Identification of a Novel microRNA Involved in Non-Specific Binding to a Decoy Transcript

by KATHRYN JOHANNA SLOWSKI B.Sc., The University of Saskatchewan, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY in THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES (Interdisciplinary Oncology)

THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver)

November 2016 © Kathryn Johanna Slowski, 2016

Abstract

MicroRNAs are known to be upregulated or downregulated in various types of cancer, leading to changes in the expression of genes involved in cellular proliferation, anti-apoptosis, migration, and invasion. To study the effects of microRNA loss or gain in different neoplasms, numerous models have been described to decrease or increase expression of microRNAs, but the offtarget effects of different methods have not been well investigated. I investigated the possibility of off-target effects in a model of miR-143 knockdown in myeloid leukemia cell lines that implemented a microRNA sponge, or decoy, as a method to reduce microRNA expression. The high expression of a sponge with repetitive sequence elements and low expression of the intended microRNA for knockdown, miR-143, created conditions with increased potential for non-specific microRNAs to bind to the sponge. Therefore, I investigated the potential binding sites present in the sponge and whether any novel microRNAs could bind to these sites. I found a number of potential candidates and eliminated them based on their likelihood of regulating protein targets and their resemblance to a microRNA in structure, leaving one potential candidate. I found genomic evidence of the existence of this novel microRNA, evolutionary conservation of function, and performed assays that confirmed the biological activity. Next, the original sponge was redesigned to inhibit the binding of the potential non-specific microRNA; miR-X, or the miR-143 binding sites were mutated to inhibit the binding of miR-143 and capture miR-X instead. This demonstrated that binding of non-specific microRNA could be abrogated and differential protein abundance specific to the knockdown of each microRNA separately was verified. I conclude that non-specific binding to the sponge is a distinct possibility in experiments using this method of microRNA knockdown, which needs to be taken into account when designing sponges in the future. This work also demonstrates that there remain novel microRNAs awaiting discovery.

Preface

Chapters 3 and 4 are based on work conducted in Aly Karsan's laboratory in the BC Cancer Research Centre. Laboratory work was conducted under the biosafety certification B13-0029 and ethical approval for working with clinical sequencing data was obtained by human ethics certification H13-02687. I completed the "Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans" Course on Research Ethics (TCPS 2: CORE). The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at http://cancergenome.nih.gov.

The proteomics samples for SILAC1 and SILAC2 datasets were prepared in Juergen Kast's laboratory and ran on the FT-ICR at UBC's Biomedical Research Centre by Nick Stoynov. I was responsible for the preparation of samples before analysis on the FT-ICR, processing of proteomics data, and all downstream testing. The proteomics samples for Selected Reaction Monitoring were prepared by myself and analyzed by LC/MS/MS by Vincent Chen. The proteomics samples in Chapter 5 were made by myself and prepared for mass spectrometry analysis, and analyzed, by Christopher Hughes and Shane Colburne of Gregg Morin's laboratory at the BC Cancer Research Centre. The Empirical Bayes analysis R script was written by Habil Zare during his PhD candidacy in Ryan Brinkman's laboratory at the BC Cancer Research Centre and I applied it to the SILAC1 and SILAC2 data from Chapter 3 as well as the proteomics data from Chapter 5. I performed filtering and correlation analysis of the proteomics dataset and the limma package from Bioconducter developed by Smyth *et al* (Smyth, Ritchie *et al* 2015) was adapted and used by Eva Yap for performing the differential expression analysis.

Table of Contents

Abstract		ii
Preface.		iii
Table of	Contents	iv
List of Ta	ables	vii
List of Fi	gures	viii
List of A	- bbreviations	x
Acknowl	edgements	xiv
Dedicatio	2n	
1. Intro	duction	
1.1 0	verview	1
1.2 m	icroBNA	
1.2.1	RNA Structural Features	2
1.2.2	Discovery and Conservation of RNAi in Other Species	3
1.2.3	Different Forms of Small Non-Coding RNA	6
1.2.4	Biogenesis of microRNA	7
1.2.5	microRNA Target Binding	11
1.2.6	microRNA Targeting and Regulation	
1.2.7	Regulation of microRNA Expression	21
1.2.8	The Role of microRNAs in Cancer	24
1.3 m	icroRNAs in Hematopoiesis and Blood Cancers	
1.3.1	Hematopoiesis	
1.3.2	Role of microRNAs in Differentiation of Blood Cells	
1.3.3	Dysregulation of microRNAs in Blood Cancers	34
1.3.4	Clinical Features and Subtypes of Myelodysplastic Syndromes	
1.3.5	The Role of Genetic Abnormalities in MDS Pathogenesis	
1.3.6	Pathogenesis of the del(5q) MDS Subtype	
1.3.7	microRNAs in Myelodysplastic Syndromes	
1.4 M	odeling the Role of a microRNA in Disease	
1.4.1	Approaches to Novel microRNA Discovery	
1.4.2	Methods of microRNA Study or Alteration	

	1.4	1.3 microRNA Knockdown Model Leading to Investigation of Non-specific Binding and	
	No	vel microRNA Discovery	51
2.	Ма	aterials and Methods	52
2	2.1	Cell Culture	52
2	2.2	Cloning of Constructs	53
2	2.3	Lentivirus Production	56
2	2.4	Reverse Transcription of RNA and Real-Time Quantitative PCR	56
	2.4	4.1 Exiqon RT-qPCR for microRNA	56
	2.4	4.2 TaqMan RT-qPCR for microRNA	57
2	2.5	Immunoblotting	58
2	2.6	Quantitative Proteomics Methodology and Statistical Analyses	59
	2.6	5.1 Stable Isotope Labelling of Amino Acids in Cell Culture and Gene Transfer	59
	2.6	3.2 Analysis of Probability of Significant Change in Protein Expression	60
	2.6	3.3 Selective Reaction Monitoring	61
	2.6	6.4 Preparation, Collection and Analysis of Proteomics Data with Modified Sponges	61
2	2.7	Flow Cytometry	63
2	2.8	Transfection of microRNA Anti-Sense Oligonucleotide Inhibitors and microRNA	
I	Mimi	ics	64
2	2.9	Luciferase Assays	65
2	2.10	Multiple Sequence Alignment and Target Prediction	65
2	2.11	Statistical Tests	65
3.	Mo	odel of miR-143 Knockdown in Blood Cancer Cell Lines	
			67
•	3.1	Introduction	67 67
:	3.1 3.2	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis	67 67 68
	3.1 3.2 3.3	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset	67 67 68 74
	3.1 3.2 3.3 3.4	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset Correlation Between SILAC Datasets and Degree of Changes Due to miR-143	67 67 68 74
	3.1 3.2 3.3 3.4 Loss	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset Correlation Between SILAC Datasets and Degree of Changes Due to miR-143	67 67 68 74 80
4.	3.1 3.2 3.3 3.4 Loss Mo	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset Correlation Between SILAC Datasets and Degree of Changes Due to miR-143	67 67 68 74 80 83
4.	3.1 3.2 3.3 3.4 Loss Ma 4.1	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset Correlation Between SILAC Datasets and Degree of Changes Due to miR-143 colecular Analysis of a Potential Novel microRNA	 67 67 68 74 80 83 83
4.	3.1 3.2 3.3 3.4 Loss Ma 4.1 4.2	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset Correlation Between SILAC Datasets and Degree of Changes Due to miR-143 colecular Analysis of a Potential Novel microRNA Introduction	67 67 68 74 80 83 83
4. 4.	3.1 3.2 3.3 Loss Ma 4.1 4.2 sequ	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset Correlation Between SILAC Datasets and Degree of Changes Due to miR-143 colecular Analysis of a Potential Novel microRNA	67 67 68 74 80 83 83 83
4. 2	3.1 3.2 3.3 3.4 Loss Ma 4.1 4.2 sequ 4.3	Introduction Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis Analysis of Significant Biological Changes in Quantitative Proteomics Dataset Correlation Between SILAC Datasets and Degree of Changes Due to miR-143 colecular Analysis of a Potential Novel microRNA	67 67 68 74 80 83 83 83

4.4	Genomic Analysis of Novel microRNA Transcripts Containing Potential Seeds 97
4.5	Biological Activity of miR-X113
4.6	Conclusions117
5. In	proving the Design of the Sponge Constructs for microRNA Knockdown119
5.1	Introduction119
5.2	Regulation of Sponge in Luciferase Reporter Vectors119
5.3	Proteomic Analysis of miR-143 or miR-X Knockdown Using Modified Sponges 121
5.4	Determination of Significant Expression Changes by Differential Expression
Ana	lysis124
5.5	Regulation of Protein Expression Specific to miR-X or miR-143 Knockdown 130
5.6	Conclusions137
6. D	iscussion, Conclusions, and Future Directions
6. D	iscussion, Conclusions, and Future Directions
6. D 6.1 6.2	iscussion, Conclusions, and Future Directions
6. D 6.1 6.2 6.3	iscussion, Conclusions, and Future Directions
 6. D 6.1 6.2 6.3 6.4 	iscussion, Conclusions, and Future Directions
 6. D 6.1 6.2 6.3 6.4 6.5 	iscussion, Conclusions, and Future Directions
 6. D 6.1 6.2 6.3 6.4 6.5 6.6 	iscussion, Conclusions, and Future Directions
 6. 6.1 6.2 6.3 6.4 6.5 6.6 6.7 	iscussion, Conclusions, and Future Directions
 6. D 6.1 6.2 6.3 6.4 6.5 6.6 6.7 6.8 	iscussion, Conclusions, and Future Directions
6. D 6.1 6.2 6.3 6.4 6.5 6.6 6.7 6.8 Refer	iscussion, Conclusions, and Future Directions

List of Tables

Table 2.1 - Cell lines used and culture conditions	52
Table 4.1 - Potential seed sites included in novel microRNA transcripts	
Table 4.2 - Hairpin structures of novel transcripts containing potential seeds	
Table 4.3 - Potential seeds in novel transcripts with microRNA hairpin structures	
Table 4.4 - Predicted and observed targets of potential seeds in SILAC1	90
Table 4.5 - Genomic locations of novel microRNA transcripts with potential seed #5	93
Table 4.6 - Percentages of human sequence identity in sequences aligning to miR	-X in other
species	
Table 4.7 - Nucleotide changes in the hairpin stem of miR-X	109
Table 4.8 - Observed vs. expected retained base pairing in nucleotide substitution	111
Table A1 - Discovery of Repetitive Sequences within Sponge	

List of Figures

Figure 1.1 - Biogenesis of microRNAs	8
Figure 1.2 - DROSHA and DICER cleavage of primary microRNA and microRNA precursors	s 10
Figure 1.3 - RISC loading	12
Figure 1.4 - Types of microRNA binding to mRNA targets	15
Figure 1.5 - Mechanism of microRNA regulated translational repression	18
Figure 1.6 - Mechanisms of mRNA degradation	19
Figure 1.7 - Regulation of DROSHA activity	23
Figure 1.8 - Simplified hematopoiesis	30
Figure 1.9 - microRNAs involved in hematopoiesis	33
Figure 1.10 - Consequences of loss of genetic material from chromosome 5	43
Figure 1.11 - Methods of microRNA study	47
Figure 2.1 - TaqMan RT-qPCR schematic	58
Figure 3.1 - Knockdown of miR-143 expression in UT-7	71
Figure 3.2 - Overlap and correlation between two miR-143KD proteomics experiments	73
Figure 3.3 - Threshold for determining significant changes in protein expression	76
Figure 3.4 - Overlapping expression of significantly changed proteins	77
Figure.3.5 - Validation of significantly differentially expressed proteins	79
Figure 3.6 - Expression levels of miR-143 in leukemic cell lines	80
Figure 4.1 - Identification of potential non-specific microRNA recognition elements in miR	-143
sponge	85
Figure 4.2 - Potential seeds with higher ratio of observed to predicted targets than random	92
Figure 4.3 - microRNA hairpin structure of Chr12 and Chr14 novel microRNA	96
Figure 4.4 - Expression of novel microRNA in AML patient small RNA-Seq libraries	99
Figure 4.5 - Detection of miR-X in other cell lines and tissues	. 101
Figure 4.6 - Conservation of miR-X sequence	. 102
Figure 4.7 - Multiple sequence alignment of miR-X hairpin and flanking regions	. 106
Figure 4.8 - Structure and putative promoter site for miR-X gene	. 107
Figure 4.9 - Overlap of predicted targets of miR-X in conserved species	. 108
Figure 4.10 - Assigning of number identities to base pairs in miR-X hairpin stem	. 109
Figure 4.11 - Cloning strategy for miR-X hairpin and reverse complement control	. 112
Figure 4.12 - Expression of mature miR-X from lentiviral hairpin overexpression	. 113
Figure 4.13 - Derepression of miR-X predicted targets observed by quantitative proteomics.	. 114

Figure 4.14 - miR-X inhibition leads to derepression of miR-X binding sites
Figure 4.15 - Inhibition of miR-X and delivery of miR-X mimics regulates predicted targets 117
Figure 5.1 - Regulation of miR-143 and miR-X specific sponges by microRNA mimics
Figure 5.2 - Modification of original sponge to achieve knockdown of specific microRNA 122
Figure 5.3 - Schematic of differential expression proteomics experiment
Figure 5.4 - Correlations of proteomics samples for SILAC differential expression analysis 126
Figure 5.5 - Linear regression analysis to find differentially expressed proteins 127
Figure 5.6 - Differentially expressed proteins between the three sponge conditions 129
Figure 5.7 - Statistical analyses applied to proteomics dataset to find expression patterns 131
Figure 5.8 - Overlapping significantly changed proteins between pLL-Orig-Spg and miR-
Xspecific-spg
Figure 5.9 - Overlapping significantly changed proteins between pLL-Orig-Spg and miR-
143specific-spg
Figure 5.10 - Overlapping expression changes in significantly differentially expressed proteins
Figure A.1 - Empirical Bayes analysis of modified sponge samples
Figure A.2 - Empirical Bayes analysis of modified sponge samples
Figure A.3 - Empirical Bayes analysis of modified sponge samples

List of Abbreviations

ADAR	adenosine deaminases
AGO	Argonaute protein
AKT	serine-threonine protein kinase
Alu	Arthrobacter luteus element
AML	acute myeloid leukemia
AML1-ETO	acute myeloid leukemia 1 - eight:twenty-one translocation
AMO	anti-microRNA oligonucleotide
ASO	anti-sense oligonucleotide
ATCC	American Type Culture Collection
ATP	adenosine tri-phosphate
AUF1	AU-Rich Element (ARE) RNA Binding Protein 1
BCL-2	B-cell lymphoma protein
BCR-ABL	"breakpoint cluster region"- Abelson murine leukemia viral oncogene homolog1
BIM	BCL-2 interacting protein
BM	bone marrow
CAF1	chromatin assembly factor 1
CCR4	chemokine (C-C motif) receptor 4
CDR	critically deleted region
CEBPA	CCAAT/Enhancer Binding Protein (C/EBP), alpha
CEBPB	CCAAT/Enhancer Binding Protein (C/EBP), Beta
chr 12	chromosome 12
chr 14	chromosome 14
c-KIT(SCFR)	tyrosine-protein kinase kit (Mast/stem cell growth factor receptor)
CLL	chronic lymphocytic leukemia
CLL	chronic lymphocytic leukemia
CLP	common lymphoid progenitors
CMP	common myeloid progenitors
CSF1R	Colony stimulating factor 1 receptor
CSNK1A1	casein kinase 1, alpha 1
CTNNA1	alpha catennin
DCP	decapping protein 1
DDX6	DEAD-box helicase 6

DEAD-box	aspartate-glutatmate-alanine-aspartate-box
DGCR8	DiGeorge syndrome chromosomal region 8
DICER1	double-stranded RNA endoribonuclease
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
DNMT3A	DNA methyl-transferase member 3A
DNMT3B	DNA methyl-transferase 3B
DROSHA	Ribonuclease 3 protein
DSMZ	Deutsche Sammlun von Mikrooganismen und Zellkulturen
dsRBD	double-stranded RNA binding domain
dsRNA	double-stranded RNA
E2F1	E2F Transcription Factor 1
EGR1	early growth response 1
elF4Al	eukaryotic initiation factor 4A1
EZH2	enhancer of zeste homolog 2
FACS	fluorescence automated cell sorting
FBS	fetal bovine serum
FISH	fluorescent in situ hybridization
FLT3-ITD	Fms-like tyrosine kinase 3 - internal tandem duplications
FOXO3	Forkhead box O3
GATA1	GATA binding factor 1
GFP	green fluorescent protein
GLRA1	glycine receptor, subunit alpha 1
GMP	granulocyte monocyte progenitor
GSC	Genome Sciences Centre
GW182	glycine-tryptophan 182 kDa protein
HDAC1	histone deacetylase 1
HG	Hoogsten
hGM-CSF	human Granulocytic-Megakaryocytic-Cytokine Stimulating Factor
hnRNP	heterogenous ribonucleoprotein
HOXA1, etc	homeobox A1
HSC	hematopoietic stem cell
HSC70	heat shock cognate 71 kDa protein
HSC90	heat shock protein 90 cognate

HSPC	hematopoietic stem progenitor cell
IDH1, IDH2	isocitrate dehyrogenase 1, 2
IDT	Integrated DNA Technologies
IPSS	International Prognostic Scoring System
JAK2	janus kinase 2
LIN28	lineage protein 28
LNA	locked nucleic acid
MAFB	V-Maf Avian Musculoaponeurotic Fibrosarcoma Oncogene Homolog B
MAPK	mitogen-activated protein kinase
MDS	myelodysplastic syndromes
MFI	mean fluorescence intensity
MID	middle domain
MLL	mixed lineage leukemia
MPL	MPL proto-oncogene, thrombopoietin receptor
MPP	multipotent progenitor
MRE	microRNA recognition element
mRNA	messenger RNA
MSA	multiple sequence alignment
MYC	Myelocytomatosis oncogene cellular homolog
NOT1	negative regulator of transcription 1
PABPC	polyadenylate tail binding protein C
PACT	Protein Activator of PKR (protein kinase R)
PAM2	PABP-interacting motif 2
PAN	Proteasome-activating nucleotidase
PAZ	Piwi Argonaute Zwille domain
PBS	phosphate-buffered saline
PDCD	pyruvate dehydrogenase complex deficiency
PI3K	phosphoinositide-3-kinase
PIWI	P-element induced wimpy testis
PSMs	peptide spectrum matches
PTEN	phophatase and tensin homolog
PU.1	PU box binding protein
RAS	Rat sarcoma
RBP	RNA binding protein

RISC	RNA induced silencing complex
RNA	ribonucleic acid
RPFs	ribosome protected fragments
RPS14	ribosomal protein S14
RT-qPCR	reverse-transcription quantitative polymerase chain reaction
RUNX1	RUNT-related transcription factor
SAMSN1	SAM-domain containing protein
SF3B1	splicing factor 3b, subunit 1
SG	shallow groove
SHIP1	SH2-containing inositol-5'-phosphatase 1
SILAC	stable isotope labeling of amino acids in cell culture
siRNA	small interfering RNA
SMAD	S mothers against decapentaplegic
SMARCA4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin,
	Subfamily A, Member 4
SP1	specificity protein 1
SPARC	secreted protein, acid, cysteine-rich
stRNA	small temporal RNA
TCGA	The Cancer Genome Atlas
TCL1	T-cell leukemia/lymphoma protein 1
TEL-AML1	Telomere length regulation protein - acute myeloid leukemia protein 1
TIRAP	toll-interleukin 1 receptor domain-containing adaptor protein
TLR	toll-like receptor
TP53	tumour protein p53
TRAF6	TNF receptor-associated factor 6
TRBP	TAR (trans-Activation Response) RNA Binding Protein
tRNA	transfer RNA
TUTase	Terminal Uridylyl Transferase
UCSC	University of California Santa Cruz
UTR	untranslated region
WC	Watson-Crick
WT1	Wilms tumor protein
XRN1	5'-3' Exoribonuclease 1
ZAP-70	Zeta-chain-associated protein kinase 70

Acknowledgements

I would like to express my deepest gratitude to my supervisor and the members of my lab past and present for helping me to develop as a scientist, for the incredible opportunities I have had in research throughout my PhD, and for supporting me in working through difficult concepts and ideas throughout this endeavor. I owe many thanks to my supervisory committee members for their advice; for questioning and challenging me, to the technicians and staff scientists who have assisted me, and to the numerous inspiring teachers and professors I have been fortunate to have learned from throughout the years.

Dedication

To my family, who have always been my teachers.

To my sister, for keeping me on my toes, and because she could teach Winston Churchill a lesson in stubborn determination.

To my mom, for sharing with me her passion for research and pursuit of the facts, and for inciting in me a love of reading and learning.

To my dad, for always believing in me, teaching me to always hustle and be versatile, and because his illness and death motivated me to study cancer and contribute my skills in research to this field.

1. Introduction

1.1 Overview

The small non-coding RNAs known as microRNAs have emerged as one of the primary forms for regulation of gene expression during the past decade and a half (Friedman et al., 2009; Guo et al., 2010; Lim et al., 2005). MicroRNAs regulate most mammalian genes through binding to the 3' untranslated region (UTR) of a messenger RNA transcript and inhibiting translation or promoting deadenylation, decapping and degradation (Chen et al., 2009; Eulalio et al., 2009b; Nishihara et al., 2013; Vella et al., 2004). A single microRNA may regulate the expression of hundreds of different genes and varies in expression in different organs or tissues. The targets of a microRNA also vary depending on tissue and as a result microRNA play an essential role in a broad range of biological functions (Eichhorn et al., 2014; Friedman et al., 2009).

The process of hematopoiesis is the production and differentiation of blood cells, which gives rise to the immune system and protection against disease, enables the exchange of oxygen and carbon dioxide in all of the body's tissues, and allows communication throughout the body (Simmons, 1997). There are a wide variety of blood cells carrying out tasks, each with distinct morphologies and functions (Simmons, 1997). In aberrant hematopoiesis and hematological malignancies, the microRNAs controlling self-renewal and differentiation are often dysregulated and contribute to the pathogenesis of the malignancy (Havelange and Garzon, 2010). The loss or gain of microRNA contributes to the etiology of many types of cancer, and can disrupt gene regulation such that oncogenic pathways are upregulated and tumour suppressive pathways are downregulated (Lu et al., 2005). In the most common subtype of myelodysplastic syndromes (MDS), del(5q) MDS, a 1.5 Mb portion of the long arm of chromosome 5 is deleted on one allele, and multiple microRNAs in the deleted region are decreased in expression (Boultwood et al., 2002; Starczynowski et al., 2010). Several features of the disease have been attributed to significantly diminished expression of microRNAs miR-143, miR-145, and miR-146a in del(5g) MDS patients, compared to MDS patients with normal karyotype (Ebert et al., 2008; Starczynowski et al., 2010).

Although it demonstrates the greatest decrease in expression in del(5q) among the three microRNAs, miR-143 has not been linked to a specific mechanism behind the clinical features of del(5q) MDS. However, it has been reported to behave as a tumour suppressor in certain

cancers. The following work began with examining the effect on the proteome following the loss of miR-143, and the development of a model of microRNA knockdown. The results raised questions about specificity of the sponge method of knockdown, as proteins were identified that changed expression in the presence of the sponge but there were few proteins undergoing consistent changes in the proteomic dataset. Given the low expression of miR-143 in the model and the high expression of a sponge transcript containing repetitive elements, there were many potential binding sites for any novel or annotated microRNA that could bind to a repetitive sequence within the sponge and regulating its own distinct set of targets. Therefore, I investigated what potential binding sites might be present in the sponge and whether any novel microRNAs might exist that could bind to these sites. I evaluated the genomic evidence and assayed the functional activity of a candidate microRNA after narrowing down a pool of potential novel microRNAs. Lastly, I sought to improve the sponge design and showed that more specific knockdown of the target microRNA can occur if 3 or 4 random non-seed nucleotides are uniquely mutated in each tandem repeat of the sponge.

1.2 microRNA

1.2.1 RNA Structural Features

Ribonucleic acid (RNA) is one of the four main categories of macromolecules in the cell: proteins, lipids, carbohydrates, and nucleic acids (Cooper and Hausman, 2007). RNA is a versatile macromolecule that carries the genetic code of the cell as messenger RNA (mRNA), which is transcribed from the genes encoded in deoxyribonucleic acid (DNA) and translated into their protein products (Cooper and Hausman, 2007). It also forms secondary or tertiary structures that incorporate with protein complexes, such as spliceosomes and ribosomes (Pikielny and Rosbash, 1986). RNA differs from DNA by one extra hydroxyl group at the 2' position on the five-member carbon ring and although RNA is a nucleic acid with four nitrogenous bases like DNA, it contains uracil instead of thymine (Markham and Smith, 1951). In RNA and DNA, planar interactions between the bases of the nucleotides are formed from two or three hydrogen bonds. The most common base pairings in RNA are Watson-Crick (WC) base pairings, A to U, G to C and the G-U wobble pair, though Hoogsteen (HG) base pairings are also infrequently found in RNA and DNA (Zwieb C et al 2014). RNA can also demonstrate base

pairing through the "sugar edge," which uses the 2' hydroxyl group of the ribose sugar and positions 2 and 3 of a purine base (Gorodkin et al., 2014; Grabow et al., 2013; Zwieb, 2014).

RNA is less stable than DNA because the hydroxyl group at the 2' position makes RNA more prone to hydrolysis. However, RNA forms many other complex secondary and tertiary structures, or combinations thereof. RNA structure can come from ordered stacking based on folding of the phosphodiester backbone or from RNA sequence motifs, which form 3D structures without Watson-Crick base pairing (Leontis and Westhof, 2003). RNA strands form various types of loops - hairpin, internal, and junction, for example - which can be seen in twodimensional representations of RNA structure (Leontis and Westhof, 2003). Three-dimensional structures of RNA are highly hierarchical and modular in nature, often using recurrent smaller structural motifs in bends or stacking to make larger structures. Secondary structure formation into a native structural state may be assisted by metal ions, increasing the similarity of RNA folding to protein folding (Grabow et al., 2013). Folding of RNA into well ordered secondary structures is found frequently and contributes to complex processes such as RNA splicing through spliceosomes, codon organization in translation through tRNA, and functioning as biological catalysts in ribozymes (Grabow et al., 2013; Leontis and Westhof, 2003). A multitude of structures created by RNA folding were observed as part of gene expression regulation even before the discovery of non-coding small RNA. The fold-back stem-loop structures in primary microRNA transcripts are well-ordered and can be statistically inferred based on thermodynamic stability (Lee et al., 2003a).

The function of RNA often relies on the proteins acting as binding partners to the RNA. RNA binding proteins (RBPs) are a common class of proteins in the cell, as RNA performs a broad variety of essential processes, and can help to stabilize RNA molecules folding into their native RNA structures. Proteins can behave as chaperones for some RNA structures and can facilitate RNA annealing and other RNA-RNA interactions, roles that are essential for many of the processes in the biogenesis and activity of microRNAs (Rajkowitsch et al., 2007).

1.2.2 Discovery and Conservation of RNAi in Other Species

The central dogma of biology describes the transcription of RNA from the DNA of the genome and the translation of the RNA into protein. However, in the mid-1990's the idea of posttranscriptional gene silencing emerged, where protein expression changed based on regulation by non-protein coding RNAs. Multiple types of non-coding RNA have been found in many species and forms, both before and since. The concept of RNA interference, or transcripts of anti-sense RNA which bind to protein-coding RNA, preceded the discovery of the posttranscriptional gene silencers in humans (Allison, 1972). Natural anti-sense transcripts were found to regulate biological processes in bacteria, and in retroviruses, anti-sense transcripts were found at genetic loci overlapping with the sense transcripts they regulated (Coleman et al., 1984; Mizuno et al., 1984; Vanhee-Brossollet et al., 1995). Later it was found that experimentally injected RNA could hybridize to endogenous messenger RNA and interfere with gene expression by an antisense mechanism (Fire and Xu, 1995). The technique of injecting double-stranded RNA (dsRNA) was used to characterize the function of genes in different pathways of Caenorhabditis elegans and Drosophila melanogaster by interfering with gene expression. DsRNA was used to interfere with genes in the wingless pathway of D. melanogaster (Kennerdell and Carthew, 1998) and also revealed the contribution of genes to essential biological processes, as in the study of the nautilus gene, where interference affected embryonic somatic muscle formation (Misquitta and Paterson, 1999). In plants, viral resistance and post-transcriptional gene silencing could be induced by transforming dsRNA into cells (Waterhouse et al., 1998).

The first microRNAs were discovered in *C. elegans*, and originally referred to as small temporal RNAs (stRNAs) because they were involved in regulation of cell fate and progression of development in a temporal manner. The first two microRNAs, *lin-4* and *let-7*, were expressed stage specifically and controlled developmental transitions (Lee and Ambros, 2001). Study of the *lin-4* gene in *C. elegans* showed two transcripts of 22 and 61 nt containing the *lin-4* sequence which did not encode a protein and demonstrated a hairpin RNA secondary structure in the 61 nt transcript. There were also complementary sequences to the *lin-4* transcripts in the 3'UTR of the *lin-14* mRNA, indicating potential anti-sense RNA-RNA regulation (Lee et al., 1993).

Other microRNAs were found using the features of *lin-4* and *let-7* as criteria - transcripts were sought with a length of around 22 nt, and precursors with a length of about 60-65 nt. Finding conserved regions of the genome between *C. elegans* and *C. briggsae* and checking for RNA secondary structures at these locations helped to determine many initial microRNA mature sequences and their hairpin precursor sequences in worms (Lee and Ambros, 2001).

Meanwhile Lagos-Quintana *et al* showed numerous microRNA mature and precursor transcripts were found in vertebrates and invertebrates, from *D. melanogaster*, *Xenopus laevis*, and *Danio rerio*, to human and mouse (Lagos-Quintana et al., 2001). These seminal discoveries prompted investigation into how many microRNAs existed and whether other non-coding RNA regulators could be found in human or other species.

One factor that supported the importance and prevalence of microRNA and post-transcriptional regulation by dsRNA across many different species was the ubiquity of the RNase III proteins of the Dicer family involved in cleavage of dsRNAs (Nicholson and Nicholson, 2002). The similarity of outcomes in microRNA function to RNA interference led to the finding that both pathways used the same Dicer and RNA-induced silencing complex (RISC) assembly proteins, but that siRNA appeared to use perfect complementarity to its mRNA target while the many of the microRNAs discovered initially did not demonstrate full complementarity with their targets (Elbashir et al., 2001; Hutvagner and Zamore, 2002). This finding expanded use of the RNAi method, from transfection of dsRNA to *in vitro* injection of synthesized short hairpin RNAs (shRNAs), as a technique to target specific genes and silence expression (Paddison et al., 2002).

