The University of Southern Mississippi

The Aquila Digital Community

Fall 10-11-2021

Human 5'-tailed Mirtrons are Processed by RNaseP

Mohammad Farid Zia

HUMAN 5'-TAILED MIRTRONS ARE PROCESSED BY RNASEP

by

M. Farid Zia

A Dissertation
Submitted to the Graduate School,
the College of Arts and Sciences
and the School of Biological, Environmental, and Earth Sciences
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved by:

Dr. Alex S. Flynt, Committee Chair
Dr. Mohammed Elasri
Dr. YanLin Guo
Dr. Fengwei Bai
Dr. Hao XuName of Committee Member

December 2021

ABSTRACT

microRNAs are pervasive small noncoding ~22 nt gene regulators that collectively control the expressions by destabilizing mRNA or repressing translation. Several thousand of miRNA genes have been reported in mammalians that mostly are generated through canonical pathway while non-canonical miRNAs are not always generated in the same pathways. Mirtrons are one of miRNA variations that bypass Drosha cleavage step and instead are produced by splicing mechanism. There are 3 classes of mirtrons: conventional, 5'- and 3'-tailed. In tailed mirtrons, splicing defines both pre-miRNA hairpin ends. Plenty of studies have been done on mirtrons but still 5'-tailed mirtrons processing has remained unclear. Here, we studied the biogenesis of human 5'-tailed mirtrons and the role of Ribonucleases in processing the tails. To address this, we engineered miR-5010 tail on the 5p end and examined their biogenesis changes. Our analysis on RNA sequencing revealed that in miR-5010, RNase P complex is involved in processing the 5p tails by its endoribonuclease activity. We also demonstrated that when Rpp30 (subunit of RNase P) is downregulated, the mirtrons undergo "Arm Switching". Besides, XRN family exoribonucleases seem to degrade the 5'-tailed mirtrons that have not been processed by the RNase P complex.

ACKNOWLEDGMENTS

The path toward this research study has been circuitous. Its accomplishment is thanks in large part to the very special people who, supported, challenged, and stuck with me along the way. I was tremendously fortunate to have Dr. Alex Flynt as my supervisor. He brought a depth of knowledge that few could match. I appreciate him for supporting and always aiming me forward and giving me such valuable opportunity to work under his supervision.

The next group of people who I wish to thank are my committee members for giving me such thoughtful feedback, always aimed at moving me forward.

A special recognition must be given to my wife Mona Fendereski whom I couldn't imagine my life without her. When I was stressed, she calmed me down and helped me when busy. When I was sad, she cheered me up. When I made a bad joke, she laughed and I know as long as she is by my side, everything will be just fine.

In addition, I would like to extend my sincere gratitude to my friends and lab mates Sweta Khanal and Jacob Peter for their friendship and supports and Dr. Yulica Santos Ortega for her helps which was pivotal in the dissertation's development.

Lastly, I would like to thank my parents for their infinite support and encouragement, and for believing in me at times when it was difficult to believe in myself. I also would like to thank my two brothers Amir and Navid for their unending support and keeping me company on long walks.

And now that I will have time to tend to relationships, I want to thank my six unborn children for being so patient."

DEDICATION

I would like to completely dedicate this work to my respectful parents and parents-in-law and my beloved wife without whose constant support this dissertation was not possible. They always inspire me, and I will always appreciate all they have done.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I - INTRODUCTION

1.1 Non-Coding RNAs

Messenger RNAs (mRNAs) are the middlemen of information flow in cells being intermediates between DNA and proteins. However, many RNA molecules are non-coding RNAs (ncRNA) functioning by regulating mRNAs (Delihas, 2015; Diamantopoulos, Tsiakanikas, & Scorilas, 2018). As the name implies, ncRNAs are RNAs that do not code for protein, splitting them from mRNA, their coding counterparts. The first ncRNAs were identified in the late 1950s with the discovery of transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), and since then other classes have surfaced (Palazzo & Lee, 2015; P. Zhang, Wu, Chen, & Chen, 2019), such as microRNAs (miRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and others (Mattick & Makunin, 2006; Shahrouki & Larsson, 2012). This ncRNA family is much larger compared to coding RNAs. Alteration in the expression of ncRNAs, may contribute to pathological phenotypes including cancers and diseases such as Alzheimer's, Autism, and heart diseases and this makes them a very broad and significantly important area of study (Knoll, Lodish, & Sun, 2015; Wang, Han, Sun, Chen, & Chen, 2019; L. Zhou et al., 2018). These findings suggest that ncRNAs are associated with the complexity of prokaryote and higher eukaryotes involved in a plethora of diverse functions (Geisler & Coller, 2013). Different ncRNAs have role in different cellular processes where mostly are involved in translation and regulations. In addition, they function in RNA splicing and DNA replication (Cheng et al., 2005; Mattick & Makunin, 2006). ncRNAs are divided into two main categories according to their transcript sizes: long non-coding RNAs (lncRNAs) which are longer than 200 nt and

small non-coding RNAs (sncRNAs) with a length typically lower than 100 nt. The small

ncRNAs are generated through a variety of pathways (Menezes, Balzeau, & Hagan,

2018; K. Miyoshi, T. Miyoshi, & H. Siomi, 2010; Okamura, Hagen, Duan, Tyler, & Lai,

2007; Ragusa et al., 2017).

1.2 MicroRNAs

miRNAs are small non-coding ~22 nucleotides RNAs which are mostly

transcribed by RNA polymerase II and processed into precursor miRNA (pre-miRNAs)

by folding into a stem- loop structure. The ubiquitous presence of different miRNAs

along with their target sites in many living organisms suggests an evolutionary ancient

origin and strong conservation (Ladewig, Okamura, Flynt, Westholm, & Lai, 2012;

Xiong, Schneider, Hulsey, Meyer, & Franchini, 2019). Thousands of miRNAs have been

discovered in plants, animals, fungi and viruses indicating an ancient origin of evolution.

Many families of miRNAs as well as their target sites are conserved among species

which demonstrates similar evolutionary pathways (Xiong et al., 2019). These small

RNAs bind to effector proteins, Argonautes (Ago), to form RNA-Induced Silencing

Complex (RISC) complexes. The miRNA guides the complex to the targets typically

found in the 3' UTR of the mRNAs and form Watson-Crick base pairing (Ma et al., 2005;

O'Brien, Hayder, Zayed, & Peng, 2018; Parker, 2010; Parker, Roe, & Barford, 2005).

Then miRNAs trigger deadenylating and decapping of the target mRNA with

collaboration of Ago and several other proteins. Therefore, the poly(A) tail of the mRNA

and the 5' cap cannot synergize which results in the translational inhibition of the

mRNAs (Okamura et al., 2007). A single microRNA can regulate the expression of

several mRNAs and a single mRNA may be regulated by more than one miRNA (Bartel, 2018; Flynt, Greimann, Chung, Lima, & Lai, 2010; Ragusa et al., 2017). miRNAs play critical roles in cellular development and processes and as such they are linked to a number of complex diseases from cancers to fertility and growth complications. For example, miR-196a, miR-124a and miR-132 have been reported to be associated with esophageal squamous cell carcinomas, neurodegenerative disease and Huntington's disease, respectively (Bloomston et al., 2007; C. Chen, Zhang, Zhang, Weakley, & Yao, 2011; Fendereski et al., 2017; Tam, de Borja, Tsao, & McPherson, 2014). Additionally, miR- 34/449 are implicated with male and female infertility as well as growth retardation. Therefore, it is inferred that miRNAs greatly influence the expression patterns of many genes with different functionalities (Ardekani & Naeini, 2010; K. Miyoshi et al., 2010; Xu et al., 2013).

1.2.1 miRNAs: A Historical Perspective

miRNAs were first noted in early 1990s when a 22 nucleotide RNA transcript of the lin-4 gene in nematode *Caenorhabditis elegans* was discovered to be controlling another gene's expression, lin-14. Without coding for any protein, lin-4 was decreasing the level of lin-14 expression. After that, Lee et al. discovered that there is a sequence on lin-4 which is complementary to the 3' Untranslated Region of the lin-14 mRNA. Then, they postulated that an endogenous small RNA transcript is decreasing gene expressions (R. C. Lee, Feinbaum, & Ambros, 1993). Therefore, lin-4 was named the first miRNA capable of regulating the gene expression via partial base pair complementary with mRNAs. However, this was dismissed due to lack of homologs in other organisms

concluding that probably this has been an isolated example in worms resulting in gene knockdown. However, several years after that when the same mechanisms were detected extensively distributed in eukaryotes, these small RNAs were categorized as a class of regulatory molecules (Lau, Lim, Weinstein, & Bartel, 2001; R. C. Lee & Ambros, 2001; Wightman, Ha, & Ruvkun, 1993). Another miRNA (21 nt) was then discovered in 2000 in *C. elegans.* This small RNA was named let-7 and since it is highly conserved across species, impacted the study of this specific miRNA (Kaufman & Miska, 2010). Let-7 was also detected in human and is associated in the late larval stages in adult transition (Reinhart et al., 2000) (Pasquinelli et al., 2000). Since then, several other short RNAs were discovered and interests in this field have soared paralleled with increasing number of miRNAs annotations across all species.

1.3 tRNAs

Transport RNAs (tRNA) are another class of noncoding RNAs that are involved in the translation of mRNA coding sequence into proteins. There are more than 500 tRNA genes in the human genome and they are initially transcribed by RNA Polymerase III in the nucleus to produce pri-tRNAs that undergo maturation process before they form "Cloverleaf" conformation. At this point, they bear 5p and 3p tails as "Leader" and "Trailer" sequences, respectively (Woese, Olsen, Ibba, & Soll, 2000). To become mature, tRNAs encounter processing events including 5p leader processing performed by RNase P complex and 3p trailer trimming performed by RNase Z family. A non-templated "CCA" motif is added to their 3' end of the resulting pre-tRNA from previous step by CCA-adding enzyme. The yielded end-matured tRNAs are detected by "Exportin-t" to be

exported to the cytoplasm. It is vital for cell to generate sufficient mature tRNAs to support the protein synthesis efficiently especially in fast growing cells (Ohira & Suzuki, 2011, 2016).

1.4 Mechanisms of microRNA Biogenesis

The biogenesis of most miRNA starts with Drosha a ribonuclease III which has endoribonucleolytic activity and participates in RNA maturation in the cell nucleus. miRNAs biogenesis is initiated by the transcription of pri-miRNAs (Fortin, Nicholson, & Nicholson, 2002; O'Brien et al., 2018). Drosha is in association with a protein called DGCR8 which is essential for the activity of Drosha. A "UGU" motif sequence in ~60% of hairpin loops guides DGCR8 to find hairpin loop. Then, Drosha recruits "UG" and "CNNC" motifs sequences on 5' and 3' flanking regions to situate on the hairpin and cut nearly one helical turn into the stem 16-18 nt upstream of "CNNC" motif. The released pre-miRNA is 55-80 nt long and leaves a 2-nucleotide overhang at the 3' end, which makes it available for Exportin-5 (XPO5) identification (Denli, Tops, Plasterk, Ketting, & Hannon, 2004; Han et al., 2006; X. Zhang & Zeng, 2010). XPO5 is thought to be the major gate for transporting the pre-miRNAs to the cytoplasm. It has been reported that many miRNAs can be processed without XPO5. However, miR-320 requires XPO1 instead of XPO5 (Westholm & Lai, 2011). After exported to the cytoplasm, pre-miRNA is further cut by Dicer, another ribonuclease III in associated with a partner protein TRBP. Both strands are cut close to the hairpin loop and a miRNA duplex is generated which contains the guide strand paired to the passenger strand with ~2 nt 3' overhang on each end. Subsequently, the miRNA duplex is loaded into the AGO to form the core of

the RISC complex (R. Li, Jia, Zou, & Zhao, 2014; K. Miyoshi et al., 2010; Ragusa et al., 2017). There are 8 members of the AGO family in humans. AGO1-4 have the ability to load miRNAs but only AGO2 has endonuclease activity. It has indicated that mice with catalytically inactive AGO2 die shortly after birth (K. Miyoshi et al., 2010). Compared to
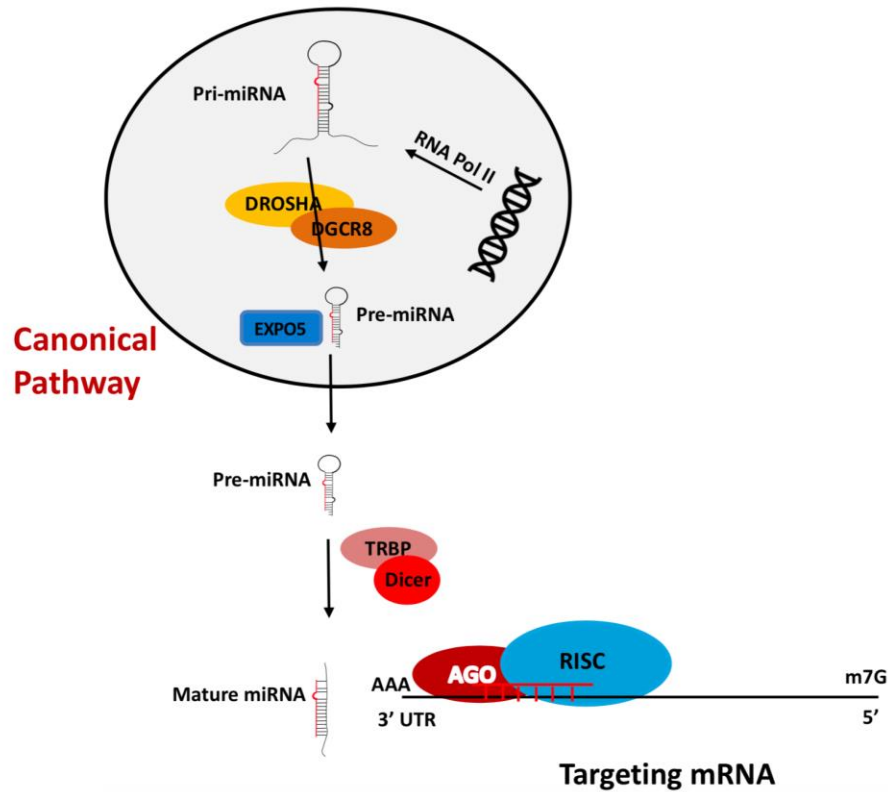
Figure 1.1: Canonical miRNAs Biogenesis Pathway. *Canonical miRNAs are processed by Drosha and Dicer to produce miRNA:miRNA* duplex.*

other Argonaute members, AGO2 is different in N-terminus and the PAZ-PIWI domain linkage space (Meister et al., 2004). The process starts by loading small RNA from the Dicer activity into the AGO2. Then the PAZ domain recognizes the 3' end of miRNAs (Tang, 2005). The 5' phosphate of miRNA is placed between PIWI and the Mid domains of the AGO protein. The enzymatic activity of the PIWI domain cuts the passenger

6

strand. Argonaute then takes the miRNA to the target mRNA and induces mRNA degradation or translational repression usually by binding to the 3' UTR region. The miRNA strands are selected as passengers or guides depending on the thermodynamic stability of each end of the duplex. The strand with more stability on its 5' terminal nucleotides is selected as the passenger strand and is degraded, while the other strand becomes mature miRNA (Figure 1) (Hutvagner & Simard, 2008; Y. K. Kim, Kim, & Kim, 2016).

1.5 Non-Coding miRNAs

In 2007, 14 short introns were identified as an alternative pathway for miRNA biogenesis. Small RNA clusters that originated from the outer edges of 56-nt introns. These sets of reads were very similar to the sets observed for the miRNA:miRNA* duplex. The sets were complementary to each other and that their 5' ends were more than 3' ends. In addition, the majority of the sets contained 2-nt 3' overhangs which mimicked them as Drosha products. Moreover, the predicted structure and the sequence were conserved in a pattern similar to that of pre-miRNAs. Despite the similarities with canonical miRNAs, pairing was stopped at the hairpin base and did not go over miRNA:miRNA* duplex which is necessary for Drosha to cleave the pri-miRNA. In the site that the pairing was extended, the reads showed conserved splice sites. This led them to the idea that these miRNAs did not originate from the canonical miRNA biogenesis pathway. Instead, an alternative pathway in which splicing defines pre-miRNA processing rather than Drosha cleavage is in place (Figure 2.1). As in the structure predictions, the terminal nucleotides characterize the first or last nucleotides of intronic

regions. They named their introns "Mirtron"s (Okamura et al., 2007; Ruby, Jan, & Bartel, 2007; Sibley et al., 2012). Moreover, tRNAs, rRNAs and some snoRNAs provide another sources of non-canonical pre-miRNAs that are independent of microprocessor-mediated processing. The initial excised structure when spliced forms a "Lariat" which is linearized by Ldbr (Lariat Debranching) enzyme (Han et al., 2004; K. Miyoshi et al., 2010; Seow, Sibley, & Wood, 2012; Zia & Flynt, 2018). Subsequently, the Ldbr ribonucleolytic activity processes the 3'-lariat tail, linearizes the intron and forms pre-miRNA hairpins that are suitable targets for Dicer cleavage. The spliceosome is then disassembled and recycled (Arenas & Abelson, 1997). Mirtrons then enter the canonical miRNA biogenesis pathway and then are exported to the cytoplasm by XPO-5. However, some studies have shown that mirtrons are not XPO-5 dependent and they utilizing other mechanisms for exporting to cytoplasm besides XPO-5 (Seow et al., 2012). Three types of mirtrons are conventional, 5'-tailed and 3'-tailed and splicing defines both ends of the hairpins in the conventional mirtrons and both 5' and 3' tailed hairpins undergo debranching by Ldbr and the 3'-tailed mirtrons are trimmed by the RNA exosome (Figure 2.1) (Ladewig et al., 2012; Rorbach, Unold, & Konopka, 2018; Westholm & Lai, 2011). The importance of the splice sites is well-studied in a research group that demonstrated artificial mirtrons have much higher knockdown efficiency than shRNAs in targeting Myotonic Dystrophy Protein Kinase (DMPK). Additionally, they illustrated that their knockdown efficiency is strongly related to their splice sites precision while Drosha dependent miRNAs are less efficient in knocking down their targets because Drosha and Dicer show fewer accuracy compared to spliceosome activity. This brings an advantage

for mirtrons by having less off-targets (Seow et al., 2012; J. O. Westholm, E. Ladewig, K. Okamura, N. Robine, & E. C. Lai, 2012).

