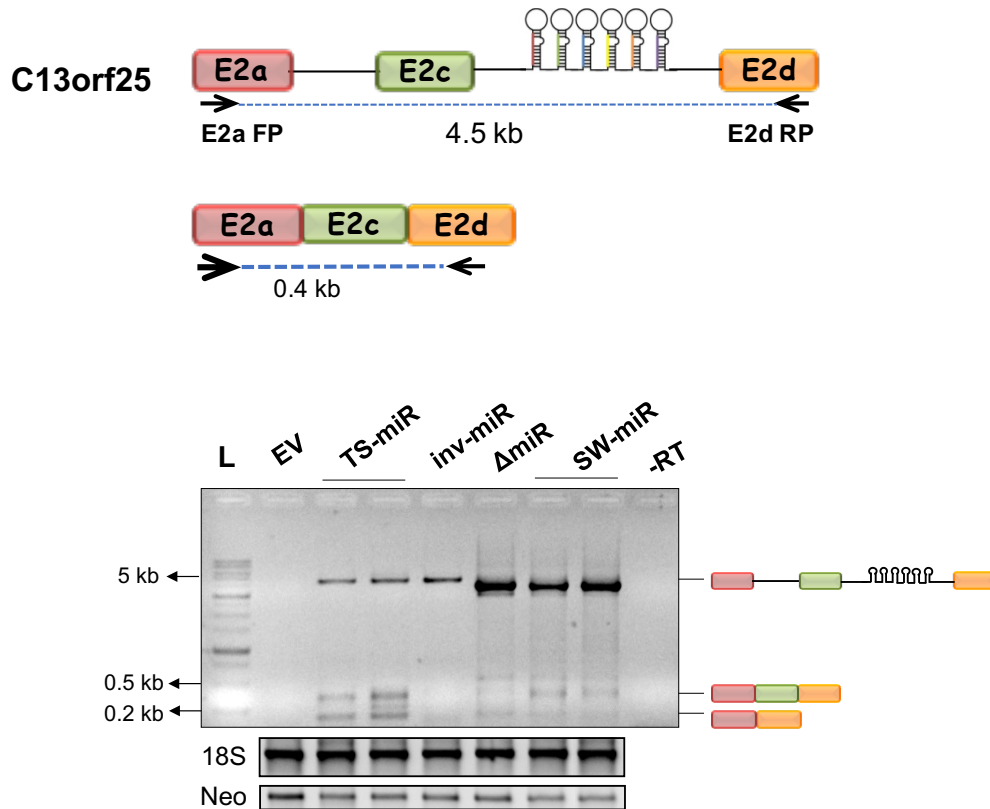
Figure 3.28: *In cellulis* splicing of C13orf25 and its related transcripts. RT-PCR analysis to address the steady levels of splicing from total RNA isolated from HeLa cells transfected with C13orf25 transcripts. 2 μg of total RNA was used for reverse transcription with E2d RP and the cDNA was subsequently used in PCR with E2a FP and E2d RP primers as shown in the schematic. 0.8% agarose gel showing RT-PCR products- EV: empty vector, lane TS-miR: C13orf25-TS (two plasmid clones were used for transfection), lane inv-miR: reverse complementary strand of TS-pri-miR, lane SW-miR: C13orf25-SW, ΔmiR: C13orf25 without the pri-miR sequence. -RT indicates reactions without reverse transcriptase.

The identity of the amplicons shown schematically was confirmed by sequencing. C13orf25-inv (lane inv-miR) shows no relevant amplicon. C13orf25-ΔmiR (lane ΔmiR) also shows faint amplicons (sequencing of these bands showed unclear identities). This suggests that TS-pri-miR influences the formation of different splice variants of C13orf25 transcript and I speculate that this effect could arise due to its altered kinetics of splicing that alters the preference of the splice site.

3. D: Conclusions

Splicing is a process that demands remarkable accuracy, and it is intriguing to contemplate the mechanisms that enable this accuracy: to remove introns ranging from as small as 30 nt to as large as 100 kb. Most mammalian microRNA (miRs) are expressed from the introns of protein-encoding and non-coding genes, and are transcribed by RNA Polymerase II along with their host transcripts (Rodriguez et al. 2004; Kim and Kim 2007). miRNA containing hairpins are cropped from pri-miRs by the Microprocessor complex (MPC) containing the nuclear RNase III enzyme Drosha and DGCR8 (Denli et al. 2004; Lee et al. 2006; Han et al. 2004). It is known that pri-miR processing is physically coupled to transcription and splicing (Shiohama et al. 2007; Agranat-Tamir et al. 2014) and that the processing is more efficient if pri-miR transcripts are retained at the transcription site (Pawlicki and Steitz 2008). Further, the clearance of introns following microprocessing of pri-miRs enhances the efficiency of splicing of the host transcript. Intronic miRNAs first undergo microRNA processing and then splicing suggesting that microprocessing precedes the completion of splicing (Kim and Kim 2007; Kataoka et al. 2009). Experiments in which self-cleaving RNA elements were included in introns demonstrate that fast cleavage leads to impaired splicing, while slow cleavage allows for proper exon tethering and effective pre-mRNA splicing (Fong et al. 2009). Additionally, splicing mutants in fission yeast reduce normal processing of centromeric transcripts into siRNAs and impair centromere silencing, suggesting that the spliceosome provides a platform that promotes siRNA biogenesis (Bayne et al. 2008).