MicroRNAs and the cellular machinery they employ are evolutionarily conserved throughout bilateria, and after their initial discovery in *C. elegans* they were found across numerous other phyla - chordates, hemichordates, echinoderms, mollusks, annelids, and anthropods. They are not found in basal metazoans, such as cnidarians and poriferans, and were originally thought to be absent in single-celled organisms (Pasquinelli et al., 2003). However, they were later found in single-celled algae, *Chlamydomonas reinhardtii*, and despite being found in smaller levels in simple organisms, it is likely that they played a role in gene regulation in early evolution (Molnar et al., 2007; Zhao et al., 2007a).

It was noted in the initial discovery of microRNA by different groups that the expression of a particular microRNA could be found in one cell type in a species but not in others (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). In a study identifying tissue-specific microRNA in the mouse, expression of one particular microRNA could dominate the microRNA population of a tissue, for instance, it was found that miR-1 accounted for 45% of all mouse microRNAs found in the heart, 72% of all cloned microRNAs in the liver consisted of miR-122 and variations, and miR-143 had the highest frequency in the spleen (Lagos-Quintana

et al., 2002). As well, expression of some microRNAs only occurs in one stage of cell differentiation while other microRNAs are expressed throughout development (Lagos-Quintana et al., 2002; Lau et al., 2001; Lee and Ambros, 2001).

1.2.3 Different Forms of Small Non-Coding RNA

Aside from microRNA, other types of non-coding RNA play critical roles in human cells. A survey of the human genome by the GENCODE project revealed that there are an estimated 9078 small non-coding RNA genes, only 3086 of which are microRNA (Hrdlickova et al., 2014). As discussed, many microRNAs are tissue-specific or temporally regulated, and certain subsets of microRNA are only expressed during a particular developmental stage. Other types of small non-coding RNAs, such as small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs), perform functions that are very separate and distinct from microRNA, but can be confused for microRNA in sequencing experiments due to their similar size.

One type of small non-coding RNA is the small nuclear RNAs family (snRNAs). The snRNAs are involved in splicing pre-mRNA, by making complexes with proteins termed small nuclear ribonucleic proteins (snRNPs). Five snRNPs form the spliceosome, which removes introns and joins exons from pre-mRNA to form mRNAs. Small nucleolar RNAs facilitate ribosome biogenesis and share a similar protein-binding component and structure as snRNAs (Peculis, 2000).

A small ncRNA family closely related to the snRNAs is the small nucleolar RNAs (snoRNAs) group. These are small RNAs involved in site-specific methylation and pseudouridylation of other RNAs in the nucleus, where they are localized (Holley and Topkara, 2011). SnoRNAs perform over 200 modifications on ribosomal RNA (rRNA), allowing it to be processed from the nucleus and form the ribosome complex with their ribosomal proteins counterparts (Filipowicz et al., 1999; Liang et al., 2009). Other snoRNAs are retained in nucleoplasmic domains called Cajal bodies, leading to the designation "scaRNAs," and guide the processing of spliceosome snRNAs. In higher organisms, almost all snoRNAs are encoded by introns and form complexes with proteins in the nucleus, small nucleolar RNA ribonucleoprotein (snoRNP) (Matera et al., 2007).

In the earliest stages of development, microRNAs exert an effect through the embryonic stem cell cycle (ESCC) microRNAs (Asikainen et al., 2015). These microRNAs sensitively regulate developmental gene expression and afford a greater degree of plasticity and robustness to cell fate determination (Asikainen et al., 2015). ESCC microRNAs primarily control G1 checkpoint regulation, and often share common seed sequences (Asikainen et al., 2015).

Another type of microRNA found in the embryonic stage is the microRNA-offset-RNAs (moRNAs) (Asikainen et al., 2015; Zhou et al., 2012). The function of the moRNAs still remains unknown, but they are located adjacent to mature microRNA sequences in the 3p and 5p arm of normal microRNA hairpins (Langenberger et al., 2009). Though they have low expression, they are present in higher fractions in hESCs and appear important in early stages of development (Shi et al., 2009). Interestingly, these microRNA are cleaved from a different arm of the hairpin than the canonical mature microRNA, and 5p arm moRNAs commonly come from a hairpin where the mature microRNA is on the 3p arm (Asikainen et al., 2015).

There are also piwi-interacting RNAs (piRNAs), which are longer, 27-30 nucleotides with a 5'terminal uridine. It was formerly thought that their expression was restricted to the germ line cells, since one class of piRNAs appears highly expressed at the pachytene stage of sperm development (Aravin et al., 2007; Brennecke et al., 2007). However, piRNAs can be expressed in different types of progenitors, such as cardiac cell progenitors, and play an important role in regulation of transposable elements in gametogenesis, embryogenesis, and stem cell maintenance (Toth et al., 2016; Watanabe and Lin, 2014).

1.2.4 Biogenesis of microRNA

The DICER1 and Argonaute proteins were identified as part of the biogenesis machinery shortly after the discovery of microRNA, but the full characterization of microRNA biogenesis remained to be elucidated and is still ongoing. The mechanisms at each stage of microRNA processing have been studied extensively in the past decade and a half. MicroRNAs are processed from primary-microRNA transcribed from DNA by RNA polymerase II (Lee et al., 2003b; Lee et al., 2002; Lee et al., 2004). An exception is the microRNAs transcribed from the C19BC cluster, which are interspersed between Alu repeats and are transcribed by RNA polymerase III.

Genomic analysis revealed that ~50 microRNAs are found in Alu element areas and may also be transcribed by RNA polymerase III (Borchert et al., 2006).

The primary microRNA transcript, or pri-miRNA, can be several kilobases long and contain a 3' polyadenylated tail as well as 5' 7-methyl guanylate (m7G) cap (Cai et al., 2004; Lee et al., 2004; Smalheiser, 2003). Lau NC *et al* originally noted that microRNAs found in the same genomic cluster appeared to be concordantly expressed (Lau et al., 2001) and Lee Y *et al* found that polycistronic microRNA in tandem in the genome were processed from a single transcript into multiple 70 nt microRNA precursors (Lee et al., 2002). Primary microRNA transcripts can function as both pri-miRNA and mRNA, as they share similar structural features, including polyadenylation and 5' capping. A transcript can produce both mRNA and microRNA, with the pre-microRNA cleaved from the 3'UTR or intronic regions of the transcript while the remaining transcript is processed as mRNA (Figure 1.1) (Cai et al., 2004).



Figure 1.1- Biogenesis of microRNAs.

In the nucleus, the primary microRNA transcript is transcribed by RNA polymerase II and cleaved into a precursor microRNA hairpin by the microprocessor complex composed of DROSHA and DGCR8. The precursor microRNA is transported to the cytoplasm by EXP5 and loop of the hairpin is cleaved by DICER, followed by loading and unwinding of the microRNA duplex in AGO. AGO-microRNA interacts with GW182 and other sets of proteins to form RISC and regulate bound mRNA targets.

The RNase III enzyme DROSHA cleaves primary microRNA transcripts into microRNA precursors, or pre-microRNA, in the nucleus with the interaction of the RNA binding protein DiGeorge syndrome chromosomal region 8 (DGCR8), which together with DROSHA makes the microprocessor complex (Figure 1.1) (Gregory et al., 2004; Han et al., 2004; Lee et al., 2003b). DROSHA and the DICER1 protein, which is involved at a later step in biogenesis, are both endoribonucleases of the RNase III family that bind and cleave dsRNA (Lamontagne et al., 2001). DGCR8 acts as a molecular ruler, interacting with and spacing DROSHA 11-bp from the base of the microRNA stem, positioning the two RNase domains of DROSHA to cleave both strands of the microRNA duplex. The typical structure of pri-miRNA includes the hairpin loop, stem, and base, a junction between the stem and flanking single strands of RNA upstream and downstream (Figure 1.2a). The average length of the hairpin stem is 33-bp, and the distance between the base of the stem and DROSHA cleavage is 11-bp or one helical turn of a nucleic acid duplex, which produces the ~22-bp RNA hairpin (Figure 1.2a) (Han et al., 2006; Zeng and Cullen, 2005; Zeng et al., 2005). Pre-microRNAs are not always produced by the microprocessor complex - some are created from introns and released from transcripts during the process of splicing (Berezikov et al., 2007; Okamura et al., 2007; Ruby et al., 2007). The microRNA precursors are exported from the nucleus by Exportin-5, a Ran-dependent nuclear transport receptor family member (Bohnsack et al., 2004; Kim, 2004; Lund et al., 2004; Yi et al., 2003). The protein recognizes pre-miRNA based on the structure of the stem, sensing duplex RNA of a defined length rather than the sequence or the hairpin loop structure, and uses the 1-4 nucleotide overhang at the 3' end to bind and export the precursor (Figure 1.1) (Lund and Dahlberg, 2006; Zeng and Cullen, 2004).

The pre-miRNA is subsequently processed by another RNase III enzyme, DICER1, which cleaves the loop of the hairpin and generates a 19-25 nucleotide duplex with two nucleotide overhangs at each of the 3' ends (Figure 1.2b) (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001). Two proteins, TRBP and PACT, have dsRNA binding domains and bind similarly and by mutual exclusion to DICER1 and the pre-miRNA. TRBP and PACT stabilize the DICER1-pre-microRNA complex during cleavage of the hairpin before loading of the microRNA duplex onto the RNA induced silencing complex (RISC) (Figure 1.1) (Chendrimada et al., 2005; Haase et al., 2005; Lee et al., 2006; Wilson et al., 2015).

a. Primary MicroRNA Transcript Cleavage



Figure 1.2 - DROSHA and DICER cleavage of primary microRNA and microRNA precursors

a. DROSHA and DGCR8 compose the microprocessor complex, which cleaves primary microRNA transcripts about 11nt or one helical turn from the bottom of the hairpin stem. b. DICER protein cleaves the precursor microRNA at the base of the apical loop with the two RNase III domains positioned at the top of the hairpin stem.

1.2.5 microRNA Target Binding

The function of microRNAs, to bind and regulate targets, begins with the process of RNA induced silencing complex (RISC) loading. The microRNAs associate with the Argonaute proteins to form the core of RISC, undergo duplex unwinding, and bind to mRNA targets. To mediate silencing and control expression of microRNA targets, AGO recruits other proteins after binding the target.

In the first stage of microRNA target binding, small RNA-duplex intermediates processed by DICER1 are taken up by one of four human Argonaute (AGO) proteins to form pre-RISCs (Kawamata et al., 2011). The microRNA duplexes are loaded into RISC in an ATP-dependent manner and require functional HSC70 and HSC90 chaperone machinery to incorporate into an AGO protein (Iki et al., 2010). The AGO protein family members consist of four conserved proteins domains, the terminal N domain, the carboxy-terminal PIWI domain with a RNase Hlike fold, the PAZ domain, and the MID domain in the middle of the protein (Figure 1.3). The MID domain contains a basic 5' binding pocket at its interface with the PIWI domain, which recognizes the 5' terminal phosphate group of the microRNA and structurally favors uracil in terminal nucleobase binding (Boland et al., 2011; Frank et al., 2010; Ma et al., 2005; Song et al., 2004; Wang et al., 2008a; Yuan et al., 2005). The PAZ domain anchors the 3' end of the microRNA in a hydrophobic cavity containing an oligonucleotide binding fold (Chen et al., 2008; Landthaler et al., 2008; Ma et al., 2005; Wang et al., 2008b). Each of the human AGO proteins regulate gene expression through mRNA degradation or translational repression of targets, however, unlike the other family members, the AGO2 protein retains catalytic RNase activity in the RNaseH fold of the PIWI domain and plays a role in RNAi as well (Jonas and Izaurralde, 2015).

Each of the four human AGO proteins is functionally equivalent at binding bulged microRNA duplexes and accept microRNA bulged duplexes with equal affinity, while AGO1 and AGO2 accept perfectly matched siRNAs with greater affinity than AGO3 and AGO4 (but only AGO2 has cleavage activity due to its RNaseH fold) (Gan and Gunsalus, 2013; Su et al., 2010). The strand of the microRNA duplex with a lower thermodynamic stability in the first 1-4 bases of the 5' end is selected from the duplex and incorporated while the other strand is degraded after unwinding (Khvorova et al., 2003; Schwarz et al., 2003). The incorporated strand, known as the guide strand, will later bind to target mRNAs. The unwinding of a microRNA duplex is driven by

the N-domain of Argonaute and does not require energy input (Kawamata et al., 2009; Kwak and Tomari, 2012). For duplexes that are highly complementary, AGO2 cleaves the phosphodiester bond of the passenger strand between nucleotides 10 and 11 of the guide strand, leading to instability and what is termed "slicer-dependent unwinding" (Leuschner et al., 2006; Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). For AGO1, 3, and 4, the mismatches in the microRNA duplex in the 3' region or mid region create instability and unwinding is independent of cleavage activity (Kawamata et al., 2009; Yoda et al., 2010).



Figure 1.3 - RISC loading

a. Argonaute proteins have four domains, the N-terminal domain (N), the PAZ domain which binds the 3' end of the guide microRNA, the MID domain, and the PIWI domain. The latter two form a hydrophilic interface that bind the 5' end of the guide microRNA, favouring uracil. b. MicroRNA duplexes are made by DICER and based on their complementarity (extensive, or nearly perfect, or imperfect) they are loaded into AGO2 or AGO1, 3, 4. The red part of the duplex symbolizes the seed sequence and the blue the rest of the microRNA duplex or hairpin. The cap structure is symbolized by a black dot.

Once the duplex is unwound and the strands separated, the AGO with a single-stranded small RNA bound is referred to as mature RISC, holo-RISC, or simply, RISC. In the Argonaute protein, microRNA binding the mRNA target involves a 2-step process and relies on a feature of the microRNA known as the "seed." Early studies in microRNA research showed that although microRNAs did not rely on perfect complementarity to inhibit their targets, the 5' end of the microRNA showed conserved pairing at nucleotide positions 2-7 or 2-8 of the sequence, a motif that became known as the "seed" site (Lewis et al., 2003). In RISC, the Argonaute protein anchors the phosphate backbone of the microRNA, ordering the conformation of the microRNA so that the bases in the seed region are exposed to solvent (Wang et al., 2008a; Wang et al., 2008b). This lowers the entropy of the binding and increases affinity for the mRNA target. The binding of the target sequence to the seed in the 5' region of the microRNA is the first step, followed by annealing to the 3' region of the microRNA (Cevec and Plavec, 2010; Cevec et al., 2008).

The microRNA seed can include a match at the 8th position, the 7mer-m8, which is the most common, or the target can have an A across from position one, the 7mer-1A site. Some pairings have both the match at the 8th position and the A in positon 1, making what is known as an 8mer site (Figure 1.4) (Bartel, 2009). In an *in vivo* study of microRNA targeting in *Drosophila*, the functional importance of the first eight nucleotides of the 5' end of the microRNA was confirmed (Brennecke et al., 2005b). In most microRNAs, multiple mismatches in the 3' end or disruption of base pairing outside the 5' conserved seed did not decrease the regulatory ability of the microRNA, and mismatches at position 1, 9, or 10 had little to no effect on repression (Brennecke et al., 2005b). The interaction between the 5' seed region and the microRNA's target depends on as few as four base pairs in positions 2-5 for effective target regulation. Interestingly in this study, complementarity between the target and bases 1-4 of the microRNA was ineffective in repressing expression of the target. Seed lengths of 4 nucleotides (4-mer), 5 nt, or 6 nt beginning at position 3 were also less effective (Brennecke et al., 2005b). This is likely due to the structural conformation the microRNA seed takes once bound to the Argonaute protein.

However, it was observed that some microRNA-target interactions have conserved pairing in the 3' region of the microRNA to increase the efficacy of binding to the target (Grimson et al., 2007). The conserved pairing between the 3' region of a microRNA and target is given the term "3'- supplementary site" and relies on Watson-Crick base pairing between nucleotides 13-16 of

the microRNA (Friedman et al., 2009). As well, about 5% of the 44,000 seed-target pairings analyzed in one study showed conserved pairing in the 3' region (Friedman et al., 2009). Some microRNAs rely on conserved base pairing in the 3' region to compensate for weaker interaction between the seed and target when single-nucleotide mismatches or bulges in the seed occur, since a bulge in the seed binding region would be less thermodynamically favourable (Friedman et al., 2009). In a more recent study, the physical base pairing interactions between microRNA and their targets were assessed using a crosslinking immunoprecipitation method and covalent ligation of endogenous Argonaute-bound RNAs (CLEAR-CLIP) (Moore et al., 2015). This study revealed that microRNAs can have many binding sites in introns as well as the 3'UTR and CDS binding sites. It was found that microRNA pairing uses seed-pairing as well as auxillary pairing in the non-seed regions (Moore et al., 2015). Instead of conflicting with past results, this may a more specific type of microRNA regulation. The non-seed pairing can differ between microRNA family members with shared seed sequences, and appear to give more distinct silencing functions, which means the 3' affects binding specificity in many cases as well (Moore et al., 2015).

Conserved microRNA Seed Binding to Target

7mer-1A		
5'-CDS	NNNNNNAPoly(A)-3'	mRNA target
3' - NNNNNNNNNNNN	NNNNNNN - 5' microRNA 87654321	
7mer-m8	Seed	
5'-CDS	NNNNNNNPoly(A)-3'	mRNA target
3' - NNNNNNNNNNN	NNNNNNN - 5' microRNA 87654321 Seed	
3' - NNNNNNNNNNNN	NNNNNNNAPoly(A)-3 IIIIIII NNNNNNN - 5' microRNA 87654321 Seed	mRNA target

Conserved 3' Region Involvement in microRNA Binding

3' supplementary site

3' compensatory site

Seed mismatch

5'-CDS......NNNN.......NNN^NNNN......Poly(A)-3' mRNA target 3' - NNNNNNNNNNNNNNNNNN - 5' microRNA 16151413 8 7 6 5 4 3 2 1 Seed

Figure 1.4 - Types of microRNA binding to mRNA targets

The seed of the microRNA is formed primarily by nucleotides 2-7 (7mer-m8), but variations of this exist. The AGO proteins favour a uracil base in the first position of the microRNA, so a corresponding A in the mRNA target is common (8-mer) or can compensate for no match at position 8 (7mer-1A). There are also more rarely cases where the 3' region of the microRNA has complementarity for compensating a mismatch in the seed region binding.

Cleavage by AGO2 is thought to proceed at a high catalytic rate, while "bulged" microRNA target sequences with imperfect complementarity proceed with a slower rate of catalysis (Cevec and Plavec, 2010; Cevec et al., 2008; Cevec et al., 2010).

1.2.6 microRNA Targeting and Regulation

In their initial discovery let-7 and lin-4 appeared to regulate the expression of protein without degradation of mRNA species, thus it was thought that microRNAs were primarily repressors of protein translation (Olsen and Ambros, 1999; Seggerson et al., 2002). However, later studies showed that for \geq 84% of proteins undergoing some degree of translational repression, mRNA destabilization occurred as well, leaving a smaller fraction of proteins regulated independently by microRNAs than previously thought (Guo et al., 2010). Genome-wide measurements of the effects of microRNAs on protein and mRNA levels, combined with ribosome profiling experiments, have shown that the degradation of mRNA targets accounts for 66-90% of microRNA mediated repression at steady state (Baek et al., 2008; Eichhorn et al., 2014; Guo et al., 2010; Hendrickson et al., 2009; Selbach et al., 2008; Subtelny et al., 2014). As well, the degree of protein expression change was not as large as that of the mRNAs, in one study, proteins showed reductions no larger than four-fold while the mRNA changes were considerably larger (Baek et al., 2008; Selbach et al., 2008). Thus, there has been revision of the original idea that microRNA regulation is primarily through translational inhibition, though it still occurs for a smaller proportion of microRNA targets.

Inhibition of gene expression can occur in a number of ways, and involves groups of proteins that seem to continually increase in size and complexity. The most studied and essential AGO partners are GW182 proteins - AGO-bound mRNA targets interact with effector complexes through the GW182 proteins. GW182 proteins co-purify with AGO proteins based on the strong interaction of the Argonaute PIWI domain to GW182. GW182 proteins function as flexible scaffolds and recruit other protein complexes to AGO that carry out deadenylation or destabilization of mRNA (Fabian and Sonenberg, 2012). GW182 proteins have two functional domains, the amino-terminal AGO-binding domain and the carboxy-terminal silencing domain (SD), which is predicted to be mainly unstructured and disordered. GW182 proteins feature multiple tryptophan (W) and glycine (G) containing motifs and are 182 kDa in size - hence the name GW182 (Eulalio et al., 2009a; Eulalio et al., 2009c; Eulalio et al., 2009d). AGO2 can cleave mRNA targets by catalytic RNase H activity, but the other three AGO proteins rely on cooperation with GW182 to recruit other protein cofactors and carry out mRNA degradation or translational inhibition. If the residues involved in GW182 interaction with AGO1 are mutated, the ability of AGO to regulate targets is abolished (Eulalio et al., 2008).

The proposed method of translational inhibition is through inhibiting the eIF4F complex association with the 43S ribosome subunit at the beginning of initiation. As the messenger RNA is transcribed from DNA, two structures are added to the mRNA, a poly-adenylate (polyA) tail at the 3' end and methyl-7-guanosine (m7G) cap at the 5' end (Colgan and Manley, 1997; Cowling, 2010). At the beginning of translation, the 43S pre-initiation complexes begin to associate, and the core initiation factors contain the eIF4F 5' cap-binding complex. This complex consists of three subunits - the scaffold protein eIF4G, the DEAD-box RNA helicase eIF4A that unwinds the 5' mRNA secondary structure, and the eIF4E protein that interacts with the 5' cap (Edery et al., 1983; Grifo et al., 1983; Sonenberg et al., 1979). The polyA binding protein (PABPC) interacts with eIF4G in the eIF4F complex and enhances translation through the circularization of mRNA (Gallie, 2014; Sachs and Varani, 2000). The circularization of mRNA stabilizes the eIF4E interaction with the cap and increases the rate of translational initiation (Figure 1.5) (Culjkovic et al., 2007; Kahvejian et al., 2005). Recent studies have determined that the block in initiation is due to the interference in assembly of eIF4F complex caused by microRNA RISC (miRISC) triggering dissociation of eIF4AI and eIF4AII (Figure 1.5) (Fukao et al., 2014; Fukaya et al., 2014). From there, the binding of the 40S subunit to the eIF4F complex may be hindered, or the initiation may be halted at the formation of the 80S ribosomal unit (Mathonnet et al., 2007).





The GW182 protein has large disordered regions with Glycine (G) and Tryptophan (W) residues that bind to AGO and other proteins through their W-binding pockets. PABPC interacts with the eIF4F cap-binding complex and brings its interaction partner, RISC, closer to the eIF4A protein that recruits the 40S subunit. The proximity of RISC interferes with assembly of the other ribosome subunits.

In the mRNA deadenylation, decapping, and degradation pathway, GW182 recruits complexes for different forms of inhibition. GW182 is essential to the silencing mechanism of microRNA because it binds through multiple W-containing motifs not only to AGO proteins, but to the deadenylase complexes, PAN2:PAN3 and CCR4:NOT1, which carry-out deadenylation and degradation of mRNA targets (Figure 1.6) (Behm-Ansmant et al., 2006; Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011). Crystal structures of these protein interactions show that the deadenylase complexes PAN2:PAN3 and CCR4:NOT1, as well as AGO, have hydrophobic W-binding pockets exposed on their surfaces, into which the W residues of GW182 are inserted (Chen et al., 2014; Christie et al., 2013; Mathys et al., 2014). The PABPC domain MLLE binds to the PABP-interacting motif 2 (PAM2) of GW182 and the C-terminal domain of PAPBC binds to a PAM2 in PAN3 of the PAN2:PAN3 complex (Figure 1.6) (Jinek et al., 2010; Kozlov et al., 2010; Siddiqui et al., 2007; Wolf and Passmore, 2014). Cleavage of the polyA tail is performed by PAN2 in the PAN2:PAN3 complex and by CCR4 and CAF1 of the CCR4:NOT1

complex, with PAN2 initiating deadenylation by cutting the tail to ~110nt and CCR4:NOT1 performing cleavage of the final 20-25 nucleotides (Figure 1.6) (Boeck et al., 1996; Brown and Sachs, 1998; Tucker et al., 2001; Wolf and Passmore, 2014; Yamashita et al., 2005). The mRNA can be degraded 3' to 5' by an exonuclease after deadenylation, or the decapping protein complex DCP1:DCP2 removes the 5' m7G cap and the mRNA is degraded then by XRN1, a 5' to 3' exonuclease (Figure 1.6) (Coller and Parker, 2004; Grosset et al., 2000; Yamashita et al., 2005). Chen and Mathys also found that DDX6 binds to the NOT1 MIF4G domain and interacts with the DCP1:DCP2 complex, providing the missing link between deadenylation and decapping (Chen et al., 2014; Mathys et al., 2014).



Figure 1.6 - Mechanisms of mRNA degradation

The mRNA is degraded by recruitment of deadenylation complexes to the polyA tail through GW182 and PABPC, and the mRNA species is decapped by the CCR4:NOT1 deadenylation complex interaction with the decapping complex DCP1:DCP2 through the DDX6 protein. Decay of the mRNA occurs 5' to 3' by the XRN1 nuclease.

In addition to a revision of the dominant mode of microRNA regulation, our understanding of certain microRNA rules have undergone significant changes since their inception over a decade ago. The number of genes regulated by an individual microRNA was initially estimated to be low. Lim LP *et al* calculated that 15-30% of the genome was regulated by microRNAs and that approximately 100 transcripts were controlled by one microRNA using microarray data from transfection of microRNA into HeLa cells (Lim et al., 2005). Development of microRNA target prediction algorithms based on seed-binding sequence motifs in the 3'UTR of mRNA transcripts provided larger estimates, predicting hundreds of targets per vertebrate microRNA (Krek et al., 2005). An overhaul of the computational tools used to find microRNA targets predicted over 60% of human protein-coding genes contained sequences pairing to miRNAs in their 3'UTRs (Friedman et al., 2009). However, computational prediction programs may overestimate the number of targets for a microRNA.

Another paradigm that has been altered in recent years is the binding position of the microRNA within its mRNA target and the base pairing involved in stabilizing this interaction. As previously discussed, the most common type of microRNA target regulation appeared to occur by the 5' seed region of the microRNA binding to sequence motifs in the 3'UTR of the mRNA target, but new evidence has include the non-seed regions as involved in binding as well (Lai, 2002). Further genomic analysis of seed pairing patterns found that microRNAs could have target sites in the coding domain sequence (CDS) and 5' untranslated region (5'UTR) of a transcript as well as the 3'UTR (Lewis et al., 2005). The microRNA binding sites in the CDS of the transcript appear to have less effective regulation of targets than binding sites in the 3'UTR, however (Selbach et al., 2008). The 5'UTR sites targeted by microRNAs often have a higher degree of local secondary structure in their 5'UTR (Gu et al., 2014). Additionally, microRNA binding sites in the 5'UTR have been shown to enhance gene expression as well as downregulate it (Master et al., 2016; Zhou and Rigoutsos, 2014). The seed to 3'UTR pattern is still used extensively in prediction of microRNA targets, though other tools have developed to incorporate prediction of other target sites in other regions of the mRNA target.

As mentioned, in previous studies the microRNA seed to target interaction appeared to depend largely on the conservation of the sequence in the 5' region of the microRNA, with disruption of the 3' region having little effect on regulation. However, in a smaller subset of target sites, it appeared that conservation in the 3' region of the microRNA could supplement or compensate the 5' seed region pairing. In another exception to the 5' conserved seed rule, microRNAs can

sometimes target transcripts by their central nucleotides instead of seed sequence (Martin et al., 2014). This is a conserved class of microRNAs which implements the conserved region of 11-12 base pairs in the middle of the microRNA and represses protein output without cleavage catalyzed by Argonaute proteins (Shin et al., 2010).

MicroRNA were once thought to be faster regulators of transcription than a regular transcription factor protein because transcribing a pri-miRNA would take less time than transcribing, processing, and translating a protein (Lee and Ambros, 2001). However, a mathematical, kinetic model of microRNA regulation taking into account dissociation, binding, and decay rates of microRNA and RISC found that the rate of regulation was slower than previously considered. The bottleneck in time-scale may be due to slow decay of proteins compared to mRNAs or the time required for microRNA to bind to an Argonaute protein (Hausser et al., 2013). As the understanding of microRNA regulation has developed over the past decade and a half, the short non-coding RNAs seem to continually produce surprises and grow in complexity, both in terms of their regulatory action, and how the expression of mature microRNA is controlled in the cell, as discussed in the next section.

1.2.7 Regulation of microRNA Expression

MicroRNA control expression of the majority of genes in mammalian systems, but how are microRNAs regulated themselves? The first microRNAs discovered in *C. elegans, lin-4* and *let-7* were controlled temporally through *cis*-acting elements (Lee et al., 1993). Heterochronic genes that are turned on at different stages of development often control microRNA genes in *C. elegans* (Reinhart et al., 2000). Similarly, expression of microRNAs may be regulated in response to various stimuli and dependent on the tissue or transcription factors expressed. Expression may be regulated by increases or decreases in methylation of the surrounding genome, and microRNAs can also potentially act to inhibit or promote their own epigenetic regulation through the proteins they target (Rouhi et al., 2008).

MicroRNA are regulated at various levels of their biogenesis. The proteins involved in biogenesis and RNA binding proteins (RBPs) that interact with the microRNA hairpin structure keep the expression of microRNA carefully controlled and are under strict control themselves. In the nucleus of the cell, the primary microRNA transcript is stabilized by capping and adenylation
in a similar fashion as a normal mRNA, and therefore is sensitive at this level to any mutations or expression changes in the proteins performing these steps. When DROSHA and DGCR8 form the microprocessor complex together with the hairpin structure of the pri-miRNA, other proteins may bind to the hairpin to coordinate or interfere with the microprocessor. Two homologs of the microRNA binding protein LIN28, LIN28A and LIN28B, bind to the pri-let-7 or pre-let-7 terminal loop and prevent cleavage by DROSHA and DICER (Figure 1.7a) (Wang et al., 2015a). DROSHA is also partially modulated by RBPs regulating a specific subset of microRNAs, such as DDX5 (p68) and DDX17 (p72), which control a number of miRNAs related to cell growth and proliferation (Fukuda et al., 2007; Fuller-Pace and Moore, 2011; Obernosterer et al., 2006). SMAD transcription factor proteins also regulate a specific subset of microRNAs in response to transforming growth factor-beta (TGF β) and bone morphogenetic protein (BMP) growth factor activation, which cause SMAD proteins to complex with DDX5 and increase DROSHA cleavage of pri-miR-21 to pre-miR-21 (Figure 1.7b) (Davis et al., 2008). DROSHA and DGCR8 are carefully auto-regulated, since DGCR8 stabilizes DROSHA in the microprocessor complex, but in its mRNA form DGCR8 includes a hairpin loop in its second exon that is cleaved by DROSHA and decreases DGCR8 expression. This feedback loop is maintained throughout evolution (Figure 1.7c) (Han et al., 2009; Kadener et al., 2009).