Recently, 478 human loci have been identified as diced small RNA duplexes derived from splicing which 242 of them are novel. Interestingly, most of the annotated introns (86% of the whole mirtrons population) are 5'-tailed and 3'-tailed mirtrons only have 3.6% of the totals. This represents the 5'-tailed mirtrons as, by far, the dominant class of mirtrons in humans (J. Wen, E. Ladewig, S. Shenker, J. Mohammed, & E. C. Lai, 2015). Comparing expression of human canonical miRNAs (deep conserved or newly evolved) to mirtrons inferred from small RNA sequencing data meta-analysis reveals that mirtrons have on average several fold lower abundance (Griffiths-Jones, Grocock, van Dongen, Bateman, & Enright, 2006). Furthermore, the distribution of mirtrons is bimodal, showing that many species accumulate at the lowest levels of expression observed for miRNAs (Figure 2.2). The type of mirtron in the lowest abundance population are tailed varieties, consistent with the requirement for additional processing during tail removal.

1.6 Regulation of Mirtrons Function by Uridylation

Several terminal modifications have been identified in pre- and mature miRNAs. There are many reports about the effects of nucleotide addition on miRNA biogenesis, functionality and stability (J. O. Westholm et al., 2012). Untemplated addition of adenine and uridine is the most common modification in miRNAs at their 3' end especially in miRNAs which end in "guanine". 5'-tailed and 3p-arms of conventional mirtrons demonstrate more accumulations of this untemplated addition (J. Wen et al., 2015; J. O.

9

Westholm et al., 2012) and this addition is performed by a class of noncanonical Poly (A)

Polymerases (PAP) which later was discovered to be a Terminal Uridyltransferase

(Tailor) (Figure 1.2). Mirtrons demonstrate a bios over uridines and adenines additions to

the 3'-end. However, other nucleotides may be added but in a very lower rate (J. O.

Westholm et al., 2012). According to the organism and the examined miRNA, 3'

untemplated nucleotide addition shows various impacts on biogenesis and functionality

of mirtrons and miRNAs (Bortolamiol-Becet et al., 2015; Gutierrez-Vazquez et al.,

2017).
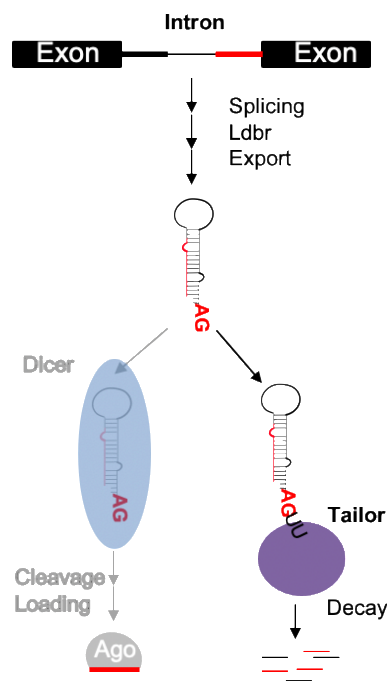
Figure 1.2: TUTase. *Mirtrons are uridylated due to their 3' terminal Guanyl which makes them good substrates for Tailor.*

Uridylation mostly inhibits the biogenesis of mirtrons Altogether, 3' nucleotide addition has vital regulatory role in gene expression (Yashiro & Tomita, 2018). But if this pathway suppresses their biogenesis, what is its utility? The answer could be found by considering evolutionary context. Tailor prefers splicing derived miRNAs, and this makes mirtrons more accessible for uridylation than canonical miRNAs particularly since the 3' terminal "AG" which is mirtron's 3' splice site nucleotides is Tailor sequence preference as well. This preference is regardless of being lower or higher expressed, well conserved or newly evolved. Therefore, it is possible that the mirtrons uridylation suppresses their biogenesis to make them less predisposed to deletions or mutations excluding them from the evolution. This can make extra evolutionary obstacles for mirtrons incorporation into conserved networks (Bortolamiol-Becet et al., 2015; Mohammed, Flynt, Siepel, & Lai, 2013; J. O. Westholm et al., 2012).

1.7 Hypothesis

In previous studies, it was inferred that tail processing of mir-1017, which is a 3'-tailed mirtron, occurs through exonuclease activity. Although we don't overlook the probable role of exoribonucleases in this process, we wanted to determine if any other mechanisms such as endoribonuclease activities are involved in 5'-tailed mirtrons biogenesis. Since it has been reported that exoribonucleases and specifically exosome is trimming the 3'- tailed mirtrons (3' to 5' processing), it is inferred that there might be an endoribonuclease activity involved in the process from 5' to 3' (Flynt et al., 2010; Westholm & Lai, 2011). Furthermore, in several studies, it has been demonstrated that

endoribonucleases play key roles in biogenesis or repair systems in tRNA processing on their 5' end and it appears that RNase P is the most important contributor in maturation of tRNAs by its endonuclease activity (Cho et al., 2019; Dutta, Malhotra, & Deutscher, 2012; Spath, Canino, & Marchfelder, 2007; Takaku, Minagawa, Takagi, & Nashimoto, 2003; Willkomm & Hartmann, 2007; Yip, Savickas, Gygi, & Shao, 2020). Among several well-known endoribonucleases, RNase P has been one of the most well-studied complexes and its contribution in 5' to 3' RNA processing at the junction between the 5' end of double-stranded and single- stranded region of RNA is proved frequently in different studies (Altman, 2011; N. Jarrous, 2002; Willkomm & Hartmann, 2007). Consequently, all these led us to hypothesize that an endoribonuclease is in charge of 5'-tailed mirtrons processing and it appears reasonable to focus the efforts on RNase P.

1.8 Significance

There are a few obscurities in the 5'-tailed mirtron processing mechanism that requires illumination. One of the most important questions that we are trying to answer in this study is to understand the mechanism by which 5'-tailed mirtrons are processed. Specifically, it is important to identify the enzymes that are involved in their maturation. Since mirtrons bypass Drosha cleavage, clarifying whether 5'-tailed mirtrons undergo individual events and enzymatic reactions to process their tails appears to be an important step towards unravelling details about their biogenesis. This unsolved issue demonstrates the significance of the study specifically because there are no published studies reporting the responsible factor(s) involved in 5'-tailed mirtrons maturation. This is while these mirtrons are the most abundant mirtrons. Moreover, knocking down or replacement of the

pathogenic alleles by RNAi are attracting more attentions in therapeutic interventions recently. Mirtrons are reported to be promising as alternatives to siRNAs, shRNAs and artificial miRNAs in this field. Therefore, exploitation and development of their unique properties will advantage future studies and therapeutic applications (Seow et al., 2012).

1.9 Objective

Mirtron 5010 comes from introns of Vacuolar -ATPase (V-ATPase) domains which are multi-subunit enzymes that mediate acidification in eukaryotes for zymogen activation, protein sorting and other intracellular processes (Gorelick, Shugrue, Kolodecik, & Thrower, 2006; Holliday, 2017). mir-5010 located on chromosome 17 is annotated on the basis of ~ 2200 small RNA reads present in miRbase.org database. miR-5010 regulates many genes that have key roles in cell cycles (Flynt et al., 2010). Also, miR-668 a 3'-tailed mirtron located on chromosome 12 of mouse genome was selected as a control.

To perform the experiments to achieve our goal, we designed changes in the tail regions of the two mirtrons (one 5'- and one 3'-tailed) and then transfected the HEK cells with the constructs. Repetitive G-rich sequences are most-likely forming "G quartets" which cause the concluded hairpins become more stable than other structures without poly G tract (Lerner & Sale, 2019). Therefore, our first objective would be uncovering the mechanism (endoribonuclease or exoribonuclease) involved in processing the 5'-tailed mirtrons. Afterwards, determining which exact ribonuclease(s) is(are) processing the 5' tails in mirtrons will be our next objective. For this purpose, we will knockdown selected ribonucleases that have this potential using sequencing. The last step will be

confirming our hypothesis through RNA sequencing data analysis along with enzyme

assay and Immunoprecipitation on determined ribonuclease.

CHAPTER II – DETECTION OF TAILED MIRTRONS BY NORTHERN BLOTTING

2.1 Abstract

microRNAs (miRNAs) have vital roles in regulating gene expression–contributing to major diseases like cancer and heart disease. Over the last decade thousands of miRNAs have been discovered through high throughput sequencing-based annotation. Different classes have been described, as well as a great dynamic range of expression levels. While sequencing approaches provide insight into biogenesis and allow confident identification, there is a need for additional methods for validation and characterization. Northern blotting was one of the first techniques used for studying miRNAs and remains one of the most valuable as it avoids enzymatic manipulation of miRNA transcripts. Blotting can also provide insight into biogenesis by revealing RNA processing intermediates. Compared to sequencing, however, northern blotting is a relatively insensitive technology. This creates a challenge for detecting low expressed miRNAs, particularly those produced by inefficient, non-canonical pathways. In this chapter, we describe a strategy to study such miRNAs by northern blotting that involves ectopic expression of both miRNAs and miRNA-binding Ago proteins. Through use of epitope tags, this strategy also provides a convenient method for verification of small RNA competency to be loaded into regulatory complexes.

Keywords

miRNA; mirtron; tailed-mirtron; low-abundance detection; RNA-immunoprecipitation; northern blot

2.2 Introduction

Small non-coding RNAs are common features of eukaryotic transcriptomes that require specialized detection methods due to their small size. Moreover, many species are present in low abundance, which further complicates detection (Q. Huang, Mao, Li, Hu, & Zhu, 2014). microRNAs are an ~23 nucleotide (nt) class of small RNAs that have vital roles in gene regulation and their dysregulation is associated with numerous diseases (Fendereski et al., 2017; Westholm & Lai, 2011). Animal microRNAs are a highly diverse class of genes. Some miRNAs are ancient, predating the birth of metazoans, while other are newly evolved and species specific (Meunier et al., 2013). Accordingly, microRNAs are a highly diversified class of transcript. High throughput sequencing has been instrumental in small RNA discovery. Indeed, thousands of miRNAs have been discovered and deposited into databases (Berezikov, 2011; Witten, Tibshirani, Gu, Fire, & Lui, 2010; J. Zhou et al., 2015). Yet, compared to the number of miRNAs annotated, there have been very few functional studies, suggesting there is a need for additional strategies for detection.

miRNA biogenesis typically involves cleavage from hairpin RNAs precursors (premiRNAs). For most miRNAs the Drosha/DGCR8-containing microprocessor complex excises pre- miRNA from primary transcripts. Final processing is then carried-out by Dicer, yielding miRNA/miRNA* duplexes (Ketting et al., 2001; Y. Lee et al., 2003). One strand then is loaded into Argonuate (Ago) proteins where it participates in targeting and repression of transcripts (J. E. Babiarz, Ruby, Wang, Bartel, & Blelloch, 2008; Song, Smith, Hannon, & Joshua-Tor, 2004). Beyond this canonical pathway, there are many atypical pathways. Multiple varieties of miRNAs have been found to bypass

16

Drosha/Dgcr8 processing like mirtrons, shRNAs, snoRNA-derived miRNAs, tRNA-derived miRNAs, and rRNA-derived miRNAs (Figure 2.1) (J. E. Babiarz et al., 2008; Ender et al., 2008; Maute et al., 2013; Ruby et al., 2007; Yoshikawa & Fujii, 2016)

In this article, we will focus on the detection of mirtrons. They are derived from small introns that after splicing and debranching can fold into small hairpins (Curtis, Sibley, & Wood, 2012; Macfarlane & Murphy, 2010; Westholm & Lai, 2011). Hundreds of mirtrons have been discovered in the human genome through inspecting hairpins coincident with splice sites from which small RNA sequencing reads are derived (Ladewig et al., 2012). Mirtrons are classified in three categories: conventional, 3'-tailed, and 5'-tailed. In conventional, both hairpin ends correspond to splice sites and after splicing can become dicer substrates. In "tailed" varieties only one side of the hairpin is
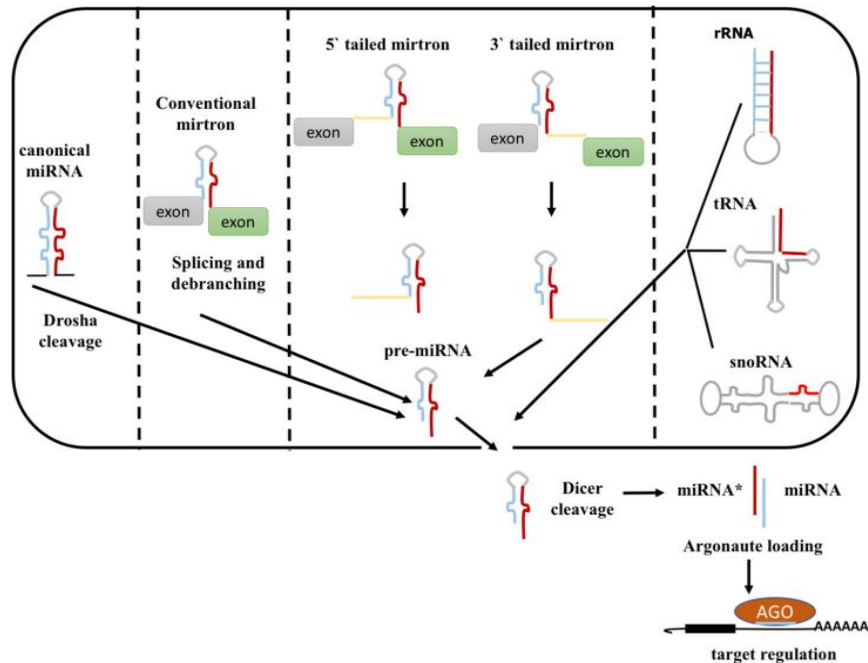
Figure 2.1: Classes of miRNA defined by biosynthesis pathway. *rRNAs, tRNAs and snoRNAs can be sources of miRNAs as non-canonical pathways.*

at a splice site, while the other side has a stretch of nt's, or tail, between the hairpin base

and splice site (J. Wen et al., 2015; Westholm & Lai, 2011). Studies in Drosophila have

also found that mirtrons are commonly modified by Tailor, a TUTase (terminal

Uridyltransferase), which uridylates hairpins to inhibit their biogenesis by abrogating

Dicer processing (Norbury, 2013). Thus, mirtrons more so than other types of miRNAs

are less robustly expressed than canonical miRNAs.

Comparing expression of human canonical miRNAs (deep conserved or newly

Expression of Human miRNAs by Class
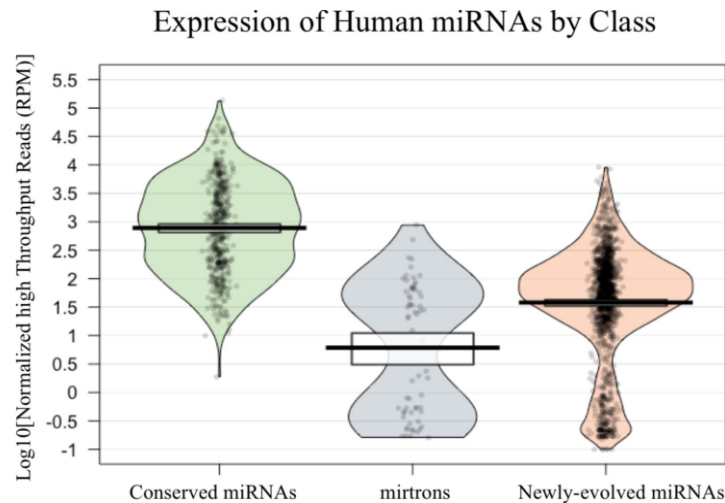
Figure 2.2: RDI Plot.

Showing the distribution of normalized (Reads per million) miRNA expression
separated by class. Values were taken were taken from miRbase.

evolved) to mirtrons inferred from small RNA sequencing data meta-analysis reveals that

mirtrons have on average several fold lower abundance (Figure 2.2) (Griffiths-Jones et

al., 2006). Furthermore, the distribution of mirtrons is bimodal, showing that many

species accumulate at the lowest levels of expression observed for miRNAs (Figure 2.2).

The type of mirtron in the lowest abundance population are tailed varieties, consistent

with the requirement for additional processing during tail removal. Therefore, detection strategies capable of detecting tailed-mirtrons via northern blotting will be applicable for studying all types of low abundance miRNAs. Northern blotting was the first technique used to detect a miRNA transcript (R. C. Lee et al., 1993). Many techniques have been developed in the interim which frequently rely on reverse transcription and PCR, which can introduce bias, artifacts, and mostly can't distinguish between biogenesis intermediates (Sharbati-Tehrani, Kutz-Lohroff, Bergbauer, Scholven, & Einspanier, 2008; Varkonyi-Gasic, Wu, Wood, Walton, & Hellens, 2007). The challenge in using northern blotting is relatively low sensitivity, making detection of tailed-mirtrons problematic. Despite this difficulty, it is possible to detect tailed-mirtrons using ectopic expression of the miRNA and Ago proteins (Ladewig et al., 2012). In this chapter, we describe a protocol for applying this technique to study mammalian tailed-mirtrons. This process involves transfection of two constructs, one to express the mirtron and the other to express a myc-tagged human Ago2 (Figure 2.3 a). In this system, the mirtron is cloned into the open-reading frame of GFP, producing a fluorescent reporter for verification that splicing has occurred efficiently. This strategy also takes advantage of a cell culture phenomena where increased rates of miRNA biogenesis are observed in cells grown to high confluency (Hwang, Wentzel, & Mendell, 2009). Thus, it is critical to use a cell line that does not exhibit contact- inhibition. The protocol described here uses Hela cells. Following transfection RNAs can either be bulk isolated by a routine RNA purification method like TriZol extraction or through immunoprecipitation targeting the myc epitope fused to Ago2. Northern blotting, as previously described, involving urea-PAGE

separation and transfer to nylon membranes can then be used to visualize RNAs in
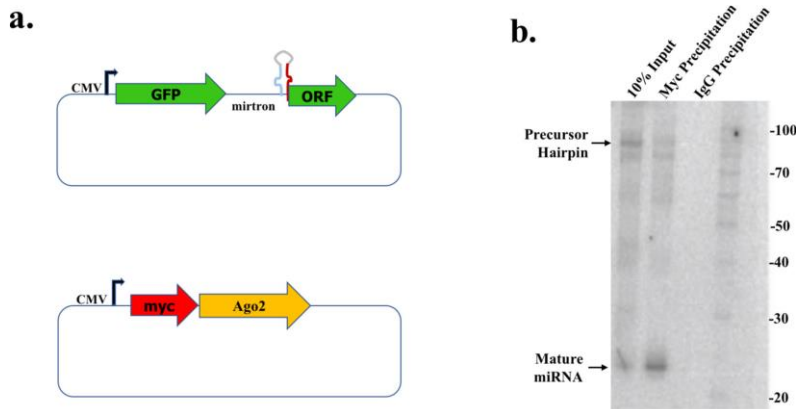
Figure 2.3: System for detecting and verifying mammalian mirtrons.

a) Constructs for co-transfection. Top vector configured to couple mirtron splicing to GFP expression, by interrupting the GFP ORF with the mirtron-containing intron. Bottom vector for ectopic expression of Ago2-myc b) Northern Blotting reveals expression of miR-5010, a 5'-tailed mirtron, using the vectors in (a) in conjunction with an immunoprecipitation procedure.

conjunction with oligonucleotide probes (Figure 2.3b) (Rio, 2014).