We and others showed that tertiary structure formation by pri-miR transcripts influences their processing into pre-miRs (Chakraborty et al. 2012; Chaulk and Fahlman 2014; Chaulk et al. 2011) Our previous work found that autoregulation of pri-miR-17-92a is imposed early on by a

distinct tertiary structure in the RNA scaffold (See 2.C) A shuffled or swapped pri-miR construct with the structure of inter pre-miR elements (such as the region between pre-miR-19a and pre-miR-20a) disrupted, resulted in a transcript lacking this distinct tertiary structure that showed altered processing efficiency *in cellulis* and *in vitro*. Tertiary structured pri-miR clusters may also influence their processing not only at the individual pri-miR processing step by masking or unmasking sites for recognition and binding of MPC but also by targeting the unprocessed transcripts to sub nuclear domains. When overexpressed, nascent pri-miR transcripts which lack 3′cleavage and polyadenylation signal are found to be tethered to the DNA template and processed more efficiently than pri-miRs that are cleaved, polyadenylated and released (Pawlicki and Steitz 2008). Also, pri-miRs that contain RNA stabilizing elements like nuclear retention elements (ENE) are accumulated in the nucleoplasm and fail to generate high levels of its precursors. Intronic miRNAs are spliced more slowly compared to introns devoid of miRNAs, suggesting that increased residence of such pri-miRs at the sites of transcription may allow efficient miRNA processing from such introns before they are spliced (Kim et al 2007). Thus, I considered the possibility that tertiary structure of intronic pri-miR could regulate splicing via the MPC.

First, using a constitutively spliced β-globin reporter gene with a TS-pri-miR engineered into its intron, I was able to study its effect on splicing. RT-PCR analysis of *in cellulis* spliced products clearly revealed that the introduction of tertiary structured pri-miR impeded the removal of introns from its host transcript. This impedance was more pronounced on the upstream intron compared to the intron harboring it, as indicated by accumulation of unspliced transcript corresponding to intron1 and decreased levels of spliced product. A swapped pri-miR of the same length, also a substrate for MPC but devoid of such a tertiary structure, has no significant effect

on splicing of its host transcript.

To determine the effect of TS-pri-miR on the kinetics of intron removal, I needed a reliable *in vitro* splicing assay. *In vitro* splicing assays have been well established for mini genes and genes containing introns of less than 1 kb (Rooke and Underwood 2001; Mayeda and Krainer 2012), but the larger introns involved here required a modified assay. To this end, I optimized an *in vitro* splicing assay for these large RNAs. The principle strategy involved the use of RT-PCR based assay incorporating 5′ end radiolabeled primers to measure the *in vitro* splicing of the 1.7 kb intron2 of β-globin with pri-miR sequences. This method gave much more reliable and consistent data for *in vitro* spliced products compared to using internally labeled RNA transcripts resolved on denaturing gels after splicing or primer extension. The kinetic analysis of these *in vitro* splicing reactions clearly suggested that the transcript containing the TS-pri-miR has a slower rate of splicing. This effect was more pronounced at the upstream intron which demonstrated ~3.2 times slower splicing compared to the WT. An interesting observation from these studies was that the delay in the spliced product appearance, the lag was much higher for intron1 of the β-globin TS-pri-miR transcript. This lag for product appearance has been assigned to the time taken for the *in vitro* assembly of the spliceosomal components on the intron which would occur co-transcriptionally *in vivo* (Mueller and Hertel 2014). It is hypothesized that pre-mRNA substrates that contain weaker splicing signals or less efficient reactions usually display longer lags. However, the factors that account for this lag in the splicing reaction are still not clear. It is quite reasonable to suggest that the longer lag with the intron1 of TS-pri-miR could be due to inefficient assembly of the spliceosomal components. Successful microprocessing could lift this inhibition and aid in competent assembly of the spliceosome.

The tertiary structured pri-miR 17-92a is located in the third intron of its endogenous non-coding primary transcript known as *C13orf25*, which is alternatively spliced. Despite the remarkable conservation of the miRNA sequences, the exonic sequences of *C13orf25* are not measurably constrained between species, suggesting that the sole function of this transcript is to produce the relevant miRs in a regulated manner. *In cellulis* splicing studies on *C13orf25* host transcript indicated that the TS-pri-miR can regulate the proportion of different spliced forms indicating altered balance of the preferred splice site.

Our studies reveal that the host transcript containing TS-pri-miR in the intron can function as a splicing regulator by impeding splicing, consistent with the current model where microprocessing precedes splicing. The slow kinetics of splicing with the TS-pri-miR containing RNA also supports the proposed feed-forward regulation between microprocessing and splicing, where 5′ SS recognition by U1 RNP promotes microprocesssing of intronic miR-211 which in turn enhances splicing of its *melastatin* gene. It was proposed that recruitment of factors like KH-type splicing regulatory protein (KSRP) and heterogenous nuclear ribonucleoprotein A1 (hnRNPA1) could mediate such a mechanism. KSRP and hnRNP A1 are splicing regulatory proteins that are reported to bind to specific loop sequences that remodel the RNA secondary structure of the pre-miR hairpins (Trabucchi et al. 2009).