DROSHA regulation of microRNA expression may be affected by a large array of other proteins, as it is found in a small complex with DGCR8 but also in a large multi-protein complex containing E-wing sarcoma family proteins, dsRBPs, hnRNPs and RNA helicases. Knockdown of any of these interacting proteins can lead to increased microRNA and dysregulation of their targets (Figure 1.7d) (Gregory et al., 2004; Kim et al., 2014a), though whether the mechanism is through binding DROSHA or RNA has not been elucidated for all cases. When DROSHA is regulated by hnRNPA1 the protein binds specifically to the hairpin loop of pri-miR18a. Interestingly, hnRNPA1 does not bind other hairpins in the miR-17-92 cluster, but binding the hairpin loop of miR-18a makes a more favourable conformation for Drosha to cleave (Guil and Caceres, 2007; Michlewski et al., 2008).



Figure 1.7 - Regulation of DROSHA activity

a. DROSHA or DICER cleavage can be inhibiting by LIN28 binding to the pri-let7 or pre-let7 microRNA hairpin. b. Numerous RBPs bind to primary microRNAs and regulate DROSHA to enhance microRNA expression. c. DROSHA and DGCR8 proteins auto-regulate, as DGCR8 stabilizes DROSHA in the microprocessor complex and DGCR8 is limited by DROSHA due to a hairpin in the second exon of the DGCR8 mRNA, which is cleaved by DROSHA. d. Various proteins form large complexes with DROSHA and decrease expression of certain microRNAs.

Since DICER1 cleavage of the loop of pre-miRNA is also an essential step in microRNA biogenesis, it behaves as a sensitive regulator of microRNA expression. *DICER1* itself can be downregulated by the RNA binding protein AUF1 binding to *DICER1* mRNA, reducing its stability by binding to sites in the 3'UTR and coding sequence (Abdelmohsen and Gorospe, 2012). The *DICER1* mRNA also contains binding sites for let-7 microRNA, which decreases expression of DICER1 and creates a negative feedback loop (Forman et al., 2008; Tokumaru et al., 2008). DICER1 interacts with various dsRBD proteins that bind to the pre-miRNA, one of the most prominent being TRBP, which has three dsRBDs. A subset of microRNAs is highly dependent on TRBP interaction with DICER1 because of conformational changes in DICER1 induced by TRBP binding. Lack of conformational change in DICER1 disrupts its interaction with AGO, and decreases expression of a subset of microRNAs when TRBP is mutated (Chendrimada et al., 2005; Lee et al., 2013; Wilson et al., 2015). The efficiency of pre-miRNA processing is modulated through MAPK-ERK phosphorylation of TRBP, which promotes its stabilization and conformational changes that allow stronger interaction with DICER1 and greater stability of the dsRNA microRNA precursor (Paroo et al., 2009).

Ultimately, the regulation of microRNA is intricately controlled at each biogenesis level and at the transcriptional levels by different mechanisms, therefore expression of a portion or all microRNAs can be inhibited at multiple stages. This makes pathways and processes that are strictly regulated by microRNAs susceptible to damage in diseases not just at the genomic, transcription factor or epigenetic level, but the microRNA biogenesis level as well.

1.2.8 The Role of microRNAs in Cancer

Given the nature of microRNAs as transcriptional regulators involved in a multitude of processes throughout the cell, normal cellular behavior can become malignant due to changes in expression or mutations in microRNA or microRNA processing pathways. Many microRNAs play important functional roles in cancer, but global microRNA profiling can also provide useful diagnostic or prognostic information about a malignancy. In a study evaluating 217 microRNAs in 334 samples of various human cancers, the microRNA profiles could distinguish between tumour classifications, even in cases of poorly differentiated tumours, and even potentially identify mechanisms of transformation, through differences in microRNA profiles for samples with *BCR-ABL*, *TEL-AML1*, or *RUNX1*, and *MLL* rearrangements (Lu et al., 2005). General

downregulation of microRNAs was observed for the tumour tissues compared to normal tissues, emphasizing their importance in resisting cancer and suggesting possible impairment of microRNA biogenesis (Lu et al., 2005; Thomson et al., 2006).

Additionally, numerous studies have examined incidence of cancer in the event of the loss or amplification of microRNAs located in genomic regions associated with fragile sites or those containing copy number alterations. In chronic lymphocytic leukemia (CLL), miR-15 and -16 are found in the chr13q14 frequently deleted region and regulate the anti-apoptotic protein BCL-2. Deletion of these two microRNAs therefore leads to an increase in the BCL-2 target and decrease in apoptosis (Calin et al., 2002). In some cases, the microRNA's genomic region is deleted due to location at a fragile site, as with the critical let-7 family of microRNAs, where all 12 members are found at fragile sites linked to human cancers (Calin et al., 2004). In the reverse case, the genomic amplification of the microRNA cluster miR-17-92 is found in diffuse large B cell lymphoma and mantle cell lymphoma, and is a major factor in tumourigenesis, demonstrating that increased expression of a microRNA can be as detrimental as deletion (He et al., 2005; Tagawa and Seto, 2005).

Aberrations in microRNA expression affect tumour development because microRNAs control many basic pathways in the cell that are also functionally necessary to cancer, such as cell proliferation, differentiation, survival, metabolism, genome stability, inflammation, invasion, and angiogenesis (Lin and Gregory, 2015). When microRNAs are downregulated or upregulated due to transcriptional activation or genomic amplification/deletion/translocation they can behave as tumour suppressors or oncogenes according to the targets that they regulate. Since oncogenic microRNAs often have multiple tumour suppressors as gene targets and exhibit significantly different expression in cancer, they are often drivers of tumourigensis and referred to as "oncomiRs" (Hammond, 2006). One of the first microRNA to be discovered, miR-21, is a well-known onco-miR, regulating the proteins PTEN (Meng et al., 2007), PDCD4 (Asangani et al., 2008), SMARCA4 (Schramedei et al., 2011), MEF2C (Yelamanchili et al., 2010), and numerous other tumour suppressors. This microRNA is highly evolutionarily conserved and dysregulated in many different types of cancer. Another onco-miR is the miR-17-92 cluster, which as stated is often amplified in cancers, and decreases expression of proapoptotic genes such as BIM, PTEN and E2F1 (Ventura et al., 2008). MYC, a transcription factor controlling numerous cell cycle progression and cellular transformation genes, and the miR-17-92 cluster are concomitantly expressed in cancer, therefore disruption of the microRNA transcriptional start site leads to

25

increased microRNA and MYC expression (Claveria et al., 2013; He et al., 2005; O'Donnell et al., 2005). MYC also activates transcription of the miR-17-92 cluster, demonstrating that oncogenic or tumour-suppressing microRNAs are often involved in feedback mechanisms (Dews et al., 2006; O'Donnell et al., 2005). The transcriptional activators or promoters that control pri-miRNA transcription can also be targets of the microRNA whose expression they promote. A tumour suppressor microRNA, miR-34a, regulates TP53, an integral protein in driving apoptosis in the cell, so survival and protection of tumour cells occurs when the microRNA expression is abrogated. The expression of miR-34a is also driven by p53, leading to diminished regulation if either gene is mutated or dysregulated (Bommer et al., 2007; Chang et al., 2007; He et al., 2007; Raver-Shapira et al., 2007; Tarasov et al., 2007).

Let-7 is a tumour suppressor, controlling RAS expression through multiple let-7 binding sites in the 3'UTR (Johnson et al., 2005). However, in neuroblastoma, T-cell lymphoma, intestinal adenocarcinoma and a variety of other cancers, let-7 can be downregulated by reactivation of the LIN28A and LIN28B proteins, which regulate let-7 during embryonic development (Heo et al., 2008; Madison et al., 2013; Newman et al., 2008; Urbach et al., 2014; Viswanathan et al., 2008; Viswanathan et al., 2009).

Mutations or disruption of microRNA biogenesis is widespread in cancers as well. The levels of DROSHA and DICER1 are down in lung, ovarian, and neuroblastoma, and in some cancers decreased expression is associated with more advanced tumourigenesis or poor prognosis (Karube et al., 2005; Lin et al., 2010; Merritt et al., 2008). Similarly, *DICER1* expression may be decreased in cancer through transcription dysregulation since many oncoproteins and dysregulated tumour suppressors, such as p53 family member TAp63, regulate cancer progression by binding to the promoter of *DICER1* (Su et al., 2010). Mutations in *DROSHA*, XPO5, and DICER1 lead to loss of function and dysregulation of microRNA expression, as found in samples from patients with Wilms tumour, where DROSHA is frequently mutated at E1147K (Rakheja et al., 2014; Torrezan et al., 2014; Walz et al., 2015). In cancers with microsatellite instability, where inactivating mutations affect the XPO5 gene, transfer of the precursor microRNAs into the nucleus by EXP5 is impaired (Melo et al., 2010). In ovarian cancer Sertoli-Leydig cell tumors, many of the loss of function mutations in DICER1 occur in the RNase IIIb metal-binding domain. These hotspot mutations led to loss of cleavage activity and led dramatically decreased production of microRNAs from the 5p side of the microRNA precursor (Heravi-Moussavi et al., 2012; Wang et al., 2015b). A mutation in DICER1 is

considered the cause of a tumour predisposition syndrome known as *DICER1* syndrome (Slade et al., 2011).

The interference of epigenetic markers, through modification of histones or DNA, can also affect microRNA regulation in cancer phenotypes (Han et al., 2007). Hypermethylation of CpG islands at the promoters of tumour-suppressive microRNAs leads to epigenetic silencing (Saito and Jones, 2006). The alteration in expression levels due to epigenetic mechanisms has been seen of miR-127 in bladder cancer cells and miR-9-1 in breast cancer (Lehmann et al., 2008). Overall, microRNAs fulfill a critical regulatory function in many cellular processes and the numerous errors which can lead to dysregulation of their expression can lead to the initiation or progression of cancer.

1.3 microRNAs in Hematopoiesis and Blood Cancers

1.3.1 Hematopoiesis

Hematopoiesis is the production of blood cells. In human adults the multitude of different blood cell types in the body derive from the hematopoietic stem cell (HSC) found in the bone marrow, but blood cells are produced from a different source in early human development. The first blood cells of the mammalian embryo begin to form from blood islands in the yolk sac, and different organs become the site of production as development progresses. Blood islands form in the third week of embryogenesis, and gradually develop tubular structures that become the vascular system (Golub and Cumano, 2013). These blood islands produce primitive erythroid cells (EryP), a larger, enucleated type of erythroid cell, and small amounts of megakaryocytes and macrophages (Baron, 2013; Kingsley et al., 2013). During this yolk sac stage, blood cells develop from transient multipotent precursors distinct from the hematopoietic stem cell (HSC), and the blood cells derived at this stage do not have the same features as their adult counterparts (Bertrand et al., 2010; Palis et al., 1999). The EryP participate in oxygen delivery and vascular development through generation of sheer forces (Baron, 2013; Lucitti et al., 2007). The first cells with HSC properties of long-term renewal and repopulation occur at around 5 weeks in development of the human embryo and are found in the aorta-gonad-mesonephros (AGM) region (Ivanovs et al., 2011; Medvinsky and Dzierzak, 1996; Muller et al., 1994), major blood vessels (Orkin and Zon, 2008), and placenta (Gekas et al., 2005; Lee et al., 2010; Ottersbach and Dzierzak, 2005).

The liver becomes the primary site of blood cell formation in the third month of embryo development, producing the majority of erythrocytes and granulocytes (Clapp et al., 1995; Kelemen and Janossa, 1980). As the hepatic phase begins, the liver is assisted by lymphogenesis in the spleen, thymus and lymph nodes (Tavassoli, 1991). The liver remains an active site of blood production until just after birth and lymphogenesis continues in the spleen into adulthood (Tavassoli, 1991). Meanwhile in the fourth month of embryogenesis, the bone marrow begins to produce granulocytes and erythroid cells, and after birth becomes the main producer of blood cells for the human lifespan (Chagraoui et al., 2003; Kikuchi and Kondo, 2006).

In hematopoiesis, the HSC gives rise to all of the differentiated blood cell types in the body (Kiel et al., 2005; Morrison and Weissman, 1994). The HSC is extremely rare and has the capability of long-term self-renewal. The first differentiated cell, the multipotent progenitor (MPP) cell, has short-term self-renewal but not long-term self-renewal, and differentiates into the multitude of blood cell types, from erythrocytes to platelets to leukocytes. The process of producing many different blood cells with unique functions is under the control of various regulators of gene expression - transcription factors, epigenetics, and post-transcriptional silencing. The classical model of hematopoiesis with differentiation from a HSC to various cell fates is in the process of revision, with the former framework consisting of multiple oligopotent progenitors branching out of the MPP and gradually differentiating to their specific blood cell types. In the classical model, common lymphoid progenitors (CLPs) develop from the MPP and differentiate into natural killer, T, and B cells (Adolfsson et al., 2005; Inlay et al., 2009; Kondo et al., 1997; Schlenner et al., 2010). Also in this model, myeloid differentiation begins with the common myeloid progenitor (CMP), followed by intrinsic and extrinsic growth factors leading to the generation of granulocyte-macrophage lineage (monocytes, neutrophils, basophils, eosinophils, and mast cells) or the megakaryocyte-erythroid lineage (erythrocytes and platelets) from this progenitor (Laiosa et al., 2006). In recent studies, however, this compartmentalized view of hematopoiesis has changed. By studying the proportions of stem cells and progenitor cells in fetal liver, neonatal cord blood, and bone marrow from adults, it was discovered that there were much fewer progenitor cells in adult hematopoiesis, while progenitors at intermediate stages of differentiation were more distinct in fetal liver (Notta, Zandi et al 2016). The various cell fates of megakaryocytic, erythroid, etc., develop from multipotent cells separated as stem cells (by the marker CD34+, CD38-) (Notta, Zandi et al 2016). As I was focused on microRNAs potentially involved in the regulation of hematopoiesis and many of the studies on the expression of

microRNAs in hematopoiesis implemented the classical view in their approach, a schematic of the classical view is given as reference (Figure 1.8).

One of the most important transcription factor regulators in myeloid differentiation is *GATA1*, because it controls other transcription factors of distinct myeloid lineages. When *HDAC1* is stimulated by *GATA1*, the CMP is skewed towards megakaryocytic-erythroid differentiation, whereas when the transcription factor *CEBPB* downregulates *HDAC1*, CMP differentiate along the granulocytic lineage (Wada et al., 2009). Another layer of gene expression regulation, epigenetics, is important in myeloid differentiation, with increased DNA methylation seen in transitions from CMP to GMP, and in the silencing of pluripotent genes (Kosan and Godmann, 2016).

The hematopoietic system performs many vital functions, such as transport of oxygen and carbon dioxide between tissues of the body and the lungs, platelets for clotting and wound healing, and is a highly regenerative tissue. Additionally, it provides a good model system for stem cell research, as the HSC can be separated and studied under specific cell culture conditions or in mouse transplant models (Huntly and Gilliland, 2005; Siminovitch et al., 1963; Tanner et al., 2014; Till and Mc, 1961). In addition to controls such as transcription factors and epigenetic silencing, hematopoiesis is regulated by microRNAs, which inhibit the expression of targets as the process of self-renewal and differentiation proceeds.



Figure 1.8 - Simplified hematopoiesis

The hematopoietic hierarchy begins with the Long term self-renewing hematopoietic stem cell (LT-HSC), followed by the short-term self-renewing hematopoietic stem cell (ST-HSC), and followed by the multipotent progenitor (MPP). This cell can differentiate into the common myeloid progenitor (CMP) or common lymphoid progenitor (CLP), each of which gives rise to a unique set of cells. The CMP differentiates into the granulocyte-macrophage progenitor, which produces basophils, eosinophils, neutrophils, and monocytes/macrophages in the process of granulopoiesis, and into the megarkaryocyte-erythrocyte progenitor (MEP), which produces erythrocytes through erythropoiesis and platelets from the megakaryocyte through megakaryopoiesis. The common lymphoid progenitor produces the cells of the adaptive immune system, B-cells, Natural killer (NK) cells, and T-cells, otherwise known as the lymphocytes.

1.3.2 Role of microRNAs in Differentiation of Blood Cells

One of the first identified roles of microRNA regulation was the influence on the process of hematopoiesis in mammals. The first microRNAs that were noted as specific to blood were miR-223, mir-181, and miR-142 (Chen et al., 2004) and their functional significance in defined cells became apparent later. MicroRNAs regulate hematopoietic differentiation and function,

fluctuating greatly in expression throughout hematopoiesis as part of the complex regulation of self-renewal and proliferation, quiescence and differentiation.

The microRNAs in the HSC are particularly important as part of the program that maintains a self-renewing, multipotent stem cell. One might expect that the levels of microRNA expression would be highest here, to regulate and repress any cellular program that might run amok, but overall microRNA expression generally increases throughout the process of differentiation (Navarro and Lieberman, 2010). Although differentiated cells have overall higher microRNA expression, there are select microRNAs that are expressed in HSPCs. MicroRNA profiling of human bone marrow CD34+ cells, the subpopulation which contains HSCs, found the most abundant microRNAs to be miR-191, miR-181, miR-223, miR-25, miR-26, miR-221, and miR-222 (Georgantas et al., 2007). In microfluidics experiment studying the expression of microRNAs in progenitors, through stages of differentiation, down to terminally differentiated cells, several miRs were identified as being enriched in stem cell and progenitor populations relative to mature blood cells. These included miR-125b, miR-196a, miR-196b, miR-130a, let-7d, miR-148b, and miR-351 (Petriv et al., 2010).

Several of the first microRNA targets to be identified in humans were the homeobox proteins (Yekta et al., 2004). Expression profiling of microRNA and mRNA in the hematopoietic compartments has been performed and the data integrated to show that microRNA mediate expression of numerous homeobox genes, *RUNX1*, and *CEBPB* expressed in the stem/progenitor compartment and critically control hematopoiesis by keeping these genes in check (Georgantas et al., 2007).

MicroRNAs are significant regulators in the complex transition point of the MPP differentiating into myeloid and lymphoid lineages. In the microRNA profiles for 27 phenotypically distinct blood cell populations acquired in the microfluidics study, the CMP and CLP shared expression of some microRNAs that were not expressed in any further differentiated populations (Figure 1.9). The microRNAs most enriched in the CMP compared to the CLP were miR-130a, miR-31, and miR-203, whereas miR-126, miR-126*, and miR-23a were enriched in the CLP (Petriv et al., 2010).

The microRNAs in cells of the myeloid lineage help to transform the CMP into multiple functionally diverse mature blood cell types. The intricate processes of erythropoiesis,

megakaryopoiesis, etc, depend on microRNA regulation of gene expression (Figure 1.9). In myelopoiesis, one of the first microRNAs discovered in blood, miR-223, was found to be integral to granulocytic differentiation, increasing in expression as maturation proceeds (Fazi et al., 2005). The targets of miR-223 include transcription factors *MEF2C* and *E2F1*, which regulate cell cycle and proliferation (Fukao et al., 2007; Johnnidis et al., 2008). Garzon R *et al* found that another subset of microRNAs, let-7a-3, let-7c, let-7d, miR-15a, -15b, -16-1, -107, -223, and miR-342, were upregulated and miR-181b was downregulated during all-trans-retinoic acid (ATRA) induced differentiation to granulocytes (Garzon et al., 2007). For monocyte development, *PU.1* activates miR-424, which induces monocytic/macrophage differentiation (Rosa et al., 2007).

Erythropoiesis shows a gradual decrease in miR-221, miR-222 (Felli et al., 2005), miR150 and miR-155 (Dore et al., 2008). In the final stages of maturation there is an increase in miR-16 and miR-451, both transcribed from the same cluster under the transcriptional control of *GATA1* (Figure 1.9) (Garzon and Croce, 2008). The phospho-serine/threonine binding protein, 143-3 ζ is inhibited by miR-451, and downregulation of this microRNA allows the activation of the transcription factor *FOXO3* and the erythoid genes it controls (Yu et al., 2010).

Expression of miR-150 increases from low in megakaryocytic-erythroid progenitor cells to high in megakaryocytes (Lu et al., 2008). Meanwhile, miR-155 and miR-146a are highly expressed in early progenitors and repress megakaryopoiesis, decreasing in expression as precursor cells differentiate into megakaryocytes (Lu et al., 2008). Another distinct subset of microRNAs downregulated during the transition from CD34+ cells to megakaryocytes include miR-10b, - 30c, -106, -126, -32, and -143. Two essential microRNAs of this subset are miR-10a and miR-130a, whose decrease in expression leads to upregulation of their targets, the lineage specific transcription factors *HOXA1* and *MAFB*, during this transition (Garzon et al., 2006). Interestingly in the context of myeloid malignancies, microRNA target sites are underrepresented in 3'UTRs of cytokine and chemokine ligands and receptors relative to the genome as a whole (Asirvatham et al., 2008), though IL-2 and IL-10 are regulated by miR-181c and miR-106a, respectively.

In lymphoid differentiation, there is an initial upregulation of miR-181 and miR-17-92 cluster in the transition from MPP to CLP, and for the transition from CLP to B-cell progenitors we see a reversal of this expression, though the expression of miR-181 continues to increase in the T-cell lineage. In the *DICER1*-null mouse model, where the effect of knocking down total cellular

expression of microRNAs can be seen, B-cell differentiation and T-cell development are highly perturbed due to the loss of miR-181a expression (Kanellopoulou et al., 2005; Muljo et al., 2005). The maturation of the lymphoid lineages also shows decreasing miR-150 expression from the MPP to B-cell, though with an increase from the CLP to progenitor T-cell, and increasing expression of miR-21, -29, -223, and -221 in the transition to T-cell as well (Havelange and Garzon, 2010).



Figure 1.9 - microRNAs involved in hematopoiesis

The subset of microRNA with highest expression exclusively in the HSC includes miR-191, miR-25, miR-520, miR-26, miR-10, miR-126, miR-221/222, and miR-155. Other microRNAs decrease or increase in expression as a cell matures (symbolized by downward arrow or upward arrow, respectively).

1.3.3 Dysregulation of microRNAs in Blood Cancers

The consequences of microRNA dysregulation in hematopoietic tissues are the disruption of hematopoiesis and initiation of tumourigenesis. Calin and Croce made the first discovery of a microRNA deletion leading to cancer in an ingenious search for the missing genetic material on chromosome 13q14 (Calin et al., 2002). This region is frequently deleted in chronic lymphocytic leukemia (CLL) patients and in other types of cancer. However, none of the protein-coding genes found in the region recapitulated the symptoms of CLL or other cancers when expression studies were performed. Following the discovery of microRNA, they were able to identify miR-15 and -16 as frequently deleted in this region and that they performed regulation of the anti-apoptotic protein BCL-2 (Calin et al., 2002). Other microRNAs that were later found to be dysregulated in CLL are miR-29b and -181b, which have tumour suppressor functionality through targeting of *TCL1* and preventing upregulation of AKT (high levels of TCL1 are associated with high levels of *ZAP-70*) (Herling et al., 2006; Mott et al., 2007; Pekarsky et al., 2007).

Self-renewal, differentiation, apoptosis, proliferation and other major functions are disrupted in myeloid malignancies due to dysregulated microRNA. Three microRNAs of critical importance in these functions are miR-125, miR-155 and miR-29. Self-renewal in hematopoiesis is related to miR-125, which is highly expressed in the HSC and decreases throughout differentiation. The overexpression of miR-125 is linked to lineage bias, enhanced HSC function and the induction of leukemia, potentially due to the targets of miR-125 mediating anti-apoptotic effects (Bousquet et al., 2010; Bousquet et al., 2012; Lin et al., 2011).

MiR-155 is also expressed at high levels in HSPCs, and decreases in expression as myeloid cells or erythroblasts become more mature, but increasing in B- and T-cells during activation (Eis et al., 2005; Georgantas et al., 2007; Masaki et al., 2007; O'Connell et al., 2008). Studies have determined that the targets of miR-155 involved in leukemogenesis are *SHIP1* and *CEBPB* (Gorgoni et al., 2002; O'Connell et al., 2009). *SHIP1* blocks the signal of the *PI3K-AKT* pathway (Damen et al., 1996; Ono et al., 1996), leading to a decrease in apoptosis and leukemic progression (Khalaj et al., 2014). The miR-29 family targets the epigenetic regulators *DNMT3A*, *DNMT3B* and *SP1*, therefore downregulation of microRNAs from this family in myeloid malignancies promotes DNA hypermethylation (Garzon et al., 2009a; Garzon et al., 2009b). Epigenetic changes in DNA methylation and histone marks may also affect microRNA

expression and AML pathogenesis. *AML1-ETO* induces heterochromatin silencing of miR-223 (Fazi et al., 2007). Hypermethylation of the miR-193a promoter region leads to its downregulation in AML patients and its expression inversely correlates with its target c-KIT (Gao et al., 2011).

In normal karyotype AML patients, certain microRNA profiles are associated with different recurrent mutations. The miR-10, let-7 and miR-29 family members are upregulated and miR-204 and miR-128a are downregulated in AML patients carrying *NPM1* mutations (Garzon et al., 2009b). In patients with *FLT3-ITD* mutations, the upregulation of miR-155 may contribute to the aggressive phenotype (Jongen-Lavrencic et al., 2008). This microRNA targets two proteins involved in hematopoietic transcriptional regulation, PU.1 and CEBPB, and is increased in other blood cancers, such as CLL and B-cell lymphomas (Faraoni et al., 2009).

Overall, many signaling pathways are thought be to affected by dysregulation of microRNAs in blood malignancies, including *MAPK* and *PI3K/AKT* signaling, the Toll-like receptor pathways, and RAS downstream interactions (Stoffers et al., 2012).

1.3.4 Clinical Features and Subtypes of Myelodysplastic Syndromes

Myelodysplastic syndromes (MDS) are a group of hematological malignancies showing a wide variation in clinical features, but primarily characterized by ineffective hematopoiesis, peripheral blood cytopenias and hypercellular bone marrow (Pellagatti and Boultwood, 2015). Since myelodysplastic syndromes are a series of disorders, with the common attribute of dysplasia in blood cells, it was at first hard to distinguish and diagnose (Komrokji et al., 2010). Patients presented with anemia that was refractory to treatment and could sometimes progress to leukemia, but displayed different cytological dysplasias and severity in outcomes.

MDS remain difficult to classify, described as "a diagnosis of exclusion" (Komrokji et al., 2010). The term "myelodysplastic syndromes" originated with the French-American-British group in a series of proposals in 1976 and was further elaborated on in the 1980s (Bennett et al., 1976; Bennett et al., 1982). The International Prognostic Scoring System (IPSS) was developed in 1997 to assess the prognosis of patients and has since undergone numerous revisions such as newer cytogenetic groupings, patient comorbidities, etc. (Greenberg et al., 1997; Greenberg et

al., 2012). A study by Haase *et al* collecting morphological, clinical, cytogenetic and follow-up data from 2124 patients measured the types and frequencies of cytogenetic abnormalities among MDS patients (Haase et al., 2007).

These data and later studies gave insight into the prognostic impact of the most frequent karyotypes and how these correlated with the subtypes of MDS (Schanz J et al 2012, Braulke F et al 2013). Cytogenetic testing revealed chromosomal abnormalities in almost 50% of total patients and in approximately 70% of patients with therapy-related MDS, which agreed with the findings of previous studies (Boultwood et al., 1994a; Boultwood et al., 1994b; Van den Berghe et al., 1974). The most frequent aberrations in karyotype involved deletions of the long arm of chromosome 5, occurring in 30% of the 1080 patients with abnormalities. Other common abnormalities included monosomy 7 or deletions on 7q, trisomy 8, deletion on 20q, monosomy 18 or deletion in 18q, and monosomy Y. Aberrations occurring with less frequency in patients were monosomy 17, trisomy 21, translocation inversion of the long arm of chromosome 3, monosomy 13 or deletion on 13q, monosomy 21, a translocation in 5q, and various others (Haase et al., 2007).

The advent of sequencing of the human genome has given some insight into the differences between MDS subtypes and allowed closer examination of the underlying factors leading to the symptoms of MDS, including mutations or changes in expression of key genes in hematopoiesis and cellular proliferation processes.

1.3.5 The Role of Genetic Abnormalities in MDS Pathogenesis

The pathogeneses of MDS fit well within the multiple-hit hypothesis of cancer formation. In blood neoplasms, the disease can begin with an HSC acquiring enhanced self-renewal capability or a progenitor cell gaining self-renewal capability, followed by increased proliferation of either the abnormal clone or its progeny. Mutations in proteins such as *RUNX1* lead to a block in differentiation, followed by genetic or epigenetic destabilization by loss of function of EZH2 or similarly functioning proteins. Other mechanisms of pathogenesis may include the development of anti-apoptotic mechanisms, evasion of the immune system, or suppression of normal hematopoiesis (Bejar et al., 2011).

Chromosomal abnormalities discussed in the previous section likely contribute to the pathogenesis of MDS. Aberrations in chromosome 7 are the second most frequent karyotypic abnormality in MDS and associated with poor prognosis, with a median survival of 14 months (Pellagatti and Boultwood, 2015). When looking specifically at those with therapy-related MDS and prior use of alkylating agents, abnormalities in this chromosome occur in 50% of patients (Christiansen et al., 2004). The chromosome 7q gene, *EZH2*, participates in methylation and translational repression and is significantly underexpressed in patients with -7/del(7q) (Cabrero et al., 2016; Sashida et al., 2014). The decreased expression is associated with worse outcomes in patients (Cabrero et al., 2016). Another gene found in a deleted region of chromosome 7 that is found to be significantly decreased in patients is *CUX1* (Jerez et al., 2012). CUX1 is a transcription factor which is thought to act as a tumour suppressor (McNerney et al., 2013). In therapy-related MDS, chromosome 7 abnormalities also co-occur with chromosome 5 deletions or *RUNX1* mutations much more frequently than would be expected by random distribution of abnormalities (Bejar et al., 2011; Pedersen-Bjergaard et al., 2006).

Trisomy 8 is an intermediate-risk abnormality, occurring in roughly 8% of patients (Bejar et al., 2011). The genetic material gained by another copy of chromosome 8 may lead to higher expression of anti-apoptotic genes and higher resistance to irradiation, giving them a survival advantage over normal HSCs (Lim et al., 2007). Deletions on chromosome 20q are common in myeloid malignancies and AML, but none of the 19 genes in the commonly deleted region have been linked to the pathogenesis of these hematological malignancies as of yet (Bench et al., 2000; Wang et al., 2000). The loss of the chromosome Y may also be unrelated to pathogenesis and both of these abnormalities have favourable prognosis for patients (Lim et al., 2007).