2.3 Protocol

2.3.1 Plasmid Constructs

Mirtrons are amplified and inserted into expression vectors within opening reading frame of GFP [29]. The protocol also utilizes a myc-tagged hAgo2 (Addgene: Plasmid #19872). Plasmid DNAs are isolated via commonly-used commercial purification kits like the Qiagen®Endo-free Maxi Prep Kit.

2.3.2 RNA Immunoprecipitation

Phosphate buffered Saline, lysis buffer (5% Glycerol, 150mM sodium chloride, 2mM Magnesium Acetate, 20mM, Tris 7.5 pH, 0.5% NP-40, 2mM DTT, 40 units/mL RNase Out ® Invitrogen, and Complete mini Protease Inhibitor Cocktail ® Roche). Lysis buffer should be used within several days of mixing and kept at 4°C.

Protein G Dynabeads ® Invitrogen, magnetic stand, and tween-20.

Anti-myc (clone 9E10, Thermofisher) and a control IgG (Jackson Immunoresearch) antibodies.

Bead preparation and RNA isolation: TRIzol ® Invitrogen, or 400 mM sodium chloride, acid phenol chloroform (Ambion), glycogen co-precipitant, isopropanol, 70% ethanol, and sodium acetate pH 5.3.

Hela cells are maintained in DMEM (Dulbecco's Modified Eagle's Medium) supplemented with 10% FBS and 1% penicillin/Streptomycin using standard cell culture methodology. Hela cells are efficiently transfected with Lipofectamine 3000 (Invitrogen) to deliver mirtron and myc-Ago2 encoding plasmids, as described above. Transfection and splicing efficiency is determined by visualizing GFP fluorescence.

a. Grow cells to 100% confluency, and change media 3 hours before transfection with antibiotics free media.

b. Warm OptiMEM to room temperature. Make a mixture of OptiMEM, Plasmids, and Lipofectamine 3000 based on the scale of transfection in accordance with manufacturer recommendations. For this protocol six well plates are typically used.

c. Incubate the mixture at room temperature for 20–25 minutes.

d. Add the transfection mixture to the cells dropwise, without touching the plate

wall. 3.1.5) Swirl the dish gently and incubate for 2 days (37°C, 5% CO2).

e. Check transfection efficiency two days after transfection.

2.3.3 RNA Isolation

Following the two-day incubation RNA is harvested. This could be done using the

TRIzol method, following manufacturer protocols, or immunoprecipitation to isolate

Ago-miRNA complexes.

a) TRIzol extraction (optional)

b) Add TRIzol directly to tissue culture plates. Cells should immediately become

opaque, and easily dislodged by micropipetting. Remove cells and place in

microcentrifuge tube.

c) Incubate TRIzol and cells for 10 mins at room temperature.

d) Add 200 µl of chloroform to every 1 mL of TRIzol and incubate at room

temperature for 5 mins.

e) Separate phases at 4°C for 15 mins by spinning at 13,000 rpm.

f) Transfer the clear aqueous phase to an equal volume of isopropanol and

incubate overnight at −20°C.

g) Spin precipitations at 4°C for 30 mins, resuspend pellets in RNase free water,

and measure concentration via spectrophotometry.

h) Immunoprecipitation of miRNA-Ago Complexes (optional).

i) Wash transfected cells with ice-cold PBS twice while in culture dish. Add

lysis buffer to cells and incubate 20 mins on ice. Cell layer should begin to

detach from the bottom of the dish with gentle micro pipetting. Transfer to microcentrifuge tube, mix well with pipette to eliminate clumps, and spin down at 13000 rpm for 20 minutes at 4°C to clarify lysate. Transfer supernatant to new tube.

j) Prepare beads for incubation with lysate. Add protein G magnetic beads to tubes (two tubes for each sample). Place the tubes on a magnetic stand and remove the supernatant. Wash tubes twice with PBS containing 0.02% tween-20. In PBS-tween add Anti-myc or control IgG antibodies.

k) Remove PBS tween from beads and add cell lysate from step 3.2.1.1. Reserve 10% of the volume to be used as input control. Place on rocker for 2 h, 4°C.

l) Wash beads with lysis buffer five times using magnetic stand to remove beads from solution.

m) Add equal volumes of 400mM NaCl and acid phenol chloroform to beads. Aggressively vortex. This procedure should also be performed on the reserved 10% volume of lysate set aside for input control.

n) Separate phases at 4°C for 15 mins by spinning at 13,000 rpm.

o) Add clear aqueous phase to an equal volume of isopropanol and incubate overnight at −20°C.

p) The following day spin precipitations at 4°C for 30 mins, resuspend pellets in RNA II loading buffer.

2.3.4 Northern Blotting

a) Sample Preparation. If immunoprecipitation was performed samples are ready to load on the gel. If TRIzol extraction was used, 20μg of RNA should be loaded in each lane.

b) Reprecipitate 20μg of RNAs by adding the appropriate volume of RNA to 100% ethanol containing 300 mM sodium acetate.

c) Incubate the RNA at −20°C for at least one hour.

d) Spin precipitations at 4°C for 30 mins, resuspend pellets in RNA II loading buffer.

e) RNAs dissolved in RNA II loading buffer are denatured by incubating at 94°C for 3 minutes and then immediately placed on ice.

f) Prepare and Run Gel using a vertical gel rig and the UreaGel system. 3.3.2.1) Mix UreaGel diluent, buffer, and concentrate, then add APS and TEMED.

g) Quickly, using a large volume pipette pour gel mixture between glass plates, insert comb, and lay horizontally until gel is set.

h) Place gel on apparatus and fill buffer tanks with TBE. Pre-run the gel at 250 V for 30 minutes.

i) Using long gel loading tips transfer RNA dissolved in RNA II loading buffer to the wells of the gel. Prior to loading samples, wells should be washed out to remove urea that has diffused into the space. Radio-labeled decade marker should also be loaded on the gel.

j) Run the gel at 250–350 V until the fastest moving, bromophenol blue dye band has migrated approximately eight inches.

k) Blotting of RNAs onto nylon membranes.

l) Soak filter paper in 0.5X TBE buffer and place onto wet-dry transfer rig.

m) Crack the gel using a separator wedge, and cut the region of the gel to be blotted. The lower dye band corresponds to ~10 nt RNAs. The gel should be cut approximately one inch below the lower band in order to capture all the RNA molecules.

n) Place the gel onto the 0.5X TBE soaked filter paper.

o) Prepare the membrane by quickly rinsing in water, followed by briefly soaking in 0.5 X TBE. Once the membrane sinks in the buffer it is ready to be placed on top of the gel.

p) Place another presoaked piece of filter paper on top of the membrane.

q) Place the lid on the rig. The filter paper/gel/membrane should be sandwiched tightly between the electrodes.

r) Transfer the RNAs to the membrane for 30 mins at 10V at 4°C.

s) Remove membrane from transfer apparatus, and air dry briefly on a paper towel.

t) Crosslink the RNAs to the membrane with UV exposure.

u) Bake the blot at 80°C for 30 mins 3.3.4) Probing and visualizing membranes.

v) Blots are place into cylindrical hybridization tubes along with enough hybridization buffer to keep the blot wet.

w) Prehybridize blots for >1 hour at 40–45°C, in a rotating hybridization oven.

x) Prepare radio-labeled probes by mixing water, complementary DNA oligo, 10X PNK buffer, PNK, and gamma-P32ATP 6000uCi/mL.

y) Incubate the labeling reaction at 37°C for > 1hour.

z) Stop the reaction by adding EDTA.

aa) Remove unincorporated radionucleotides using G-25 sephadex columns. Remove buffer from column by spinning at 2000rpm for 4 min. Transfer the probe-labeling reaction to the column and spin again for 2000rpm, 4 min. The probe is in the column flowthrough.

bb) Add the radiolabeled probe to the hybridization solution. Avoid touching the membrane with concentrated probe.

cc) Incubate the membrane overnight at 40–45°C in a rotating hybridization oven.

dd) Wash and expose the blot.

ee) Wash the membrane twice with high stringency wash buffer at the hybridization temperature for one hour.

ff) Wash the membrane twice with low stringency wash buffer at the hybridization temperature for one hour.

gg) Remove membranes from tubes, and dab on filter paper. Do not over-dry the membrane. The objective is to remove excess liquid.

hh) Wrap the membrane with cling wrap and expose to X-ray film or to phosphorimager screens.

ii) Develop films or scan screens after one day.

2.3.5 Reprobing Membrane

a) Signal can be stripped from blots by boiling membranes in 0.1% SDS. Blots should be placed in boiling solution and left to cool for three hours. This should be repeated once.

b) Following the stripping procedure, blots are placed into hybridization buffer and the procedure repeated starting at step w.

c) Blots can be subject to multiple reprobings before the integrity of the immobilized RNA is compromised.

2.4 Notes

A) The protocol described here relies on a radiolabel to visualize the interaction of probes with RNAs immobilized on blots. There are alternatives to this type of detection, which are untried with mirtrons [30]. These procedures involve an epitope labeled probe such as biotin followed by antibody detection leading to visualization through chemiluminescence. For labs that want to avoid isotope this could be a viable alternative.

B) This protocol calls for DNA oligonucleotide probes, but is also appropriate for use with LNA-based probes. These molecules offer greater specificity compare to DNA probes. To use this alternative chemistry the hybridization buffer should be substituted with a formulation containing formamide.

C) LNA hybridization buffer: 50% Formamide, 5X SSPE, 5x Denhardt's, 0.5% SDS, 20ug/ml ssDNA.

D) Avoid touch nylon membranes with gloves, and then only on the edges. This will avoid the appearance of blotting artifacts that might obscure signal.

E) Care should be taken to load the optimal amount of RNA ladder. The Decade Marker protocol calls for a single μl of gamma-P32ATP 6000uCi/ml, but this could lead to innapropriatly intense bands that obscure signal if not properly diluted. When fresh

ATP is used, the marker should be serially diluted in RNA II loading buffer. 1 μl of reaction should be diluted in 30 μl followed by a second dilution in of 1 μl in 30 μl. One μl of the final dilution should be loaded on the gel.

F) Following the final wash step it is typical to observe very little signal emanating from the blot via Geiger counter. However, when signals are very weak longer exposure times may be required. Experimenters could extend exposure times to a week for the lowest signals.

G) When removing combs from UreaGels the wells should be washed aggressively with water to prevent residual acrylamide from polymerizing and causing unevenness in the bottom of the well which will result in RNA migration artifacts.

H) Moderation of voltage used during PAGE separation of RNAs is recommended. This prevents bending of bands.

CHAPTER III – HUMAN 5'-TAILED MIRTRONS ARE OROCESSED BY RNaseP

3.1 Abstract

Approximately a thousand microRNAs (miRNAs) are documented from human cells. A third appear to transit non-canonical pathways that typically bypass processing by Drosha, the dedicated nuclear miRNA producing enzyme. The largest class of non-canonical miRNAs are mirtrons which eschew Drosha to mature through spliceosome activity. While mirtrons are found in several configurations, the vast majority of human mirtron species are 5'-tailed. For these mirtrons, a 3' splice site defines the 3' end of their hairpin precursor while a "tail" of variable length separates the 5' base of the hairpin from the nearest splice site. How this tail is removed is not understood. Here we examine sequence motifs in 5'-tailed mirtrons and interactions with RNA turnover processes to characterize biogenesis processes. Through studying the high confidence 5'-tailed mirtron, hsa-miR-5010, we identify RNaseP as necessary and sufficient for "severing" the 5' tail of this mirtron. Further, depletion of RNaseP activity globally decreased 5'-tailed mirtron expression implicating this endoribonuclease in biogenesis of the entire class. Moreover, as 5'-tailed mirtron biogenesis appears to be connected to tRNA processing we found a strong correlation between accumulation of tRNA fragments (tRFs) and 5'-tailed mirtron abundance. This suggests that dysregulation of tRNA processing seem in cancers also impact expression of the ~400 5'-tailed mirtrons encoded in the human genome.

3.2 Introduction

microRNAs (miRNAs) are small non-coding ~22 nucleotide (nt) RNAs processed from stem-loop structures (Bartel, 2018). Biogenesis is initiated by the microprocessor which contains Drosha, an RNase III enzyme, that crops ~70 nt precursor hairpin from primary transcripts and leaves 2-nucleotide 3' overhang. The overhangs are recognized by Exportin-5 (XPO5), which carries pre-miRNAs to the cytoplasm for final processing by Dicer, another RNase III, after which they are loaded into effector Argonaute (Ago) proteins. Once recruited by Ago, miRNAs typically base pair with mRNAs via their 5' nucleotides between position 2-9, inducing degradation or translational repression(Grimson et al., 2007).

While most miRNAs mature through the standard Drosha-Dicer-Ago pathway, numerous additional routes have been identified that take advantage of alternate RNA processing enzymes (Keita Miyoshi, Tomohiro Miyoshi, & Haruhiko Siomi, 2010). Indeed, miRNAs have been found to reside in most major non-coding RNA (ncRNA) varieties such as tRNAs, rRNAs, and snoRNAs. Each of these ncRNAs are cut from longer transcripts, cuts which are repurposed for creation of non-canonical miRNAs. Among these the most prevalent variety is tRFs, which are found in numerous cell types as a sign of stress and are diagnostic of an oncogenic phenotype (Guzzi & Bellodi, 2020; Y. Zhang, Qian, He, & Gao, 2020). This makes many tRNAs dual functional molecules with both tRNA and miRNA capabilities such as tRNA-Ile which can form a 110 bp hairpin encoding hsa-miR-1983 which is also a mature tRNA molecule (Hasler et al., 2016).

Unsurprisingly, the most prevalent RNA processing event in eukaryotes, splicing, also generates miRNAs outside the canonical pathway (Okamura et al., 2007; Ruby et al., 2007; Sibley et al., 2012). miRNAs where the hairpin base is defined by splicesome-mediated cuts are called mirtrons and were the first recognized class of non-canonical miRNA. The initial set of mirtrons were minute introns where the precursor fold comprised the length of the entire intron. After splicing, lariat intermediates are linearized by Lariat Debranchase (Ldbr) permitting folding into structures that are suitable for export, Dicer cleavage, and Ago loading. After this discovery, investigation of microprocessor deficient ES cells showed additional types of mirtrons existed where only one side of the hairpin corresponded to a splice site (Joshua E. Babiarz et al., 2011). On the hairpin terminus that was not a splice site a "tail" of nucleotides extended to the scissile base of a distant splice site. Studies in *Drosophila* revealed that removal of a 3' tail is carried out by the 5'-3' RNA exosome, which was previously found to be confounded by terminal hairpins on substrates (Flynt et al., 2010). Critically, this established a paradigm that generic turnover enzymes can participate in productive miRNA biogenesis.

In addition to the alternate method of biogenesis, mirtrons also show a different conservation profile relative to miRNAs. Approximately 80 animal miRNAs are conserved in all metazoans while mirtrons, as well as other non-canonical miRNAs, are rarely found shared at the order taxonomic level (Ladewig et al., 2012; Jiayu Wen, Erik Ladewig, Sol Shenker, Jaaved Mohammed, & Eric C. Lai, 2015). Moreover, mirtron biogenesis modes are not consistently represented. Three types of mirtrons are observed: the original strict mirtrons where splicing generates both ends of the hairpins, 3'-tailed

31

mirtrons where a 5' splice site coincides with the hairpin base and a tail of nucleotides extends to a distal 3' splice site, and the 5'-tailed mirtrons where the hairpin base is a 3' splice site with a tail between the hairpin and a 5' splice site. Strict mirtrons are found throughout animal genomes with only a handful of examples. Likewise, 3'-tailed mirtrons are rare genomic features, which nevertheless are found in invertebrate and vertebrates at approximately the same rate. Unlike the other mirtron configurations, 5'-tailed mirtrons are extremely abundant in mammalian genomes, while essentially absent or present at the same rate as 3'-tailed mirtrons in lower organisms. Approximately 450 human loci have been identified as sources of diced small RNA duplexes derived from splicing, remarkably 86% are 5'-tailed mirtrons (Ladewig et al., 2012; Jiayu Wen et al., 2015).