In a polycistronic miRNA like pri-miR-17-92a, as all the miRNAs in the cluster are co-transcribed, the spatio-temporal release of the individual pre-miRs from the cluster could potentially be achieved by modulating the *trans* binding proteins. For example, miR-18a of this pri-miRNA transcript is not a substrate for Drosha processing until it is locally remodeled by the binding of hnRNPA1 to its stem regions (Guil and Cáceres 2007). Binding of Lin28 to pre-let-7

pre-miRNA domain has been shown to substantially bend the helix and facilitate its processing (Newman et al. 2008). The tertiary structure of pri-miR-17-92a, as reported in the previous chapter (See 2.C) might mask recognition sites for the MPC, thus creating the observed kinetic barrier for its processing. This would then explain our observation of the slow kinetics of splicing for the TS-pri-miR. Importantly the lag in the appearance of the product corresponding to intron1 removal that was observed with the tertiary structure suggests that the reaction for the removal of intron1 removal is not efficient until the remodeling of the pri-miR-17-92a occurs. This could occur through *trans* binding proteins that allow microprocessing, which would then make it a better substrate for splicing. Thus, RNA tertiary structure could interfere with spliceosomal assembly by concealing splice sites or enhancer binding sites within stable helices, indicating the importance of the RNA structure as an active regulator of splicing **(Figure 3.29).**
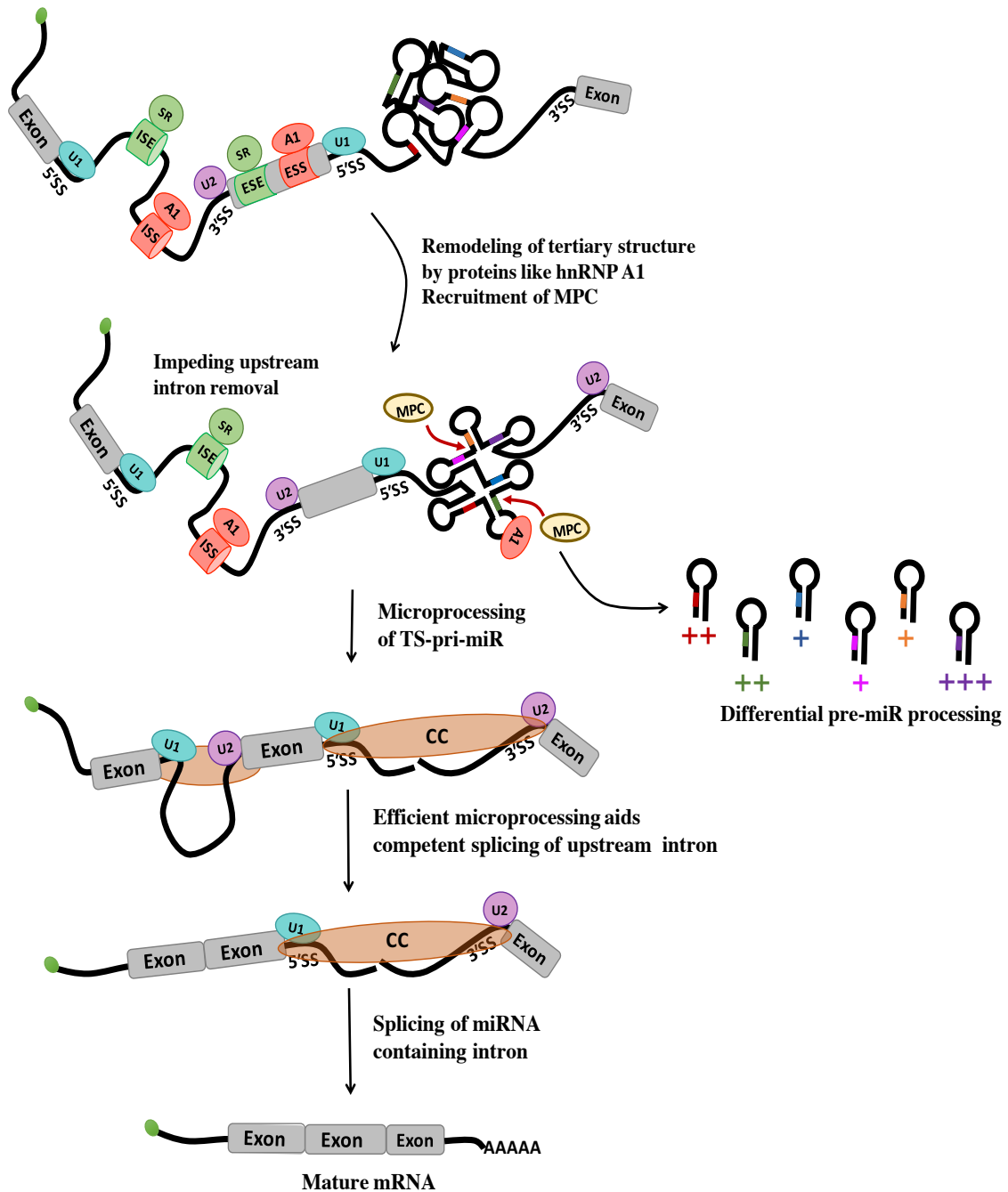
Figure 3.29: A model for splicing of tertiary structured intronic pri-miR transcript. Presence of tertiary structured pri-miR impedes splicing of its upstream intron possibly by masking the recognition sites for binding of *trans* regulatory proteins. Remodeling of the tertiary structure by proteins like hnRNP A1 result in efficient microprocessing of the intron harboring the pri-miR cluster. This then aids the splicing of upstream intron and because the exons flanking the intronic pri-miR have already been paired and tethered by the commitment complex (CC), splicing of this intron would still occur efficiently despite the discontinuity of the intron.

157

When it is harbored in a host transcript that undergoes alternative splicing the TS-pri-miR changes the relative abundance of different spliced products, which indicates its influence on splice site choice. C13orf25, the host transcript of pri-miR17-92a, gives two spliced variants one fully spliced with all its exons ligated and the other where intron2 and intron3 with the pri-miR 17-92a are retained. Given that microprocessing is co-transcriptional, one can envisage that expressing the miRs from this cluster could generate the smaller isoform, while the larger isoform with retained intron could be generated and stored in the sub nuclear domains until needed. Alternative splicing adds another dimension to the regulation of splice site selection as it involves modulating the pairing of selected splice sites (SS) that is determined by several parameters, including the proximity and strength of splicing signals. Exon and intron sequences which repress the use of nearby splice sites by forming of secondary structures that restrict the availability of SS have been described (Estes et al. 1992).