While genetic lesions such as amplifications, deletions, and balanced translocations are commonly found in MDS, mutations in individual genes regulating key cell processes can also contribute to the disease. Mutations that are associated with poor prognosis or increased progression to AML include *RUNX1*, *TP53*, *NRAS/KRAS*, *EZH2*, *IDH1*, *IDH2*, *CEBPA*, *FLT3*, *CSF1R*, and *CKIT* (Bejar et al., 2011). *RUNX1*, found as mutated in 15-20% of patients, is a key transcription factor in the differentiation of hematopoietic stem cells, leading to mutations in MDS that occur in the DNA- or protein-binding domains impair the protein's function (Chen et al., 2007; Growney et al., 2005; Harada and Harada, 2009; Steensma and List, 2005; Tang et al., 2009). Mutations in another transcription factor, *CEBPA*, are rarer, found in 2-5% of patients, and lead to defects in differentiation of granulocytes (Bejar et al., 2011). *EZH2* is necessary for

epigenetic regulation and mutation of the gene in roughly 6% of MDS patients leads to the loss of critical histone-3-lysine 27-methyltransferase activity in the cell (Ernst et al., 2010; Makishima et al., 2010; Morin et al., 2010). *IDH1* and *IDH2* mutations are involved in the catalytic reaction of converting alpha-ketoglutarate to 2-hydroxyglutarate, and the accumulation of 2-hydroxygluatarate that results from mutations inhibits other dioxygenases requiring alpha-ketoglutarate as a cofactor (Dang et al., 2009; Kosmider et al., 2010; Paschka et al., 2010; Tefferi, 2010; Thol et al., 2010; Ward et al., 2010). Gain of function mutations in *NRAS* and *KRAS*, which are GTPases, in approximately 10% of patients leads to constitutive activation of serine/threonine kinases, while *FLT3*, *CSF1R*, and *CKIT* mutations lead to constitutive activation of tyrsosine kinase signaling, but in less than 2% of patients (Paquette et al., 1993; Sargin et al., 2007).

In a study of 944 MDS patients with various subtypes, the most frequently mutated genes were *TET2, SF3B1, ASXL1, SRSF2, DNMT3A*, and *RUNX1* (Haferlach et al., 2014). SF3B1 mutation were frequently found in the isolated del(5q) subtype and associated with a better clinical outcome (Bacher et al., 2008). Even in subtypes of MDS with normal karyotype, mutations were found in 73% of cases (Haferlach et al., 2014). The multitude of single gene mutations at low frequencies supports the heterogeneity of MDS, and while it is difficult to attribute a single gene mutation to the pathogenesis of the disease, patterns in abnormalities are still emerging and may share similar underlying genetic backgrounds (Abdel-Wahab et al., 2010; Bacher et al., 2007; Mardis et al., 2009).

1.3.6 Pathogenesis of the del(5q) MDS Subtype

The most common chromosomal abnormality in MDS is the interstitial deletion on the long arm of chromosome 5 in *de novo* MDS (Haase et al., 2007). The classification of the del(5q) MDS subtype in patients is applied when del(5q) is the sole karyotypic abnormality and there is a blast count in the bone marrow of less than 5% (Bernasconi et al., 2005). The deletion on 5q is heterozygous, and until recently no mutations in genes on the intact allele were found so the disease was thought to result from only the haploinsufficiency of one or more genes (Gondek et al., 2008; Graubert et al., 2009; Heinrichs et al., 2009). However, a recent study examining *CSNK1A1*, an gene located in the 5q32 CDR, found recurrent mutations in 7% of del(5q) MDS patients (Schneider et al., 2014). Furthermore, while heterozygous expression of *CSNK1A1* led

to HSC expansion, homozygous deletion of the gene led to HSC failure (Schneider et al., 2014). While patients with del(5q) syndrome have a good prognosis, in 10% of patients the illness transforms to AML (Boultwood et al., 2010).

The extent of the deleted region on 5q was identified by Boultwood J et al using molecular mapping and fluorescent *in situ* hybridization (FISH), and narrowed down to a 1.5Mb interval at 5q32-33 bordered by the DNA marker *D5S413* and the *GLRA1* gene (Boultwood et al., 1994a). This is the deleted region associated with the del(5q) subtype, but deletion in the proximal 5q31 region also occasionally occurs and is associated with therapy-related or more aggressive MDS or AML phenotypes (Boultwood et al., 2002; Horrigan et al., 2000; Van den Berghe et al., 1974). The 5q31 region includes the early growth response (*EGR1*) and alpha catenin (*CTNNA1*) genes, the former of which has been posited to increase stem cell renewal and the hypermethylation of the latter on the remaining allele is associated with transformation to AML (Ye et al., 2009).

The clinical features of del(5q) are macrocytic anemia, hypolobulated megakaryocytes, and normal or high platelet count (Vardiman et al., 2009). The haploinsufficiency of chromosomal material is attributed to the cause of the illness, and the loss of *RPS14* and other genes have been related to the pathogenesis. In examination of the CD34+ BM from del(5q) patients, *RPS14* and other genes from the CDR such as *CSNK1A1* were downregulated (Pellagatti et al., 2008). A thorough study of the consequences of loss for each of the 40 genes in the del(5q) commonly deleted region (CDR) was performed by Ebert *et al* using an RNA interference screen (Ebert et al., 2008). Partial loss of *RPS14* led to a block in erythroid differentiation and northern blotting of rRNA showed decrease of 18S/18SE RNA species concurrently with the accumulation of 30S species, indicating that the defect in ribosomal processing affects erythroid production (Ebert et al., 2008). Ribosome deficiencies were also implicated in a congenital disease with a similar phenotype, Diamond-Blackfan anemia (Horos et al., 2012).

There are a number of p53 activators observed in cases of stress in ribosomal biogenesis (Golomb et al., 2014). A mouse with haploinsufficiency in a syntenic region to the CDR demonstrated anemia, as well as monolobulated megakaryocytes, accumulation of p53, and increased apoptosis in the bone marrow (BM) (Barlow et al., 2010a). The defects in hematopoiesis could be partially eliminated by crossing the del(5q) mouse with a p53 deficient mouse (Barlow et al., 2010a; Barlow et al., 2010b). Loss of the entire CDR does not appear

necessary for this phenotype, as haploinsufficiency of RPS14 causes erythroid/MK defects and p53 accumulation as well, and heterozygous p53 KD relieves these defects (Schneider et al., 2014).

Haploinsufficiency of multiple protein-coding genes in the CDR accounted for a portion of the symptoms in the del(5q) disease but other features were not linked pathogenically. This hinted at the potential dysregulation of a number of non-coding genes found in the CDR. This led to the investigation of the microRNAs located in the CDR and how their loss of expression contributed to the pathogenesis of the del(5q) subtype of MDS.

1.3.7 microRNAs in Myelodysplastic Syndromes

Dysregulation of microRNA expression in hematopoiesis and other integral cellular process has been discussed as a factor in multiple forms of blood cancer in previous sections. This highlights the severe impact the loss of microRNAs from the CDR in del(5q) may have on the pathogenesis of the disease. Previous studies from our lab determined that more than 70% of miRNAs are located in regions of recurrent copy number alterations in MDS and AML, indicating the dysregulation of other microRNAs is correlated with these illnesses and may influence the pathogenesis in progression of del(5g) MDS (Starczynowski et al., 2011). Multiple microRNAs involved in the regulation of hematopoiesis appear dysregulated in MDS, though they do not distinguish different disease subtypes. In patients with AML evolved from MDS, decreases in expression of miR-221 were seen, and miR-150 was increased. Given the integral role that these two microRNAs play in erythroid and megakaryocytic differentiation, it is not surprising that inhibition of erythroid proliferation is seen (Hussein et al., 2010a; Hussein et al., 2010b). Increased miR-150 expression is observed in del(5q) MDS patients as well, which may explain increased platelet levels. Other MDS patients show decreased expression of miR-150, as well as miR-146a and let-7e, in mononuclear BM cells, while demonstrating high levels of miR-222 and miR-10a (Sokol et al., 2011). The levels of miR-146a in CD34+ BM cells from del(5g) patients are also decreased amounts (Votavova et al., 2011). The expression patterns of let-7a, miR-17-5p and miR-20a in CD34+ HSCs from the bone marrow of 43 MDS patients and the peripheral blood of 18 healthy donors were recently evaluated, and found to be up-regulated in low risk MDS patients but down-regulated in high- risk MDS patients (Vasilatou et al., 2013).

The *PI3K/AKT* pathway, whose function is inhibited in myeloid malignancies associated with dysregulated microRNA, regulates transcription of miR-22. The overexpression of miR-22 in HSPC cells gives enhanced proliferation and defective differentiation, sometimes leading to an MDS-like phenotype over the course of time (Bar and Dikstein, 2010; Polioudakis et al., 2013; Song and Pandolfi, 2014). Increased expression of miR-22 is also associated with aberrant hypermethylation in MDS patients with poor survival, and has been considered as a novel therapeutic target for MDS and myeloid leukemia, perhaps in combination with a disrupter of hypermethylation (Song et al., 2013).

For the microRNAs located in or around the critically deleted region (CDR) of chromosome 5 in del(5q) MDS patients, previous studies in our lab examined expression of each microRNA in del(5q) MDS patients compared to MDS patients with normal karyotype. The three microRNAs with significantly decreased expression in del(5q) were miR-143, miR-145 and miR-146a. The two microRNAs miR-143 and miR-145 are within the CDR, and though miR-146a is found slightly outside of the region, it is often downregulated in del(5q) patients (Kumar et al., 2011).

To elucidate the role of these microRNAs in MDS, our lab used retroviral decoys to knock down miR-146a and miR-145 in mouse HSPCs and transplanted them into lethally irradiated mice (Starczynowski et al., 2010). Eight weeks following transplantation, the recipient mice exhibited thrombocytosis, mild and variable neutropenia, and hypolobated megakaryocytes in the bone marrow (Starczynowski et al., 2010). To determine the mechanism through which the microRNA loss was leading to this phenotype, the predicted targets for each microRNA were analyzed and the innate immune signaling pathway was enriched with protein targets (Starczynowski et al., 2010). The proteins TIRAP and TRAF6 are members of the Toll-like receptor (TLR) pathway of innate immunity, and were confirmed through luciferase assays to be targeted by miR-145 and miR-146a, respectively (Starczynowski et al., 2010). TIRAP and TRAF6 interact together to activate downstream signaling, according to immunoprecipitation experiments, and TIRAP fails to activate nuclear factor-kappaB (NF-kB) signaling in the presence of a dominant-negative TRAF6 (Starczynowski et al., 2010). The TRAF6 protein, an E3 ubiquitin ligase that directs downstream activation of NF-kB, was overexpressed in mouse bone marrow cells and transplanted into recipient mice. The recipient mice developed neutropenia, thrombocytosis, and increased hypolobated megakaryocytes in the bone marrow by 12 weeks, and over half progressed to bone marrow failure or AML at ≥5 months posttransplantation (Starczynowski et al., 2010). The similarity between the induced-TRAF6

41

phenotype and the features of del(5q) MDS in humans suggests a potential mechanism by which miR-145 and miR-146a drive the del(5q) phenotype.

The megakaryocyte and erythroid regulatory transcription factor, *FLI-1*, was also identified as a target of miR-145 involved in the pathogenesis of del(5q) (Kumar et al., 2011). The 3'UTR of *FLI-1* contains numerous miR-145 binding sites and overexpression of miR-145 leads to a decrease in megakaryocytic cell production relative to erythroid, while inhibition of miR-145 leads to a reciprocal effect (Kumar et al., 2011). The combined loss of miR-145 and RPS14 alter megakaryocyte and erythroid differentiation in an analogous manner to del(5q) and, together with our lab's findings on miR-145 regulation of *TIRAP*, demonstrate the multi-faceted impact that loss of a microRNA may have in the course of a disease. The targets of miR-143 have yet to be explored for links to part of pathogenesis in del(5q), and given the large number of targets predicted for each microRNA, further exploration of the regulation by miR-145 and miR-146a is likely to be required as well.



Figure 1.10 - Consequences of loss of genetic material from chromosome 5

In addition to the CDR at 5q33.1 (shown in bracket), chromosome 5 has additional mutated or deleted genes that may be important in oncogenesis (shown in red), including miR-146a at 5q33.3, which is often deleted in del(5q) patients (shown in red). The loss of SPARC has been attributed to thrombocytopenia and erythrocytopenia (purple, Lehmann S et al 2007), and the knockdown of RPS14 by shRNA leads to blocks in erythropoiesis (orange, Ebert BL et al 2008). The lost of miR-145a leads to upregulation of TIRAP, which in combination with the upregulation of the target TRAF6 from loss of miR-146a, leads to the activation of IL-6 signaling and increased, defective megakaryopoiesis (blue, Starczynowski DT et al 2009). The upregulation of FLI-1 is also a consequence of miR-145 loss and leads to increased megakaryopoiesis relative to erythroid cell production (green, Kumar MS et al 2011).

1.4 Modeling the Role of a microRNA in Disease

There are numerous ways of modeling the gain or loss of microRNA in a disease and the methodology of the greatest focus for this thesis is the sponge or decoy method. The sponge method functions as a competitive RNA at work in the cell and binds the microRNA in a similar manner as endogenous targets. The regulation performed by microRNA binding to target mRNAs is taken advantage of by the sponge method, which offers a high number of binding sites for the microRNA of interest through the sponge transcript, which acts as a decoy and inhibits binding to other mRNA.

1.4.1 Approaches to Novel microRNA Discovery

Following the discovery of the microRNAs let-7 and lin-4 in *C.elegans*, initial screening for other microRNA species used a cloning approach for discovery of new sequences (Lagos-Quintana et al., 2002). As the number of microRNAs grew, the expression of microRNAs in different samples could be measured by microarray, which was useful for detecting differences in the microRNA expression profiles of, for example, tumour or normal tissue (Calin et al., 2002; Calin et al., 2004; Nelson et al., 2004). However, microarray profiling of microRNA expression left out the potential for discovery of novel microRNAs, and could experience background or cross-hybridization problems (Creighton et al., 2009). (Berezikov et al., 2006a)With the advent of more affordable next-generation sequencing methods, small RNA-Seq was developed and used to find the expression levels of microRNAs, both annotated and novel (Berezikov et al., 2006a).

As the array of tools for determining novel microRNAs widened, the two most common approaches to microRNA discovery developed, the experimentally-driven approach, initiated by the physical detection of a small RNA, and the computationally-driven approach, which uses *in silico* methods. These approaches commonly overlap or are used for cross validation. In experiment driven methods, the expression of a microRNA is established first, often by small RNA-seq experiments, followed by prediction of hairpin secondary structure and phylogenetic conservation using bioinformatics techniques (Berezikov et al., 2006b).

In the small-RNA sequencing method of discovery, the presence of a small RNA read is one criterion, but characteristics of the read, the location it is found in, and other properties are used

to decide whether the read is a potential microRNA or not. These criteria are computationally validated, rather than manually, due to the huge number of reads sequenced in small RNA-seq experiments (Chiang et al., 2010). In one study, the potential novel microRNA sequences were mapped to the genome, and the sequence of the transcript along with 100 bases flanking either side was fetched so that 220bp sequences could be tested for microRNA-like hairpin. If there was a stem-loop structure, the sequences were trimmed down and tested for folding correctly again (Creighton et al., 2009). There are numerous computational tools for microRNA discovery to apply after small-RNA sequencing experiments, with the intent of novel microRNA discovery, such as miR-Deep, CD-miRNA, and MiRank (Chiang et al., 2010).

For computationally driven methods, the potential microRNAs were first predicted in whole genome sequences on the basis of features such as structure and conservation. After their computational prediction, experimental techniques such as northern blotting, RNA-primed Arraybased Klenow Extension (RAKE) approaches, or primer extension assays were used to validate the theoretical microRNA (Berezikov et al., 2006b). Various computational methods were originally developed to predict microRNA hairpin structure. A study implementing an *ab initio* method used the property of microRNAs sometimes being found in clusters and searched within 20 kb of flanking region for potential fold-back RNA structures (Sewer et al., 2005). Another method found homologs in different species based on a query sequence, and used secondary structure, free energy, sequence alignment and conservation to find conserved microRNA genes (Artzi et al., 2008).

1.4.2 Methods of microRNA Study or Alteration

There are many ways of achieving altered expression of a microRNA, whether the setting is for genetic therapy purposes or modeling a disease *in vitro* or *in vivo*. The various methods to attain lowered microRNA expression include anti-sense oligonucleotides (ASOs), decoys or sponges designed to bind the microRNA of interest, or genetic knockout in animal models. Methods of mimicking microRNA gain or the effects thereof include transfection of microRNA hairpins, and transduction of viral vectors stably expressing the microRNA hairpin, while germline transgenic approaches with animal models tend to focus on modifying expression of the specific protein targets of the microRNA (Issler and Chen, 2015). Our work focuses on the use of sponges in decreasing the expression of microRNAs that show reduced expression in del(5q) MDS, though

microRNA inhibition using ASOs, and overexpression using microRNA mimics and hairpin expression vectors are explored as well.

Loss of microRNA leads to upregulation of numerous targets normally regulated by the small RNA. The change in expression of targets may be measured at the mRNA level by RNA-Seq, microarrays, or RT-qPCR and at the protein level by immunoblotting and quantitative proteomics. Another method of measuring the regulation by a particular microRNA is to adapt the 3'UTR of a luciferase gene by insertion of a microRNA binding site, either a reverse complement sequence or the 3'UTR of an endogenous target, in whole or only a section surrounding the sequence to which the microRNA binds. Since microRNA behavior is usually to inhibit or repress expression of a gene, the upregulation of the target following the loss of microRNA is referred to as "derepression." Upregulation of predicted protein targets or phenotypic changes are often used as markers of knockdown or knockout, and luciferase assays with the insertion of a microRNA binding site in the luciferase 3'UTR are often used to measure derepression and assess whether loss of microRNA expression has occurred.



Figure 1.11 - Methods of microRNA study

a. The loss of microRNA expression can be modeled using small anti-sense oligonucleotides as inhibitiors, through transduction of a virally expressed sponge that can bind microRNA and act as a decoy, or by genetic knock-outs of the microRNA, such as in mice. b. microRNA expression can also be increased if a gain of effect model is sought. This can be done through transfection of microRNA hairpin mimics into cells, or transduction with a viral vector expressing the microRNA hairpin. c. Sometimes the increases in microRNA are studies by genetic knockout of their protein targets, which would decrease with increased expression of the microRNA. However, effects of increases of protein targets, following knockdown, could also be examined.

The field of microRNA research has used siRNA, dsRNA, and shRNA species delivery into cells to assess the role of microRNAs in gene function, and methods employing other forms of small RNA species have been used for modeling the loss of microRNA. Studies of microRNA expression and regulation have implemented small oligonucleotide inhibitors such as ASOs, known alternately as anti-microRNA oligonucleotides ("anti-miRs" or AMOs). ASOs are a diverse group of microRNA inhibitors, which block the mature microRNA by WC base pairing in the cytoplasm from binding to their targets and may sequester the microRNAs for degradation (Davis et al., 2006; Krutzfeldt et al., 2005). ASOs are nucleic acids with a wide variety of chemical modifications, such as phosphorothioate linkages between nucleotides, locked nucleic

acids (LNAs), or 2'-O-methyl RNA, which affect potency, nuclease sensitivity, and thermodynamics of duplex formation. The activity of ASOs depends on their design, since some chemical modifications, such as 2'-O-methyl bases, promote the half-life of the ASO by nuclease resistance but fail to fully repress the microRNA (Lennox and Behlke, 2010; Meister et al., 2004), and others such as LNAs increased the binding stability but decrease specificity. The degree of complementarity to the target also affects the activity of the ASO, as mismatches in binding can lend different degrees of stability as the oligonucleotide duplex is incorporated into RISC (Brennecke et al., 2003; Hutvagner and Zamore, 2002).

Another method of modeling loss of microRNA in disease is use of sponges or decoys. Sponges have tandemly arrayed microRNA binding sites positioned in the 3'UTR of a reporter gene such as a fluorescent protein. There are RNA polymerase II sponges, which use typical eukaryotic promoters to express the sponge transcript, or RNA polymerase III sponges which drive very strong expression of abundant cellular RNA about <300 nucleotides which do not require an open reading frame. The latter allows microRNA to bind without affecting the translational process (Ebert et al., 2007). Sponges prevent regulation of the microRNA's natural targets through acting as competitive inhibitors. Supraphysiological levels of sponge transcripts are produced in the cell due the stability of transcript and strong promoters, which saturate the microRNA binding in the cell and inhibit regulation of natural target transcripts. The sponge can demonstrate active microRNA regulation by the mean fluorescence intensity (MFI) of the fluorescent marker protein encoded by the sponge (Gentner et al., 2010). The ratio of vector control MFI to the microRNA knockdown MFI gives the fold repression by microRNAs, and high MFI for a reporter protein in the microRNA knockdown cells indicates that the high levels of sponge transcript have bound all available microRNA and the remaining sponge transcripts are translated into the reporter protein (Brown and Naldini, 2009). In some cases the microRNA levels are far in excess of the transcript vector copies and continue to function against their natural targets as well (Almeida et al., 2012). However, several studies also found the number of sponge transcript copies in microRNA inhibited cells were several fold in excess of the microRNA copies per cell, so it appears the stoichiometry of this reaction is dependent on the abundance of the microRNA of interest in the tissue (Ebert et al., 2007; Gentner et al., 2009). It was expected using a sponge for knockdown would not alter the amount of endogenous microRNA in the cell, only prevent its activity by sequestering the bound microRNA, but the levels of microRNA decreased two-fold according to Northern blot analysis (Ebert et al., 2007).

In luciferase experiments comparing the knockdown of a microRNA with a sponge to ASOs and locked nucleic acids (LNAs), the sponge performed more effectively than ASOs and equal to LNAs (Brown and Naldini 2009). The cross-reactivity of the sponge is theorized to bind multiple members of a microRNA family, compared to ASOs that are usually designed to only target one member of a family (Brown and Naldini 2009). LNAs have less binding specificity due to their nucleotide base chemistry and could be a match for sponges in potency, but they have the disadvantage of more susceptibility to nuclease cleavage (Lennox and Behlke, 2010). As well, LNA delivery is limited to transfection into cells and have limited periods of activity, while sponges can be transduced and stably expressed over time via viral constructs and genomic integration.

Genetic knockout models of microRNA loss are highly effective in decreasing the amount of microRNA in the cells and demonstrating the importance of select microRNA in various biological processes. The ablation of certain microRNA can reveal their essential roles in embryonic development, such as deletion of the miR-1-2 gene leading to 50% embryonic lethality with cardiac defects (Zhao et al., 2007b), or deletion of the miR-17-92 cluster resulting in perinatal death with lung and heart defects (Ventura et al., 2008). Conditional knockout models can examine microRNA knockout in a tissue-specific or time dependent manner and isolate the microRNA targets involved in a particular tissue or system. The genetic knockout technique is highly versatile and allows broader physiological perspective of the mechanism behind the phenotype related to microRNA loss. However, genetic knockout models have a number of disadvantages when studying microRNA. Nearly 40% of microRNA could lead to inhibition of protein function through truncation of the protein amino acid chain, or by a decrease or lack of expression (Brown and Naldini, 2009; Kim et al., 2009).

While genetic knockouts can demonstrate the role of a microRNA in different systems, such as abnormal B-cell development in mice with knockout of miR-17-92 indicating the regulation of immunity-related proteins, sponges may provide greater opportunity to study the spatial and temporal regulation by a microRNA. In one study, Loya *et al* implemented a miR-9 sponge transgene in mouse that could be induced in different tissues of *D. melanogaster* to demonstrate the necessity of miR-9 in different stages of development (Loya et al., 2009). In another case, Gentner *et al* selectively allowed expression of *GALC* at later stages of hematopoiesis while inhibiting expression in HSPCs where the presence of the protein is toxic

49

by inserting a sponge sequence into the 3'UTR of *GALC* that binds HSPC-specific microRNA. The GALC activity in later stages of hematopoietic differentiation effectively treated globoid cell leukodystrophy, a rare metabolic disease, in a mouse model by using sponges and microRNA regulation as gene therapy (Gentner et al., 2010).

Genetic knockout and sponge knockdown models of microRNA loss can be similarly controlled through inducible and/or tissue-specific promoters in transgenic animals. Complete knockdown of the microRNA is more assured in genetic knockout models, and for microRNA loss on only one allele, haploinsufficient models can be achieved through use of animals with heterozygous knockout. ASOs are not ideal for study of microRNA roles in developmental and physiological models because of their short half-life and the difficulty in delivery to certain tissues, but the level of microRNA inhibition can be somewhat controlled through dosage in *in vitro* and *in vivo* models.

The use of microRNA sponges in genetic therapy is promising but still high-risk. The feasibility of the sponge strategy works partially because the change in expression mediated by microRNAs on natural targets is carefully regulated. A vector encoded transgene or genetic knockout may be difficult to regulate and stable gene transfer may be inhibited by trans-gene specific immunity (Brown et al., 2006), even in the previous example, where microRNA regulation is used to fine-tune expression of a therapeutic protein by suppressing translation in particular tissue (Gentner et al., 2010). This would be an ingenious use of microRNA sponges in a clinical setting, and could become more feasible in the future as knowledge of microRNA regulation in different tissues is expanded and the sponge can be designed to control for vector dose, transduction level, and promoter activity. However, it is highly difficult to be able to predict off-target effects in gene therapy. While sponge expression can be controlled in particular tissues by exploiting the regulation of endogenous microRNA and using a specific genomic promoter site for expression dependent on tissue, the sponge expression may be inhibited or increased by extenuating factors within the cell that come from incomplete knowledge of the genetic and environmental factors (Brown and Naldini, 2009).

1.4.3 microRNA Knockdown Model Leading to Investigation of Non-specific Binding and Novel microRNA Discovery

The initial hypothesis leading to this work was that loss of miR-143 in del(5q) MDS leads to dysregulation of proteins involved in hematopoiesis or pathways leading to oncogenic signaling. To test this hypothesis my first aim was to identify proteins with significant changes in expression due to knockdown of miR-143 and next to conduct pathway analysis and functional validation. My project began by investigating genes targeted by the putative tumor suppressor miR-143 for involvement in the pathogenesis of del(5q) MDS. A model of miR-143 knockdown in del(5q) MDS was developed in blood cancer cell lines using the sponge method for inhibition of miR-143 activity and expression, which was used to investigate the proteins undergoing change in response as described in Chapter 3. However, due to the small number of overlapping, significant changes in protein expression and the discovery of the chosen cell line expressing lower amounts of miR-143 than previously thought, the issue of non-specific binding to the miR-143 knockdown sponge was introduced.

My new hypothesis which was investigated through Chapters 4 and 5 of this thesis, is that unknown annotated or novel microRNAs could bind to a repetitive 7 nt sequence in the sponge through its seed site and regulate a set of protein targets distinct from that of miR-143. The ability of the sponge to bind unintended microRNAs as well as miR-143 leads to the possibility of two different sets of targets being differentially regulated simultaneously and leading to interference with each other. In Chapter 4 I describe my experiments to identify potential microRNA recognition elements in the sponge and searched within novel microRNA transcript sequence data to find any novel microRNAs with matching seed sites. Next, I characterized the potential novel microRNA that I discovered by analysis of expression in small RNA sequencing libraries, conservation of RNA secondary structure and microRNA targets in different species, and assays for biological activity. My last aim, discussed in Chapter 5, demonstrated that the non-specific binding effect could be eliminated by strategic mutation of nucleotides in the sites of the sponge binding to microRNA seeds, thereby providing a potential solution to reduce offtarget effects of this sponge knockdown approach to target microRNAs.

2. Materials and Methods

2.1 Cell Culture

Expression of miR-143 was examined by massively parallel small RNA sequencing and by quantitative real-time polymerase chain reaction (qRT-PCR) in a variety of cell lines. The initial cell lines described as used in small RNA sequencing were HL-60, MDS-L, UT-7, and THP-1. UT-7 was used for further study, and later the expression of miR-143 was examined by qRT-PCR in the AML-5, K562, MDS-L, MOLM-13, TF-1, THP-1, and UT-7 cell lines. The expression of miR-X was examined in all of the aforementioned cell lines, as well as KG-1, M07e, U937, SKBR3, COLO205, and HeLa cell lines. The examination of myeloid leukemia cell lines was explored because miR-X had been detected in myeloid leukemia patient samples, and the cell lines deriving from other tissues were selected to explore whether there was expression in other types of tissue. All cell lines were cultured in a humidified atmosphere containing 5% CO₂ at 37°C.

Table 2.1 - Cell lines used and culture conditions

Cell lines were acquired from the Deutsche Sammlun von Mikrooganismen und Zellkulturen (DSMZ), American Type Culture Collection (ATCC), and Kaoru Tohyama's lab. The media used are alpha-Modified Eagle Medium (alpha-MEM), Roswell Park Memorial Institute (RPMI) 1640, and Dulbecco's Modified Eagle Medium (DMEM). Supplements to the media include heat inactivated fetal bovine serum (FBS), 2 mM glutamine, 2 U/mL each of penicillin and streptomycin, and recombinant human proteins interleukin-3 (IL-3) and granulocyte macrophage colony-stimulating factor (GM-CSF).

Cell Line	Source	Culturing Media Conditions
AML-5	DSMZ	α-MEM, 20% FBS, 1% PSG, 5 ng/mL
(OCI-AML5)		
COLO-205	ATCC	RPMI-1640, 10% FBS, 1% PSG
K562	ATCC	RPMI-1640, 10% FBS, 1% PSG
KG-1a	ATCC	RPMI-1640, 10% FBS, 1% PSG
M07e	DSMZ	RPMI-1640, 10% FBS, 1% PSG, 10ng/mL hGM-CSF
MDS-L	K.Tohyama Lab	RPMI-1640, 20% FBS, 1% PSG, 10 ng/mL hIL-3, 0.5 mM β-
		Mercaptoethanol
MOLM-13	DSMZ	RPMI-1640, 20% FBS, 1% PSG
OCI-AML3	DSMZ	α-MEM, 20% FBS, 1% PSG
TF-1	ATCC	RPMI-1640, 10% FBS, 2 ng/mL hGM-CSF
THP-1	ATCC	RPMI-1640, 10% FBS, 1% PSG
U937	ATCC	RPMI-1640, 10% FBS, 1% PSG
UT-7	DSMZ	DMEM, 20% FBS, 1% PSG, 5 ng/mL hGM-CSF

For the culture conditions, the media alpha-Modified Eagle Medium (alpha-MEM) was obtained from Sigma-Aldrich (St. Louis, MO), the heat inactivated fetal bovine serum (FBS) from Hyclone (Logan, Utah), the glutamine, penicillin and streptomycin from Gibco, Invitrogen (Carlsbad, CA), and hGM-CSF and IL-3 from Stem Cell Technologies (Vancouver, BC, Canada).