Other features further separate mirtrons from standard miRNAs. Comparing expression of human canonical miRNAs to mirtrons shows mirtrons have on average several fold lower abundances (Ladewig et al., 2012; Jiayu Wen et al., 2015; Zia & Flynt, 2018). This is potentially linked to a higher rate of terminal modifications identified on pre- and mature miRNAs that occur on mirtrons (De Almeida, Scheer, Zuber, & Gagliardi, 2018; Le Pen et al., 2018; J. O. Westholm et al., 2012). Mirtrons are highly modified by untemplated addition of adenine and uridine carried out by noncanonical Poly (A) Polymerases (PAP) or Terminal Uridyltransferase (TUTases) (Bortolamiol-Becet et al., 2015; Jakub O. Westholm, Erik Ladewig, Katsutomo Okamura, Nicolas Robine, & Eric C. Lai, 2012). Nucleotide addition seems to lead to variety of impacts on miRNA biogenesis and functionality; however, effects mostly appear to be inhibitory (Bortolamiol-Becet et al., 2015; Gutierrez-Vazquez et al., 2017; Heo et al., 2008).

Modification can make hairpins poor substrates for Dicer or shift the Dicer cleavage site on the pre-miRNA by altering thermodynamic features and thereby modulating strand selection by Ago (B. Kim et al., 2015; H. Kim et al., 2020; Yang et al., 2019). Strict and 5'-tailed mirtrons appear to be more prone to uridylation due to the 3' terminal "AG" of consensus splice sites which is Tailor's, a terminal uridylyl transferase, preferred substrate. This bias suggests that uridylation of mirtrons is used to specifically suppress their biogenesis. Thus, not only is there selective pressure working to limit mirtron function, but also a targeted modification is applied that further reduces functionality.

Here, we investigate these seemingly contradictory miRNA species through probing biogenesis mechanism of tailed mirtrons in humans. Despite their seeming disfavored biogenesis, mirtrons remain abundant (Ladewig et al., 2012; Jiayu Wen et al., 2015). We show that 3'-tailed mirtrons in mammals likely follows what was found in *Drosophila*. 5'-tailed mirtrons, on the other hand, are produced by the activity of an endoribonuclease, which we identified to be RNaseP. This is the enzyme that is responsible for removal of 5' leader segments from tRNA precursors, a structure that mirrors the base of 5'-tailed mirtron hairpins (Nayef Jarrous & Reiner, 2007). We also find that dysregulation and accumulation of tRFs seen in cancer correlates with a similar accumulation of 5' tailed mirtrons. This further intertwines processing of 5'-tailed mirtrons with tRNA processing and suggests that the importance of mirtrons might not lie in normal cell physiology, but rather is more impactful during disease processes.

3.3 Results

3.3.1 *5'-Tailed Mirtrons Are Characterized by G-quartet Containing Precursors*

To investigate biogenesis of 5' tailed mirtrons we first sought to identify shared motifs in these miRNAs. To do this, we applied the seqlogo algorithm to the first 20 bases of small RNAs produced from human 5'-tailed mirtron hairpins (Fig 3.1A-B) (Crooks, Hon, Chandonia, & Brenner, 2004). Comparing all species, similar motifs are apparent. 5' arms are comprised primarily of "G" residues while 3' arms are typically C and U residues, likely representing intronic polypyrimidine (ppy) tracts. This situation was likewise noted in prior efforts to assess sequence elements in mammalian mirtrons (Jiayu Wen et al., 2015). This suggests that G rich stretches arising adjacent to ppy
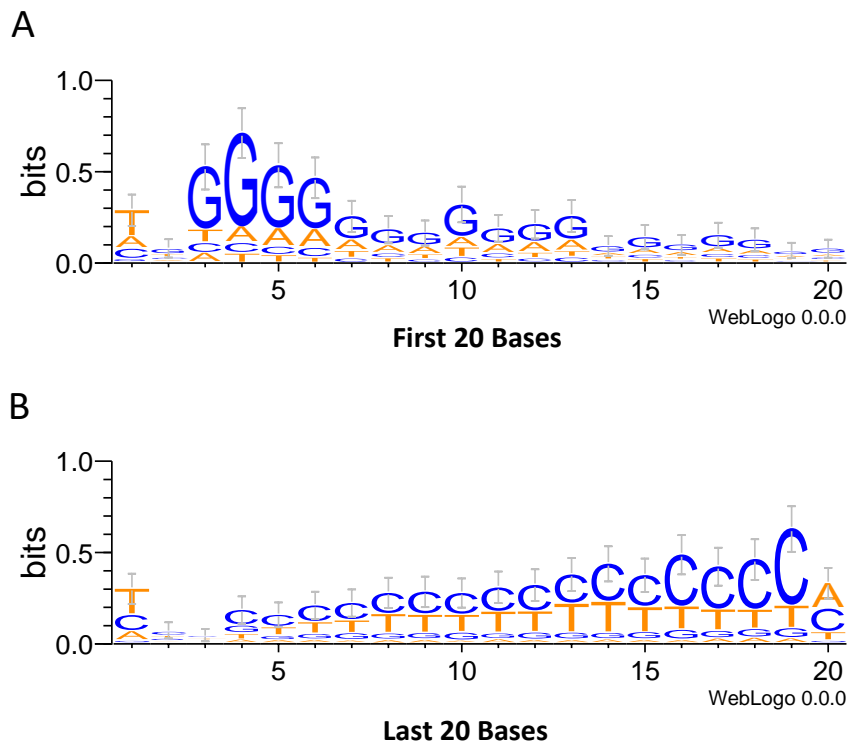
First 20 Bases

Last 20 Bases

Figure 3.1:5'-Tailed Mirtrons Elements. *A) Mirtrons Have High "G" content on 5p arms and B) high "CT" content on 3p arms.*

elements might be sufficient to lead to 5'-tail mirtron formation. Also, the higher

confidence G quartet at the beginning of the 5' arm suggests a major role for hairpin

structure in biogenesis of 5'-tailed mirtrons. To confirm this, we assessed the prevalence

of G quartets in all human introns relative to 5' tail mirtron encoding introns, showing

significantly higher G quartet occurrence in mirtrons (Fig 3.2). Thus, biogenesis of 5'-

tailed mirtrons appears linked to RNA structure, specifically polyG tracts. To further

connect 5'-tailed mirtron biogenesis to nucleotide content we compared the free energy

(ΔG) and GC content of 767 canonical miRNAs

and 404 5'-tailed mirtrons (Fig 3.3). As a class,

5'-tailed mirtrons show much high GC content

relative to most miRNAs. Interestingly, this does

not seem to correlate with a substantially

different energy profile. If mirtron hairpin

energy is greatly divergent from conserved

miRNAs this might lead to incompatibilities in

Ago loading. Considering the bias of 5' tailed

mirtrons for G quartets and high GC content this

suggested that part of their biogenesis might be

linked to being challenging substrates for

RNases. To examine this possibility, we
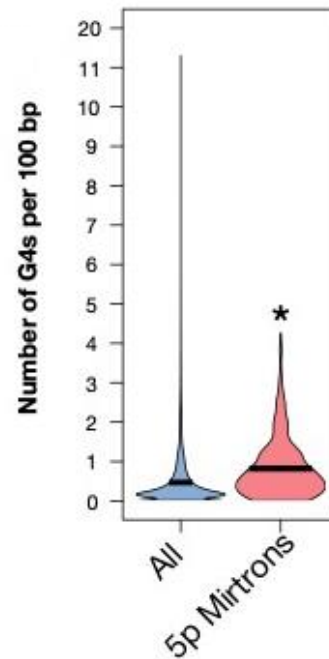

Figure 3.2: G-Quartets in All Introns and 5'-Tailed Mirtrons. *5'-tailed mirtrons have significantly more G quartets compared to all introns. (*p≤0.05).*

investigated the impact of major 5'-3' turnover RNases, Xrn1 and Xrn2, in 5'-tailed

mirtron production (Valen et al., 2011).

Using public data from XRN 1/2 depleted cells, 5'-tailed mirtrons are expressed at a significantly higher level compared to control (Fig 3.4 A-B). A higher level of mirtron expression was observed as well as species that were completely absent in control conditions appeared after XRN1 1/2 knockdown. Initial reporting of these datasets revealed a similar situation, however, this study benefits from a formalized annotation of 5'-tailed mirtrons (Jiayu Wen et al., 2015).

These data suggest that exoribonuclease activity antagonizes mirtron production, which led us to investigate the characteristics of mirtrons that only accumulate in Xrn1/2 knockdown datasets. A notable feature of XRN-eliminated mirtrons is the presence of heterogeneous isoforms, where over half were represented by two or more isoforms (Fig 3.5).
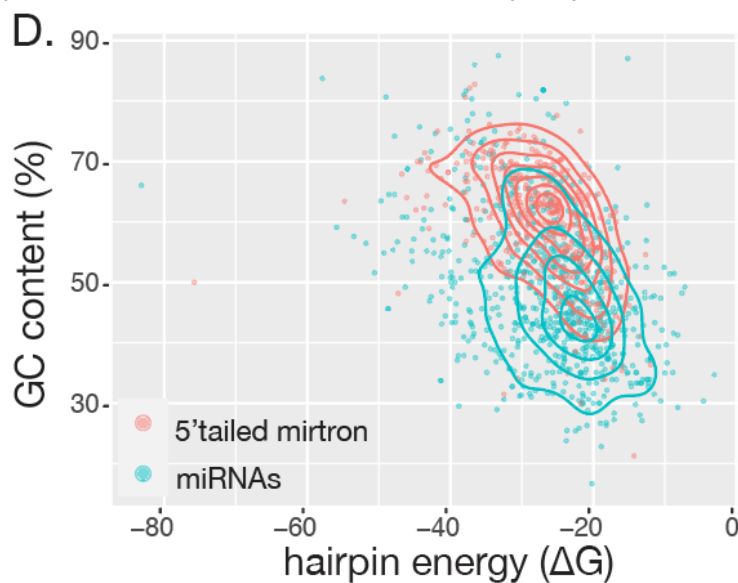

Figure 3.3: Hairpin Energy vs GC Content. *Comparison of ΔG (X axis) and GC (Y axis) contents in 767 canonical miRNAs and 404 5'-tailed mirtrons. Density overlays show separation of the two groups based on GC content, but not ΔG.*

This Imprecise processing is consistent with a biogenesis mechanism that is not dedicated to mirtron production, unlike the reliability of Drosha to generate precise hairpin ends. Focusing on a well-expressed mirtron, hsa-miR-3620, greater heterogeneity is seen for 5p arm small RNAs (Fig 3.6). Moreover, the greatest isoform diversity is seen at the 5' end of the 5p read, which is problematic as this shifts the identity of a miRNA's targets. It also hints that 5'-tail removal is not occurring in a deliberate fashion, suggests a role for XRN activity in limiting 5'-tailed mirtron accumulation, and that this class of miRNA requires features that allow evasion of 5'-3' exoribonuclease activity.
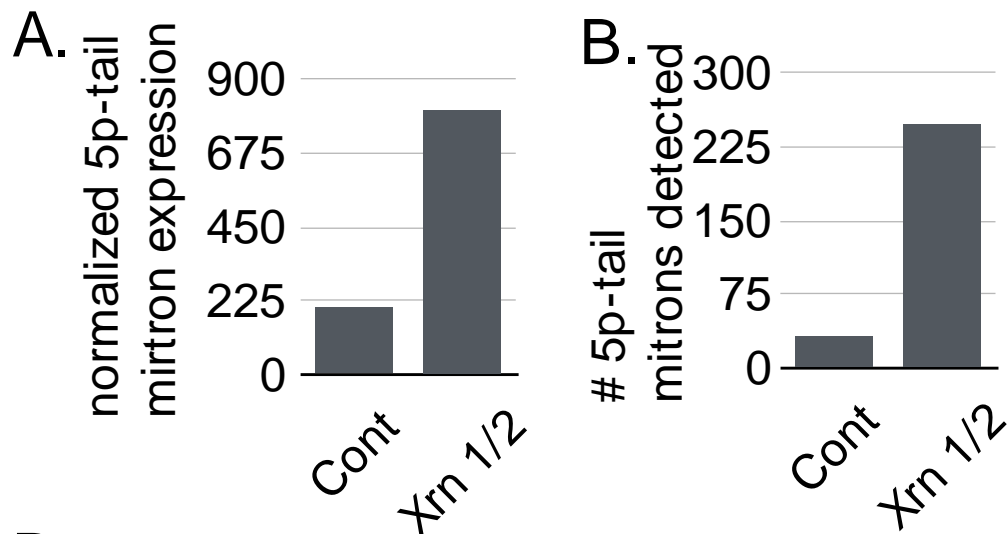
Figure 3.4: XRN 1/2 Degrade 5'-Tailed Mirtrons. *A) Normalized accumulation of all combined mirtron mapping reads after XRN 1/2 knockdown and in control datasets. B) Number of individual 5'-tailed mirtrons detected in control and XRN 1/2 knockdown libraries.*
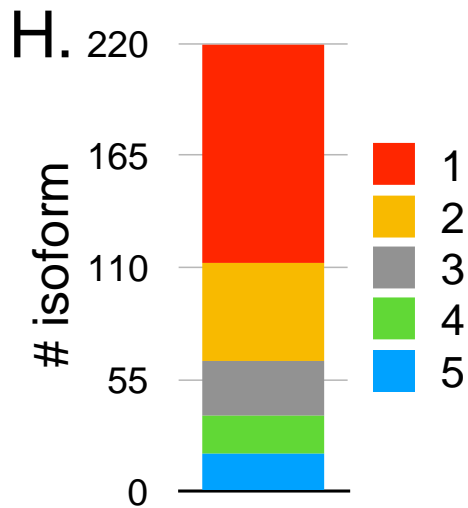
Figure 3.5: Number of isoforms counted *for each of the mirtrons only found in XRN 1/2 knockdown libraries.*
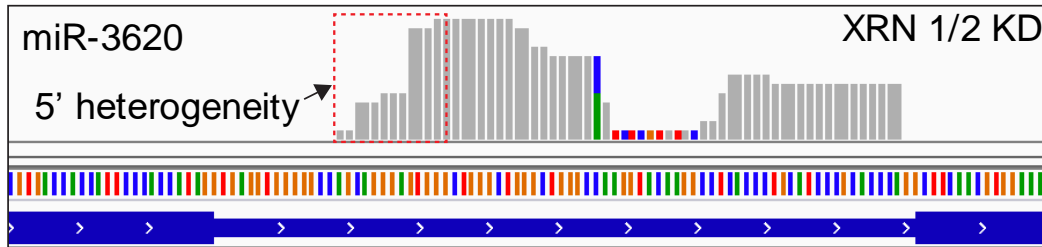
Figure 3.6: Alignment of Reads at the miR-3620 Locus Found in XRN 1/2 Knockdown Libraries. *Red box highlights 5p arm reads differing at their 5' position.*

3.3.2 Processing of the '-tailed Mirtron, has-miR-5010

To dissect the biogenesis of 5' tailed mirtrons we examined hsa-miR-5010 which is located in ATP6V0A1 (Fig 3.7). hsa-miR-5010 is one of the most highly expressed 5'-tailed mirtrons and exhibits relatively consistent 5p arm processing. As a point of

comparison, we also examined the mouse 3'-tailed mirtron mmu-miR-668, which is

likewise a high confident tailed mirtron (Fig 3.7) (Ladewig et al., 2012; Seitz et al.,

2004). While there is a homolog of mmu-miR-668 in humans that likewise resides in a
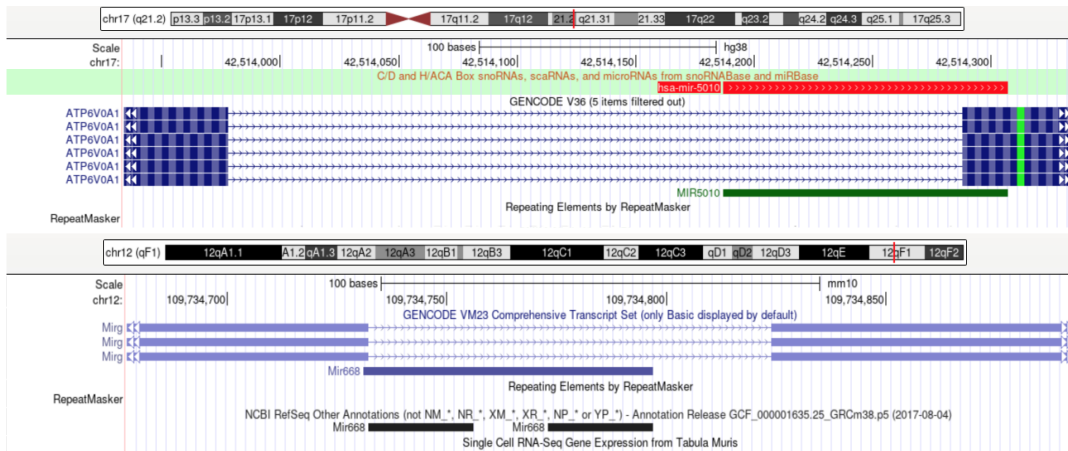
Figure 3.7: UCSC Tracks of has-miR-5010and mmu-miR-668.

large miRNA cluster it is only a 3'-tailed mirtron in mice (Seitz et al., 2004). Based on

the tendency for 5p-tailed mirtrons to harbor polyG tracts we investigated the effect of

adding polyG tracts into the tails of hsa-miR-5010 and mmu-miR-668. After each mirtron

was cloned into an expression vector, either a 12G element or a mixed identity 20-

nucleotide insert was placed in the tail of each mirtron (Fig 3.8 A-C). Wildtype, polyG,

and insert constructs were transfected into HEK 293 cells and expression of the mirtron

assessed by small RNA sequencing. Consistent with our observations regarding 5'-tailed

mirtrons, addition of the polyG tract led to greater accumulation of hsa-miR-5010 reads

while the mixed insert did not (Fig 3.8 A-B). hsa-miR-5010 biogenesis was not perturbed

by insert, and in all cases terminal modifications were not dramatically altered in

ectopically expressed hsa-miR-5010 reads (Fig 3.8B, Fig 3.9). The opposite was seen for

mmu-miR-668 (Fig3.10 A-B). Both insertions led to decreased expression.

Increased hsa-miR-5010 expression after polyG insertion was verified using a

luciferase assay to assess target silencing (Fig 3.8C, 3.12). For the polyG construct we

observed significant reduction in expression, while very little was seen for WT and insert.