Implicated in miRNA biogenesis (Davis and Hata 2009), hnRNP A1, is an abundant protein which shuttles between the nucleus and the cytoplasm and has been extensively characterized for its role in splicing *in vivo* and *in vitro* (Del Gatto-Konczak et al. 1999; Cáceres et al. 1994). A constitutive role for hnRNP A1 in all steps of spliceosome assembly has been well established and it is essential importantly to help the splicing machinery discriminate between cryptic and functional 3' SS by forming a ternary complex with U2AF (Zhou et al. 2002; Jurica et al. 2002; Will and Lührmann 2011). While hnRNP A1 is known to bind to the exonic splicing silencer (ESS) and cause splicing repression, SR proteins bind to exonic splicing enhancer (ESE) to activate splicing (Simard and Chabot 2000; Del Gatto-Konczak et al. 1999; Paradis et al. 2007; Zhou and Fu 2013). The role of hnRNP A1 as a modulator of alternative splicing has been widely studied

(Jean-Philippe et al. 2013). Binding of hnRNPA1 to various *cis*-elements on the RNA regulates splicing by different mechanisms: (i) by binding to ESS overlapping an ESE, thus displacing the SR proteins and promoting exon skipping from the mRNA (Zahler et al. 2004; Rooke et al. 2003; Venables et al. 2005) (ii) by binding to a high affinity site which functions as an ESS, resulting in cooperative binding of other hnRNPA1 molecules on the transcript. This results in inhibition of SR protein binding thus promoting exon exclusion (Okunola and Krainer 2009). (iii) It can also bind to an intronic splicing silencer (ISS) overlapping with an intronic splicing enhancer (ISE) or BP, displacing SR proteins and U2 snRNP and thus inhibiting splicing of downstream exon (Tange et al. 2001; Guo et al. 2013) (iv) The interaction between hnRNPA1 proteins bound to the ISS of an upstream and downstream alternately spliced exon promotes looping out and exclusion of exon (Hutchison et al. 2002; Blanchette and Chabot 1999).

Interestingly hnRNP A1 pre-mRNA undergoes alternative splicing into two isoforms, hnRNP A1 and hnRNP A1b isoforms. The intron separating exon7 and hnRNP A1b-specific exon 7B has a highly-conserved element CE1, which is a binding site for hnRNP A1 protein which modulates the frequency of exon7B inclusion by looping out the internal 5′ SS and activating the distal site (Yang et al.1994). Changes in the expression of hnRNP A1/A1b have been reported to antagonize the activity of SR proteins by modulating splice site selection (Mayeda et al. 1993). Further, formation of intra-intronic competing RNA secondary structures has been reported to be crucial in mutually exclusive splicing that is a strictly regulated form of alternative splicing where one of two or more exclusive exons are part of the mature mRNA isoform (Graveley 2005; Yang et al. 2011).

A kinetic role for transcription in splicing was originally suggested by Eperon *et al* who found that the rate of RNA synthesis may affect its secondary structure, which in turn affects splicing (Eperon et al. 1988). Several studies thereafter indicated that alternative pre-mRNA splicing is modulated by regulating the kinetics of RNA Pol II wherein slow Pol II elongation allows weak splice sites to be recognized leading to higher rates of inclusion of alternate exons (Shukla and Oberdoerffer 2012; Dujardin et al. 2014; Brody et al. 2011; Roberts et al. 1998). Chakraborty *et al* showed that folding of pri-miR-17-92a into a tertiary structure masks Drosha binding sites and thus influences its processing into pre-miRNAs. I now propose that by folding into 3D-structure it could also mask or reveal accessibility to key splice sites and regulate splicing. Its removal by processing would also then be expected to lift this splicing inhibition.

Thus far, the role of the RNA structure in splicing has been limited to the formation of local secondary structures, which could influence splice site choice by masking or unmasking key regulatory sites required for binding to *trans*-regulatory proteins. The present study shows that the slow kinetics of intron removal associated with the TS-pri-miR may play a role in facilitating proper exon tethering and determining splicing choices of the host transcript. Further, splicing of *Dscam* gene hints an unusual mode of modulating long distance interactions in introns to generate different isoforms. Thus, it is not trivial to consider that *in vivo* RNA pairing could be possible over a long range between different introns of a given transcript. In such a situation, formation of tertiary structure could result in either masking the *cis*-regulatory elements, or in inter-intronic RNA pairing that could assist in bringing the distantly spaced elements together to form active splicing complex. Our data with TS-pri-miR-17-92a suggests that tertiary structured intronic miRNA clusters could interfere with the spliceosome assembly. This could be possibly mediated

by hnRNPA1 and other proteins to first allow efficient microprocessing, that is followed by splicing, adding another layer of gene regulation mediated by the introns. Understanding how molecular signatures on the intronic RNA can modulate its structural plasticity resulting in the host RNA auto-regulating its processing is an area of RNA-based gene regulation that is now ripe for investigation.

3. E: References

Agranat-Tamir, L., Shomron, N., Sperling, J. and Sperling, R. 2014. Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. *Nucleic Acids Research* 42(7), pp. 4640–4651.

Aitken, S., Alexander, R.D. and Beggs, J.D. 2011. Modelling reveals kinetic advantages of co-transcriptional splicing. *PLoS Computational Biology* 7(10), p. e1002215.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J. and Frey, B.J. 2010. Deciphering the splicing code. *Nature* 465(7294), pp. 53–59.

Bayne, E.H., Portoso, M., Kagansky, A., Kos-Braun, I.C., Urano, T., Ekwall, K., Alves, F., Rappsilber, J. and Allshire, R.C. 2008. Splicing factors facilitate RNAi-directed silencing in fission yeast. *Science* 322(5901), pp. 602–606.