2.2 Cloning of Constructs

The first miR-143 sponge in the pLentiLox3.7 (pLL3.7) vector was constructed from a pIDTSmart-KAN vector with a minigene containing the four-repeat sponge sequence (Integrated DNA Technologies (IDT), Coralville, IA). The four repeat sponge sequence consisted of four reverse complement sequences to miR-143 with 3 nucleotides in the middle, positions 10-12, changed to their complement bases (indicated in **bold**) and 1 nucleotide in position 13 removed from the sequence to create a gap (indicated by hyphen). These modifications were to produce a bulge in the middle of a miR-143/sponge RNA duplex and enhance stability. Each reverse complement was interspersed by 4-5 random nucleotides, and restriction enzyme digestion sites were added to the ends of the sequence (EcoRI and SphI in **Iowercase bold** at the 5' end and SphI, KpnI, and EcoRI in **Iowercase bold** at the 3' end). The sequence following the GFP coding region in pLL3.7 was as follows:

miR-143 sponge

5'-gaattc gcatgc GAGCTACAG-CGA<u>TCATCTCA</u>GCTA GgGCTACAG-CGA<u>TCATCTCA</u>TTAAG GAGCTgCAG-CAA<u>TCATCTCA</u>AACCT GAGCTACgG-CGA<u>TCATCTCA</u>atCA gcatgc ggtacc gaattc-3'

The sponge sequence was digested from its original vector backbone using EcoRI restriction enzyme (New England Biolabs, Ipswich, MA) and ligated into the 3'UTR of the GFP in the pLentiLox3.7 vector using T4 DNA Ligase (New England Biolabs, Ipswich, MA). This vector was designed and cloned by Dr. Joanna Wegrzyn-Woltosz.

The miR-X hairpin of the MND-miR-X vector, which encompassed 119 bases downstream of the miR-X mature sequence and 159 bases upstream (approximately 100-120 nucleotides on either side of the microRNA hairpin structure) was ordered as a gBlocks fragment (IDT, Coralville, IA). Restrictions enzymes sites for the enzymes AscI and PacI were added to the end of gBlocks fragment and the MND-miR-X vector was created by cloning of the gBlocks fragment into the MND-SVMCS-GFP vector. The hairpin insert and the MND-SVMCS-GFP vector were digested with AscI and PacI and ligated together with T4 DNA Ligase (New England BioLabs, Ipswich, MA).

Dual-luciferase reporter vectors used the pmiR-Glo vector as a backbone, with a variety of 3'UTRs cloned into the Multiple Cloning Site (MCS) (Promega, Madison, WI). For these constructs, 50 bp oligonucleotides containing the microRNA binding site and approximately 5-10

bp of upstream and downstream flanking sequence in the 3'UTR were annealed with their reverse complement oligonucleotides to create a short DNA insert sequence (IDT, Coralville, IA). The restriction enzyme digest site Sacl (lowercase) was added to the 5' end of the insert and the restriction site for Xbal (lowercase) was added to the 3' end of the insert through the design of the oligonucleotides. The oligonucleotide sequences were as follows:

miR-Xrevcomp-fwd	5'-TACgagctcCCAAGCAGAAGAGGCCTACAGAGTAAGGACtctagaCGA-3'
miR-Xrevcomp-rev	5'-TCGtctagaGTCCTTACTCTGTAGCCTCTTCTGCTTGGgagctcGTA-3'
miR-Xrevcomp-mut-fwd	5'-TACgagctcCCAAGCAGAAGAGGCCTAgtcAGTAAGGACtctagaCGA-3'
miR-Xrevcomp-mut-rev	5'-TCGtctagaGTCCTTACTgacTAGCCTCTTCTGCTTGGgagctcGTA-3'
TNPO1-3UTR-fwd	5'-TACgagctcGCAAGTTAAAAGCTACAGAGTGAAAGTtctagaTCG-3'
TNPO1-3UTR-rev	5'-CGAtctagaACTTTCACTCTGTAGCTTTTAACTTGCgagctcGTA-3'
TNPO1-3UTR-mut-fwd	5'-TACgagctcGCAAGTTAAAAGCTAgtcAGTGAAAGTtctagaTCG-3'
TNPO1-3UTR-mut-rev	5'-CGAtctagaACTTTCACTgacTAGCTTTTAACTTGCgagctcGTA-3'
SAMSN1-3UTR-fwd	5'-TTAgagctcACGCATTCCCAACTATATATCTCCAGACTGCATTCCATtctagaTCG-3'
SAMSN1-3UTR-rev	5'-CGAtctagaATGGAATGCATCTGTAGATATATAGTTGGGAATGCGTgagctcTAA-
	3'
SAMSN1-3UTR-mut-fwd	5'-TTAgagctcACGCATTCCCAACTATATATCCATCCATtctagaTCG-3'
SAMSN1-3UTR-mut-rev	5'-CGAtctagaATGGAATGCATgacTAGATATATAGTTGGGAATGCGTgagctcTAA-3'

The annealed DNA sequences containing the reverse complement of miR-X, the miR-X binding site in the SAMSN1 3'UTR, and binding site in the TNPO1-3'UTR were digested with SacI and XbaI and ligated into pmiR-Glo vector by T4 DNA ligase (New England Biolabs, Ipswich, MA). Three nearly identical vectors were also cloned by the same method, differing by 3 bp (in **lowercase bold underline**) in nucleotide positions 3-5 of the microRNA recognition element (MRE, in **bold underline**) to simulate mutated microRNA binding sites.

Three dual-luciferase reporter vectors were constructed using the pmiR-Glo backbone and annealed oligonucleotide DNA fragments containing two tandem repeats of the original sponge sequence, with two reverse complements to miR-143 with the aforementioned bulge in the middle and 4-5 random nucleotides in between (IDT, Coralville, IA). The oligonucleotides were designed to have sticky ends after annealing, without need for restriction enzyme digestion, and were ligated into the pmiR-Glo vector digested by SacI and XbaI with T4 DNA ligase (New England Biolabs, Ipswich, MA). In one construct, the sponge sequence had no mutations (Luc-2T-nomut), in one construct both miR-143 seed binding sites had mutations in nucleotides corresponding to positions 3-5 of the 5' end of the microRNA (Luc-2T-miR143mut), and in another construct both miR-X seed binding sites had mutations in nucleotides corresponding to positions 3-5 of the 5' end of the microRNA (Luc-2T-miR143mut), and in another construct both miR-X seed binding sites had mutations in nucleotides corresponding to positions 3-5 of the 5' end of the microRNA (Luc-2T-miR143mut), and in another construct both miR-X seed binding sites had mutations in nucleotides corresponding to positions 3-5 of the 5' end of the microRNA (Luc-2T-miR143mut), and in another construct both miR-X seed binding sites had mutations in nucleotides corresponding to positions 3-5 of the formation in nucleotides corresponding to positions 3-5 of the microRNA (Luc-2T-miRXmut). The oligonucleotide sequences are as follows:

Luc-2T-nomut- FWD	5'-cTCGAATGG CTACAGA CAA <u>TCATCTC</u> AGCCAGTG <u>CTACAGA</u> CGA <u>TCATCTC</u> ATCAAGTt-3'
Luc-2T-nomut- REV	5'- tctagaACTTGAT <u>GAGATGA</u> TCG <u>TCTGTAG</u> CACTGGCT <u>GAGATGA</u> TTG <u>TCTGTAG</u> CCATTCGAgagct c-3'
Luc-2T-miR143 mut- FWD	5'-cTCGATGAG <u>CTACAGA</u> CAA <u>TCATgag</u> AGCCAGTG <u>CTACAGA</u> CGA <u>TCATgag</u> ATGTAGTt-3'
Luc-2T-miR143 mut- REV	5'- tctagaACTACAT <u>ctcATGA</u> TCG <u>TCTGTAG</u> CACTGGCT <u>ctcATGA</u> TTG <u>TCTGTAG</u> CCATTCGAgagctc-3'
Luc-2T-miRX mut- FWD	5'-cTCGAATGG <u>CTAgtcA</u> CAA <u>TCATCTC</u> AGCCAGTG <u>CTAgtcA</u> CGA <u>TCATCTC</u> ATCAAGTt-3'
Luc-2T-miRX mut- REV	5'- tctagaACTT <u>GATGAGA</u> TGATCG <u>TgacTAG</u> CACTGGCT <u>GAGATGA</u> TTG <u>TgacTAG</u> CCATTCGAgagctc- 3'

The fixed sponges were ordered as gBlocks Gene Fragments of the original sponge sequence with the four tandem repeats of microRNA binding sites as shown previously (bulge nucleotides in **bold** and hyphen(-) to symbolize gap), the EcoRI, KpnI, and SphI restriction enzyme sites flanking the sequence (in **bold lowercase**), and three to four nucleotide substitutions in the seed sites of miR-143 for the miR-X fixed sponge or in the seed sites of miR-X for the miR-143 fixed sponge (all mutations to decrease non-specific binding in lowercase). The gBlocks fragments and pLL3.7 vector were digested with EcoRI restriction enzyme and ligated with T4 DNA Ligase.

miR-143specific sponge:

5'-**gaattc gcatgc** GAcgTAgtG-**TGA**<u>TCATCTCA</u>GCTA GgGCTtgAG-**CTA**<u>TCATCTCA</u>TTAAG GAagTgCAG-**CAA**<u>TCATCTCA</u>AACCT GAGCatCgG-AGA<u>TCATCTCA</u>atCA **gcatgc ggtacc gaattc**-3'

miR-Xspecific sponge:

5'-gaattc gcatgc CATCgCAcgT-AGG<u>GCTACAGC</u>TATa tTCgCtcTAA-GGA<u>GCTACAGC</u>AATC gTCagAAACC-TGA<u>GCTACAGAGA</u>TG AaCTCAatC-AGA<u>GCTACAGC</u>GAT gcatgc ggtacc gaattc-3'

The accuracy of the cloning was assessed by restriction digest accompanied by visualization of DNA fragments on agarose gel and by Sanger Sequencing at Genome Quebec in McGill University.

2.3 Lentivirus Production

293T cells were seeded on plasma-treated 100 mm tissue culture plates (Sarstedt, Newton, NC), at a density of 5-6.5 x 10^6 cells per plate in DMEM media supplemented with 10% heat inactivated FBS, 2 mM glutamine and 100 U each of penicillin and streptomycin (Gibco, Invitrogen, Carlsbad, CA). On the second day the media was changed 2-4 hours before transfection, adding 7.0 mL to each plate. The transfection was performed by combining the following amounts of packaging and envelope vectors with H₂O up to 450 µL: 6.5 µg of Rev Repsonse Element (RRE), 2.5µg of REV viral protein (REV), 3.5µg of Vesicular stomatitis virus G protein (VSVG) and 10µg of the lentiviral construct being packaged. The 450 µL mixture of H_2O and DNA plasmids was gently combined with 50 μ L of 2.5 M CaCl₂ in a polypropylene tube. The DNA-CaCl₂ mixture was left to sit for 2-3 minutes and then added dropwise to 2x HBS solution (2.8 mL 5 M NaCl, 0.15 mL 0.5 M Na2HPO4, and 5 mL 1 M HEPES) to make a 1:1 solution and left to sit for five minutes. The CaPO₄-DNA solution was added to the 293T cells and incubated for 18-20 hours. On the third day of the procedure, 5.0 mL of fresh media per plate was exchanged for the old media. On the fourth day of the procedure, the viral supernatant was collected and filtered through a 0.45µm low protein binding Millipore filter. The viral supernatant was centrifuged at 25,000 rpm for 90 minutes at 4°C in a Beckman Coulter L8-60M ultracentrifuge. Aspiration of the supernatant isolated the viral pellet, which was suspended in 100 µL of DMEM media with 5% DNase. The viral pellet was shaken in this solution for 45 minutes, then divided into 25 µL aliquots and stored at -85°C. On the fifth day, this viral collection was repeated.

2.4 Reverse Transcription of RNA and Real-Time Quantitative PCR

2.4.1 Exigon RT-qPCR for microRNA

Total RNA was extracted using TRIzol (Invitrogen) and 200ng reversed transcribed using the Exiqon miRCURY LNA[™] Universal cDNA Synthesis Kit, both according to manufacturer's instructions (Exiqon, Copenhagen, Denmark). Synthesized miR-143 oligonucleotides from IDT were used to make five standards of 100, 10,000, 100,000, and 1,000,000 copies. The miRCURY LNA[™] microRNA PCR primers for miR-143 and control set (Exiqon, Copenhagen,

Denmark), and SYBR Green (Invitrogen) were implemented for the RT-qPCR according to the manufacturer's instructions.

2.4.2 TaqMan RT-qPCR for microRNA

Total RNA was extracted using TRIzol (Invitrogen) and 100 ng used for quantification of each sample. Quantification of microRNA species using the TaqMan® Small RNA Assays was performed with two-step RT-qPCR. In the reverse transcription (RT) step, cDNA was reverse transcribed from total RNA sample using small RNA-specific, stem-loop RT primers and reagents from the TaqMan® Reverse Transcriptase kit according to manufacturer's instructions (ThermoFisher Scientific). The looped RT primer bound the 3' end of a mature microRNA, allowing the synthesis of the cDNA strand to be carried out in the 5' to 3' direction (Figure 2.1).

A forward primer specific to the small RNA of interest was extended by DNA polymerase to form the second cDNA strand. During PCR amplification, the TaqMan MGB probe anneals to a complementary sequence on the small RNA, between the forward and reverse primer sites. The proximity of the reporter dye to the quencher dye was disrupted by DNA polymerase cleavage of the probe, leading to fluorescent signal of specific amplification.

The TaqMan® U47 small RNA Assay, one of the recommended probes for normalization, was used as a control for the amount of input RNA in each TaqMan assay. To find the number of copies for miR-143 or the novel microRNA, miR-X, synthetic RNA oligonucleotide copies of the microRNA sequences with concentrations of 1 nmol/L, 0.1 nmol/L, 0.001 nmol/L, 0.0001 nmol/L, and 0.00001 nmol/L were used as standards. The slope of the linear Ct to -log (nmol/L) and the molecular weight of the oligonucleotides were used to calculate the number of copies per 100 ng of total RNA input.
Step One: Reverse Transcription





Looped-primer RT-qPCR is used to detect mature microRNA. A stem-loop primer binds to the 3' end of the microRNA, the 3' of the primer is extended to make cDNA in the reverse transcription step. In real-time PCR, the forward primer complementary to the microRNA sequence is used to synthesize the second cDNA strand. In PCR amplication, the forward primer binds to the microRNA sequence, the reverse primer binds to a sequence in the loop primer cDNA region, and the TaqMan probe binds to a sequence between. The DNA polymerase cleaves the TaqMan probe when the strand is extended, which separates the quencher (Q) from the fluorescent probe (F) and causes fluorescence signal.

2.5 Immunoblotting

Cells were lysed in RIPA buffer (PBS containing 1% NP-40 (Sigma-Aldrich, St. Louis, MO), 0.5% sodium deoxycholate, and 0.1% SDS (Sigma-Aldrich, St. Louis, MO)) with addition of fresh protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN). 50 µg of total protein, as measured using Bio-Rad DC Protein Assay System (Bio-Rad Laboratories, Hercules, CA),

were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis, transferred to nitrocellulose membranes (Bio-Rad Laboratories, Hercules, CA), and developed by enhanced chemiluminescence (PerkinElmer Life Sciences, Boston, MA). Membranes were probed with the following antibodies: 1:1000 rabbit anti-SMAD3, ab28379 (Abcam, Cambridge, MA), 1:10,000 mouse anti-GAPDH (Sigma-Aldrich, St. Louis, MO), 1:2000 rabbit anti-NUMA1, ab36999 (Abcam, Cambridge, MA), 1:2000 rabbit anti-PSME1, PW8185 (Enzo Life Sciences, Farmingdale, NY), 1:1000 mouse anti-KRAS, ab55391 (Abcam, Cambridge, MA), 1:1000 rabbit anti-HK2 (C64G5), 2867 (Cell Signaling Technology, Danvers, MA), 1:1000 rabbit anti-DNAJA2, 12236-1-AP (Proteintech, Chicago, IL), and 1:1000 rabbit anti-STAG2, ab155081 (Abcam, Cambridge, MA).

2.6 Quantitative Proteomics Methodology and Statistical Analyses

2.6.1 Stable Isotope Labelling of Amino Acids in Cell Culture and Gene Transfer

The stable isotope labeled amino acids used for this experiment were deuterium labeled lysine and ¹³C labeled arginine, which were added to lysine and arginine free DMEM or RPMI 1640 media. The two cell lines grown in Stable Isotope Labelling of Amino Acids in Cell Culture (SILAC) media were UT-7 and KG-1a. The cells were first tested for their rate of incorporation of the labeled amino acids into protein chains by taking samples of cell culture on the first, third, fifth, and seventh day in culture of the labeled media and running these samples on the nanoelectrospray-MS/MS (NanoES-MS/MS). The labeled amino acids have a higher mass and cause a shift in the mass/charge ratio of the peptide from the heavy labeled media, so there are two peaks or fragmentations patterns for the same peptide. Four or five peptides each from different common proteins in each sample increased in their heavy-to-light ratio over time, and with the light peptide's near disappearance on the seventh day, the incorporation was considered complete.

A range of 200,000-500,000 UT-7 cells were seeded in labeled DMEM media and non-labeled DMEM media and grown for ten days. The non-labeled cells were infected with the pLL-GFP lentivirus and the labeled cells were infected with pLL-miR-143spg lentivirus. Lentiviral infections were performed in 6-well suspension culture plates, with $1x10^6$ cells/well, 1.5 mL media with 8 µg/mL polybrene, and 2-3 µL of superconcentrated virus. Fresh media was exchanged for the viral media 16 hours later, and 48 hours following this, the infected cells were

sorted and collected by the population expressing GFP. The sorted cells were replaced into SILAC media and allowed to readjust and expand for 24-48hrs.

The cells were lysed in RIPA buffer (PBS containing 1% NP-40 (Sigma-Aldrich, St. Louis, MO), 0.5% sodium deoxycholate, and 0.1% SDS (Sigma-Aldrich, St. Louis, MO)) with addition of fresh protease inhibitor cocktail (Roche Applied Science, Indianapolis, IN). The protein concentrations of the heavy miR-143spg lysate and the light GFP lysate were checked and 75 µg of each were combined together and run on a 10% polyacrylamide gel. The gel was stained with Coomassie blue to visualize the protein bands, and the gel was cut into 23 separate bands. Each band was chopped up into 1 mm by 1 mm cubes, and the gel pieces were washed and treated with 10 mM DTT in 100 mM NH₄HCO₃ solution to reduce the proteins and with 55 mM iodoacetamide in 100 mM NH₄HCO₃ solution to block the thiol group of cysteine and prevent disulfide bonds. Any remaining Coomassie Blue stain was washed out, and the pieces were treated with 1 μ g/uL trypsin overnight at 37°C in the Eppendorf Thermomixer Comfort (1.5 mL) at 600 rpm. The peptides from the digested proteins were cleaned and concentrated using stopand-go-extraction tips (Stage-tips), which consist of reverse-phase C18-bonded silica disks within 200 µL pipette tips. The concentrated peptide samples were run on the Fourier Transform-Ion Cyclotron Resonance mass spectrometer at the Biomedical Research Centre at the University of British Columbia. The proteomics data was analyzed using MaxQuant, Version 1.0.13.13, which identified and quantified the H/L ratio for each protein.

2.6.2 Analysis of Probability of Significant Change in Protein Expression

An Empirical Bayes method used multivariate statistics that integrated the results of 3 replicate experiments to find the differentially abundant proteins and their statistical significance, and evaluate the non-Gaussian tails or regions of data sparsity within the proteomics data. It modeled log₂ SILAC protein ratio values from replicate experiments and inferred the full shape of class-conditional probability distributions for biologically relevant proteins versus background. This method fitted a flexible model to the dense central data region, and constrained the tails to be fit by a parametric model. For the data, each experiment measured ratio values from three sets of samples, the SILAC1 dataset, the SILAC2 dataset and a set of proteins from untransduced cells grown in heavy and light media and combined. According to the Empircal Bayes method implemented, each experiment was modeled separately and the statistics were

subsequently integrated. There is assumed to be a 3-class model of SILAC ratio values, representing upregulated, downregulated and not differentially regulated, upon microRNA overexpression or repression.

2.6.3 Selective Reaction Monitoring

UT-7 cells were grown and lentiviral transductions were performed as previously described, with pLL-GFP and pLL-miR-143spg lentiviruses. The cells were lysed in RIPA buffer (PBS containing 0.5% sodium deoxycholate, and 0.1% SDS (Sigma-Aldrich, St. Louis, MO)) without NP-40 and protease inhibitors. The protein concentrations of the pLL-miR-143spg lysate and pLL-GFP lysate were assayed using Bio-Rad Protein Assay and 75 µg of each were diluted in 4 times in 100% ethanol. The solution was brought to 50 mM NaCH₃COO, pH 5 by using 2.5 M NaCH₃COO and 20 μ g of glycogen was added to each sample. The solutions were mixed and allowed to stand at room temperature for 3 hours. The pellet was precipitated by centrifuging for 10 minutes at 17000 x g and the supernatant removed by aspiration. The pellet was resuspended in digestion buffer (1% sodium deoxycholate, 50 mM NH₄HCO₃), and heated to 99°C for 5 minutes. This was followed by addition of DTT to a concentration of 10 mM with incubation at 37°C for 30 minutes and iodoacetamide to a concentration of 55 mM with incubation at 37°C for 30 minutes. An amount of 1 ug of trypsin per 35 ug of protein was added and the digestion was carried out overnight at 37°C. Samples were stage-tipped and run on 4000 QTRAP LC/MS/MS instrument (MDS SCIEX, Applied Biosystems) in the Charles and Elaine Shnier Mass Spectrometry Instrument Suite at the GSC. The area under the peak for ACTNA1 was used to normalize input between the two samples and the fold-change for each peptide was calculated using the area under the peak.

2.6.4 Preparation, Collection and Analysis of Proteomics Data with Modified Sponges

UT-7 cells were cultured in normal media and SILAC media, with D_4 -L-Lysine and ${}^{13}C_6$ -L-Arginine (see Section 2.6.1). However, in this case, the cells grown in normal media were transduced with one of four lentiviruses, and the cells grown in SILAC media were untransduced. For the transduced cells in normal media, two samples were transduced with

pLL-GFP, two with the original pLL-miR-143spg (called pLL-orig-spg), three with pLL-miR-Xfixedspg, and three with pLL-miR-143fixedspg. The untransduced UT-7 cells grown in heavy media were grown alongside the transduced cultures and added in equal amounts to the lysate from the transduced samples. The heavy, untransduced lysate was used as input normalization for the protein samples from transduced cells grown in light media.

The protein samples were prepared using the Single-Pot Solid-Phase-enchanced Sample preparation (SP3) protocol developed by Christopher Hughes, Lars M. Steinmetz, and Jeroen Krijgsveld (Hughes et al., 2014). The cells are lysed with 50mM HEPES (pH 8.5) with 1% SDS, 1X cOmplete Protease Inhibitor Cocktail (Roche), and treated with benzoase to shear chromatin, protein lysates are treated with 200 mM dithiothreitol in 50 mM HEPES at pH 8.5 and incubate for 30 minutes at 45°C, then 400mM iodoacetamide in 50mM HEPES pH8.5 with incubation for 30 minutes at 24°C in the dark. The protein lysate was then digested by binding to Sera-Mag Speed Beads and acidification with 1% formic acid. The samples were treated with trypsin in a 1:25 enzyme to substrate ratio and incubated for 14 hours at 37°C. The peptides were recovered from the beads by reconstitution of beads in 2% DMSO in water and sonication, followed by collection of the supernatant.

Ten samples were processed, run on the Orbitrap Fusion (ThermoScientific) in the Charles and Elaine Shnier Mass Spectrometry Instrument Suite at the GSC, and analyzed using Proteome Discoverer, Version 1.4. Peptides that were identified as contaminants (according to the library of common contaminant peptides in Proteome Discoverer) and/or non-unique were filtered out by the software, those with more 5 or more blank values out of the ten samples were removed, the peptides with multiple peptide spectrum matches (PSMs) were aggregated, the abundance values of each peptide in each sample were normalized, and these values aggregated by the protein accession numbers. Using the expression data for each protein in the samples and controls, the ratio of expression in each sample compared to the average expression in the two controls, pLL-GFP1 and pLL-GFP2 was determined. The $log_2(x)$ function of these ratio values was then calculated and returned to give the log_2 fold-change in expression in each sample compared to the average in each sample compared to the average of the average of the controls. The abundance value given after these are aggregated is the average of the abundance values from the respective PSMs.

The normalization of the abundance values accounts for the differences in sample amounts so, for example, if one sample was twice the amount of another I would expect to see a higher abundance of all peptides in that sample. To normalize, the abundance of all of the peptides in each sample are summed to give a Total Peptide Abundance value. A normalization factor is calculated (the highest Total Peptide Abundance value is found and then that value is divided by the Total Peptide Abundance value for each sample). The abundance of each peptide in each of the samples is then multiplied by the corresponding normalization factor to bring them up to the abundance I would expect to see if all samples had the same amount as the one with the highest Total Peptide Abundance. So the abundance values for each peptide Abundance because, by definition, this sample had a normalization factor of 1.

After the log₂ fold-change for each protein in the 8 sponge-infected samples was calculated, the dataset was checked for correlation using ggplots and gGally packages in R. The datasets showed low Pearson correlation coefficients, though two samples, pLL-miR-Xfixed3 and pLLmiR-143fixed3 showed higher correlation. Since these samples were grown in culture at an earlier time point, using media that had a longer shelf life than the other six samples and their heavy-labeled normalization controls, these samples were excluded from further analysis. The genes in the remaining 6 samples were filtered for high levels of variance. Genes with a log₂ fold-change difference of greater than 0.5 in the pLL-Orig.Spg duplicates, or greater than 0.3 difference in pLL-miR-Xfixedspg or pLL-miR-143fixedspg duplicates were excluded. The correlation coefficients were considerably improved after filtering, and a linear regression model was applied to the 3000 proteins passing this filter using the limma R package. The genes with a significant difference between the pLL-Orig Spg log₂ fold-change and pLL-miR-Xfixedspg log₂ fold-change or between the pLL-Orig.Spg log₂ fold-change and pLLmiR-143fixedspg log₂ foldchange were determined from this linear regression model. Proteins with relevant expression patterns between the three different sponge conditions were selected using the R subset function and Venn diagram overlap.

2.7 Flow Cytometry

For the pLL-GFP and pLL-miR-143spg infected UT-7 cells cultured in SILAC media in Pt I, fluorescence activated cell sorting (FACS) was performed on the FACS 440 instrument from

Becton Dickinson. The cells positive for the GFP marker and negative for propidium iodide (PI) a fluorescent molecule that permeates non-viable cells and intercalates into DNA - were collected in phosphate-buffered saline solution (PBS), and resuspended in regular UT-7 growth media. Sorting to separate fluorescently labeled cells was later performed on the Influx II and Aria III BD FACS instruments in the FlowCore Facility in the Terry Fox Laboratory at the BC Cancer Research Centre.

Transfection efficiency of the microRNA hairpin mimics was tested by a fluorescent control mimic, Dy547. The percentage of fluorescent cells was determined by setting a gate according to fluorescence of the negative, untransfected cells and then running the sample containing cells transfected with the fluorescent control mimic and observing how many cells from this sample exhibited fluorescence greater than the gating around the negative population.

2.8 Transfection of microRNA Anti-Sense Oligonucleotide Inhibitors and microRNA Mimics

IDT MicroRNA Inhibitors are steric blocking oligonucleotides designed to hybridize to mature microRNAs, containing 2'-O-methyl residues with ZEN[™] modifications, were ordered from Integrated DNA Technologies. When used in luciferase assays measuring the regulation of Firefly luciferase transcripts, the UT-7 or AML5 cells were co-transfected with the miR-X(revcomp) pmiR-Glo vector and microRNA inhibitors using TransIT-X2 reagent and protocol (Mirus Bio).

The microRNA mimics are commercially synthesized microRNA hairpins (ThermoFisher Scientific). For measurement of luciferase regulation by mimics in 293T cells, the mimics were transfected into the cells using the TransIT-SiQuest reagent 24 hours post-transfection of the Luc-2T-nomut construct by TransIT-LT1 reagent (Mirus Bio). For transfection into K562, AML5, UT-7, and OCI-AML3 cell lines to assay target protein regulation by immunoblotting, the TransIT SiQuest reagent and protocol were used. For measurement of luciferase regulation in COLO205 and AML5 cells, the Luc-2T-nomut, Luc-2T-miR143mut, and Luc-2T-miRXmut were cotransfected with the control, miR-X, and miR-143 mimics using the TransIT-X2 reagent and protocol (Mirus Bio).

2.9 Luciferase Assays

All of the luciferase assays were performed using modifications of the pmiR-Glo vector and the Dual-Luciferase Reporter Assay kit (Promega, Madison, WI) and the luminescence values were gathered on the Victor-X3 Plate Reader (Perkin Elmer). The derepression of pmiR-Glo vectors containing the miR-X reverse complement, SAMSN1 or TNPO1 binding sites were transfected using TransIT-2020 (Mirus Bio) into UT-7 cells which had previously been transduced with pLL-GFP or pLL-miR-143spg lentivirus.

2.10 Multiple Sequence Alignment and Target Prediction

The multiple sequence alignment (MSA) was performed using sequence for selected mammalian genomes acquired from the UCSC Genome Browser, Multiz Alignment track (Kent et al., 2002). The multiple sequence alignment algorithm T-Coffee was used for examining the conservation and sequence identity between genomes, using the webserver for T-Coffee (Di Tommaso et al., 2011; Notredame et al., 2000). Promoter analysis for transcription start sites for RNA polII was performed using the algorithm and webserver Promoter 2.0 (Knudsen, 1999). The prediction of targets of microRNAs was achieved using the TargetScan, version 5.1, algorithm, or for the potential microRNA seed sites, TargetScan Custom, version 5.0, algorithm was used (Chiang et al., 2010).