Together these data suggest that mammalian 5'-tailed and 3'-tailed mirtrons are produced
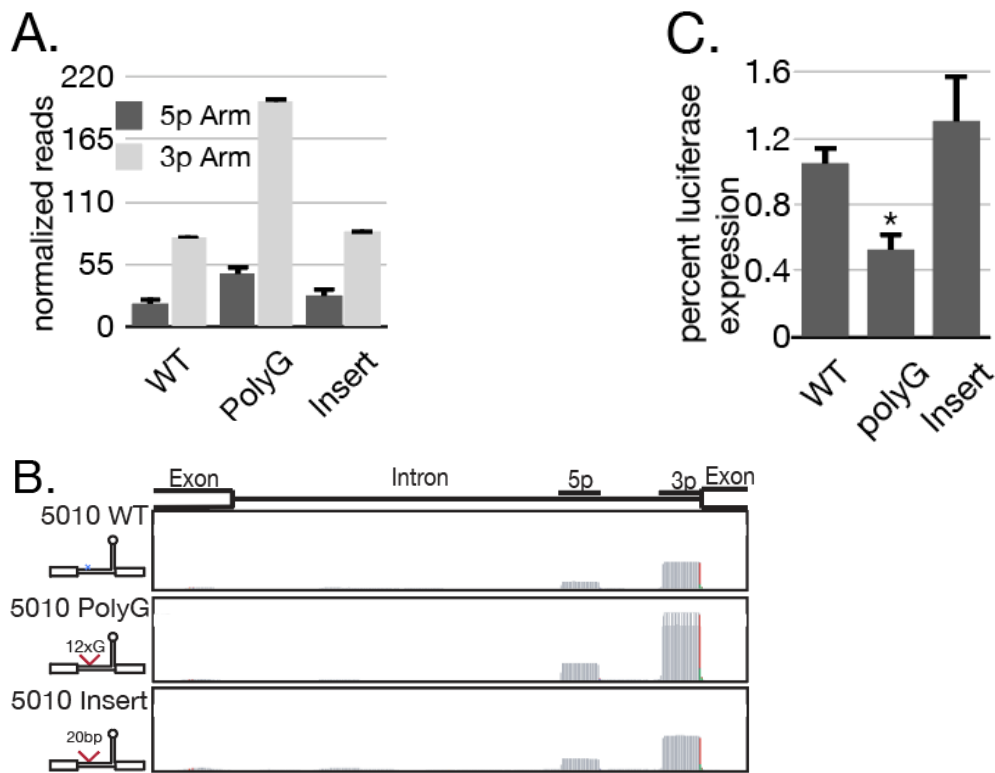
Figure 3.8: Endoribonucleases Process miR-5010. *A) Accumulation of expression, determined by small RNA sequencing, from three mirtron constructs (WT, PolyG and Insert) transfected into HEK 293T. Visualization of small RNA read alignments from transfection using the "IGV" software. Left shows the modifications made to the different mirtrons. C) Quantification of luciferase activity from a hsa-miR-5010 sensor following cotransfection with WT, Poly G and Insert hsa-miR-5010 expression plasmids.*

by very different mechanisms. The presence of additional nucleotides in mmu-miR-668

(3'-tailed) resulted in inactivation, which is like what was observed for miR-1017 in

Untemplated Nucletides
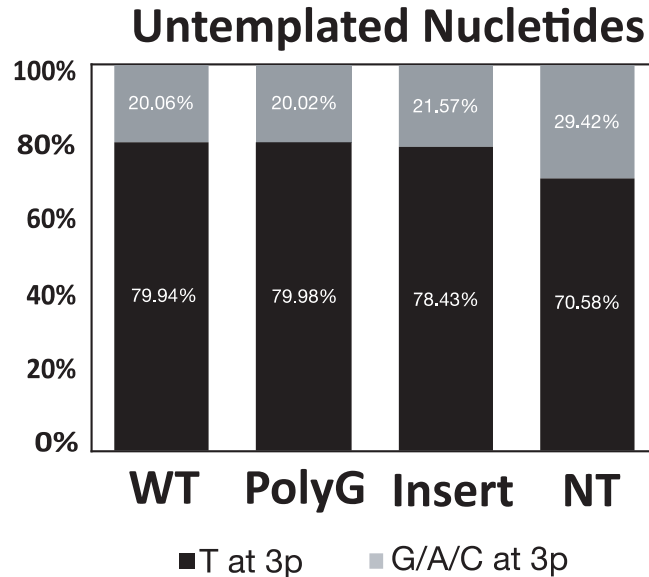
Figure 3.9: Bar Graph Showing Rates of Different Nucleotide Additions.
Quantification of 3' modifications from transfected hsa-miR-5010 constructs shows similar profiles to endogenous expressed hsa-miR-5010.

drosophila (Flynt et al., 2010). Inhibition of 3'-5' exoribonucleases by tail inserts negatively impacted biogenesis. hsa-miR-5010, in contrast, is enhanced specifically by a sequence element inhibitory to exoribonucleases. This is consistent with the effects of Xrn1/2 knockdown on 5'-tailed mirtron expression and suggests that this type of mirtron is produced by endoribonuclease-mediated severing of tails from hairpins. Considering the structure of tailed mirtrons, where a double-stranded RNA is connected to an unpaired region, it is reminiscent of immature tRNAs. Excision of tRNAs from precursors is carried out by two endoribonucleases: the 5' Leader cut by RNaseP and the 3' Trailer by RNaseZ (Hopper & Nostramo, 2019). In the case of 5'-tailed mirtrons the clear fit would be RNaseP. This enzymatic activity is carried out by a multi-subunit complex that coordinates the activity of a ribozyme. To assess the role of RNaseP in

mirtron expression, siRNAs targeting the Rpp30 subunit of RNaseP were transfected into

HEK cells followed by a second transfection with WT hsa-miR-5010 constructs (Fig 3.10
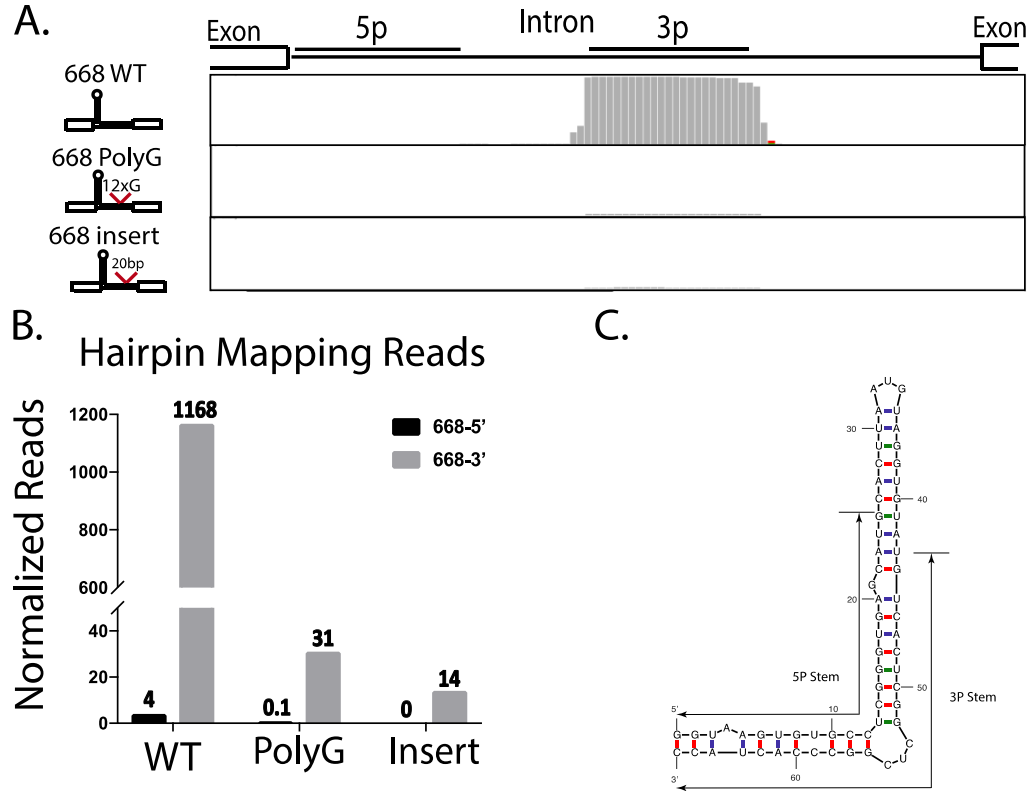
Figure 3.10: Exoribonucleases Process miR-6668. *A) Three mirtron constructs (WT, Poly G and Insert) were transfected into HEK 293T cells. In the Poly G, 12 "G"s and in the "Insert" 20 random nucleotides were inserted on the 3p tail of mmu-miR-668. Reads from RNA sequencing analysis were visualized using "IGV" software. Mirtrons with Poly G tract showed higher mature mirtrons. Untemplated "T" nucleotide reads at the 3' end are observed mostly on 3p end of the 3p arms but less than that in 5'-tailed mirtrons. B) Hairpin mapping reads were normalized based on the total number of reads in each library. WT showed much higher matured mirtrons.*

A-B). After this, small RNA sequencing libraries were created to assess differences in

expression and processing. The data shows that the expression of hsa-miR-5010 in Rpp30

knockdown is decreased by half compared to control libraries implicating RNaseP as a

factor necessary for 5'-tailed mirtron expression (Fig 3.11 A). This led us to evaluate

differences in RNAs produced from hsa-miR-5010 in control and Rpp30 knockdown

42
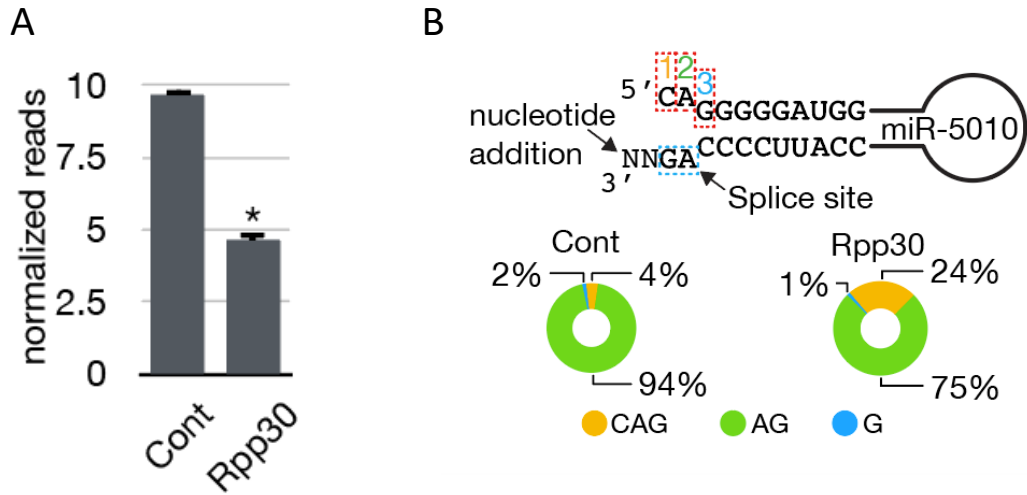
Figure 3.11: miR-5010 Sequencing. *A) Quantification of hsa-miR-5010 by high throughput sequencing after transfection of control or Rpp30 siRNAs followed by WT hsa-miR-5010 expression construct. B) Effects of Rpp30 knockdown on 5' arm heterogeneity. Dominant position found are listed as 1,2,3 as show on the hairpin.*

libraries. We detected three different types of endings at the 5' end of a mature hsa-miR-5010 (Fig 3.11B). Most of the reads (94%) in control started with "AG" (position 2) and when a "U" is added to the 3' end by uridylation, a 2 nt overhang is obtained which resembles a perfect Drosha product. This is while only 4% of the reads start with "CAG".
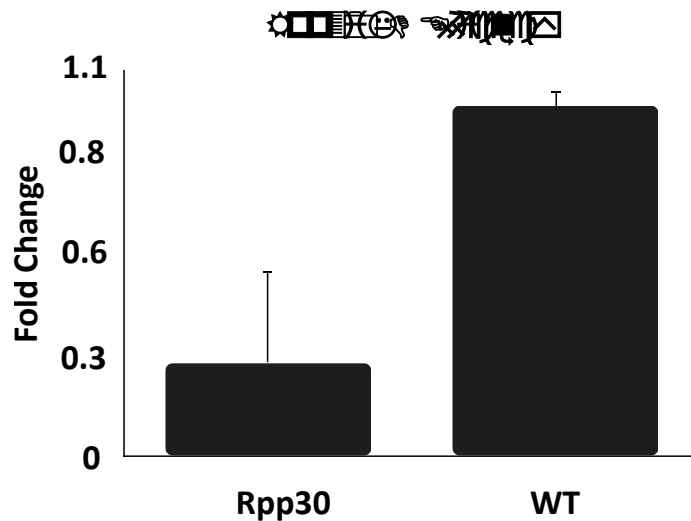
Figure 3.12: Knockdown efficiency of Rpp30 dsiRNAs.

43
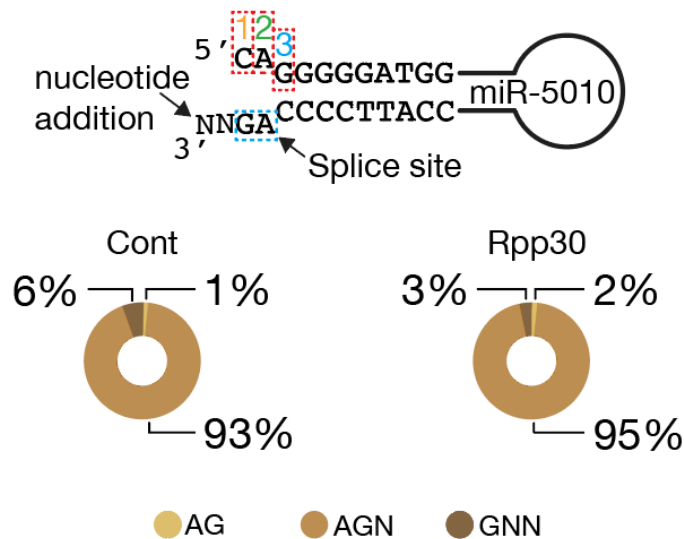
Figure 3.13: No change in hairpin 3' residue identity seen after Rpp30 knockdown. *Single nucleotide addition, which when combined with cleavage at the "2" 5' position yields an ideal Drosha product mimic.*

On the other hand, in Rpp30 knockdown libraries, 5' processing has been perturbed and a shift in the first base is observed and the rate of the "AG" is reduced to 75%. This suggests that Rpp30 knockdown has affected the 5' end processing of hsa-miR-5010. A similar analysis was applied to isoforms differing at the 3' end but there was no significant difference detected, further implicating Rpp30 in shaping the 5p arm of 5'-tailed mirtrons (Fig 3.13).

To verify the role of RNaseP in processing hsa-miR-5010, we sought to reconstitute 5'-tail removal using immunopurified complexes (Fig 3.14 A-B). Antibodies against Rpp20, another RNaseP subunit, were used to isolate complexes. *In vitro* synthesized hsa-miR-5010 primary intron transcripts were incubated with these isolated

complexes or with beads bound to mock anti-rabbit IgG antibodies. Incubation with

isolates led to significant accumulation of the ~85 nt hairpin of hsa-miR-5010 in the

RNaseP IP, but not the IgG control. In whole lysate incubation, the RNA is degraded by
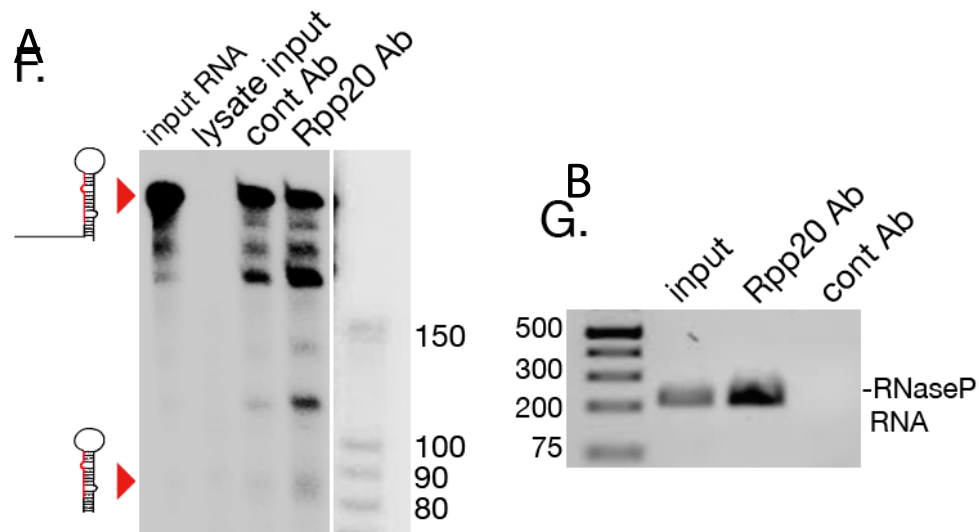
Figure 3.14: RNaseP Immunoprecipitation. *A) Incubation of radiolabeled hsa-miR-5010 primary intron with immunoprecipitated RNaseP (Rpp20 Ab), mock pull down (cont Ab), or whole cell lysate. Processing products were separated on a PAGE-urea gel. B) RT-PCR of RNaseP complex RNA subunit to verify complex isolation.*

cellular RNases. Immunoprecipitation efficiency was verified by RT-PCR of the RNA

subunit of RNaseP, which was greatly enriched in the Rpp20 IP condition relative to

control IgG (Fig 3.11B). These results indicate that not only is RNaseP necessary for

processing of hsa-miR-5010, but that it is also sufficient to recapitulate removal of the

tail of this 5'-tailed mirtron.