Bird, G., Zorio, D.A.R. and Bentley, D.L. 2004. RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Molecular and Cellular Biology* 24(20), pp. 8963–8969.

Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* 72, pp. 291–336.

Blanchette, M. and Chabot, B. 1999. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *The EMBO Journal* 18(7), pp. 1939–1952.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.-M., Neugebauer, K.M., Darzacq, X. and Shav-Tal, Y. 2011. The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biology* 9(1), p. e1000573.

Cáceres, J.F., Stamm, S., Helfman, D.M. and Krainer, A.R. 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* 265(5179), pp. 1706–1709.

Chakraborty, S., Mehtab, S., Patwardhan, A. and Krishnan, Y. 2012. Pri-miR-17-92a transcript

folds into a tertiary structure and autoregulates its processing. *RNA (New York)* 18(5), pp. 1014–1028.

Chatzispyrou, I.A. and Houtkooper, R.H. 2014. The two-faced progeria gene and its implications in aging and metabolism. *EMBO Reports* 15(5), pp. 470–471.

Chaulk, S.G. and Fahlman, R.P. 2014. Tertiary structure mapping of the pri-miRNA miR-17~92. *Methods in Molecular Biology* 1182, pp. 43–55.

Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E.M., Veldhoen, R.A., Khanna, S.K., Goping, I.S., MacMillan, A.M., Mendell, J.T., Young, H.S., Fahlman, R.P. and Glover, J.N.M. 2011. Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biology* 8(6), pp. 1105–1114.

Chen, M. and Manley, J.L. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews. Molecular Cell Biology* 10(11), pp. 741–754.

Chen, Y. and Stephan, W. 2003. Compensatory evolution of a precursor messenger RNA secondary structure in the Drosophila melanogaster Adh gene. *Proceedings of the National Academy of Sciences of the United States of America* 100(20), pp. 11499–11504.

Chillón, I., Marcia, M., Legiewicz, M., Liu, F., Somarowthu, S. and Pyle, A.M. 2015. Native purification and analysis of long rnas. *Methods in Enzymology* 558, pp. 3–37.

Chu, C.S., Trapnell, B.C., Curristin, S., Cutting, G.R. and Crystal, R.G. 1993. Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nature Genetics* 3(2), pp. 151–156.

De Conti, L., Baralle, M. and Buratti, E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley interdisciplinary reviews. RNA* 4(1), pp. 49–60.

Davis, B.N. and Hata, A. 2009. Regulation of MicroRNA Biogenesis: A miRiad of mechanisms. *Cell Communication and Signaling* 7, p. 18.

Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F. and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* 432(7014), pp. 231–235.

Dujardin, G., Lafaille, C., la Mata, M. de, Marasco, L.E., Muñoz, M.J., Le Jossic-Corcos, C., Corcos, L. and Kornblihtt, A.R. 2014. How slow RNA polymerase II elongation favors alternative exon skipping. *Molecular Cell* 54(4), pp. 683–690.

Edery, I. and Sonenberg, N. 1985. Cap-dependent RNA splicing in a HeLa nuclear extract. *Proceedings of the National Academy of Sciences of the United States of America* 82(22), pp. 7590–7594.

Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q.,

O'Hanlon, D., Kim, P.M., Wrana, J.L. and Blencowe, B.J. 2012. Tissue-specific alternative splicing remodels protein-protein interaction networks. *Molecular Cell* 46(6), pp. 884–892.

Eperon, L.P., Graham, I.R., Griffiths, A.D. and Eperon, I.C. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *Cell* 54(3), pp. 393–401.

Erster, S.H., Finn, L.A., Frendewey, D.A. and Helfman, D.M. 1988. Use of RNase H and primer extension to analyze RNA splicing. *Nucleic Acids Research* 16(13), pp. 5999–6014.

Fletcher, S., Meloni, P.L., Johnsen, R.D., Wong, B.L., Muntoni, F. and Wilton, S.D. 2013. Antisense suppression of donor splice site mutations in the dystrophin gene transcript. *Molecular genetics & genomic medicine* 1(3), pp. 162–173.

Fong, N., Ohman, M. and Bentley, D.L. 2009. Fast ribozyme cleavage releases transcripts from RNA polymerase II and aborts co-transcriptional pre-mRNA processing. *Nature Structural & Molecular Biology* 16(9), pp. 916–922.

Del Gatto-Konczak, F., Olive, M., Gesnel, M.C. and Breathnach, R. 1999. hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer. *Molecular and Cellular Biology* 19(1), pp. 251–260.

Ghetti, A., Piñol-Roma, S., Michael, W.M., Morandi, C. and Dreyfuss, G. 1992. hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Research* 20(14), pp. 3671–3678.

Goldstrohm, A.C., Greenleaf, A.L. and Garcia-Blanco, M.A. 2001. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* 277(1-2), pp. 31–47.

Graveley, B.R. 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123(1), pp. 65–73.

Green, M.R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annual review of cell biology* 7, pp. 559–599.

Green, M.R., Maniatis, T. and Melton, D.A. 1983. Human beta-globin pre-mRNA synthesized in vitro is accurately spliced in Xenopus oocyte nuclei. *Cell* 32(3), pp. 681–694.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432(7014), pp. 235–240.

Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research* 36(Database issue), pp. D154–8.

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff,

K. and Hutton, M. 1999. 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *The Journal of Biological Chemistry* 274(21), pp. 15134–15143.

Guil, S. and Cáceres, J.F. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Structural & Molecular Biology* 14(7), pp. 591–596.