2.11 Statistical Tests

To determine significant variation between pLL-GFP and pLL-miR-143 samples in TaqMan RTqPCR, unpaired t-tests with Welch's correction were performed using GraphPad Prism, version 6 (Chapter 3.2 and 3.3). Variation between luciferase derepression in pLL-GFP and pLL-143spg samples was also determined using unpaired t-test with Welch's correction. Variation between samples treated with the miR-X inhibitor and control inhibitor in UT-7 cells was also assessed using an unpaired t-test with Welch's correction. The significance in variation between miRmimics regulation of Luc-2T-nomut in 293T cells was assessed by unpaired t-test with Welch's correction. In COLO205 and AML5 cells, the significance of variation between regulation of Luc-2T constructs was measured by one unpaired t-test per row (for each of the three constructs) for treatment with control (CTR) mimic versus miR-X mimic (ie: whether there was significant difference between the CTR mimic and miR-X mimic conditions in Luc-2T-nomut, the Luc-2T-miRXmut, or the Luc-2T-miR143mut cells), and by one unpaired t-test per row for treatment with CTR mimic versus miR-143 mimic. This was a multiple t-test for each miR-mimc treatment comparison, performed with False Discovery Rate (FDR) approach and the desired FDR (Q) set to 1%. This was also performed using GraphPad Prism. Linear regression analysis of the proteomics datasets in Chapter 4 is described above in Section 2.6.4, Preparation, Collection and Analysis of Proteomics Data with Fixed Sponges.

3. Model of miR-143 Knockdown in Blood Cancer Cell Lines

3.1 Introduction

MicroRNA are essential regulators of gene expression and pair with conserved sequence motifs found in over 60% of 3'UTRs in coding genes, as well as the protein coding domains and 5'UTR of many transcripts (Friedman et al., 2009; Lewis et al., 2005). There are over 2500 microRNA genes within the human genome and their loss or gain through genomic alterations is found in numerous diseases, as the large number of genes targeted by a single microRNA can lead to the disruption of many cellular functions (Bartel and Chen, 2004; Hsu et al., 2006; Lim et al., 2005). One of the many complex biological processes controlled by microRNA is hematopoiesis, where coordinated networks of microRNAs and transcription factors regulate gene expression programs in each stage of blood cell differentiation.

According to previous studies performed in our lab, loss of microRNAs in the critically deleted region of chromosome 5 contribute to the pathogenesis of del(5q) MDS and dysregulation of their targets TIRAP and TRAF6 lead to abnormal hematopoiesis (Starczynowski et al., 2010). The co-transcribed miR-143 and miR-145 are tumour suppressors and studies in other forms of cancers have found lower expression of miR-143 and -145 compared to normal tissue (Akao et al., 2007a; Akao et al., 2007b). My lab has shown that loss of miR-143 and -145 leads to downregulated expression in the del(5q) subtype of myelodysplastic syndromes, suggesting they may play a similar role in the development of del(5q) MDS. I initially posited that haploinsufficiency of miR-143 in MDS leads to derepression of oncogenic signaling pathways and/or genes involved in hematopoiesis. In the process of finding targets of miR-143 and determining whether any of the derepressed targets were involved in oncogenic pathways or dysregulation of hematopoiesis, an interesting question arose about the method of microRNA knockdown.

Initial findings based on this experimental design showed that the degree of changes and correlation of changes between two replicates was low. This is likely due to biological and technical variation between proteomics experiments. However, this method of finding microRNA targets was somewhat unique in that it looked at targets, or proteins that were upregulated, after knockdown of a microRNA, rather than searching for downregulated proteins after microRNA transfection, as is commonly done. The knockdown was performed by a lentiviral sponge, which

had four binding sites of the reverse complement of miR-143. However, microRNA binding is often dependent on matches between the target and the first eight nucleotides of the 5' region of the microRNA and complementarity in the 3' region is not essential. During the course of these studies it was found that levels of miR-143 in the model cell line were lower than initially gauged, and led to the hypothesis that conditions were optimal for non-specific microRNAs to be bound by the lentiviral sponge. This chapter introduces our model of miR-143 knockdown in a leukemic cell line and the preliminary data collected, and the impetus to investigate non-specific binding of novel microRNAs.

3.2 Modeling Loss of miR-143 and Differential SILAC Proteomic Analysis

Sponges, or decoys, against microRNAs have been used to reduce expression of microRNAs in a variety of cell types in an analogous manner to how microRNAs regulate gene expression. As discussed in the Introduction, sponges give similar inhibition of microRNA expression as methods utilizing transfected anti-sense oligonucleotides and are stably expressed in the cell after retroviral or lentiviral transduction and genomic integration.

A lentiviral sponge was designed with the 3'UTR of GFP consisting of four tandem repeats of a miR-143 binding site, a target sequence complementary to the miR-143 sequence in miRBase, with spacers of four to five random nucleotides between repeats (Figure 3.1a). The most effective sponges used in previous studies implemented microRNA binding sequences with mismatched bases opposite to position 9-12 of the microRNA, called bulged sites (Ebert et al., 2007; Gentner et al., 2009). In humans, most microRNA duplexes have central mismatches or imperfect complementarity to facilitate recognition by Argonaute proteins and RISC-loading (Liu et al., 2004; Xia et al., 2013a; Xia et al., 2013b).

At the time of the project's outset, there was controversy over the primary mode of microRNA regulation, through translational inhibition or mRNA degradation, and the degree of change observed at the protein level or mRNA level. Proteomics was the chosen method of evaluating the effects of loss of miR-143, because changes in protein expression could be a consequence of translational inhibition and mRNA degradation, and effects due to post-transcriptional silencing or regulation could be seen.

The UT-7 blood cancer cell line was chosen to look for targets of miR-143 because it allowed evaluation of expression changes by quantitative proteomics in an abundant tissue similar to the cellular environment of interest and as a more homogenous system compared to primary tissues (Nagaraj et al., 2011). Stable Isotope Labeling of Amino Acids in Cell Culture (SILAC) was used to compare protein expression between miR-143 knockdown cells and cells with the GFP control. SILAC is a quantitative proteomics technique in which cells are grown in either normal, or "light," media or in "heavy" media where some of the amino acids have been replaced with stable-isotope labeled amino acids. Cells grown in light media were transduced with pLL-GFP control and cells grown in media with D₄-L-Lysine and ¹³C₆-L-Arginine amino acids were transduced with pLL-miR-143spg. The protein lysates from the control and miR-143 knockdown cells were combined together in equal amounts, trypsin digested, and the concentrated peptides were run an FT-ICR mass spectrometer at the Biomedical Research Centre at UBC. The differences in protein expression between the control and knockdown cells were compared by the ratio of peptides abundances in the light versus heavy forms. The SILAC method was chosen because the samples are combined together at the protein lysate stage, before treatment to reduce disulfide bonds or the trypsinization of proteins into peptides, and the method therefore lessens the amount of variation between the two conditions due to any difference in technical treatment.

I transduced the UT-7 cell line with pLL-miR-143spg or pLL-GFP lentivirus, measured knockdown of miR-143 by TaqMan microRNA qPCR, and detected decreased levels of miR-143 in cells transduced with pLL-miR143spg compared to cells transduced with pLL-GFP, as expected (Figure 3.1b). The decrease in expression was almost 70%, indicating that knockdown of miR-143 had occurred. The remaining detection of miR-143 in the pLL-miR-143spg transduced cells could also potentially occur due to the sequestering of the AGO-GW182 complex with the sponge and miR-143 species into P-bodies. Repressed target mRNAs that are not degraded directly after incorporation into RISC sometimes accumulate in these cytoplasmic foci, where enzymes involved in mRNA degradation are concentrated (Jackson and Standart, 2007). The isolation of RNA from the transduced cells could free the miR-143 sequestered in P-bodies and contribute to the miR-143 still seen in the qPCR experiment.

Before performing SILAC proteomics to evaluate the proteome for changes in expression, a marker protein was needed to demonstrate that knockdown of miR-143 causes changes in the proteome. Numerous changes in protein expression occur after knockdown of a microRNA, one

being the upregulation of direct, MRE-containing targets of the microRNA. A variety of tools exist to predict direct targets of microRNAs through computational algorithms and multiple databases with validated predicted targets from the literature. Following measurement of miR-143 knockdown, we performed immunoblotting for SMAD3, which was a predicted target of miR-143 according to TargetScan, and a member of the TGF β signaling pathway and a potentially relevant pathogenesis-related target. SMAD3 was increased in UT-7 cells transduced with pLL-miR-143spg compared to cells transduced with pLL-GFP according to measurement by immunoblotting (Figure 3.1c,d). Another predicted target of miR-143 was the protein KRAS, which also increased in expression in UT-7 cells transduced with the miR-143 sponge. The levels of miR-143 in UT-7 were greater than 90,000 RPM in preliminary data from Solexa small-RNA sequencing performed at the Genome Sciences Centre in the BC Cancer Research Centre. The high levels of expression were thought to lead to greater changes in the proteome when the microRNA was knocked down. Since qPCR experiments showed significant knockdown in microRNA expression and upregulation of predicted protein targets was observed, the SILAC experiment to measure changes in protein expression was carried out using UT-7 cells.





a. Schematic of miR-143 knockdown lentiviral vector, pLL-miR-143spg. The sponge has four tandem repeats of miR-143 binding sites, separated by 4-5 different random nucleotides after each repeat. b. RT-qPCR of miR-143 in UT-7 cells virally transduced with the miR-143 knockdown construct (mean ± SD, n=3, unpaired t-test with Welch's correction). c. Immunoblotting of lysates from UT-7 cells transduced with pLL-miR-143spg revealed upregulation of SMAD3 and KRAS. d. Immunoblots from three replicate experiments were quantified by densitometry. (Bars show mean ± SD).

Two SILAC datasets were collected, with 1561 proteins identified in the first data set, SILAC1, and 946 proteins in the second, SILAC2. Each protein was identified and quantified in the dataset by detection of two or more unique peptides with sequences matching part of the

protein sequence. The change in expression was denoted by the heavy-to-light ratio, in which the ratios of peptide intensities (or abundances) in heavy form were compared to intensities in light form and averaged for each protein (Figure 3.2a). The SILAC1 and SILAC2 datasets had different numbers of proteins undergoing changes in expression at higher levels (Figure 3.2b). There was considerable overlap in the proteins identified between these two proteomics datasets, with 902 proteins found in both (Figure 3.2c), and many proteins had variation in the degree and direction of expression changes between datasets. The linear association of the 902 common proteins reflects this, with a positive slope but a Pearson correlation coefficient of 0.259 (Figure 3.2c). The low correlation value derives from proteins with high variation in expression between the two datasets.

Proteomics data may suffer from low correlation between biological replicates due to sample heterogeneity. Proteins have higher stability and longer half-life on average than mRNAs, so measuring proteins has some advantages in consistency. The pLL-GFP and pLL-miR143-spg cells are serum starved following transduction to provide some synchronization in cell cycle and decrease heterogeneity in this respect, but serum starvation is only performed for 1 hr to minimize cell death, and many proteins with longer half-lives may not be affected by this treatment. There may also be differences between cells, such as energy state or concentration of regulatory molecules (Viney and Reece, 2013).

There may also be variation between replicates due to technical variance such as instrument variation, peptide digestion and extraction, and long-term drift of the instrument. One of the reasons for the low correlation between our samples was likely due to the peptide digestion/ extraction and analysis by the FT-ICR instrument at time intervals that were several weeks apart, and possible fluctuation in response of the instrument (Piehowski et al., 2013).



Figure 3.2 - Overlap and correlation between two miR-143KD proteomics experiments a. Schematic of SILAC quantitative proteomics experiment finding proteins with differential expression between pLL-GFP and pLL-miR-143spg - two SILAC datasets were collected. b. The proteins in SILAC1 (purple) and SILAC2 (orange) undergoing different degrees of expression changes vary between datasets. c. Of the 1562 proteins observed in SILAC1 (purple) and the 946 proteins observed in SILAC2 (light orange), 902 proteins were found in both datasets (dark orange). d. The Pearson correlation coefficient for log₂ fold changes of the 902 overlapping proteins between SILAC1 (y-axis) and SILAC2 (x-axis) is 0.259, with a p-value of 5.11×10⁻¹⁵.

The key to finding proteins changed in expression due to the knockdown of miR-143 in the cells was to look for proteins that shared the same expression trend (upregulation or downregulation) between datasets and had significant changes in expression. Our task was simultaneously to

find proteins with similar trends in expression in both SILAC1 and SILAC2, while also establishing the proteins with significant differential expression, excluding those with natural variation due to protein turnover rates, translational efficiency, etc. Assessing significant changes in protein abundance between conditions is often performed by a standard 2-sample t-test to compare relative or absolute abundance for each peptide or protein. However, this type of testing does not lend itself to the variation often seen in proteomics data, or to the range of changes in differential expression analysis. I will discuss how similar methods used in determining differential expression of transcripts in microarrays can be implemented in proteomic differential expression, and how this led to our statistical analysis of choice.

3.3 Analysis of Significant Biological Changes in Quantitative Proteomics Dataset

Early assessment of differential expression in microarrays used fold-change alone (DeRisi et al., 1996; McCarthy and Smyth, 2009; Schena et al., 1996), but these methods had low reproducibility and the use of Wilcoxon tests, Empirical Bayes, and other statistical means were developed (Baldi and Long, 2001; Efron and Tibshirani, 2002; Smyth, 2004; Tusher et al., 2001; Wright and Simon, 2003). With these new methods, small fold changes could be considered statistically significant, and fold-change cut-offs were implemented as criteria as well. The biological significance of a fold-change is likely to depend on the gene and on the experimental context, so differentially expressed genes with a required p-value (<0.01 or <0.05) were ranked by fold-change cut-offs of 1.5, 2, or 4 (Patterson et al., 2006)(Patterson et al 2006) or in some experiments, a fold-change of 1.3 and a p-value of less than 0.2 were required (Huggins et al., 2008)(Huggins et al 2008).

Proteomics experiments can be analyzed in a similar way to microarray or transcriptome experiments. However, the degree of change may be slightly less depending on the cause of differential expression. To find which H/L log₂ fold changes constituted significant expression changes and determine the biologically relevant proteins within the two datasets, an Empirical Bayes method developed by Margolin A *et al* was implemented (Margolin et al., 2009). This method accounted for non-Gaussian tails in the distribution of H/L log₂ fold changes and used multivariate statistics to find the differently abundant proteins. A flexible Gaussian model was fitted to the dense central data region and the tails were constrained by a parametric model to

find the significant proteins. I found the posterior probabilities for proteins found in both SILAC1 and SILAC2 experiments of being differentially regulated (Figure 3.3a). This analysis found the posterior probability for each protein of being upregulated or downregulated due to the knockdown of microRNA and found that proteins with higher magnitudes of change had a greater posterior probability of being differentially regulated. This distinguished proteins that were more likely changing expression due to microRNA knockdown rather than technical or biological variation. The proteins with posterior probabilities of 50% or higher of being differentially regulated in the proteomics experiment had log₂ fold changes of greater than or less than 0.25 or -0.20 in SILAC1 and 0.30 or -0.30 in SILAC2 (Figure 3.3b).

As described with microarray and transcriptome analyses, finding biologically significant changes in gene expression often implements a combination of fold-change cut-off and p-value requirements. In this case the posterior probability from the Empirical Bayes analysis and a fold-change cut-off was used. Numerous proteins had a high posterior probability of being upregulated in one SILAC experiment, but lower posterior probability of being upregulated in the other experiment. For example, some proteins might have had over 85% posterior probabilities in SILAC1 and between 50-85% posterior probability in SILAC2, or vice versa. In order to keep some proteins for analysis, a low threshold of posterior probability was set. The proteins with greater than 50% posterior probability in SILAC1 and SILAC2 were selected. Next, the log₂ fold-change cut-off was set at +/- 0.3, because this incorporated proteins in both SILAC1 and SILAC2 with over 50% posterior probability of change, and was supported by literature that proteins with this fold-change and greater were potentially biologically important in microRNA knockdown experiments (Guo et al., 2010).



Figure 3.3 - Threshold for determining significant changes in protein expression a. The posterior probability (measured on the right y-axis) of a protein being upregulated (green dotted line) or downregulated (pink dotted line) was determined by Empirical Bayes analysis of the 902 overlapping proteins in SILAC1 (left) and SILAC2 (right). b. Proteins with a \log_2 fold-change of less than -0.3 and greater than 0.3, and a posterior probability of greater than 55% (blue dotted lines), met the criteria for further analysis.

Next I wanted to find which of these significantly changed proteins supported by Empirical Bayesian analysis and fold-change cut-off had consistent expression trends between the two SILAC experiments. Significantly differentially expressed proteins following the same expression trend were kept and proteins with high variation between SILAC1 and SILAC2 were filtered out using the UpSet R Package. This package was used to find the intersecting proteins between four subsets - significantly upregulated in SILAC1, significantly upregulated in SILAC2, significantly downregulated in SILAC1 and significantly downregulated in SILAC2 (Figure 3.4a).





a. Significantly changed proteins were separated into four subsets - downregulated in SILAC1, downregulated in SILAC2, upregulated in SILAC1, and upregulated in SILAC2, which were evaluated for overlapping significant proteins. The orange bar represents the proteins downregulated in both datasets; the blue bar represents proteins that are upregulated in both datasets. The Venn diagram (top left) also shows overlap between the four subsets. b. Histogram of SILAC1 and c. SILAC2 protein expression changes, with proteins downregulated in both SILAC1 and SILAC1 and SILAC2 marked as orange, and proteins upregulated in SILAC1 and SILAC2 marked as blue.

For the consistently downregulated proteins in both datasets with significant change, the proteins CORO1A, DDX21, DYNC1H1, FUS, GLUL, GRSF1, HIST2H2AB, LCP1, MRTO4, PDLIM1, PEBP1, PSPH, RSU1, SNRPD1, SRSF2, SRSF3, TMPO, and VIM were found (orange bars, Figure 3.4b and c). For upregulated proteins in both datasets with significant change, the proteins ACAT1, ACOT13, CALB2, DCXR, DFFA, DSTN, EIF3M, FAH, GMFG, HSPB1, ITGA2B, MCM4, NUMA1, PSME1, SPC24, and TIPRL were found (blue bars, Figure 3.4b and c).

Among the 16 significantly upregulated proteins, there are no conserved predicted targets of miR-143 supported by the TargetScan algorithm. Analysis of the 3'UTRs for the mRNAs of these proteins was also performed, to look for miR-143 seed binding sites. The four types of miR-143 binding sites are a 6 nt match in the mRNA to nucleotide positions 2-7 of the microRNA (6mer), a 7 nt match to positions 2-8 (7mer-m8), a 7 nt match with an A at the 3' end to match positions 1-7 (7mer-1A), and an 8 nt match with nucleotide positions 1-8 (8mer). There were no 8mer or 7mer-m8 binding sites in the 16 proteins, but EIF3M had 1 7mer-1A site and 4 6mer sites. For 6mer sites, ACAT1 had 1, DFFA had 2, HSPBP1 had 1, MCM4 had 2, and SPC24 had 3 sites. However, the 6mer binding sites have a weaker effect on regulation, and are more likely to be found by random change due to their shorter length (Nielsen et al., 2007).

I performed pathway analysis of the 16 significantly upregulated proteins, to determine if miR-143 knockdown had an effect on a particular biological function. The 16 proteins upregulated in SILAC1 and SILAC2 were analyzed for enrichment of genes in biological process using Database for Annotation, Visualization and Integrated Discovery (DAVID) (Figure 3.5a). The biological process with the highest statistical significance and lowest p-value (0.04) was the mitotic cell cycle, and the three proteins from our filtered set that were enriched in this process were NUMA1, PSME1, and SPC24 (Figure 3.5b). Knockdown of miR-143 in UT-7 cells was performed again, and immunoblotting was conducted on PSME1 and NUMA1, two of the proteins with high probability of upregulation. NUMA1 had a 0.96 probability of being upregulated in SILAC1 and a 0.55 probability in SILAC2, while PSME1 had a 0.85 probability in SILAC1 and a 0.91 probability in SILAC2. Derepression of NUMA1 and PSME1 was observed by immunoblotting after pLL-miR-143spg transduction (Figure 3.5c and d).





a. The DAVID alrgorithm was used to search for biological processes enriched for GO terms associated with the 16 significantly upregulated proteins in both SILAC datasets. b. NUMA1, PSME1, and SPC24 were enriched for GO terms in the biological process of the mitotic cell cycle (p-value = 0.041). c. Immunoblotting of lysates from UT-7 cells transduced with pLL-miR-143spg confirmed upregulation of NUMA1 (top) and PSME1 (bottom) compared to pLL-GFP control. d. Immunoblots from three replicate experiments were quantified by densiometry. Bars show mean ± SD.

The relationship of miR-143 to mitotic cell cycle can also be observed in the literature, since a few studies have linked miR-143 loss to increased cell proliferation. In one study in colorectal cancer, it was demonstrated that miR-143 regulates IGF1R, and loss of miR-143 leading to derepression of IGF1R, which increases cell proliferation (Su et al., 2014). Knockout studies of both miR-143 and miR-145 in mice studying effects in vascular smooth muscle cells showed that the smooth muscle layers of aorta were noticeably thinner, due to decreased actin-based stress fibers. It was found using this model that these miRNAs modulate actin dynamics and

cytoskeletal assembly and act through regulators such as myocardin, ADD3, TPM4 and SRF (Lai et al., 2012). There is a relationship between actin cytoskeleton dynamics and the mitotic cell cycle, since the actin cytoskeleton is involved in cleavage of the cell during cytokinesis and in the separation of centrosomes during metaphase (Heng and Koh, 2010).

3.4 Correlation Between SILAC Datasets and Degree of Changes Due to miR-143 Loss

The expression of miR-143 in UT-7 cells appeared high according to preliminary small RNA sequencing data collected in the lab, and this was consistent with the knockdown of the mature microRNA measured by qPCR and derepression of SMAD3 in immunoblotting. However, the initial high expression value of miR-143 in UT-7 was due to misalignment of the reads, and remapping of microRNA sequencing data found that expression of miR-143 was lower in UT-7 cells than expected (Figure 3.5a). This was confirmed in microRNA qPCR by using synthetic miR-143 oligonucleotides to make standards and give absolute quantification (Figure 3.5b).



Figure 3.6 - Expression levels of miR-143 in leukemic cell lines

a. Solexa platform small RNA sequencing was performed on leukemic cell lines and there were initially ~10,000 RPM for miR-143 in the UT-7 library, but after remapping, very few reads were observed at the miR-143 locus. b. Standard curve RT-qPCR was used for absolute quantification of miR-143 in leukemic cell lines. UT-7 cells showed 0-20 copies per 100ng of total RNA in three replicates (red, black, and blue coloured symbols represent three biological replicates).

The low levels of miR-143 in UT-7 cells were interesting in light of the low coverage and low overlap in expression changes seen in our proteomics experiments The degree of changes seen in the SILAC1 replicate was similar to the range of expression changes seen in other SILAC proteomics studies designed to find targets of a particular microRNA (Baek D et al 2008, Selbach M et al 2008). SILAC2 had fewer proteins and did not have as many proteins undergoing the same extent of change, making the extent of expression changes in the overlapping proteins between SILAC1 and SILAC2 smaller as a result. There were larger non-Gaussian tails in the SILAC1 histogram representing the range of protein expression changes, but they were not chosen for validation due to lack of replication in the second dataset. Since the coverage of the proteome was greater for SILAC1 compared to SILAC2, and there were almost four hundred proteins without measurement in SILAC2, it leaves the possibility that the protein may have had consistent differential expression in both datasets if it had been detected in both.

While knockdown of miR-143 expression and upregulation of certain proteins was consistent in UT-7 cells, numerous proteins varied widely in expression between the two SILAC datasets. The potential reasons for the differences in expression between the two SILAC datasets are numerous. Many proteins were likely undergoing change stochastically and as discussed in Section 3.2, there was most likely technical variation in preparation of the two replicates for mass spectrometry, differences in cell cycle states due to complete synchronization, variation in half life of protein and other reasons that commonly contribute to noise in proteomics experiments between replicates. However, the size of the group of overlapping expression changes, 16 upregulated proteins and 21 downregulated proteins, seemed small compared to other experiments (Baek et al., 2008; Selbach et al., 2008). As well, the lack of conserved predicted targets or stronger seed binding sites is concerning.

The low degree of overlap between the two datasets combined with the low levels of miR-143 in UT-7 cells led to the idea of other microRNA species than miR-143 potentially binding to the sponge transcript and regulating a different set of proteins than miR-143. The non-seed region of the microRNA is known to be mutation tolerant and contain many Watson-Crick pairing mismatches, which allows it to bind to many targets with varying degrees of complementarity. In my model, the sponge was designed with four tandem repeats to bind multiple copies of miR-143, and the transcript was highly expressed from a strong, lentiviral CMV promoter after cells

were transduced with a highly concentrated lentivirus. Given the high levels of sponge transcript with repetitive sequences and the comparatively low levels of the microRNA, as well as the mutation tolerance of the non-seed region in binding targets, there is potential for non-specific binding by novel or annotated microRNAs.

4. Molecular Analysis of a Potential Novel microRNA

4.1 Introduction

MicroRNA hairpins and microRNA-mRNA target complexes are forms of RNA that use selfcomplementarity and sequence complementarity to make stable secondary structures. MicroRNAs demonstrate stable binding to mRNA with varying degrees of sequence complementarity, affecting the microRNA-mRNA duplex structure and determining the incorporation into the RISC machinery, thereby influencing the outcome of microRNA targeting. The earliest studies in *C. elegans* found that seven nucleotides in the 5' end of the microRNAs were complementary to conserved sequences of the 3'UTR of their lin-14 targets (Wightman et al., 1993). In massive parallel sequencing of the mouse small RNA transcripts less than 30 nucleotides, the most conserved areas were nucleotides 3-7 and 10-15 (Reid et al., 2008). In a study by Brennecke *et al*, as few as four basepairs in positions 2-5 of the microRNA was less effective (Brennecke et al., 2005b). The 3' end of the microRNA, the non-seed, is required for duplex thermodynamic stability but does not usually require the same extent of reverse complement base pairing with the mRNA target, and has been shown to be highly tolerant to mutation (Doench and Sharp, 2004).

This has the capacity for problems when the sponge method is used to knockdown microRNA expression, due to the highly expressed, repetitive sequence creating an abundance of positions where part of the sequence is capable of binding the seed site of a non-specific microRNA. The UT-7 cell line used in Chapter 3 for knockdown has low levels of endogenous miR-143 and high expression of the miR-143 sponge. There are potentially other annotated or novel microRNAs that may bind to a repeated element in the sponge's tandem reverse complement sequences. My hypothesis is that in addition to miR-143, non-specific microRNAs bind to the miR-143 sponge and lead to derepression of a distinct, separate target set. This activity makes the targets of the intended microRNA more difficult to discern. More importantly, sponges used in applications such as gene therapy may also be subject to novel or annotated microRNAs binding non-specifically to the sponge, that compete with knockdown of the intended microRNA and derepression of its targets.

To assess if non-specific microRNAs bind to the miR-143 sponge and regulate distinct target sets, I identified microRNA recognition elements within the miR-143 sponge sequence, analyzed novel microRNA transcript data containing matching seed sites to the MREs, and evaluated the potential of non-specific binding by the novel transcripts containing seed sites. This was followed by genomic and molecular analysis of the candidate novel microRNA to confirm its presence and functionality.

4.2 Identifying potential microRNA recognition elements within the sponge sequence

The sponge designed for knockdown of miR-143 in leukemic cell lines is full of repetitive sequences due to the four tandem reverse complement binding sites of miR-143. While the sequence includes 4-5 different random nucleotides between each of the four repeats, there are numerous 7 nucleotide sequences repeated throughout the sequence. The first step in identifying non-specific microRNAs that could bind to the sponge was to scan the sequence of the sponge for potential microRNA seed binding sites, MREs (Figure 4.1a).

The initial discovery of microRNA regulation led to a multitude of mRNA target prediction algorithms. Discovery of the requirement for Watson-Crick base-pairing in the 5' region of the microRNA at positions 2-7 limited the number of false-positive predictions and assisted the development of these algorithms (Brennecke et al., 2005b; Krek et al., 2005; Lewis et al., 2003). The 5' region is the most conserved portion of metazoan microRNAs and often only 7 nt matches were found in aligned sequences of vertebrate 3'UTRs (Lewis et al., 2003; Lim et al., 2003b). This 5' region of the microRNA was called the seed and binds to complementary sequences known as microRNA recognition elements (MREs) in target transcripts. Many of the target prediction algorithms have overlapping sets of predicted targets and use similar methodology, finding seed pairing in conserved regions of target 3'UTRs for each annotated microRNA. TargetScan Custom is a modification of the TargetScan algorithm, where instead of selected an annotated microRNA of interest, a theoretical seed site can be entered and a set of the seed's predicted targets generated. The required input size is seven nucleotides only, so the sequence of the pLL-miR-143spg was analyzed using a scanning window of seven nucleotides to find repeats potentially acting as MREs for microRNAs aside from miR-143 (Figure 4.1b) (Bartel, 2009). In the analysis of the miR-143spg sequence, any heptamer sequences repeated

more than once were treated as an MRE for a potential seed site in a novel or annotated microRNA. Based on evaluation of each seed by TargetScan Custom, none of the potential seeds belonged to a known, annotated microRNA.



Figure 4.1 - Identification of potential non-specific microRNA recognition elements in miR-143 sponge a. Schematic of miR-143 binding to miR-143 MRE (pink) in sponge sequence, and causing derepression of mir-143 targets (top), versus hypothetical binding of novel microRNA seeds to non-specific MREs on sponge (in orange and blue), regulating different sets of targets. b. A scanning window of 7-nt was used to search for MREs belonging to novel or annotated microRNAs. The potential seeds that would bind to these MREs are pictured (bottom half). The species binding non-specifically to the sponge sequence were theorized to be microRNAs, due to the ability of the seed to form the strongest interaction with the target. Since the potential seeds did not correspond to any known microRNA, they were searched for within a set of potential novel microRNA transcripts. Recently, whole transcriptome RNA sequencing and microRNA sequencing was performed at the Genome Sciences Centre on samples collected from AML patients for The Cancer Genome Atlas (TCGA), a large-scale pancancer study. The unclassified sequences within the microRNA libraries were designated as potential novel microRNAs because of alignment of reads with the genome (for a single library, regions that had more than one small RNA read mapping to the same locus), and formation of RNA secondary structure with the flanking genomic sequence according to RNALfold. The novel microRNA transcripts found by this pipeline were searched for transcripts containing the potential seed sites binding to my miR-143spg construct. Out of the 12 potential seeds binding to repetitive sequences in the sponge, 9 of these seeds were found in putative novel microRNA transcripts (Table 4.1).

 Table 4.1 - Potential seed sites included in novel microRNA transcripts

The potential seeds binding to theoretical microRNA recognition elements in the sponge sequence, found by scanning window, are numbered 1-12, and the predicted targets from TargetScan Custom 4.0 for each potential seed are given (third column). The first selection criteria for finding a novel microRNA containing the potential seed was to find a transcript containing the seed in small RNA-Seq libraries (fourth colum, blue). The next criterion was that the novel transcript and flanking sequence forms a hairpin structure (green).