3.3.3 *5'-Mirtron expression is linked to RNaseP activity*

To assess the effects of Rpp30 loss and its correlation with 5'-tailed mirtrons, we

compared the expression of all 5'-tailed mirtrons after Rpp30 knockdown to levels found

in control libraries (Fig 3.15). We find that after knockdown of Rpp30 statistically

significant changes in 5'-tailed mirtron expression are almost entirely seen in down

regulated species. This is apart from ΔG values which were not predictive of significant

down regulation of 5'-tailed mirtrons in Rpp30 knockdown libraries. This reinforces our
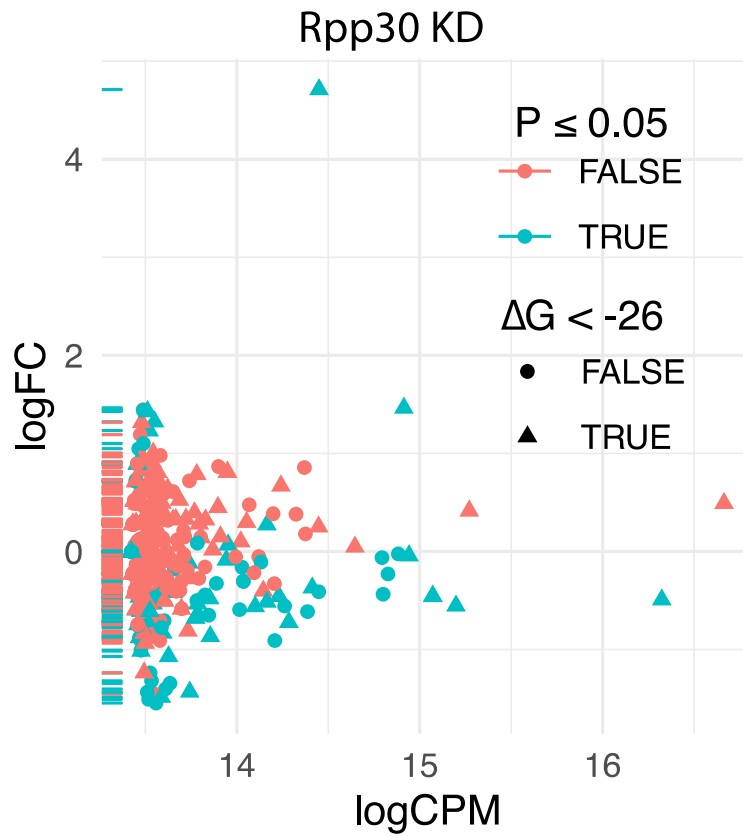
Figure 3.15: RNaseP processing and global 5'-Tailed Mirtron expression. *Scatter plot showing expression of 5'-tailed mirtrons following Rpp30 knockdown. Blue colored data points represent statistically significant changes. Red represents non-significant changes. Data point shape denotes stable (ΔG < -26) or less stable (ΔG > -26).*

observation that strength of RNA fold does not typically track with expression. The exception is the handful of outliers, which do exhibit lower ΔG values. However, outliers do not seem to respond consistently to Rpp30 knockdown, suggesting that these changes might be more related to host gene expression changes than instructive of changes in biogenesis.

Given the apparent connection of 5'-tailed mirtrons to RNaseP activity we sought to correlate changes in mirtron expression with biogenesis of tRFs (Fig 3.16 A-B). In small RNA populations, a significant group appear to be derived from pieces of tRNAs (Maute et al., 2013). tRFs are not simply degradation products, but are actively loaded into Ago proteins (Kuscu et al., 2018; Lyons, Fay, & Ivanov, 2018; Shigematsu & Kirino, 2015) Production of tRFs occurs through heterogeneous pathway with most segments of tRNA clover leaf structures giving rise to small regulatory small RNAs (Yu et al., 2020). While the function of tRFs is controversial, their expression has been correlated with an oncogenic phenotype (Yu et al., 2020). This suggests that cancer cells have a dysregulated tRNA biogenesis, which might also impact 5'-tailed mirtrons. Indeed, after surveying public small RNA sequencing from several cancer types we generally observe greater expression of 5'-tailed mirtrons relative to control tissues (Fig 3.16A). In all cases the mirtrons were either more abundant in cancer or equally distributed between cancer and norm. Decrease in cancer was not observed. Intriguingly, the degree to which 5'-tailed mirtrons are elevated, tracks almost exactly with the abundances of tRFs. The exception in this situation is lung cancer where mirtrons were enriched to the same degree as tRFs. Even with this outlier, correlation between 5'-tailed mirtrons and tRFs

47

was clear (Fig 3.16B). These results suggest a compelling correlation between changes in tRNA processing and 5'-tailed mirtrons, reinforcing a shared biogenesis mechanism.
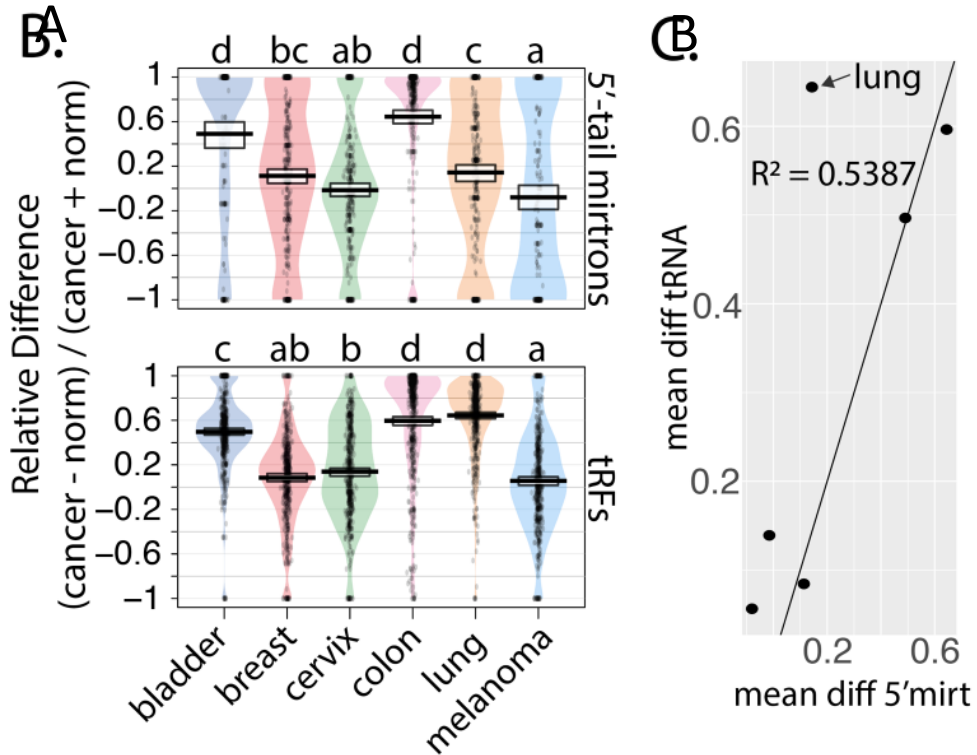
Figure 3.16: tRFs and 5'-Tailed Mirtrons Expressions. *A) Expression of 5'-tailed mirtrons and tRFs in cancers listed on x-axis. Values are relative difference of RPM values (cancer-norm)/(cancer + norm). Letters above plots are significance groups determined by TukeyHSD. B) Plot of Mean relative difference of cancer vs norm in 5'-tailed mirtrons and tRFs plotted by tissues in "B". R2 value indicated.*

3.4 Discussion

In this study, we identified how a large class of human non-canonical miRNAs, 5'-tailed mirtrons, are processed (Fig 4) (Holzmann et al., 2008; Zia & Flynt, 2018). This class of miRNA exhibits distinct features such as high G content on 5' arms that appears

to be a complement to 3' splice site adjacent ppy tracts. Prior reports on mirtron features report high GC content (J. Wen et al., 2015; J. O. Westholm et al., 2012). Their accumulation appears to be highly antagonized by the activity of 5'-3' exoribonucleases such that when this enzymatic activity is genetically depleted expression of 5'-tailed mirtrons increases ~4 fold. Moreover, following XRN 1/2 knockdown 5 times as many mirtron species become detectable. Linked to inhibition of 5'-3' processing we also observe extreme 5' arm read heterogeneity. Together these data suggest an antagonistic role for 5' turnover pathways in 5'-tailed mirtron biogenesis. Consistent with this, insertion of a 12G tract into a high-confidence human 5'-tailed mirtron, hsa-miR-5010, led to a nearly 4-fold increase in mature RNA expression. This is the opposite result to what was observed when testing sequence requirements for miR-1017, a 3'-tailed mirtron encoded in the *Drosophila* genome. For miR-1017, exoribonuclease activity of the RNA exosome was essential for 3' tail removal. Interestingly, when we perform a similar test with a mammalian 3'-tailed mirtron, mmu-miR-668, a similar result was found. Thus, it would seem both vertebrate and invertebrate 3'-tailed mirtrons likely share a biogenesis mechanism–exoribonuclease processing. They also suggest for mammalian 5'-tailed mirtrons tail removal is likely carried out by an endoribonuclease. Together these results indicate a fundamentally different biogenesis mechanism underlies 5'-tail and 3'-tail mirtrons.

In our studies we implicate the activity of RNaseP in the processing of 5'tailed

mirtrons. This ribozyme containing complex has a role in the maturation of tRNAs by

removing 5' leader sequences from precursors. However, studies have found RNaseP to

interact and cleave a variety of additional transcripts such as pre-rRNA, snoRNAs, along

with additional nuclear RNAs such as the HRA1 antisense RNA (Chamberlain, Pagan,

Kindelberger, & Engelke, 1996; Coughlin, Pleiss, Walker, Whitworth, & Engelke, 2008;

Marvin et al., 2011). Thus, it is not unreasonable to expect this complex to moonlight

with mitron primary intron transcripts. Moreover, binding affinity of RNaseP is biases to
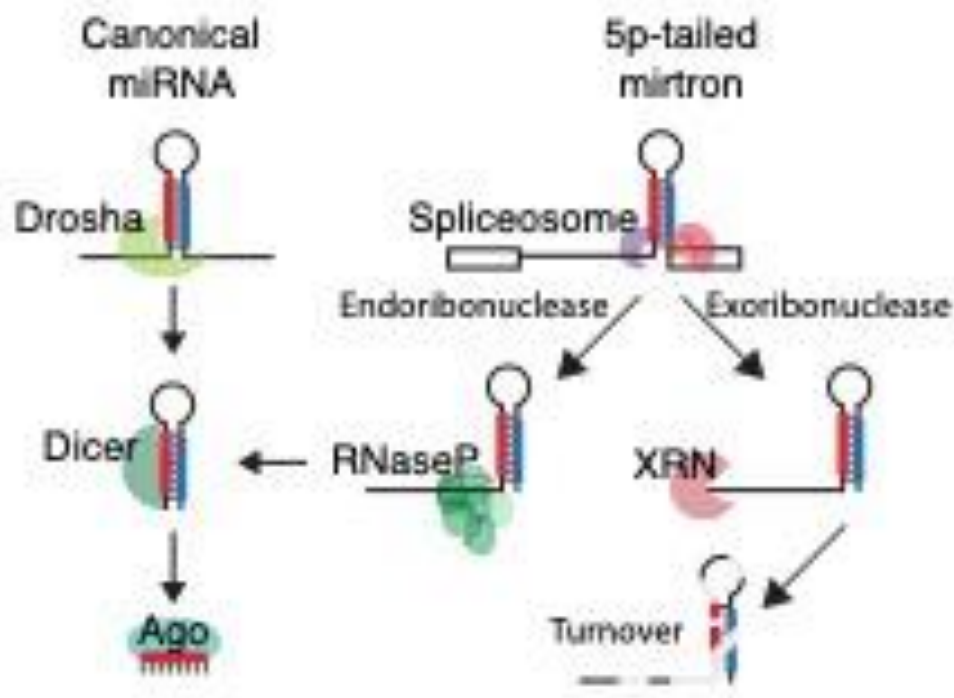
Figure 4: 5'-tailed mirtrons can undergo two different processes. *They either are processed by RNaseP and produce miRNAs or undergo turnover by XRN exoribonuclease activity and will not be targeted by RNaseP.*

nucleotide homopolymers, with greatest preference for polyG, a defining characteristic of 5'-tailed mirtrons. This with the tRNA-like single-stranded to double-stranded structure of 5'-tailed mirtrons further reinforces confidence of RNaseP as the likely 5'-tail "severing" enzyme.

One unclear aspect of 5'-tailed biology is their hyper abundance in mammalian genomes. RNaseP is a universal enzyme that is found in all cells and would be able to participate in 5'-tailed mirtron biogenesis. Likewise, as 5'-tailed mirtrons appear to arise from accumulation of polyG tracts adjacent to the universal eukaryotic ppy tract splicing element they should arise at the same rate across many kingdoms (Jacob et al., 2018). What then leads to the radical accumulation in mammalian genomes? The answer is likely related to the extreme intron length seen in mammals, and the more elaborate alternative splicing patterns (Bell, Cowper, Lefranc, Bell, & Screaton, 1998; Hube & Francastel, 2015; Roy, Kim, Xing, & Lee, 2008). We propose two possible scenarios. First, as do the massive length of mammalian introns, exact placement of strong splicing elements such as the ppy are highly favored. A polyG tract could then form through neutral evolutionary process to the point that a viable pre-mirtron hairpin arises, which then become subject to negative selection. Unfortunately, there is significant constraint on the ppy tract such that 5'-tail mirtron 3p arms are unable to acquire hairpin disrupting mutations. Harboring the 5'-tailed mirtron is preferable to degrading the strength of the ppy track. In the second scenario, the 5' arm polyG sequence has a role in modulating splicing as a cis-element. Here by forming a hairpin that occludes the ppy tract, it might lead to alternative 3' splice choice, and thereby contributing to the greater splicing complexity favored by mammalian gene expression programs. The G-quartets could also

51

be cis-elements that influence alternative splicing through recruitment of hnRNPF (H. Huang, Zhang, Harvey, Hu, & Cheng, 2017). In this second scenario 5'-tailed mirtron expression is an unintended outcome of a different, favorable arrangement.

Considering this case, perhaps 5'-tailed mirtrons can be written off as little more than genomic detritus. They may shape genomes through disfavoring polyG tracts in splice adjacent ppy elements, and thereby apply constraints on ppy neighborhoods. However, it might not be a role in normal cell physiology that should draw the attention of 5'-tailed mirtron function. As potential regulatory molecules that are diverted from Ago loaded by multiple mechanisms such as XRN activity and inhibitory nucleotide addition, these mechanisms could become corrupted to serve the subversive genetics of cancer cells. Indeed, we find a trend towards increased 5'-tailed mirtron expression in cancer-derived small RNA datasets. A similar, and possibly mechanistically related, situation is seen with tRFs, which are also amplified in cancer cells. Further, components of RNaseP are upregulated in nearly all cancer types (Uhlén et al., 2015; Uhlen et al., 2017). Both types of non-canonical miRNAs are RNaseP substrates with unclear roles in regulation of gene expression. They are excellent candidates for exploitation by cancer biology to induce survival promoting gene expression changes as well as excreted RNAs could be used to modulate the function of stromal cells within the tumor microenvironment.

3.5 Materials and Methods

3.5.1 *Cell Culture, cloning and transfection.*

Human embryonic kidney (HEK-293T, ATCC) cells were grown in Dulbecco's modified Eagle's medium (DEMEM, GIBCO) supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicilline-Streptomycine (GIBCO). Cell cultures were maintained at 37^0 C in 5% CO_2, 95 % air humidified atmosphere. The constructs to over express the mirtrons: mmu-miR-668 and hsa-miR-5010; were cloned in the pcDNA 3.1 vector. All the clones were transfected into the HEK-293T cells at 90% confluency using FuGene 6 as transfection reagent according to the manufacturer's instructions (Promega). dsiRNAs targeting Rpp30 (10 nM) were transfected into HEK 293 cells using Lipofectamine 3000 reagent according to the manufacturer's instructions (Invitrogen). Target are available on www.idtdna.com. Two days after transfection, the total RNA was extracted and used for qRT-PCR or RNA sequencing (Sup Table 1).

3.5.2 Luciferase Assay

To evaluate functionality of mirtrons constructs a dual reporter luciferase assay was used. Briefly, the hsa-miR-5010 hairpin was cloned downstream of Renilla into the psiCHECK2 vector (Table 1) Then the plasmid constructs (150 ng) were transfected into HEK 293T cells at their 90% confluency using Dual-Luciferase Reporter Assay System (Promega). The Luciferase activity was measured 72 h (48h for DsiRNA and 24 h for plasmid constructs) according to manufacturer's instructions. Four replicates were used to measure the Luciferase activity and calculated as the ratio of Renilla:Firefly.

3.5.3 *RNA Isolation, qRT-PCR, Sequencing and Analysis pipeline*.

Total RNA was collected from the transfected HEK cells by Trizol extraction

protocol (Invitrogen,Carlsbad, CA) and 2 μg of each sample was used to make small

RNA libraries for sequencing. cDNAs was prepared using RevertAid First Strand cDNA

Synthesis Kit (Thermoscientific) according to the manufacturer's recommendation to

perform qRT-PCR (BioRad) (Sup Table 1). Sequencing was performed on the Illumina

NextSeq platform, and adaptor were clipped using fastx_clipper (Hannon, 2010). Bowtie

was used for alignments to hg19 or mm10 (Langmead, Trapnell, Pop, & Salzberg, 2009).

Next, samtools and bedtools were used to quantify reads and examine alignments (H. Li

et al., 2009; Quinlan & Hall, 2010). The EdgeR R package was used to assess 5'-tailed

mirtron expression in libraries (Robinson, McCarthy, & Smyth, 2010).

3.5.4 In Vitro Transcription

Synthetic full length hsa-miR-5010 intron RNA was generated by *in vitro*

transcription with MEGAscript® T7 Kit (Invitrogen) using full length intron PCR

amplification with a T7 encoding forward primer (Sup Table 1). RNAs were radiolabeled

by addition of alpha-P^{32} UTP to the reaction.