Guo, R., Li, Y., Ning, J., Sun, D., Lin, L. and Liu, X. 2013. HnRNP A1/A2 and SF2/ASF regulate alternative splicing of interferon regulatory factor-3 and affect immunomodulatory functions in human non-small cell lung cancer cells. *Plos One* 8(4), p. e62729.

Hang, J., Wan, R., Yan, C. and Shi, Y. 2015. Structural basis of pre-mRNA splicing. *Science* 349(6253), pp. 1191–1198.

Han, J., Lee, Y., Yeom, K.-H., Kim, Y.-K., Jin, H. and Kim, V.N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Development* 18(24), pp. 3016–3027.

Han, J., Xiong, J., Wang, D. and Fu, X.-D. 2011. Pre-mRNA splicing: where and when in the nucleus. *Trends in Cell Biology* 21(6), pp. 336–343.

Heckman, K.L. and Pease, L.R. 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols* 2(4), pp. 924–932.

Heyd, F. and Lynch, K.W. 2011. Degrade, move, regroup: signaling control of splicing proteins. *Trends in Biochemical Sciences* 36(8), pp. 397–404.

Hicks, M.J., Lam, B.J. and Hertel, K.J. 2005. Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. *Methods* 37(4), pp. 306–313.

Hoskins, A.A. and Moore, M.J. 2012. The spliceosome: a flexible, reversible macromolecular machine. *Trends in Biochemical Sciences* 37(5), pp. 179–188.

Hutchison, S., LeBel, C., Blanchette, M. and Chabot, B. 2002. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *The Journal of Biological Chemistry* 277(33), pp. 29745–29752.

Janas, M.M., Khaled, M., Schubert, S., Bernstein, J.G., Golan, D., Veguilla, R.A., Fisher, D.E., Shomron, N., Levy, C. and Novina, C.D. 2011. Feed-forward microprocessing and splicing activities at a microRNA-containing intron. *PLoS Genetics* 7(10), p. e1002330.

Jean-Philippe, J., Paz, S. and Caputi, M. 2013. hnRNP A1: the Swiss army knife of gene expression. *International Journal of Molecular Sciences* 14(9), pp. 18999–19024.

Jiang, Z., Cote, J., Kwon, J.M., Goate, A.M. and Wu, J.Y. 2000. Aberrant splicing of tau pre-mRNA caused by intronic mutations associated with the inherited dementia frontotemporal dementia with parkinsonism linked to chromosome 17. *Molecular and Cellular Biology* 20(11),

pp. 4036–4048.

Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N. and Moore, M.J. 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA (New York)* 8(4), pp. 426–439.

Kar, A., Fushimi, K., Zhou, X., Ray, P., Shi, C., Chen, X., Liu, Z., Chen, S. and Wu, J.Y. 2011. RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Molecular and Cellular Biology* 31(9), pp. 1812–1821.

Kataoka, N. and Dreyfuss, G. 2008. Preparation of efficient splicing extracts from whole cells, nuclei, and cytoplasmic fractions. *Methods in Molecular Biology* 488, pp. 357–365.

Kataoka, N., Fujita, M. and Ohno, M. 2009. Functional association of the Microprocessor complex with the spliceosome. *Molecular and Cellular Biology* 29(12), pp. 3243–3254.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M. and Stamm, S. 2013. Function of alternative splicing. *Gene* 514(1), pp. 1–30.

Kim, Y.-K. and Kim, V.N. 2007. Processing of intronic microRNAs. *The EMBO Journal* 26(3), pp. 775–783.

Konarska, M.M., Padgett, R.A. and Sharp, P.A. 1984. Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* 38(3), pp. 731–736.

Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. 1984. Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* 36(4), pp. 993–1005.

Lai, D., Proctor, J.R. and Meyer, I.M. 2013. On the importance of cotranscriptional RNA structure formation. *RNA (New York)* 19(11), pp. 1461–1473.

Leach, K.M., Vieira, K.F., Kang, S.-H.L., Aslanian, A., Teichmann, M., Roeder, R.G. and Bungert, J. 2003. Characterization of the human beta-globin downstream promoter region. *Nucleic Acids Research* 31(4), pp. 1292–1301.

Lee, Y., Han, J., Yeom, K.H., Jin, H. and Kim, V.N. 2006. Drosha in primary microRNA processing. *Cold Spring Harbor Symposia on Quantitative Biology* 71, pp. 51–57.

Lee, Y. and Rio, D.C. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual Review of Biochemistry* 84, pp. 291–323.

Liu, W., Zhou, Y., Hu, Z., Sun, T., Denise, A., Fu, X.-D. and Zhang, Y. 2010. Regulation of splicing enhancer activities by RNA secondary structures. *FEBS Letters* 584(21), pp. 4401–4407.

Lopez-Mejia, I.C., de Toledo, M., Chavey, C., Lapasset, L., Cavelier, P., Lopez-Herrera, C., Chebli, K., Fort, P., Beranger, G., Fajas, L., Amri, E.Z., Casas, F. and Tazi, J. 2014. Antagonistic functions of LMNA isoforms in energy expenditure and lifespan. *EMBO Reports* 15(5), pp. 529–539.

Lynch, K.W. 2007. Regulation of alternative splicing by signal transduction pathways. *Advances in Experimental Medicine and Biology* 623, pp. 161–174.

Marotta, C.A., Forget, B.G., Weissman, S.M., Verma, I.M., McCaffrey, R.P. and Baltimore, D. 1974. Nucleotide sequences of human globin messenger RNA. *Proceedings of the National Academy of Sciences of the United States of America* 71(6), pp. 2300–2304.

Matlin, A.J., Clark, F. and Smith, C.W.J. 2005. Understanding alternative splicing: towards a cellular code. *Nature Reviews. Molecular Cell Biology* 6(5), pp. 386–398.