Seed #	Repeated Seed (12)	Predicted Targets	Does the seed exist in a transcript in the small-RNA Seq library? (9)	Does the transcript form a hairpin structure? (6)
1	AGAUGAU	128		
2	AUCGCUG	5	Seed #2	
3	AUGAUCG	39	Seed #3	Seed #3
4	CGCUGUA	7	Seed #4	Seed #4
5	CUGUAGC	128	Seed #5	Seed #5
6	GAUCGCU	0	Seed #6	Seed #6
7	GAUGAUC	72	Seed #7	
8	GCUGUAG	119	Seed #8	Seed #8
9	GUAGCUC	90	Seed #9	
10	UCGCUGU	0		
11	UGAUCGC	0		
12	UGUAGCU	90	Seed #12	Seed #12

The first criterion for a novel microRNA transcript to meet to determine if it was a true microRNA was the formation of a hairpin structure by the transcript and flanking sequence (Table 4.1). The

novel transcripts were found in a variety of different lengths, some only 16-17 nt in length, and others exceeding the length of a typical microRNA at 50-60 nt in length. The shorter transcripts were treated as microRNA with potential degradation and the longer transcripts were considered potential pre-miRNA transcripts. To assess the potential hairpin structure, the sequence surrounding the transcript was included to make each transcript 90 nt in total length. DROSHA cleavage of the microRNA precursor hairpin requires ~40 nt of flanking ssRNA to either side of the base of the hairpin, and the hairpin precursor is typically ~60 nt following DROSHA cleavage (Starega-Roslan et al., 2011)(Auyeung et al., 2013), so the length of 90 nt was chosen to accommodate these features. Each novel microRNA transcript was mapped to its genomic locus and the sequence of the transcript as well as its flanking genomic sequence was acquired from UCSC Genome Browser. A window of 90 nt containing the transcript was positioned with the transcript at the 5' edge of the window, and moved in 3-5 nt increments until the transcript was at the 3' edge of the window. The 90 nt sequence was taken at each increment and assessed for secondary structure formation by RNAfold. Some transcripts did not form stable hairpins when the sequence was extended into the flanking genomic region in a variety of iterations, and some transcripts demonstrated circular or tripod RNA secondary structure instead of hairpin structure (Table 4.2).

Coordinates	Potential Seed to MREs in Sponge (DNA form)	MicroRNA Hairpin Structure?			
(hg18)	Transcripts				
Seed #2	AUCGCUG (ATCGCTG)				
Chr17: 38794153- 38794239	GATACAGTGGAAGGCTTGTGGTGTTGCTGGCCCTTGAT No hairpin, circular RNA secondary CGCTGGAAGGATTCCGAGGTGTAGTTTTCGAAGCGGGA structure GTTTTGTTCG				
Seed #3	AUGAUCG (ATGATCG)				
Chr 14: 34095057- 34095160	CTGTTATT <mark>ATGATCG</mark> GCGCTGGGTCTGGATGTGTGGTG TTCAAAACACGGGCTGCTGGGCAGTTCGCTTTCGTTTTC ACGTTTTTGTGGGGGGTAGGGCGATTG	Forms hairpin, but transcript/ seed not appropriately positioned for miRNA processing			
Seed #4	CGCUGUA (CGCTGTA)				
Chr10: 731649- 731665	CGGG <mark>CGCTGTA</mark> GGCTG	Forms hairpin, but transcript/seed not appropriately positioned for miRNA processing			
Seed #5	CUGUAGC (CTGTAGC)				
Chr 12: 4817268 - 4817288	CTGTAGCCTCTTCTGCTTGG	Forms hairpin, seed out of position in transcript.			
Chr 14: 20170361- 20170382	TCTGTAGCCTCTTCTGCTTGG	Forms hairpin, but poor base pairing probability.			
Chr 3: 61703306 - 61703359	CAAGTTCCAAATGAAGAAGGTGTTATGTCTGG <mark>CTGTAGC</mark> TGTTGGTCACGTGA	No hairpin , circular RNA secondary structure			
Chr X: 66265775 - 66265792	GAAGCA <mark>CTGTAGC</mark> TCTC	No hairpin , circular RNA secondary structure			
Seed #6	GAUCGCU (GATCGCT)				
Chr17: 38794153- 38794239	GATACAGTGGAAGGCTTGTGGTGTTGCTGGCCCTT <mark>GAT CGCT</mark> GGAAGGATTCCGAGGTGTAGTTTTCGAAGCGGGA GTTTTGTTCG	Forms hairpin with flanking sequence, seed site near appropriate position.			
Seed #7	GAUGAUC (GATGATC)				
Chr 3: 182119123- 182119144	ACCTGGATGATCCTGCCAGTT	No hairpin, circular RNA secondary structures			
Seed #8	GUAGCUC (GTAGCTC)				
Chr 1: 237594754 -237594777	GGGGT <mark>GTAGCTC</mark> AGTGGCAGAGC	No hairpin, forms tripod RNA secondary structure			
Chr X: 66265775 - 66265792	GAAGCACT <mark>GTAGCTC</mark> TC	No hairpin, seed site not appropriately positioned.			
Seed #9	GCUGUAG (GCTGTAG)				
Chr 8: 86629462- 86629478	GGGG <mark>GCTGTAG</mark> GCTTA	Forms hairpin, but transcript/ seed not appropriately positioned for miRNA processing			
Chr 14: 20169139- 20169159	CG <mark>GCTGTAG</mark> GAATACTTTTC	No hairpin , circular RNA secondary structures			
Chr 3: 61703306- 61703359	CAAGTTCCAAATGAAGAAGGTGTTATGTCTG <mark>GCTGTAG</mark> C TGTTGGTCACGTGA	Forms hairpin, but transcript/seed not appropriately positioned for miRNA processing			
Seed #12	UGUAGCU (TGTAGCT)				
Chr 1: 237594754 -237594777	GGGG <mark>TGTAGCT</mark> CAGTGGCAGAGC	No hairpin, forms tripod RNA secondary structure, or hairpin with low probability base pairing			
Chr 3: 61703306 - 61703359	CAAGTTCCAAATGAAGAAGGTGTTATGTCTGGCTGTAGC TGTTGGTCACGTGA	Forms hairpin, but transcript/ seed not appropriately positioned for miRNA processing			
Chr 8: 92714127 - 92714146	TTTTA <mark>TGTAGCT</mark> TACCTCA	Forms hairpin, low probability base pairing, not appropriately positioned for processing			

Table 4.2 - Hairpin structures of novel transcripts containing potential seeds

Another criterion was added for hairpin formation in the next stage of selection. In some cases where a hairpin structure was created, the transcript and potential seed was not in a position that was favourable for further microRNA processing (such as positioned in the apical loop of the hairpin). These transcripts were eliminated as potential microRNAs due to demonstrating poor microRNA secondary structure. At this selection stage, the hairpins were kept if the transcript containing the seed was in the 5p-arm or 3p-arm of the hairpin stem. Out of the six seeds found in sequences that formed a hairpin RNA secondary structure, three seeds were found in transcripts that formed correct hairpin structures (Table 4.3).

Table 4.5 - Polential seeds in novel transcripts with microrian hairpin structures	Table 4.3	- Potential	seeds in r	novel transc	ripts with	microRNA	hairpin structures
--	-----------	-------------	------------	--------------	------------	----------	--------------------

The criteria for potential seed sequences and the small RNA sequencing transcripts containing them (blue) to be considered potential novel microRNAs was first that the transcripts containing the seed formed hairpin structures (green). Seeds contained in novel transcripts that are in appropriate positions for DICER/DROSHA cleavage are shaded in yellow.

Seed #	Repeated Seed (12)	Predicted Targets	Does the seed exist in a transcript in the small-RNA Seq library? (9)	Does the transcript form a hairpin structure? (6)	Is the seed/ transcript appropriately positioned for microRNA processing? (3)
3	AUGAUCG	39	Seed #3	Seed #3	
4	CGCUGUA	7	Seed #4	Seed #4	
5	CUGUAGC	128	Seed #5	Seed #5	Seed #5
6	GAUCGCU	0	Seed #6	Seed #6	Seed #6
8	GCUGUAG	119	Seed #8	Seed #8	
12	UGUAGCU	90	Seed #12	Seed #12	Seed #12

For the final criterion to determine if any of the transcript containing potential seeds were true microRNA candidates, the number of predicted targets for each seed and the number of these predicted targets upregulated in our proteomics data was examined (Table 4.4). I conducted this screen using the upregulated proteins (those with a log₂ fold-change greater than 0.3) from the SILAC1 dataset, as there were a greater number of proteins for screening and larger changes in expression than the SILAC2 dataset. The predicted targets for each seed were searched for within the upregulated proteins of the proteomics data collected for the pLL-miR-143spg. Out of twelve seeds, nine seeds had predicted targets that were found within the upregulated proteins dataset, while three had no matching predicted targets in the dataset. The number of predicted targets (pred.) was given as a ratio for each potential seed (Table 4.4).

	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5	Seed 6	Seed 7	Seed 8	Seed 9	Seed 10	Seed 11	Seed 12
Potential Seeds	AGAUGAU 128 targets	AUCGCUG 5 targets	AUGAUCG 39 targets	CGCUGUA 7 targets	CUGUAGC 128 targets	GAUCGCU 0 targets	GAUGAUC 72 targets	GCUGUAG 119 targets	GUAGCUC 90 targets	UCGCUGU 0 targets	UGAUCGC 0 targets	UGUAGCU 93 targets
	ANK1	COROC1	STAU1	VAPA	PITPNB		FXR1	MAPK1	CALM1			ELL2
5	OCIAD1		HMGB3		HNRNPA3		VAPA	PPP3CA	PITPNB			RAB6A
AC	CA8		ANP32E		OLA1		UBE2N	GORASP2	ELL2			HNRNPA1
SIL	RBM22				MAT2A		CUL3	CD36	METAP2			BZW1
s in	HNRNPU				HNRNPA1			TRIP12	PAK2			DNAJA2
rget	CUL3				OGT			METAP1	RAB24			TNPO1
Таі	AHCYL2				BZW1			ACTR1A	HNRNPA1			STAG2
Ited	PPT1				ANP32E			YWHAZ	TNPO1			
gula	NCBP1				DNAJA2				DNAJC19			
brei					C12orf23							
Ωp					TNPO1							
S					SMSN1							
bse					DDX6							
0					STAG2							
					CDK6							
Obs./ Pred	0.094	0.077	0.200	0.143	0.110	0.000	0.056	0.067	0.089	0.000	0.000	0.078

Next, to find the chance of these observed to predicted ratios occurring by chance, a set of 300 scrambled microRNA sequences was taken and the observed to predicted ratio was found for each seed in the 300 scrambled sequences (scrambled seeds). This set of ratios for 300 scrambled seeds gave a distribution range for comparison with the potential seeds. The scrambled seeds were generated in silico using the total set of mature human microRNA sequences from mirBase. These 300 sequences were taken from the database with nucleotide composition as a factor. This meant the sequences were selected randomly but proportionally so that the amount of GC content per sequence would reflect the distribution of GC content (and thus nucleotide composition) in microRNA sequences found in nature. In a study by Zhang B et al looking at all metazoan microRNA sequences, the percentage composition of the nucleotides among metazoans is skewed from the equal percentage value of 25% each, and the majority of sequences, or 85.93%, have 34 - 58% GC content. Our partially random collection of human mature microRNA sequences was designed to follow a similar trend. Scrambling the sequence may have diminished certain features, such as common mutation biases or position-specific nucleotide distribution, which naturally occur in microRNAs. For example, uracil is enriched at three sites in microRNA sequences, the first, ninth, and the five terminal 3' nucleotides (Zhang et al., 2009). However, the position-specific nucleotide composition of the 12 potential seed binding sites in the sponge was also not taken into account, so the potential seeds and scrambled seeds have a similar degree of randomness. As with the 12 potential seeds, the sets of predicted targets for each of the 300 scrambled seeds were found by TargetScan Custom and searched for upregulated SILAC1 proteins. The ratio of observed to predited targets for miR-143 was 0.079 and the average ratio for the scrambled seed was 0.085 (Figure 4.2a). For the three seeds with transcripts that formed microRNA hairpins appropriate for DROSHA and DICER processing, the observed to predicted targets ratios were 0.000 for Seed #6, 0.078 for Seed #12, and 0.109 for Seed #5 (Figure 4.2b). According to a one-sample z-test comparing the potential seeds to the population of scrambled seeds, neither Seed #5, #6, nor #12 have a pvalue of less than 0.05 at the 95% confidence interval, but Seed #5 has the lowest p-value of the three. Does Seed #5 have more predicted targets upregulated in the SILAC1 dataset than would be expected due to chance? Seed #5 has a greater observed/predicted ratio than the average of the 300 scrambled seeds, or any of the other potential seeds found in hairpins. Taking the following analysis of hairpin structure and genomic expression into account, and considering that the ratio of Seed #5 is higher than miR-143 as well, Seed #5 is the most likely candidate to bind to the sponge non-specifically (Figure 4.2c).



Figure 4.2 - Potential seeds with higher ratio of observed to predicted targets than random a. Comparison of the observed to predicted targets ratios for seeds of 300 scrambled microRNA sequences (blue bars) versus the ratios for each of the potential seeds and the miR-143 seed (red squares). b. The number of predicted targets for Seed #5 and miR-143, and their ratios of observed/predicted targets. c. Seed #5 (CUGUAGC), meets each selection criteria (circled in red).

The application of these criteria narrowed the potential seeds down to a likely candidate. Seed #5 was the only seed with a higher observed to predicted ratio than miR-143 and the scrambled seed average, which also met the criteria of being found in a novel microRNA transcript and forming an appropriate microRNA hairpin structure. The sequence of this transcript matched two loci, one on chromosome (chr) 12 and the other on chromosome (chr) 14 (Table 4.5).

Coordinates **	Potential Seed to MREs in Sponge Sequence (Matching DNA Sequence)				
	Transcripts				
Seed #5	CUGUAGC (CTGTAGC)				
Chr 12: 4817268 -4817288	CTGTAGCCTCTTCTGCTTGG				
Chr 14: 20170361-20170382	T <mark>CTGTAGC</mark> CTCTTCTGCTTGG				

Table 4.5 - Genomic locations of novel microRNA transcripts with potential seed #5

4.3 RNA Structure Analysis of Novel microRNA Transcripts Containing Potential Seed Sequences

The flanking sequences surrounding each of the Seed #5 candidate transcripts are different at each location and each genomic region forms a distinct microRNA hairpin structure. RNA forms hairpin secondary structures frequently, but not all RNA hairpins are processed into microRNAs. There are a variety of features that make a microRNA hairpin distinct, and comparison to these features was used to evaluate the hairpins of both transcripts.

The primary feature was the length of ~11bp between the basal junction and DROSHA cleavage site. The Microprocessor complex recognizes the junction between the stem of the hairpin and the flanking ssRNA, and positions DROSHA one helical turn away from the base of the hairpin. The hairpin stem length needs to be roughly 31-33 bp to accommodate the cleavage machinery of the DROSHA-DGCR8 complex. The basal stem of the chromosome (chr) 12 hairpin is 29 nt in length while the chromosome (chr) 14 hairpin is only 24 nt. This created an issue for the chr 14 hairpin, as the hairpin stem was not long enough for a helical turn between the base of the flanking ssRNA on the 3p arm of the hairpin (Figure 4.3e). In the chr 12 pri-microRNA structure, there are 10 nt between the end of the microRNA sequence where DROSHA cleavage occurs and the base of the hairpin, with 10nt on the 5p arm and 8 nt on the 3p arm (Figure 4.3c). This provides enough length for DROSHA positioning and cleavage.

Additionally, a recent study distinguished certain sequence motifs that are frequently present in and unique to human microRNA hairpin secondary structure (Auyeung et al., 2013). It was found that the flanking sequence to either side of the hairpin usually contained at least nine
unstructured nucleotides, though predicted pairing was tolerated in one flank provided the other flank contained at least 5-7 unpaired bases. As well, three motifs are commonly found in conserved human microRNAs, with at least one of the three found in 79% of sequences (Auyeung et al., 2013). One motif was base pairing at the first base of the stem loop shows a preference for G:C pairing, and the nucleotide preceding the G (on the 5' side of the stem loop) is often a U (Figure 4.3a). Secondly, two C residues separated by two intervening nucleotides were found at 17-18 nt downstream of the DROSHA cleavage site on the 3p arm of the hairpin. This site was discovered to bind the SRp20 protein by an RNA-recognition motif and may bind other proteins which enhance pri-miRNA recognition and processing. Another important motif present in human and *D. rerio* microRNAs was either enrichment of UGU or GUG in the apical stem loop (Auyeung et al., 2013). The chr 12 and chr 14 hairpins were evaluated for these common motifs. The chr 12 hairpin included a GUG motif in the apical loop of the hairpin (Figure 4.3c), which promotes pri-microRNA processing for chr 12 hairpin and increases its likelihood as a true microRNA. The chr 14 hairpin structure had a U at the first position of the flanking sequence at the base of the stem on the 5p side (Figure 4.3e), but considering that the basal junction at this position would also lead to truncation of the mature microRNA sequence, this motif may not strengthen the case for this microRNA.

Based on length of hairpin and sequence motifs, the chr 12 hairpin showed more features consistent with a true microRNA. The minimum free energies of the hairpin structures were -21 kcal/mol for the chr 14 hairpin structure and -25 kcal/mol for the chr 12 hairpin structure (Figure 4.3c,e). In numerous novel microRNA discovery methods (Ambros et al., 2003; Chiang et al., 2010), the highest free energy for the hairpin structure was set to -25 kcal/mol. The chr 12 transcript is more stable according to this criteria, and the chr 14 does not meet the minimum free energy requirement. The higher instability of the chr 14 transcript is likely due to the six G-U base pairings in the hairpin stem, which are not as stable as A-U or G-C base pairing (Figure 4.3e).



Figure 4.3 - microRNA hairpin structure of Chr12 and Chr14 novel microRNA

a. The characteristic microRNA hairpin structure has an apical loop of >10nt, a double-stranded RNA stem spanning ~33bp, and single-stranded RNA flanking sequences upstream and downstream of the basal junction. DROSHA cleavage usually occurs about ~11 nt from the basal junction, and a UGU or GUG motif at the apical junction is found in ~25% of human pri-miRNAs. b. High probability of base pairing was observed in the stem of the chr12 microRNA hairpin by minimum free energy analysis in RNAfold. c. The transcript at chromosome 12 forms a hairpin with canonical structure, with a 29-nt hairpin stem and 8-10 nt from the basal junction to putative DROSHA cleavage site, and a GUG motif in the apical loop. The potential novel microRNA sequence is coloured red, and the seed sequence of of the microRNA is in bold. d. The transcript at chromosome14 forms a hairpin, but with a shorter length hairpin stem than would be allowable for DROSHA cleavage. The novel microRNA sequence is coloured red, and the seed sequence is in bold. e. The probability of base pairing observed in the stem of the chr14 microRNA hairpin was ~0.5 according to minimum free energy analysis performed in RNAfold.

Both potential microRNA transcripts are found on the 3p arm of the microRNA hairpin. Studies have found that many of the novel microRNA candidates that have been discovered more recently are poorly conserved and lower expressed than the collection of microRNAs discovered after the initial discovery of these gene regulators (Chiang et al., 2010; Gong et al., 2013). Most canonical microRNAs produce one dominant species, with a tendency to come from the 5p arm, as noted in a study where expression of a set of canonical microRNAs was assessed and 202 microRNAs originated from 5p and 141 species from 3p arm (Chiang et al., 2010). It was also found that selection of one arm over the other was less pronounced for non-conserved microRNAs (Chiang et al., 2010).

4.4 Genomic Analysis of Novel microRNA Transcripts Containing Potential Seeds

The twelve potential seeds were narrowed down to two potential novel microRNA transcripts found in the output of a novel microRNA discovery algorithm developed at the Genome Sciences Centre (Morin et al., 2008). The two novel transcripts were almost identical in sequence and contained the sequence of the potential seed, Seed #5. According to the output of the discovery algorithm, one transcript aligned uniquely to a genomic locus on chr 14, and the other transcript, identical in sequence except missing the first base of the chr 14-aligned transcript, aligned to genomic loci at chr 12 and chr 14. However, investigation into the microRNA libraries and the reads aligning to these two genomic loci found that instead of two transcripts mapping to the two genomic locations, there were actually three transcripts mapping to two genomic locations. One transcript, CUGUAGCCUCUUCUGCUUGG, aligned to the genome perfectly at chr 12 and chr 14 (known as "crossmapping"), while a transcript with the same sequence but beginning with A aligned uniquely to chr 12 and a transcript with the same sequence but beginning with U aligned uniquely to chr 14 (Figure 4.4a). This was slightly different from the two forms the novel transcripts had shown in the output from the novel microRNA discovery algorithm. This was explained by an error at a step known as adaptor trimming during the processing of microRNA sequencing data, which was corrected in the libraries after the novelty algorithm was run.

The novel microRNA transcripts were initially found in microRNA libraries isolated from AML patients in the TCGA project. In addition to the data from small RNA sequencing performed for

the TCGA project, our lab also has performed small RNA sequencing on nearly 200 patients with MDS or AML (known henceforth as the "in-house" dataset). This provided two large datasets to examine the genomic evidence and determine which of the three sequences - the crossmapped read, the chr 12 transcript, or the chr 14 transcript - produced the true microRNA involved in non-specific binding to the sponge.

It should be noted that in the creation of microRNA sequencing libraries, when the algorithm processes and aligns small RNA reads to the genome, reads that map to multiple locations (known as "crossmapping reads") are randomly assigned by the algorithm to one of the locations. The reads with the sequence crossmapping to chr 12 and chr 14 were randomly chosen by the algorithm to map to chr 14 in this case. The expression levels of each of the three transcripts were calculated and compared separately, rather than with the cross-mapped reads grouped with the chr 14 reads (Figure 4.4b and c).

I first searched for the three transcripts within the microRNA libraries isolated from AML patients generated in our lab and within the microRNA libraries from the TCGA. In both the TCGA and in-house microRNA libraries there are reads mapping to the chr 12 and chr 14 locations, as well as many cross-mapping reads. In TCGA libraries, 192 patients out of 195 had expression of the chr 12 transcript, 185 had expression of the cross-mapped transcript, and 153 patients had expression of the chr 14 transcript (expression in this case being >1 read per library). The mean expression for the cross-mapped transcript was 20 RPM, 3 RPM for the chr 14 transcript, and 167 RPM for the chr 12 transcript. In our in-house AML dataset, out of 185 patient libraries, 146 patients had expression of the chr 12 transcript, and 37 patients had expression of the chr 14 transcript. In the in-house dataset, the mean expression for cross-mapped read was 4 RPM, 0.5 RPM for the chr 14 transcript, and 19 RPM for the chr 12 transcript (Figure 4.4b and c).

The expression levels of the cross-mapped read, chr 14 transcript, and chr 12 transcript fall within the typical expression range for a microRNA. This was demonstrated by finding the range of average expression for each individual microRNA found in the in-house libraries. When the expression of each microRNA species is averaged across patient libraries, the majority of microRNAs fall into the expression range of 0-10 RPM (Figure 4.4d).

a. Transcripts Containing Seed #5





a. Three potential novel microRNA transcripts were found containing Seed #5 (underlined). One transcript crossmaps to chr 12 and chr 14. The unique chr 14 transcript (blue) has a U (dotted line) first nucleotide, and unique chr 12 transcript (red) has an A (orange) as first nucleotide at its genomic location. b. Normalized expression of all three transcripts in the TCGA AML patient libraries, (n= 192 patients, mean ± SD Reads Per Miillion (RPM), p-value by non-parametric t-test). c. Normalized expression of all three transcripts in in-house data set in de novo AML patient libraries (n=146 patients, mean ± SD RPM, p-value by non-parametric t-test). d. The number of known microRNAs with average expression in the 0-10 RPM range, the 10-100 RPM range, etc., across AML libraries (in-house data). The expression of the chr 14 transcript was significantly lower than the expression of the chr 12 transcript in both the TCGA and in-house datasets (Figure 4.4b and c). The transcript that uniquely aligns to chr 14 was found in 83% of the TCGA libraries and in 29% of the in-house libraries, but the majority of reads aligning to the chr 14 location in those libraries were cross-mapped reads. If the sequence of the cross-mapped transcript were a mature microRNA sequence, Seed #5 would be in nucleotide positions 1-7 instead of 2-8 (Figure 4.4a). This would change the seed site from Seed #5 to a different sequence and would affect the entire predicted target set of this transcript. As well, there is no available method of determining which genomic loci the cross-mapped reads come from, and so it is unknown if the cross-mapped reads come from the genomic location forming a proper microRNA hairpin structure or not (chr 12).

The number of reads for the potential microRNA at chr 12 was significantly higher than the cross-mapped reads or the reads at chr 14, which would lend preference to the chr 12 transcript as the more likely microRNA (Figure 4.4b and c). The sequence of the chr 12 reads would also put Seed #5 in nucleotide positions 2-8, the correct position of a microRNA seed site (Figure 4.4a). Based on the stronger evidence for the chr12 transcript because of correct microRNA hairpin structure, higher expression, and appropriate seed position within transcript, the chr 12 transcript was taken as the most likely true microRNA candidate for non-specific binding to the sponge.

The sequence of the chr 12 transcript and hairpin was submitted to miRBase and was named miR-X for further discussion while awaiting a number designation.



Figure 4.5 - Detection of miR-X in other cell lines and tissues

a. Absolute quantification of miR-X using stardard curve TaqMan RT-qPCR in various cell lines (Three different colours represent three biological replicates, individual symbols represent technical replicates, mean ± SD). b. The normalized expression of miR-X in libraries from the TCGA pancancer study (Red line is mean of each set of libraries. Each symbol is a microRNA library from a different patient sample, libraries with less than 1 RPM were excluded for clarity. BLCA = bladder urothelial carcinoma, BRCA = breast invasive carcinoma, CESC = cervical squamous cell carcinoma, COAD = colon adenocarcinoma, DLBC = lymphoid neoplasm diffuse large B-cell lymphoma, HNSC = head and neck squamous cell, KICH = kidney chromophobe, KIRC = kidney renal clear cell carcinoma, LGG = brain lower grade glioma, LIHC = liver hepatocellular carcinoma, LUAD = lung adenocarcinoma, LUSC = lung squamous cell carcinoma, OV = ovarian serous cystadenocarcinoma, SARC = sarcoma, SKCM = skin cutaneous melanoma, STAD = stomach adenocarcinoma, THCA = thyroid carcinoma, UCEC = uterine corpus endometrioid carcinoma).

I performed microRNA RT-qPCR and detected the mature form of miR-X in multiple cell lines representing leukemic, breast cancer, colorectal cancer, and HeLa cells (Figure 4.5a). I also performed a pancancer search within data collected by the TCGA project for expression of miR-X. The microRNA had expression greater than 1 RPM in patient libraries from 22 types of cancer, demonstrating particularly high expression in a number of cancers specific to females. including breast invasive carcinoma (BRCA), cervical squamous cell carcinoma (CESC). ovarian serous cystadenocarcinoma (OV), and uterine corpus endometrioid carcinoma (UCEC); and in gastrointestinal cancers such as colon adenocarcinoma (COAD) and rectum adenocarcinoma (READ) (Figure 4.5b). There was no expression of miR-X in the following ten types of cancer in the TCGA: adrenocortical carcinoma (ACC), cholangiocarcinoma (CHOL), esophageal carcinoma (ESCA), glioblasoma multiform (GBM), mesothelioma (MESO), pheochromaocytome and paraglioma (PCPG), testicular germ cell tumours (TGCT), thymoma (THYM), uterine carcinosarcoma (UCS), and uveal melanoma (UVM). Since most microRNAs vary in expression between different tissues, and can be upregulated or downregulated in certain cancers because of genetic abnormalities, this pattern of expression fits with what is expected of a true microRNA.

As set out in the criteria of many initial screening methods for novel microRNA discovery, I looked at conservation of the miR-X sequence in other vertebrate species. The multiple sequence alignments of genomes from 100 vertebrate species are featured in the UCSC Genome Browser, and using the browser interface I looked at the alignment of the portion of the human genome containing miR-X with various other species (Figure 4.6a). The sequence of miR-X was conserved in primates, giving 100% sequence identity in chimp, rhesus monkey, and gorilla, and shows 75-85% sequence identity in white rhinoceros, aardvark, elephant, and manatee. Other mammals such as horse, cow, sheep, dog and cat show 61-67% sequence identities, while in rabbit and mouse there is only 48 and 51% sequence identity, respectively (Figure 4.6a). However, mouse and rabbit both have complete conservation of the seed site, while cow, sheep, dog, and cat only have half of the correct nucleotides in the seed positions. The conserved seed site means that there may still be a microRNA functionally similar to miR-X encoded from this position in mouse and rabbit.

	miR-X sequence																				
	3'																				5'
<	G	G	Т	т	С	G	т	С	т	т	С	т	С	С	G	Α	Т	G	Т	С	Α
																mi	R-X s	seed			
Human Chimp Rhesus Mouse Rabbit Cow Sheep Horse White Rhino Cat Dog	GGGAGGG-GGG	G G G G T A A G G A A	ТТТТСТТСТСТ СТ	тттссттсссс	C C C - A T T C C C A	GGG - AAAAAAAAAA	T T C T T T T T T	000-0000000	T T T T T T T T T	T T T - C C C T T T T	ссс - сссссс с	T T T C T T T T T G T	0000000000000	C C C C C C C C C T T	G G G G G A A A C A A	A A A A A A A A A A A A A A A A A A A	T T T T C G G C T T T	G G G G G A A G G A A	T T T T T C C T T T T	000000000000	A A G C A A A A A A
Elephant Manatee Aardvark Gaps	G G G	G G G 4	T T	C C	A - A	A - A	т - Т	C C C	C T T	T T T	C C C	T T T	C C C	C C T	G G G	A A A	T T T	G G G	T T T	C C C	A A A
4.88 -								Mult	iz Ali	gnme	ents o	of 100) Vert	ebrat	es						
-4.50								DI		D		0									

chr12:4,947,007-4,947,027 (21 bp)

PhyloP Basewise Conservation

Figure 4.6 - Conservation of miR-X sequence

The sequence conservation of the chr12 novel microRNA transcript across fourteen vertebrate species. The total miR-X sequence is indicated by the black bar, the arrow indicates direction of transcription. The green line indicates the miR-X seed in the human genome and the corresponding place in other mammalian genomes.

Novel microRNA discovery methods also sometimes set criteria regarding the conservation of the secondary structure of the microRNA. The conservation of the miR-X 5p arm was examined and it was found that sequence identity between human and other species was similar for the 5p arm as for the 3p arm. Chimpanzee and rhesus monkey had 100% and 95% sequence identity to human, while manatee, horse, cat, white rhinocerous, and rabbit had 75-85% sequence identity (Table 4.6). This was higher sequence identity for the 5p arm of the microRNA hairpin in cat and rabbit than had been demonstrated in the 3p arm (Table 4.6). The sequence identity was lower in the 5p arm for cow, sheep and elephant than in the 3p arm, with sequence identity of 40-65% (Table 4.6).