3.5.5 *Immunoprecipitation and In Vitro Enzymatic Processing.*

For immunoprecipitation, the HEK 293T cells were lysed in 0.1% NP-40, 800 nM

NaCl, 20 mM Tris at 8.0, 30 mM HEPES, 2 mM Magnesium Acetate, Protease Inhibitor

by sonicated. The insoluble fraction was removed by centrifugation at 14,400 rpm for 15

minutes at 4^0 C. After that the supernatant was incubated for 2 hour with 10 uL of Rpp20

Antibody (Novus Biologicals) immobilized Dynabeads™ Protein G Immunoprecipitation

Kit (Invitrogen). The beads were washed five times with IP buffer (0.1% NP-40, 20 mM

Tris at 8.0, 30 mM HEPES, 2 mM Magnesium Acetate, 150 mM Potassium Acetate),

using a magnetic stand. The enzyme assay reaction was performed at 37^0 C for 15

minutes with incubation buffer (5% PEG, 10 mM MgCl2, 50 mM Tris-HCl, 100 nM

NH4Cl) by incubating RNAs with RNaseP-bound or control IgG-beads. Beads were

pelleted and the supernatant was loaded on 8% urea-acrylamide gel and the gel was

exposed to phosphor imaging plate (Fujifilm) and was read by Typhoon FLA 7,000. In a

separate run, beads were subject to organic extraction followed by RT-PCR of RNaseP

RNA subunit.

3.5.6 Cancer Small RNA-Metanalysis.

Public data from NCBI was used to analyze 5'-tailed mirtron and tRF expression

(Table 2). Cancer-norm pairs were from same studies. Alignment and read counting was

performed as described above. tRFs coordinates were obtained for hg19

http://gtrnadb.ucsc.edu. Counts were normalized to million of reads mapping (RPM) in

the corresponding library. After, relative difference was calculated as (cancer-

norm)/(cancer+norm).

Table 1: Primers and DsiRNAs Used.

Construct	Sequence
RPP30 F	tcagcatggcggtgtttgcag
RPP30 R	gcggctgtctccacaagtccg
B-Actin F	catgtacgttgctatccaggc
B-Actin R	ctccttaatgtcacgcacgat
hsa-miR-5010 T7 F	taatacgactcactataggTGAGTACCTCTCTCCGGGC
hsa-miR-5010 R	ctggggaatgggagacacaaatcc
mmu-miR-668 F	gtaagtgtgcctcgggtgag
mmu-miR-688 R	cattcacacggagcccactc
hsa-miR-5010 + Strand, Luciferase insert	ggccgcCCATCC CCCactgacCCATCCCCCactgacCCATCCCCC actgacCCATCCCCCc
hsa-miR-5010 - Strand, Luciferase insert	tcgagGGGGGATGGgtcagtGGGGGATGGgtcagtGGGGGATGGgtcagtGGG GAT
dsiRNA used	hs.Ri.Rpp30.13.1
dsiRNA used	CD.Ri.195363.13.5
Rpph1F	ccactgatgagcttccctcc
Rpph1F	ggaggagagtagtctgaattgg

Table 2: Public Data Used From NCBI.

Cancer	Cancer	Norm
Lung	SRR14634327,SRR14634347,SRR14634359	SRR14634273,SRR14634279,SRR14634283
Melanoma	SRR14510928,SRR14510929,SRR14510930	SRR14510925,SRR14510926,SRR14510927
Bladder	SRR333655,SRR333657,SRR333659	SRR333656,SRR333658,SRR333660
Colorectal	SRR8932122,SRR8932124	SRR8932123,SRR8932125
cervical	SRR11095745,SRR11095746,SRR11095747, SRR11095748	SRR11095741,SRR11095742,SRR1109574, SRR11095744
Breast	SRR8330380,SRR8330384,SRR8330374, SRR8330376,SRR8330378,SRR8330382	SRR8330375,SRR8330377,SRR8330379, SRR8330381,SRR8330383,SRR8330385

CHAPTER IV – CONCLUSION AND FUTURE DIRECTION

Since the discovery of the mirtrons this is the first study working on the biogenesis of 5'-tailed mirtrons and 5' tail removal. We answered the most important question about their biogenesis demonstrating that 5'-tailed mirtrons are processed through "Endoribonuclease" activity. We then identified the "RNaseP" complex as the main enzyme complex involved in this process which is a remarkably influential discovery in the field of mirtrons biogenesis specifically in that 86% of the mirtrons are 5'-tailed mirtrons. Additionally, we uncovered that XRN 1/2 family enzymes contribute in the 5'-tailed mirtrons turn over. We also showed the relationship between the changes in the mirtrons biogenesis and different human cancers which emphasizes on the importance of their biogenesis and modifications on their structures especially "Uridylation" on the 3p end. By altering the involved biogenesis factors, not only the 5'-tailed mirtron biogenesis is perturbed, but also their arm bios is subjected to shift from one arm to another. This has been reported in cancer studies as well (L. Chen et al., 2018; Z. Zhang et al., 2019). Therefore, studying the relationship between uridylation on 3p end of mirtrons and their biogenesis with different cancers to elucidate the molecular mechanism in this process can be an outstanding future topic.

Moreover, it is unclear exactly how many of other subclasses of ncRNAs are expressed in all aspects of living organisms and especially higher mammalians. Therefore, we postulate that there might be more examples of alternative non-canonical pathways resulting in small RNAs. Hence, revealing more details of how and to what extent the genes are regulated through unexplored non-canonical RNAs, appears to be an enigmatic prospect of this field. This could be applied to other Drosha independent

pathways such as shRNAs in which they generate miRNAs. This can be a leading indication to investigate whether RNaseP complex is processing shRNAs in the same fashion as mirtrons are.

To date, about 500 mirtrons have been annotated in the human genome, covering almost 20% of the total canonical miRNA population. Besides, remarkable elaborations have been disclosed in understanding their biogenesis details. However, still some uncertainties regarding mirtron's turnover, export pathways and modifications requires future explorations. Additionally, experiments are needed to identify targets of these non-canonical miRNAs to address their biological function and significance. This can be a boon for researchers in the field to uncover the relevance of mirtrons to diseases. This has become more challenging in mirtrons biogenesis and functionality after we revealed that there is a strong correlation with changes in 5'-tailed mirtron processing with tRNA, proposing shared biogenesis mechanism. In addition, by revealing more structural features of mirtrons such as high G content, the path to define more reliable computational algorithms of discovering novel mirtrons seems more feasible. By this, few mirtrons that might have been incorrectly annotated as mirtrons can be removed from the list. Moreover, defining algorithms that help recognizing the cleavage site of the RNaseP complex on mirtrons seems an area in which much effort could be invested.

There has been attempts to introduce artificial mirtrons as therapeutic implements; however, more evidence and considerations seem indispensable. This becomes more crucial now that pharmaceutical companies and research laboratories are racing to pioneer RNAi-based drugs to treat diseases like cancers and Alzheimer's. This has become more challenging during the COVID-19 pandemic and global accessibility and

application of RNA-based vaccines which signifies increasing demands on the emergence of new technologies and discoveries in RNA-based drugs as a prominent approach to overcome ever-challenging and even future diseases. Our findings here in this study, regarding mirtron G content and polyG tract blocking the maturation process from 5' end might be applicable in designing better siRNAs. About half of the mirtrons with lower ΔG were downregulated in RNaseP knockdown. Therefore, better siRNAs can be designed by considering ΔG level of the designs. On the other hand, more considerations on the 5' end of the siRNAs seems reasonable to avoid being degraded by ribonucleases such as XRN 1/2 before they are delivered in the cell to target the gene of interest.

On the other hand, another area which seems essential to provide more evidence is to assess as to why mirtrons have emerged since they are reported to be poorly conserved and most of them are not retained during the evolution. Perhaps the answer could be found in the natural selection as their primary sequence has evolved quickly. This infers that while much has been done on mirtrons in recent years, there is still many studies to be accomplished in their evolutionary investigations.

Overall, our findings on RNaseP and XRN family provides influential small RNA functional applications in future studies revealing biogenesis pathways of other known and unknown ncRNAs.

REFERENCES

Altman, S. (2011). Ribonuclease P. *Philos Trans R Soc Lond B Biol Sci, 366*(1580), 2936-2941. doi:10.1098/rstb.2011.0142

Ardekani, A. M., & Naeini, M. M. (2010). The Role of MicroRNAs in Human Diseases. *Avicenna J Med Biotechnol, 2*(4), 161-179. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23407304

Arenas, J. E., & Abelson, J. N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U S A, 94*(22), 11798-11802. doi:10.1073/pnas.94.22.11798

Babiarz, J. E., Hsu, R., Melton, C., Thomas, M., Ullian, E. M., & Blelloch, R. (2011). A role for noncanonical microRNAs in the mammalian brain revealed by phenotypic differences in Dgcr8 versus Dicer1 knockouts and small RNA sequencing. *RNA (New York, N.Y.), 17*(8), 1489-1501. doi:10.1261/rna.2442211

Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P., & Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev, 22*(20), 2773-2785. doi:10.1101/gad.1705308

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell, 173*(1), 20-51. doi:10.1016/j.cell.2018.03.006

Bell, M. V., Cowper, A. E., Lefranc, M. P., Bell, J. I., & Screaton, G. R. (1998). Influence of intron length on alternative splicing of CD44. *Mol Cell Biol, 18*(10), 5930-5941. doi:10.1128/MCB.18.10.5930

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet, 12*(12), 846-860. doi:10.1038/nrg3079

Bloomston, M., Frankel, W. L., Petrocca, F., Volinia, S., Alder, H., Hagan, J. P., . . . Croce, C. M. (2007). MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. *JAMA, 297*(17), 1901-1908. doi:10.1001/jama.297.17.1901

Bortolamiol-Becet, D., Hu, F., Jee, D., Wen, J., Okamura, K., Lin, C. J., . . . Lai, E. C. (2015). Selective Suppression of the Splicing-Mediated MicroRNA Pathway by the Terminal Uridyltransferase Tailor. *Mol Cell, 59*(2), 217-228. doi:10.1016/j.molcel.2015.05.034

Chamberlain, J. R., Pagan, R., Kindelberger, D. W., & Engelke, D. R. (1996). An RNase P RNA subunit mutation affects ribosomal RNA processing. *Nucleic Acids Res, 24*(16), 3158-3166. doi:10.1093/nar/24.16.3158

Chen, C., Zhang, Y., Zhang, L., Weakley, S. M., & Yao, Q. (2011). MicroRNA-196: critical roles and clinical applications in development and cancer. *J Cell Mol Med, 15*(1), 14-23. doi:10.1111/j.1582-4934.2010.01219.x

Chen, L., Sun, H., Wang, C., Yang, Y., Zhang, M., & Wong, G. (2018). miRNA arm switching identifies novel tumour biomarkers. *EBioMedicine, 38*, 37-46. doi:10.1016/j.ebiom.2018.11.003

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., . . . Gingeras, T. R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science, 308*(5725), 1149-1154. doi:10.1126/science.1108625

Cho, H., Lee, W., Kim, G. W., Lee, S. H., Moon, J. S., Kim, M., . . . Oh, J. W. (2019).
Regulation of La/SSB-dependent viral gene expression by pre-tRNA 3' trailer-
derived tRNA fragments. *Nucleic Acids Res, 47*(18), 9888-9901.
doi:10.1093/nar/gkz732

Coughlin, D. J., Pleiss, J. A., Walker, S. C., Whitworth, G. B., & Engelke, D. R. (2008).
Genome-wide search for yeast RNase P substrates reveals role in maturation of
intron-encoded box C/D small nucleolar RNAs. *Proc Natl Acad Sci U S A,
105*(34), 12218-12223. doi:10.1073/pnas.0801906105

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence
logo generator. *Genome Res, 14*(6), 1188-1190. doi:10.1101/gr.849004

Curtis, H. J., Sibley, C. R., & Wood, M. J. (2012). Mirtrons, an emerging class of
atypical miRNA. *Wiley Interdiscip Rev RNA, 3*(5), 617-632.
doi:10.1002/wrna.1122

De Almeida, C., Scheer, H., Zuber, H., & Gagliardi, D. (2018). RNA uridylation: a key
posttranscriptional modification shaping the coding and noncoding transcriptome.
Wiley Interdiscip Rev RNA, 9(1). doi:10.1002/wrna.1440

Delihas, N. (2015). Discovery and characterization of the first non-coding RNA that
regulates gene expression, micF RNA: A historical perspective. *World J Biol
Chem, 6*(4), 272-280. doi:10.4331/wjbc.v6.i4.272

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., & Hannon, G. J. (2004).
Processing of primary microRNAs by the Microprocessor complex. *Nature,
432*(7014), 231-235. doi:10.1038/nature03049

Diamantopoulos, M. A., Tsiakanikas, P., & Scorilas, A. (2018). Non-coding RNAs: the riddle of the transcriptome and their perspectives in cancer. *Ann Transl Med, 6*(12), 241. doi:10.21037/atm.2018.06.10

Dutta, T., Malhotra, A., & Deutscher, M. P. (2012). Exoribonuclease and endoribonuclease activities of RNase BN/RNase Z both function in vivo. *J Biol Chem, 287*(42), 35747-35755. doi:10.1074/jbc.M112.407403

Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., . . . Meister, G. (2008). A human snoRNA with microRNA-like functions. *Mol Cell, 32*(4), 519-528. doi:10.1016/j.molcel.2008.10.017

Fendereski, M., Zia, M. F., Shafiee, M., Safari, F., Saneie, M. H., & Tavassoli, M. (2017). MicroRNA-196a as a Potential Diagnostic Biomarker for Esophageal Squamous Cell Carcinoma. *Cancer Invest, 35*(2), 78-84. doi:10.1080/07357907.2016.1254228

Flynt, A. S., Greimann, J. C., Chung, W. J., Lima, C. D., & Lai, E. C. (2010). MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. *Mol Cell, 38*(6), 900-907. doi:10.1016/j.molcel.2010.06.014

Fortin, K. R., Nicholson, R. H., & Nicholson, A. W. (2002). Mouse ribonuclease III. cDNA structure, expression analysis, and chromosomal location. *BMC Genomics, 3*(1), 26. doi:10.1186/1471-2164-3-26

Geisler, S., & Coller, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol, 14*(11), 699-712. doi:10.1038/nrm3679

Gorelick, F. S., Shugrue, C. A., Kolodecik, T. R., & Thrower, E. C. (2006). Vacuolar

adenosine triphosphatase and pancreatic acinar cell function. *J Gastroenterol

Hepatol, 21 Suppl 3*, S18-21. doi:10.1111/j.1440-1746.2006.04576.x

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006).

miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids

Res, 34*(Database issue), D140-144. doi:10.1093/nar/gkj112

Grimson, A., Farh, K. K.-H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel,

D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond

seed pairing. *Molecular Cell, 27*(1), 91-105. doi:10.1016/j.molcel.2007.06.017

Gutierrez-Vazquez, C., Enright, A. J., Rodriguez-Galan, A., Perez-Garcia, A., Collier, P.,

Jones, M. R., . . . Sanchez-Madrid, F. (2017). 3' Uridylation controls mature

microRNA turnover during CD4 T-cell activation. *RNA, 23*(6), 882-891.

doi:10.1261/rna.060095.116

Guzzi, N., & Bellodi, C. (2020). Novel insights into the emerging roles of tRNA-derived

fragments in mammalian development. *RNA Biology, 17*, 1-9.

doi:10.1080/15476286.2020.1732694

Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., & Kim, V. N. (2004). The Drosha-

DGCR8 complex in primary microRNA processing. *Genes Dev, 18*(24), 3016-

3027. doi:10.1101/gad.1262504

Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., . . . Kim, V. N. (2006).

Molecular basis for the recognition of primary microRNAs by the Drosha-

DGCR8 complex. *Cell, 125*(5), 887-901. doi:10.1016/j.cell.2006.03.043

Hannon, G. J. (2010). FASTX-Toolkit. *http://hannonlab.cshl.edu/fastx_toolkit*. Retrieved

from http://hannonlab.cshl.edu/fastx_toolkit.

Hasler, D., Lehmann, G., Murakawa, Y., Klironomos, F., Jakob, L., Grässer,

Friedrich A., . . . Meister, G. (2016). The Lupus Autoantigen La Prevents Mis-

channeling of tRNA Fragments into the Human MicroRNA Pathway. *Molecular

Cell, 63*(1), 110-124. doi:https://doi.org/10.1016/j.molcel.2016.05.026

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., & Kim, V. N. (2008). Lin28 mediates the

terminal uridylation of let-7 precursor MicroRNA. *Mol Cell, 32*(2), 276-284.

doi:10.1016/j.molcel.2008.09.014

Holliday, L. S. (2017). Vacuolar H(+)-ATPases (V-ATPases) as therapeutic targets: a

brief review and recent developments. *Biotarget, 1*.

doi:10.21037/biotarget.2017.12.01

Holzmann, J., Frank, P., Loffler, E., Bennett, K. L., Gerner, C., & Rossmanith, W.