Matsuo, M., Nishio, H., Kitoh, Y., Francke, U. and Nakamura, H. 1992. Partial deletion of a dystrophin gene leads to exon skipping and to loss of an intra-exon hairpin structure from the predicted mRNA precursor. *Biochemical and Biophysical Research Communications* 182(2), pp. 495–500.

Mayeda, A., Helfman, D.M. and Krainer, A.R. 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Molecular and Cellular Biology* 13(5), pp. 2993–3001.

Mayeda, A. and Krainer, A.R. 2012. In Vitro Splicing Assays. In: Stamm, S., Smith, C. W. J., and Lührmann, R. eds. *Alternative pre-mRNA Splicing*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, pp. 320–329.

May, G.E., Olson, S., McManus, C.J. and Graveley, B.R. 2011. Competing RNA secondary structures are required for mutually exclusive splicing of the Dscam exon 6 cluster. *RNA (New York)* 17(2), pp. 222–229.

McManus, C.J. and Graveley, B.R. 2011. RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development* 21(4), pp. 373–379.

Misteli.T. 2000. Different site, different splice. *Nature Cell Biology* 2, pp. E98-E100.

Movassat, M., Mueller, W.F. and Hertel, K.J. 2014. In vitro assay of pre-mRNA splicing in mammalian nuclear extract. *Methods in Molecular Biology* 1126, pp. 151–160.

Mueller, W.F. and Hertel, K.J. 2014. Kinetic analysis of in vitro pre-mRNA splicing in HeLa nuclear extract. *Methods in Molecular Biology* 1126, pp. 161–168.

Naftelberg, S., Schor, I.E., Ast, G. and Kornblihtt, A.R. 2015. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annual Review of Biochemistry* 84, pp. 165–198.

Newman, M.A., Thomson, J.M. and Hammond, S.M. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA (New York)* 14(8), pp. 1539–1549.

Niblock, M. and Gallo, J.-M. 2012. Tau alternative splicing in familial and sporadic tauopathies. *Biochemical Society Transactions* 40(4), pp. 677–680.

Okunola, H.L. and Krainer, A.R. 2009. Cooperative-binding and splicing-repressive properties of hnRNP A1. *Molecular and Cellular Biology* 29(20), pp. 5620–5631.

Ota, A., Tagawa, H., Karnan, S., Tsuzuki, S., Karpas, A., Kira, S., Yoshida, Y. and Seto, M. 2004. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. *Cancer Research* 64(9), pp. 3087–3095.

Pandya-Jones, A. and Black, D.L. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA (New York)* 15(10), pp. 1896–1908.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics* 40(12), pp. 1413–1415.

Pan, T. and Sosnick, T. 2006. RNA folding during transcription. *Annual review of biophysics and biomolecular structure* 35, pp. 161–175.

Paradis, C., Cloutier, P., Shkreta, L., Toutant, J., Klarskov, K. and Chabot, B. 2007. hnRNP I/PTB can antagonize the splicing repressor activity of SRp30c. *RNA (New York)* 13(8), pp. 1287–1300.

Pawlicki, J.M. and Steitz, J.A. 2008. Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. *The Journal of Cell Biology* 182(1), pp. 61–76.

Pereira, M.J.B., Behera, V. and Walter, N.G. 2010. Nondenaturing purification of co-transcriptionally folded RNA avoids common folding heterogeneity. *Plos One* 5(9), p. e12953.

Rahman, M.A., Masuda, A., Ohe, K., Ito, M., Hutchinson, D.O., Mayeda, A., Engel, A.G. and Ohno, K. 2013. HnRNP L and hnRNP LL antagonistically modulate PTB-mediated splicing suppression of CHRNA1 pre-mRNA. *Scientific reports* 3, pp. 2931.

Raker, V.A., Mironov, A.A., Gelfand, M.S. and Pervouchine, D.D. 2009. Modulation of alternative splicing by long-range RNA structures in Drosophila. *Nucleic Acids Research* 37(14), pp. 4533–4544.

Reichert, V. and Moore, M.J. 2000. Better conditions for mammalian in vitro splicing provided by acetate and glutamate as potassium counterions. *Nucleic Acids Research* 28(2), pp. 416–423.

Revil, T., Gaffney, D., Dias, C., Majewski, J. and Jerome-Majewska, L.A. 2010. Alternative splicing is frequent during early embryonic development in mouse. *BMC Genomics* 11, pp. 399.

Rino, J. and Carmo-Fonseca, M. 2009. The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends in Cell Biology* 19(8), pp. 375–384.

Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J. and Smith, C.W. 1998. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Research* 26(24), pp. 5568–5572.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L. and Bradley, A. 2004. Identification of

mammalian microRNA host genes and transcription units. *Genome Research* 14(10A), pp. 1902–1910.

Rooke, N., Markovtsov, V., Cagavi, E. and Black, D.L. 2003. Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. *Molecular and Cellular Biology* 23(6), pp. 1874–1884.

Rooke, N. and Underwood, J. 2001. In vitro RNA splicing in mammalian cell extracts. *Current protocols in cell biology* Chapter 11, Unit 11.17.

Schroeder, R., Grossberger, R., Pichler, A. and Waldsich, C. 2002. RNA folding in vivo. *Current Opinion in Structural Biology* 12(3), pp. 296–300.

Senanayake, S.D. and Brian, D.A. 1995. Precise large deletions by the PCR-based overlap extension method. *Molecular Biotechnology* 4(1), pp. 13–15.

Shevchuk, N.A., Bryksin, A.V., Nusinovich, Y.A., Cabello, F.C., Sutherland, M. and Ladisch, S. 2004. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Research* 32(2), pp. e19.