(2008). RNase P without RNA: identification and functional reconstitution of the

human mitochondrial tRNA processing enzyme. *Cell, 135*(3), 462-474.

doi:10.1016/j.cell.2008.09.013

Hopper, A. K., & Nostramo, R. T. (2019). tRNA Processing and Subcellular Trafficking

Proteins Multitask in Pathways for Other RNAs. *Frontiers in Genetics, 10*(96).

doi:10.3389/fgene.2019.00096

Huang, H., Zhang, J., Harvey, S. E., Hu, X., & Cheng, C. (2017). RNA G-quadruplex

secondary structure promotes alternative splicing via the RNA-binding protein

hnRNPF. *Genes Dev, 31*(22), 2296-2309. doi:10.1101/gad.305862.117

Huang, Q., Mao, Z., Li, S., Hu, J., & Zhu, Y. (2014). A non-radioactive method for small RNA detection by northern blotting. *Rice (N Y), 7*(1), 26. doi:10.1186/s12284-014-0026-1

Hube, F., & Francastel, C. (2015). Mammalian introns: when the junk generates molecular diversity. *Int J Mol Sci, 16*(3), 4429-4452. doi:10.3390/ijms16034429

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol, 9*(1), 22-32. doi:10.1038/nrm2321

Hwang, H. W., Wentzel, E. A., & Mendell, J. T. (2009). Cell-cell contact globally activates microRNA biogenesis. *Proc Natl Acad Sci U S A, 106*(17), 7016-7021. doi:10.1073/pnas.0811523106

Jacob, H., Stanisavljevic, L., Storli, K. E., Hestetun, K. E., Dahl, O., & Myklebust, M. P. (2018). A four-microRNA classifier as a novel prognostic marker for tumor recurrence in stage II colon cancer. *Sci Rep, 8*(1), 6157. doi:10.1038/s41598-018-24519-4

Jarrous, N. (2002). Human ribonuclease P: subunits, function, and intranuclear localization. *RNA, 8*(1), 1-7. doi:10.1017/s1355838202011184

Jarrous, N., & Reiner, R. (2007). Human RNase P: a tRNA-processing enzyme and transcription factor. *Nucleic Acids Research, 35*(11), 3519-3524. doi:10.1093/nar/gkm071

Kaufman, E. J., & Miska, E. A. (2010). The microRNAs of Caenorhabditis elegans. *Semin Cell Dev Biol, 21*(7), 728-737. doi:10.1016/j.semcdb.2010.07.001

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., & Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA

involved in developmental timing in C. elegans. *Genes Dev, 15*(20), 2654-2659. doi:10.1101/gad.927801

Kim, B., Ha, M., Loeff, L., Chang, H., Simanshu, D. K., Li, S., . . . Kim, V. N. (2015). TUT7 controls the fate of precursor microRNAs by using three different uridylation mechanisms. *EMBO J, 34*(13), 1801-1815. doi:10.15252/embj.201590931

Kim, H., Kim, J., Yu, S., Lee, Y. Y., Park, J., Choi, R. J., . . . Kim, V. N. (2020). A Mechanism for microRNA Arm Switching Regulated by Uridylation. *Mol Cell, 78*(6), 1224-1236 e1225. doi:10.1016/j.molcel.2020.04.030

Kim, Y. K., Kim, B., & Kim, V. N. (2016). Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc Natl Acad Sci U S A, 113*(13), E1881-1889. doi:10.1073/pnas.1602532113

Knoll, M., Lodish, H. F., & Sun, L. (2015). Long non-coding RNAs as regulators of the endocrine system. *Nat Rev Endocrinol, 11*(3), 151-160. doi:10.1038/nrendo.2014.229

Kuscu, C., Kumar, P., Kiran, M., Su, Z., Malik, A., & Dutta, A. (2018). tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA, 24*(8), 1093-1105. doi:10.1261/rna.066126.118

Ladewig, E., Okamura, K., Flynt, A. S., Westholm, J. O., & Lai, E. C. (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res, 22*(9), 1634-1645. doi:10.1101/gr.133553.111

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology, 10*(3), R25. doi:10.1186/gb-2009-10-3-r25

Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science, 294*(5543), 858-862. doi:10.1126/science.1065062

Le Pen, J., Jiang, H., Di Domenico, T., Kneuss, E., Kosalka, J., Leung, C., . . . Miska, E. A. (2018). Terminal uridylyltransferases target RNA viruses as part of the innate immune system. *Nat Struct Mol Biol, 25*(9), 778-786. doi:10.1038/s41594-018-0106-9

Lee, R. C., & Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science, 294*(5543), 862-864. doi:10.1126/science.1065329

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell, 75*(5), 843-854. doi:10.1016/0092-8674(93)90529-y

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., . . . Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature, 425*(6956), 415-419. doi:10.1038/nature01957

Lerner, L. K., & Sale, J. E. (2019). Replication of G Quadruplex DNA. *Genes (Basel), 10*(2). doi:10.3390/genes10020095

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, R., Jia, Y., Zou, H., & Zhao, R. (2014). Breed-specific expression of DROSHA,

   DICER and AGO2 is regulated by glucocorticoid-mediated miRNAs in the liver

   of newborn piglets. *Anim Genet, 45*(6), 817-826. doi:10.1111/age.12232

Lyons, S. M., Fay, M. M., & Ivanov, P. (2018). The role of RNA modifications in the

   regulation of tRNA cleavage. *FEBS Lett, 592*(17), 2828-2844. doi:10.1002/1873-

   3468.13205

Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., & Patel, D. J. (2005). Structural

   basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi

   protein. *Nature, 434*(7033), 666-670. doi:10.1038/nature03514

Macfarlane, L. A., & Murphy, P. R. (2010). MicroRNA: Biogenesis, Function and Role

   in Cancer. *Curr Genomics, 11*(7), 537-561. doi:10.2174/138920210793175895

Marvin, M. C., Clauder-Munster, S., Walker, S. C., Sarkeshik, A., Yates, J. R., 3rd,

   Steinmetz, L. M., & Engelke, D. R. (2011). Accumulation of noncoding RNA due

   to an RNase P defect in Saccharomyces cerevisiae. *RNA, 17*(8), 1441-1450.

   doi:10.1261/rna.2737511

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Hum Mol Genet, 15 Spec No

   1*, R17-29. doi:10.1093/hmg/ddl046

Maute, R. L., Schneider, C., Sumazin, P., Holmes, A., Califano, A., Basso, K., & Dalla-

   Favera, R. (2013). tRNA-derived microRNA modulates proliferation and the

   DNA damage response and is down-regulated in B cell lymphoma. *Proc Natl

   Acad Sci U S A, 110*(4), 1404-1409. doi:10.1073/pnas.1206761110

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell, 15*(2), 185-197. doi:10.1016/j.molcel.2004.07.007

Menezes, M. R., Balzeau, J., & Hagan, J. P. (2018). 3' RNA Uridylation in Epitranscriptomics, Gene Regulation, and Disease. *Front Mol Biosci, 5*, 61. doi:10.3389/fmolb.2018.00061

Meunier, J., Lemoine, F., Soumillon, M., Liechti, A., Weier, M., Guschanski, K., . . . Kaessmann, H. (2013). Birth and expression evolution of mammalian microRNA genes. *Genome Res, 23*(1), 34-45. doi:10.1101/gr.140269.112

Miyoshi, K., Miyoshi, T., & Siomi, H. (2010). Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. *Mol Genet Genomics, 284*(2), 95-103. doi:10.1007/s00438-010-0556-1

Miyoshi, K., Miyoshi, T., & Siomi, H. (2010). Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. *Molecular Genetics and Genomics, 284*(2), 95-103. doi:10.1007/s00438-010-0556-1

Mohammed, J., Flynt, A. S., Siepel, A., & Lai, E. C. (2013). The impact of age, biogenesis, and genomic clustering on Drosophila microRNA evolution. *RNA, 19*(9), 1295-1308. doi:10.1261/rna.039248.113

Norbury, C. J. (2013). Cytoplasmic RNA: a case of the tail wagging the dog. *Nat Rev Mol Cell Biol, 14*(10), 643-653. doi:10.1038/nrm3645

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne), 9*, 402. doi:10.3389/fendo.2018.00402

Ohira, T., & Suzuki, T. (2011). Retrograde nuclear import of tRNA precursors is required

    for modified base biogenesis in yeast. *Proc Natl Acad Sci U S A, 108*(26), 10502-

    10507. doi:10.1073/pnas.1105645108

Ohira, T., & Suzuki, T. (2016). Precursors of tRNAs are stabilized by methylguanosine

    cap structures. *Nat Chem Biol, 12*(8), 648-655. doi:10.1038/nchembio.2117

Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., & Lai, E. C. (2007). The mirtron

    pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell, 130*(1),

    89-100. doi:10.1016/j.cell.2007.06.028

Palazzo, A. F., & Lee, E. S. (2015). Non-coding RNA: what is functional and what is

    junk? *Front Genet, 6*, 2. doi:10.3389/fgene.2015.00002

Parker, J. S. (2010). How to slice: snapshots of Argonaute in action. *Silence, 1*(1), 3.

    doi:10.1186/1758-907X-1-3

Parker, J. S., Roe, S. M., & Barford, D. (2005). Structural insights into mRNA

    recognition from a PIWI domain-siRNA guide complex. *Nature, 434*(7033), 663-

    666. doi:10.1038/nature03462

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B.,

    . . . Ruvkun, G. (2000). Conservation of the sequence and temporal expression of

    let-7 heterochronic regulatory RNA. *Nature, 408*(6808), 86-89.

    doi:10.1038/35040556

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing

    genomic features. *Bioinformatics (Oxford, England), 26*(6), 841-842.

    doi:10.1093/bioinformatics/btq033

Ragusa, M., Barbagallo, C., Brex, D., Caponnetto, A., Cirnigliaro, M., Battaglia, R., . . . Purrello, M. (2017). Molecular Crosstalking among Noncoding RNAs: A New Network Layer of Genome Regulation in Cancer. *Int J Genomics, 2017*, 4723193. doi:10.1155/2017/4723193

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., . . . Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature, 403*(6772), 901-906. doi:10.1038/35002607

Rio, D. C. (2014). Northern blots for small RNAs and microRNAs. *Cold Spring Harb Protoc, 2014*(7), 793-797. doi:10.1101/pdb.prot080838

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Rorbach, G., Unold, O., & Konopka, B. M. (2018). Distinguishing mirtrons from canonical miRNAs with data exploration and machine learning methods. *Sci Rep, 8*(1), 7560. doi:10.1038/s41598-018-25578-3

Roy, M., Kim, N., Xing, Y., & Lee, C. (2008). The effect of intron length on exon creation ratios during the evolution of mammalian genomes. *RNA, 14*(11), 2261-2273. doi:10.1261/rna.1024908

Ruby, J. G., Jan, C. H., & Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature, 448*(7149), 83-86. doi:10.1038/nature05983

Seitz, H., Royo, H., Bortolin, M. L., Lin, S. P., Ferguson-Smith, A. C., & Cavaille, J. (2004). A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res, 14*(9), 1741-1748. doi:10.1101/gr.2743304

Seow, Y., Sibley, C. R., & Wood, M. J. (2012). Artificial mirtron-mediated gene knockdown: functional DMPK silencing in mammalian cells. *RNA, 18*(7), 1328-1337. doi:10.1261/rna.030601.111

Shahrouki, P., & Larsson, E. (2012). The non-coding oncogene: a case of missing DNA evidence? *Front Genet, 3*, 170. doi:10.3389/fgene.2012.00170

Sharbati-Tehrani, S., Kutz-Lohroff, B., Bergbauer, R., Scholven, J., & Einspanier, R. (2008). miR-Q: a novel quantitative RT-PCR approach for the expression profiling of small RNA molecules such as miRNAs in a complex sample. *BMC Mol Biol, 9*, 34. doi:10.1186/1471-2199-9-34

Shigematsu, M., & Kirino, Y. (2015). tRNA-Derived Short Non-coding RNA as Interacting Partners of Argonaute Proteins. *Gene Regul Syst Bio, 9*, 27-33. doi:10.4137/GRSB.S29411

Sibley, C. R., Seow, Y., Saayman, S., Dijkstra, K. K., El Andaloussi, S., Weinberg, M. S., & Wood, M. J. (2012). The biogenesis and characterization of mammalian microRNAs of mirtron origin. *Nucleic Acids Res, 40*(1), 438-448. doi:10.1093/nar/gkr722

Song, J. J., Smith, S. K., Hannon, G. J., & Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science, 305*(5689), 1434-1437. doi:10.1126/science.1102514

Spath, B., Canino, G., & Marchfelder, A. (2007). tRNase Z: the end is not in sight. *Cell Mol Life Sci, 64*(18), 2404-2412. doi:10.1007/s00018-007-7160-5

Takaku, H., Minagawa, A., Takagi, M., & Nashimoto, M. (2003). A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. *Nucleic Acids Res, 31*(9), 2272-2278. doi:10.1093/nar/gkg337

Tam, S., de Borja, R., Tsao, M. S., & McPherson, J. D. (2014). Robust global microRNA expression profiling using next-generation sequencing technologies. *Lab Invest, 94*(3), 350-358. doi:10.1038/labinvest.2013.157

Tang, G. (2005). siRNA and miRNA: an insight into RISCs. *Trends Biochem Sci, 30*(2), 106-114. doi:10.1016/j.tibs.2004.12.007

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., . . . Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science, 347*(6220), 1260419. doi:10.1126/science.1260419

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., . . . Ponten, F. (2017). A pathology atlas of the human cancer transcriptome. *Science, 357*(6352). doi:10.1126/science.aan2507

Valen, E., Preker, P., Andersen, P. R., Zhao, X., Chen, Y., Ender, C., . . . Jensen, T. H. (2011). Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat Struct Mol Biol, 18*(9), 1075-1082. doi:10.1038/nsmb.2091

Varkonyi-Gasic, E., Wu, R., Wood, M., Walton, E. F., & Hellens, R. P. (2007). Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods, 3*, 12. doi:10.1186/1746-4811-3-12

Wang, W. T., Han, C., Sun, Y. M., Chen, T. Q., & Chen, Y. Q. (2019). Noncoding RNAs in cancer therapy resistance and targeted drug development. *J Hematol Oncol, 12*(1), 55. doi:10.1186/s13045-019-0748-z

Wen, J., Ladewig, E., Shenker, S., Mohammed, J., & Lai, E. C. (2015). Analysis of Nearly One Thousand Mammalian Mirtrons Reveals Novel Features of Dicer Substrates. *PLOS Computational Biology, 11*(9), e1004441. doi:10.1371/journal.pcbi.1004441

Wen, J., Ladewig, E., Shenker, S., Mohammed, J., & Lai, E. C. (2015). Analysis of Nearly One Thousand Mammalian Mirtrons Reveals Novel Features of Dicer Substrates. *PLoS Comput Biol, 11*(9), e1004441. doi:10.1371/journal.pcbi.1004441

Westholm, J. O., Ladewig, E., Okamura, K., Robine, N., & Lai, E. C. (2012). Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. *RNA, 18*(2), 177-192. doi:10.1261/rna.030627.111

Westholm, J. O., Ladewig, E., Okamura, K., Robine, N., & Lai, E. C. (2012). Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. *RNA (New York, N.Y.), 18*(2), 177-192. doi:10.1261/rna.030627.111

Westholm, J. O., & Lai, E. C. (2011). Mirtrons: microRNA biogenesis via splicing. *Biochimie, 93*(11), 1897-1904. doi:10.1016/j.biochi.2011.06.017

Wightman, B., Ha, I., & Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell, 75*(5), 855-862. doi:10.1016/0092-8674(93)90530-4

Willkomm, D. K., & Hartmann, R. K. (2007). An important piece of the RNase P jigsaw solved. *Trends Biochem Sci, 32*(6), 247-250. doi:10.1016/j.tibs.2007.04.005

Witten, D., Tibshirani, R., Gu, S. G., Fire, A., & Lui, W. O. (2010). Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. *BMC Biol, 8*, 58. doi:10.1186/1741-7007-8-58

Woese, C. R., Olsen, G. J., Ibba, M., & Soll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev, 64*(1), 202-236. doi:10.1128/MMBR.64.1.202-236.2000

Xiong, P., Schneider, R. F., Hulsey, C. D., Meyer, A., & Franchini, P. (2019). Conservation and novelty in the microRNA genomic landscape of hyperdiverse cichlid fishes. *Sci Rep, 9*(1), 13848. doi:10.1038/s41598-019-50124-0

Xu, G., Zhang, Y., Wei, J., Jia, W., Ge, Z., Zhang, Z., & Liu, X. (2013). MicroRNA-21 promotes hepatocellular carcinoma HepG2 cell proliferation through repression of mitogen-activated protein kinase-kinase 3. *BMC Cancer, 13*, 469. doi:10.1186/1471-2407-13-469

Yang, A., Bofill-De Ros, X., Shao, T. J., Jiang, M., Li, K., Villanueva, P., . . . Gu, S. (2019). 3' Uridylation Confers miRNAs with Non-canonical Target Repertoires. *Mol Cell, 75*(3), 511-522 e514. doi:10.1016/j.molcel.2019.05.014

Yashiro, Y., & Tomita, K. (2018). Function and Regulation of Human Terminal Uridylyltransferases. *Front Genet, 9*, 538. doi:10.3389/fgene.2018.00538

Yip, M. C. J., Savickas, S., Gygi, S. P., & Shao, S. (2020). ELAC1 Repairs tRNAs Cleaved during Ribosome-Associated Quality Control. *Cell Rep, 30*(7), 2106-2114 e2105. doi:10.1016/j.celrep.2020.01.082

Yoshikawa, M., & Fujii, Y. R. (2016). Human Ribosomal RNA-Derived Resident MicroRNAs as the Transmitter of Information upon the Cytoplasmic Cancer Stress. *Biomed Res Int, 2016*, 7562085. doi:10.1155/2016/7562085

Yu, M., Lu, B., Zhang, J., Ding, J., Liu, P., & Lu, Y. (2020). tRNA-derived RNA fragments in cancer: current status and future perspectives. *J Hematol Oncol, 13*(1), 121. doi:10.1186/s13045-020-00955-6

Zhang, P., Wu, W., Chen, Q., & Chen, M. (2019). Non-Coding RNAs and their Integrated Networks. *J Integr Bioinform, 16*(3). doi:10.1515/jib-2019-0027

Zhang, X., & Zeng, Y. (2010). The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res, 38*(21), 7689-7697. doi:10.1093/nar/gkq645

Zhang, Y., Qian, H., He, J., & Gao, W. (2020). Mechanisms of tRNA-derived fragments and tRNA halves in cancer treatment resistance. *Biomarker Research, 8*(1), 52. doi:10.1186/s40364-020-00233-0

Zhang, Z., Pi, J., Zou, D., Wang, X., Xu, J., Yu, S., . . . Yu, J. (2019). microRNA arm-imbalance in part from complementary targets mediated decay promotes gastric cancer progression. *Nat Commun, 10*(1), 4397. doi:10.1038/s41467-019-12292-5

Zhou, J., Cheng, Y., Yin, M., Yang, E., Gong, W., Liu, C., . . . Zhang, Y. (2015). Identification of novel miRNAs and miRNA expression profiling in wheat hybrid necrosis. *PLoS One, 10*(2), e0117507. doi:10.1371/journal.pone.0117507

Zhou, L., Lim, M. Y. T., Kaur, P., Saj, A., Bortolamiol-Becet, D., Gopal, V., . . .
Okamura, K. (2018). Importance of miRNA stability and alternative primary
miRNA isoforms in gene regulation during Drosophila development. *Elife, 7*.
doi:10.7554/eLife.38389

Zia, M. F., & Flynt, A. S. (2018). Detection and Verification of Mammalian Mirtrons by
Northern Blotting. *Methods Mol Biol, 1823*, 209-219. doi:10.1007/978-1-4939-
8624-8_16