Shiohama, A., Sasaki, T., Noda, S., Minoshima, S. and Shimizu, N. 2007. Nucleolar localization of DGCR8 and identification of eleven DGCR8-associated proteins. *Experimental Cell Research* 313(20), pp. 4196–4207.

Shukla, S. and Oberdoerffer, S. 2012. Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochimica et Biophysica Acta* 1819(7), pp. 673–683.

Simard, M.J. and Chabot, B. 2000. Control of hnRNP A1 alternative splicing: an intron element represses use of the common 3' splice site. *Molecular and Cellular Biology* 20(19), pp. 7353–7362.

Sirand-Pugnet, P., Durosay, P., Orval, B.C. d', Brody, E. and Marie, J. 1995. β-Tropomyosin Pre-mRNA Folding Around a Muscle-specific Exon Interferes with Several Steps of Spliceosome Assembly. *Journal of Molecular Biology* 251(5), pp. 591–602.

Somarowthu, S., Legiewicz, M., Chillón, I., Marcia, M., Liu, F. and Pyle, A.M. 2015. HOTAIR forms an intricate and modular secondary structure. *Molecular Cell* 58(2), pp. 353–361.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A. and Soreq, H. 2005. Function of alternative splicing. *Gene* 344, pp. 1–20.

Tagawa, H. and Seto, M. 2005. A microRNA cluster as a target of genomic amplification in malignant lymphoma. *Leukemia* 19(11), pp. 2013–2016.

Tange, T.O., Damgaard, C.K., Guth, S., Valcárcel, J. and Kjems, J. 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *The EMBO Journal* 20(20), pp. 5748–5758.

Tarn, W.-Y. 2007. Cellular signals modulate alternative splicing. *Journal of Biomedical Science*

14(4), pp. 517–522.

Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A.D., Filipowicz, W., Ramos, A., Gherzi, R. and Rosenfeld, M.G. 2009. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* 459(7249), pp. 1010–1014.

Traunmüller, L., Gomez, A.M., Nguyen, T.-M. and Scheiffele, P. 2016. Control of neuronal synapse specification by a highly dedicated alternative splicing program. *Science* 352(6288), pp. 982–986.

Valadkhan, S. 2007. The spliceosome: a ribozyme at heart? *Biological Chemistry* 388(7), pp. 693–697.

Venables, J.P., Bourgeois, C.F., Dalgliesh, C., Kister, L., Stevenin, J. and Elliott, D.J. 2005. Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. *Human Molecular Genetics* 14(16), pp. 2289–2303.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456(7221), pp. 470–476.

Wang, H., Chen, Y., Li, X., Chen, G., Zhong, L., Chen, G., Liao, Y., Liao, W. and Bin, J. 2016. Genome-wide analysis of alternative splicing during human heart development. *Scientific reports* 6, p. 35520.

Wang, Z. and Burge, C.B. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA (New York)* 14(5), pp. 802–813.

Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. and Chang, H.Y. 2011. Understanding the transcriptome through RNA structure. *Nature Reviews. Genetics* 12(9), pp. 641–655.

Warf, M.B. and Berglund, J.A. 2010. Role of RNA structure in regulating pre-mRNA splicing. *Trends in Biochemical Sciences* 35(3), pp. 169–178.

Webb, C.-H.T. and Hertel, K.J. 2014. Preparation of splicing competent nuclear extracts. *Methods in Molecular Biology* 1126, pp. 117–121.

Wein, N., Vulin, A., Falzarano, M.S., Szigyarto, C.A.-K., Maiti, B., Findlay, A., Heller, K.N., Uhlén, M., Bakthavachalu, B., Messina, S., Vita, G., Passarelli, C., Brioschi, S., Bovolenta, M., Neri, M., Gualandi, F., Wilton, S.D., Rodino-Klapac, L.R., Yang, L., Dunn, D.M., Schoenberg, D.R., Weiss, R.B., Howard, M.T., Ferlini, A. and Flanigan, K.M. 2014. Translation from a DMD exon 5 IRES results in a functional dystrophin isoform that attenuates dystrophinopathy in humans and mice. *Nature Medicine* 20(9), pp. 992–1000.

Will, C.L. and Lührmann, R. 2011. Spliceosome structure and function. *Cold Spring Harbor Perspectives in Biology* 3(7).

Wong, T.N. and Pan, T. 2009. RNA Folding During Transcription: Protocols and Studies. In: *Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding: Part A*. Methods in Enzymology. Elsevier, pp. 167–193.

Xu, Q., Modrek, B. and Lee, C. 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Research* 30(17), pp. 3754–3766.

Yang, Y., Zhan, L., Zhang, W., Sun, F., Wang, W., Tian, N., Bi, J., Wang, H., Shi, D., Jiang, Y., Zhang, Y. and Jin, Y. 2011. RNA secondary structure in mutually exclusive splicing. *Nature Structural & Molecular Biology* 18(2), pp. 159–168.

Yuan, X., Liu, C., Yang, P., He, S., Liao, Q., Kang, S. and Zhao, Y. 2009. Clustered microRNAs' coordination in regulating protein-protein interaction network. *BMC Systems Biology* 3, p. 65.

Zahler, A.M., Damgaard, C.K., Kjems, J. and Caputi, M. 2004. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. *The Journal of Biological Chemistry* 279(11), pp. 10077–10084.

Zerivitz, K. and Akusjärvi, G. 1989. An improved nuclear extract preparation method. *Gene analysis techniques* 6(5), pp. 101–109.

Zhou, Z. and Fu, X.-D. 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* 122(3), pp. 191–207.

Zhou, Z., Licklider, L.J., Gygi, S.P. and Reed, R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* 419(6903), pp. 182–185.