Table 4.6 - Percentages of human sequence identity in sequences aligning to miR-X in other species

Species	3p arm	5p arm
	Percentage of	Percentage of
	Sequence Identity (%)	Sequence Identity (%)
Human	100	100
Chimp	100	100
Rhesus	100	95
White Rhino	86	85
Aardvark	81	70
Elephant	76	65
Manatee	76	75
Horse	71	75
Cow	67	40
Dog	67	70
Sheep	67	40
Cat	62	75
Rabbit	57	85
Mouse	48	45

The pairwise identity scores derived from multiple sequence alignment of the same mammalian genomes described above also demonstrated that the miR-X precursor sequence was highly conserved, particularly in the seed region (Figure 4.7a). Alignment of the flanking genomic sequences, 100bp to either side of the miR-X hairpin sequence, demonstrated a similar degree of conservation on the 5p flanking region as the miR-X hairpin (Figure 4.7a). However, the 3p flanking sequence had lower sequence identity than the miR-X hairpin sequence alignment and the 5p flanking region (Figure 4.7a). This corresponded to the Multiz alignments from UCSC Genome Browser, where downstream of the 3p end of miR-X there were no conserved sequences in the species selected except the non-human primates (Figure 4.7b). An analysis of the genomic sequence 3000bp upstream and downstream of miR-X was conducted using Promoter 2.0 to find RNA polymerase II transcriptional start sites (TSSs) (Figure 4.8). The 5p flanking sequence about 844 bp upstream of the miR-X hairpin contains a RNA polymerase II

transcriptional promoter site, where transcription of the primary microRNA transcript for miR-X would begin (Figure 4.8). It is possible the reason that the 5p flanking sequence is more conserved is that the 5p flanking sequence is important for the transcription and stability of primary miR-X transcript, while the 3p flanking sequence is less conserved because it is not important for stability, or it marks the end of the primary transcript. The primary miR-X transcript was not found in RNA-Seq data from the in-house dataset patient libraries, however, in the future experiments such as Rapid Amplification of cDNA Ends (RACE) PCR could be performed to help elucidate the primary transcript sequence length.



Figure 4.7 - Multiple sequence alignment of miR-X hairpin and flanking regions

a. Multiple sequence alignment (MSA) of miR-X flanking and hairpin sequences from various mammalian animals. In the 3p flanking region at the top, the elephant sequence was excluded because the genome did not extend to this region. The miR-X hairpin precursor sequence is in the middle MSA, and the 5p flanking region is the bottom MSA. All MSAs show sequences from right to left 5' to 3' direction. The species and total consistency value are to the left of each corresponding sequence in the MSA, the colour key for total consistency values (BAD-AVG-GOOD) is at the bottom. The key for the colour scheme shows support for a residue within the MSA on a scale of 0 (blue, poorly supported) to 9 (dark pink, strongly supported). b. UCSC Genome Browser Multiz alignments track also shows lower conservation downstream of the 3' end of miR-X hairpin.



Figure 4.8 - Structure and putative promoter site for miR-X gene

The exact length of the miR-X gene is not known. The sequence of miR-X is on the minus strand on Chr 12, and lies within the same region where KCNA6 is transcribed from the positive strand. The predicted RNA pollI site is 844 bp upstream from the miR-X sequence (middle).

The predicted targets of miR-X in humans were compared to the predicted targets of miR-X in species with conservation of its sequence. The predictions for the targets of the miR-X seed were available in eight species (Figure 4.9a). The target sets predicted for chimpanzee and rhesus monkey were the same, and had 124 overlapping predicted targets with human (Figure 4.9b). There were 100 overlapping targets between human and dog, 90 between human and cow, and 88 between human and mouse, demonstrating the number of targets common to human and the different species increases with the greater percentage of sequence identity in the mature microRNA sequence (Figure 4.9b). Conservation of microRNA sequence and numerous overlapping predicted targets between the conserved species lends even greater support to the likelihood of miR-X being a true microRNA.



Figure 4.9 - Overlap of predicted targets of miR-X in conserved species

a. Targets of miR-X as predicted by TargetScan Custom for 8 vertebrate species. Chimpanzee and Rhesus monkey have the same set of predicted targets. b. Three of the species with conserved miR-X sequence based on Multiz Alignment, and the overlapping predicted targets for each of the three species.

Another aspect of novel microRNA discovery, which can support the validity of a putative microRNA, is the conservation of RNA secondary structure, or that the base pairing in the microRNA stem is evolutionarily conserved even if the exact sequence is not. The bases of the microRNA were assigned number identities (Figure 4.10) and the base pairings between these identities were examined for nucleotide substitutions between species that led to mismatches or matches.



Figure 4.10 - Assigning of number identities to base pairs in miR-X hairpin stem Each base pairing within the miR-X hairpin is given a number for each base, to evaluate the conservation of the base pairing.

There were 49 nucleotide changes out of 74 total changes led to a mismatch, not including the unpaired nucleotides in bulges in the miR-X stem (Table 4.7). At the same time, there were 25 out of 74 total changes that led to a matched base pairing (Table 4.7). The ratio of random base pair mutations leading to the correct base pairing are 4/24, according to a previous study, and the ratio of mutations leading to mismatches would be 20/24 (Gibb et al., 2015). The changes which occurred in our data set were were compared to the expected ratio of changes leading to G:C, A:U, or G:U, which is detailed in Table 4.8. I found that nucleotide changes were significantly more likely to be compensatory (25/74 vs. 12/74, observed vs. expected) than non-compensatory (49/74 vs. 62/74, observed vs. expected) changes to base pairing in the miR-X stem (p = 0.014, Chi-squared test), as shown in Table 4.8. This supports the predicted structure of the miR-X hairpin and the identity of the microRNA as a true novel microRNA.

	- 14	uci	-01	lue	CIIC	ang	C 3 I	n u	16 1	iaii p	2111	310	7111	UI.		\-															
Position	1	63	2	62	3	61	4	60	5	59	6	6 5	8	7	57	8	56	9	55	10)54	11	53	12	252	13	51	14	50	15	49
Structure	()	()	()	•	•	•	•		•		•	•	()	()	()	()	•	·	()	()	()
Human	A	U	L	A	G	С	С	С	A	А	0	ι	J	U	U	С	G	С	G	С	G	A	U		U	G	С	С	G	Α	U
Chimp	A	U	L	A	G	С	С	С	A	А	0	ι	J	U	U	С	G	С	G	С	G	A	U		U	G	С	С	G	Α	U
Rhesus	A	U	L	A	G	С	С	С	G	U	C	ι	J	U	U	С	G	С	G	С	G	A	U		U	G	С	С	G	A	U
Mouse	A	С	L	G	С	U	G	С	U	A	0	ι	J	С	U	С	A	С	Α	U	G	G	U		C	G	-	U	-	Α	-

Table 4.7 - Nucleotide changes in the hairpin stem of miR-X

Position	1	63	2	62	3	61	4	60	5	59	6	58	7	57	8	56	9	55	10)54	11	53	12	52	13	51	14	50	15	49
Structure	()	()	()	•	•	•	•	•	•	·		()	()	()	()	•	•	()	()	()
Rabbit	A	G	G	А	-	U	A	С	G	С	С	U	U	A	С	G	С	G	С	U	A	С		С	G	А	С	-	А	С
Cow	A	U	A	G	G	U	A	С	G	С	С	U	U	U	С	А	С	G	C	A	G	U		U	G	U	С	А	Α	U
Sheep	A	U	A	G	G	U	A	С	G	С	С	U	U	U	С	А	С	G	C	A	G	U		U	G	U	С	А	Α	U
Horse	A	С	G	G	G	U	A	С	G	С	С	U	G	U	С	С	С	-	U	G	С	С		С	G	С	С	А	А	U
White	С	G	U	Α	С	U	Α	С	G	С	С	U	U	IJ	С	С	C	G	U	G	А	U		С	G	С	С	Α	Α	U
rhino											•						•			Ū						•	•			
Cat	A	U	С	A	A	U	A	С	G	С	С	U	U	U	С	А	C	G	U	A	G	С		С	С	С	С	А	А	U
Dog	A	U	U	G	G	U	A	С	G	С	С	U	U	U	С	А	U	G	G	A	G	U		С	G	А	С	А	А	U
Elephant	A	U	U	G	G	U	A	С	G	С	С	U	С	U	A	А	С	G	U	G	A	U		С	С	А	С	А	А	U
Manatee	G	U	U	G	G	U	A	U	G	С	С	U	С	U	С	А	С	G	U	G	A	-		-	С	-	С	-	А	-
Aardvark	A	U	U	A	G	U	U	С	G	С	U	U	С	U	С	А	С	G	U	G	A	U		С	С	А	С	А	А	U
Position	16	48	17	47	18	46	19	45	20	44	21	43	22	42	23	41	24	40	2	539	26	38	27	37	28	36	29	35		
Structure	()	()				•		•		•	()	()	()			()	()	()	()		
Human	G	С	A	U		U		С	U	U		С	G	С	С	G	U	A		U	С	G	A	U	G	С	U	Α		
Chimp	G	С	A	U		U		С	U	U		С	G	С	С	G	U	Α		U	С	G	A	U	G	С	U	А		
Rhesus	G	С	A	U		U		С	U	U		С	G	С	С	G	U	A		U	С	G	A	U	G	С	U	А		
Mouse	С	-	С	-		-		-	U	С		С	G	С	С	G	U	A		U	U	G	G	U	G	С	С	G		
Rabbit	G	С	A	U		С		С	G	U		С	G	С	С	G	U	A		С	С	G	A	U	G	С	С	С		
Cow	-	С	-	U		С		С	-	U		С	-	С	-	А	-	Α		G	-	А	-	С	-	С	С	А		
Sheep	-	С	-	U		С		С	-	U		С	-	С	-	А	-	А		G	-	А	-	С	-	С	С	А		
Horse	G	С	A	U		U		С	U	U		С	G	С	U	А	U	A		С	С	G	A	U	G	С	G	А		
White	G	C	Δ	U		u		С	С	U		С	G	С	С	С	U	Δ		U	С	G	Δ	IJ	G	С	G	Δ		
rhino			<i>``</i>	Ũ					Ŭ	Ũ				Ŭ	Ŭ	Ŭ							í.	Ŭ		Ŭ	Ŭ			
Cat	С	С	A	U		U		С	U	G		С	G	U	С	А	U	A		U	С	А	A	U	G	С	U	А		
Dog	С	С	A	U		U		С	U	U		С	G	U	С	А	U	A		U	С	А	G	U	G	С	U	А		
Elephant	С	С	A	С		U		С	А	U		С	G	С	С	G	U	А		U	С	G	A	U	G	С	U	Α		
Manatee	С	С	A	U		U		С	A	U		С	G	С	С	G	U	A		U	С	G	A	U	G	С	U	A		
Aardvark	С	С	A	U		U		С	A	U		С	G	U	С	G	U	A		U	С	G	A	U	G	С	U	А		
		Original stem bp									Ni	ımh	ero	f hr	ch	ana	۵۵	74												
		Original stem mismatched br) Ni	umb	er o	of mi	sma	atch	ed b	op c	hang	ges:	49									
		Ch	anę	ged	l to	a r	nat	che	ed b	р	(not including unpaired bases in bulges (blue columns))																			
		Ch	anę	ged	l to	a n	nisı	mat	ch	bp	-Number of matched bp changes: 25																			
																												1		

Original Base Pairing	lf Random Nucleotide Substitution	Retained Base Pairing	Type of Change	Observed # of Changes	Expected # of Changes
GU	GA	no	If total # of changes is:	74	74
	GG	no	Compensatory changes:	25	12.33
	GC	YES			
	G-	no	If total # of changes is:	74	74
	AU	YES	Non-compensatory	49	61.67
	CU	no	· · · · · ·	•	•
	UU	no			
	-U	no			
GC	GU	YES			
	GA	no			
	GG	no			
	G-	no			
	AC	no			
	CC	no			
	UC	no			
	-C	no			
AU	AG	no			
	AC	no			
	AA	no			
	A-	no			
	GU	YES			
	CU	no			
	UU	no			
	-U	no			
	24 Total Possible Changes	4 Changes with Retained BP			
IF	74 Possible Changes	12 Changes with Retained BP			

	Table 4.8 - Observed vs.	expected retained base	e pairing in nucleotide substitutio	n
--	--------------------------	------------------------	-------------------------------------	---

I tested whether the microRNA hairpin structure predicted by minimum-free energy analysis in RNAfold could be processed into a mature microRNA by cloning the microRNA hairpin and flanking sequence into an expression vector (Figure 4.11a). The 300 bp region containing miR-X, its putative hairpin sequence, and flanking sequence were cloned into a lentiviral expression vector after a MND promoter (Figure 4.11a). The miR-X sequence is located on the minus strand of chromosome 12, therefore the miR-X sequence consisted of the 300 bp sequence from the minus strand sequence taken in its 5' to 3' orientation, which was cloned into the expression vector in the 5' to 3' orientation (Figure 4.11a). The reverse complement sequence of miR-X and flanking regions were also cloned into the expression vector, as a control (Figure 4.11b). In MDS-L cells transduced with the vector expressing miR-X hairpin and flanking

sequence, high expression of the mature miR-X species is detected compared to the GFPcontrol, and for cells transduced with the reverse complement as a control, there was no increase in miR-X expression compared to the GFP-control (Figure 4.12).

a. Cloning Strategy for Overexpression of miR-X Hairpin:

miR-X and flanking region - Chr 12: 4946889 - 4947188 (Minus Strand) 5 ' –GCUUAUUUUCUACUCUCAGAGUGGAACACUUGGAGAGCGGUCUCUGCAGUCUAGCGCAGCAGCAGCACCCAAAU ACAAAGGCUCUUUCAAACCACCUCUGAGCUGCAGAGCCCACACUCAGGCAGUCUAUCCAUGCACUCCAGGCAGU GCUCAGUAAGUGACUGUAGCCUCUUCUGCUUGGGUUACCAUAAAAGCCUCCCACAGGUAGAGCAACGAGGAGUGUG UGGAGAGCAGUCUGUACCCACAGGGCAAUAAAGGGUUGCAUUGUCCACAGAGGACAUUAAAACAAUAAUACUGG-3 '



b. Non-microRNA hairpin structured negative control:



Figure 4.11 - Cloning strategy for miR-X hairpin and reverse complement control

a. The sequence of miR-X and its flanking regions (119nt in the 3' direction and 159nt in the 5' direction), was cloned in a 5' to 3' orientation into an overexpression vector, MSV-pMND-PGK. b. The reverse complement of the miR-X and flanking regions sequence was cloned in a 5' to 3' orientation into an overexpression vector, MSV-pMND-PGK.



Figure 4.12 - Expression of mature miR-X from lentiviral hairpin overexpression Expression of miR-X measured by TaqMan RT-qPCR in MDSL cells virally transduced with MND-GFP, MND-miR-X, and MND-non-hairpin control. (n=2 biological replicates, bars show mean ± SD of three technical replicates).

4.5 Biological Activity of miR-X

The novel microRNA, miR-X, was selected after various criteria were applied to a group of potential seeds that could non-specifically bind to a sponge intended to knock down miR-143 expression. While analysis of hairpin structure and genomic expression provided strong evidence for the presence of miR-X, activity of miR-X in regulating predicted targets demonstrates the capability of non-specific binding to give expression changes in proteins unrelated to the microRNA knockdown of interest.



Protein	Fold Change	Significance
DNAJA2	2.57	< 0.0001
STAG2 (1)	2.36	< 0.0001
STAG2 (2)	2.00	< 0.0001
MAT2A	1.75	<0.0001
DDX6	1.63	0.0005
BZW1	1.43	0.0051
HNRNPA1	1.25	0.0147
TNPO1	1.02	0.6672

Figure 4.13 - Derepression of miR-X predicted targets observed by quantitative proteomics Peptides from miR-X predicted targets have increased expression in pLL-miR-143spg transduced cells compared to pLL-GFP based on selected reaction monitoring (SRM) quantification. These expression increases are significant compared to peptides from the housekeeping protein ACTIN1, which has no increase, except for that of TNPO1. (Bars show mean ± SD, n=2 biological replicates, and two technical replicates each. Unpaired, two-tailed t-test was used for significance).

The expression changes in the predicted targets of miR-X that changed in the first miR-143KD SILAC experiment were detected through Selected Reaction Monitoring (SRM). I selected two to three peptides for each of the predicted targets, however during optimization of the peptide transitions using synthetic copies of these peptides I found that not all peptides could be used in combination with each other for detection of the twelve targets. Higher levels of the peptides from miR-X predicted protein targets compared to Actin housekeeping peptides were observed in cells infected with the pLL-miR-143spg than the pLL-GFP control vector (Figure 4.13). The confirmed the changes in the predicted targets of miR-X in the SILAC1 experiment. However, given that these effects could be due to knockdown of miR-143 as well, I needed to assess specific inhibition of miR-X separately.

Using a variety of methods, I looked for confirmation that protein targets with miR-X binding sites were upregulated while miR-X was inhibited. The 3'UTR sites of two predicted miR-X targets, SAMSN1 and TNPO1 were cloned into the 3'UTR of Firefly luciferase in a dual-luciferase reporter vector. The reverse complement of miR-X was also cloned into the 3'UTR of Firefly luciferase in a reporter vector. Following transduction with pLL-GFP or pLL-miR143spg,

UT-7 cells were transfected with the luciferase vectors containing miR-X binding sites and derepression of the miR-X luciferase 3'UTR targets was measured by luciferase activity in cells transduced with the sponge compared to pLL-GFP (Figure 4.14a). An increase in luciferase activity demonstrated derepression of the miR-X targets by the sponge acting as a decoy for miR-X. It was found that the miR-Xrevcomp and SMSN1 targets were derepressed when the pLL-miR-143spg was also present (Figure 4.14b and c). The TNPO1 construct, which was made prior to SRM results, did not show significant derepression, confirming the same observation made in the SRM experiment, that miR-X does not regulate TNPO1 (Figure 4.14d). TNPO1 may have less stability or a shorter protein half-life that appears as change in expression but is actually due to natural variation.

To test the effects on protein targets when miR-X is specifically inhibited rather than inhibited by a sponge that also inhibits miR-143, an anti-sense oligonucleotide was designed to inhibit miR-X. Anti-sense oligonucleotides (ASOs) use the reverse complement sequence of the microRNA to be inhibited, with modified nucleotide base chemistry to improve stability and potency. The inhibitor was transfected into UT-7 cells containing the miR-X(revcomp) luciferase construct, which showed significant derepression in UT-7 after 24 hours (Figure 4.15a). When miR-X was inhibited in OCI-AML3 cells by transfection of the miR-X ASO, the miR-X target STAG2 showed upregulation by immunoblotting (Figure 4.15b).

Two predicted targets of miR-X were also evaluated by immunoblotting after transfection with miR-X mimic, a commercially designed version of the microRNA hairpin with the mature microRNA sequence in a proprietary hairpin cassette. The STAG2 protein and DNAJA2 protein were downregulated by the miR-X mimic after 72 hours in K562 and after 48 hours in UT-7 and AML5 cells (Figure 4.15c and d). The difference in length of time likely varied due to protein turnover rates between the different cell lines, which vary in proliferation rate, genetic abnormalities, and epigenetic programming.



Figure 4.14 - miR-X inhibition leads to derepression of miR-X binding sites

a. Derepression of the luciferase target is measured when miR-X is bound by the sponge transcript, which is virally transduced into cells prior to transfection with the luciferase reporter plasmid. b. A luciferase reporter with the reverse complement of the miR-X sequence shows increased activity in cells transduced with the pLL-miR-143spg compared to cells with pLL-GFP control. c. A luciferase reporter with the SAMSN1 binding site of miR-X shows similar results. d. However, the luciferase reporter with the binding site for miR-X in TNPO does not show derepression (b-d, n=6, bars show mean ± SD, significance measured by unpaired t-test with Welch's correction).



Figure 4.15 - Inhibition of miR-X and delivery of miR-X mimics regulates predicted targets

a. A luciferase reporter with the reverse complement of the miR-X sequence in 3'UTR shows increased activity following transfection of miR-X inhibitor. (n=6, Mean ± SD, unpaired t-test with Welch's correction) b. Immunoblotting of OCI-AML3 cells lysate after transfection with miR-X inhibitor reveals upregulation of STAG2 c. Immunoblotting of AML5 cells lysate following transfection of miR-X mimic reveals upregulation of STAG2 at 48 hr and 72 hr. d. Immunoblotting of K562 cell lysate following transfection of miR-X mimic reveals upregulation of STAG2 at 72 hr. e. Immunoblotting of UT-7 and K562 cell lysate following transfection of miR-X mimic reveals downregulation of DNAJA2 at 48 hr and 72 hr, respectively.

4.6 Conclusions

The existence of a potential microRNA transcript containing a seed site that matches part of a repetitive sequence in the pLL-miR-143spg transcript demonstrates the potential for a novel microRNA to bind to the sponge and inhibition of its activity to occur simultaneously with inhibition of miR-143. I have provided a detailed and systematic approach to finding novel microRNA involved in non-specific binding, and have accrued significant support for the

existence of this scenario's microRNA, miR-X. There are numerous methods of discovery for novel microRNA, and the route to finding miR-X resembled more that of a forward genetics approach rather than a typical sequencing or computationally driven approach, though it employed aspects of both these methods. Further characterization of miR-X was performed to support the presence and activity of this novel microRNA, and the sequence was submitted to the microRNA database, miRBase. The implications for the sponge method of knocking down microRNA expression and the impact of the novel microRNA, miR-X, in this experiment are critical to explore further, to find potential improvements or solutions to non-specific binding so that this powerful knockdown technique can be used successfully in different applications.

5. Improving the Design of the Sponge Constructs for microRNA Knockdown

5.1 Introduction

A model of miR-143 knockdown was created using a lentiviral sponge with four miR-143 binding sites and transducing the sponge into a leukemic cell line. Investigation of non-specific binding to the sponge identified repeated sequences that could bind the seed sites of novel microRNA. I identified transcripts with the putative seeds and evaluated their properties and potential for being true microRNA. The potential transcripts were narrowed down to one candidate and its validation as a novel microRNA was carried out.

The novel microRNA, miR-X, contained a seed, which matched a repetitive sequence within the miR-143 knockdown sponge and demonstrated the potential for unknown and novel microRNAs to be inhibited by the sponge as well as miR-143. Characterization of miR-X was performed and the activity of miR-X through regulation of predicted targets was confirmed. The ability of miR-X to bind to the miR-143 knockdown sponge and regulate a set of targets unique from miR-143 demonstrates a potential problem with application of the sponge technique for microRNA knockdown in other applications. In this section, improvements or solutions to non-specific binding are presented to ameliorate off-target effects due to sequence complementarity. I studied whether the sponge could be modified to eliminate the potential for non-specific binding and inhibit miR-X or miR-143 specifically. To do this, the seed-site binding requirements for microRNA regulation of target mRNAs was examined for miR-X and miR-143, and modification of the target binding sites was performed in the 3'UTR of dual-luciferase reporter vectors and in lentiviral sponges, to test the capability to more specifically inhibit one particular microRNA.

5.2 Regulation of Sponge in Luciferase Reporter Vectors

The question that arises with validation of the genomic presence and biological activity of miR-X is whether the sponge could be modified to eliminate the potential for non-specific binding and inhibit miR-X or miR-143 specifically. The postulation was made that mutation of the microRNA

recognition elements in the sponge for either microRNA will inhibit miR-X but not miR-143 and vice versa.

Work by Brennecke J *et al* (2005) investigated the minimal requirements for a functional miRNA-target duplex and found that mismatches in the microRNA target site corresponding to the seed site of the microRNA could abolish the biological activity of the microRNA. Using this information, three vectors were designed that would measure regulation by miR-143 and miR-X when they included intact seed binding sites for both microRNAs, or when either binding site was mutated. The first vector, for binding both microRNAs, was constructed by cloning the first two-tandem repeats of the original miR-143 sponge into a dual-luciferase reporter construct (Figure 5.1a) and tested by performing double transfection of the luciferase construct with control (CTR), miR-X, or miR-143 mimics into 293T cells (Figure 5.1b). This demonstrated that the sponge luciferase vector was regulated by both miR-143 and miR-X when MREs for both microRNAs were intact (Figure 5.1b).

To inhibit binding and regulation by only one microRNA, the other two vectors had mutations in either the miR-143 or miR-X binding sites (Figure 5.1a). In one luciferase reporter vector, the nucleotides in positions 3-5 of the miR-143 binding sequence were mutated to their reverse complement nucleotides, and in the third luciferase reporter, positions 3-5 of the miR-X seed-binding site were mutated (Figure 5.1a). Each of the luciferase reporter constructs, Luc-2T-nomut, Luc-2T-miR143mut, and Luc-2T-miRXmut, were co-transfected in turn with one of three microRNA mimics - miR-143, miR-X or CTR to demonstrate abrogation of non-specific binding by either miR-143 or miR-X.

Mutation of the microRNA seed binding sites in the reporter constructs abrogated binding of the corresponding microRNA mimics in AML5 and COLO205 cell lines (Figure 5.1c, d). The Luc-2T-nomut construct had significantly lower luciferase expression when co-transfected with miR-X or miR-143 mimics compared to the control mimic, and Luc-2T-miR143mut showed significantly decreased luciferase expression when transfected with miR-X mimic, but not with the miR-143 mimic, and vice versa with Luc-2T-miRXmut (Figure 5.1c, d). When mimics were co-transfected with the mutated dual-luciferase reporter vectors, regulation of the luciferase expression was observed when an intact binding site for the microRNA seed was available, but not when the binding site is mutated. These findings demonstrate that both miR-X and miR-143 can bind to the sponge and regulate expression of the associated protein.

a.

b.



Figure 5.1 - Regulation of miR-143 and miR-X specific sponges by microRNA mimics

a. Design of three dual-luciferase reporters - Luc-2T-nomut, Luc-2T-miR143mut, and Luc-2T-miRXmut. b. Transfection of miR-X mimics and miR-143 mimics into 293T cells shows decreased activity of the Luc-2T-nomut luciferase reporter. c. Co-transfection of dual-luciferase reporters and microRNA mimics into AML5 cells and d. COLO205 cells shows repression of luciferase activity for constructs with intact binding sites for the co-transfected microRNA (n=10 for b, n=6 for c and d, Mean ± SD, unpaired t-test with Welch's correction.)

5.3 Proteomic Analysis of miR-143 or miR-X Knockdown Using Modified Sponges

The strategy of mutating MREs belonging to specific microRNAs worked to abolish regulation by miR-X or miR-143 in luciferase constructs, and consequently, this strategy was chosen to correct the potential for non-specific binding to the pLL-miR-143spg. I did this by mutating non-seed positions in the reverse complements of miR-143 in a selective manner for each repeat

within the sponge. Two new constructs were developed from the original pLL-miR-143spg - one adjusted for miR-X knockdown specifically and the other for miR-143 knockdown specifically (Figure 5.2). Three single nucleotide mutations were introduced in each tandem repeat to interrupt non-specific binding to the miR-143 seed or the miR-X seed, making two sponge constructs one of which was miR-X specific (miR-Xspecific-spg) and the other miR-143 specific (miR-143specific-spg) (Figure 5.2). The number of miR-X seed binding sites was increased to four in the miR-X specific sponge from the two in the original sponge (Figure 5.2) so that it would have an equal number of binding sites to the miR-143 specific sponge.

Reverse Comp	lemen	nt to miR-143					
		Random					
		Nucleotide	es		miR-143 speci	fic spon	ge
GAGCTACAG	CGA	TCATCTCAGCTA	GACGTAGTG	TGA	TCATCTCA GCTA	Repeat	1
GgGCTACAG	CGA	TCATCTCA TTAAG	GgGCTTGAG	СТА	TCATCTCATTAAG	Repeat	2
GAGCTgCAG	CAA	TCATCTCAAACCT	GAAGTGCAG	CAA	TCATCTCAAACCT	Repeat	3
GAGCTACgG	CGA	TCATCTCA atCA	GAGCATCGG	AGA	TCATCTCA atCA	Repeat	4
GAGCTACAG	CGA	TCATCTCAGCTA			miR-X specifi	c sponge	
GgGCTACAG	CGA	TCATCTCA TTAAG	CATCGCACGT	AGG	GCTACAGCTATA	Repeat	1
GAGCTgCAG	CAA	TCATCTCAAACCT	TTCGCTC TAA	GGA	GCTACAGCAATC	Repeat	2
GAGCTACgG	CGA	TCATCTCA atCA	GTCAG AAACC	TGA	GCTACAGAGATG	Repeat	3
	Bulge		AACTCAATC	AGA	GCTACAGCGAT	Repeat	4
				Bulge			
Green - 1	miR-2	X seed					
Pink – m	iR-14	43 seed					

Blue - Bases to Interrupt Non-specific Binding

Figure 5.2 - Modification of original sponge to achieve knockdown of specific microRNA

The original sponge contains four repeats of miR-143 binding sequences with 4-5 random nucleotides interspersed. In the repeats of the original construct on the left, the green nucleotides indicate the seed binding site of miR-X, and the pink nucleotides, that of miR-143. In the repeats for the constructs on the right, in the top the miR-X seed site has been mutated by 3-4 nucleotides (blue) per repeat. In the bottom construct, the miR-143 binding sites have been similarly disrupted.

The next step to find distinct sets of proteins regulated by inhibition of miR-X or miR-143 alone was to evaluate the expression of multiple proteins at once through quantitative proteomics. UT-7 cells were chosen for transduction by the modified sponges based on higher likelihood to detect the previously identified significant proteins from the SILAC miR-143KD experiments in Chapter 3 and predicted miR-X targets in SRM, but in future experiments a number of cell lines should be screened to find one that has higher expression of both miR-X and miR-143. The expression of miR-X in different TCGA cancer samples can serve as a guide for which tissues or types of cancer cell lines should be used. The SILAC method was used to quantify proteins in the modified sponges samples (Figure 5.3a). However instead of the pLL-GFP transduced cells being cultured in "light," unlabeled media and mixed with pLL-OrigSpg transduced cells cultured in "heavy," stable-isotope labeled media as in my previous SILAC experiments, the new experiment had untransduced cells cultured in heavy media and cells transduced with pLL-GFP or one of the three sponge viruses were all cultured in light media (Figure 5.3a).



Figure 5.3 - Schematic of differential expression proteomics experiment

a. UT-7 cells are grown in normal, non-isotope labeled media and transduced with microRNA knockdown vectors or a GFP control. The protein lysate is combined with an equivalent amount of protein from untransduced UT-7 cells grown in labeled media. The heavy labeled peptides provide an input normalization for the transduced samples. b. 10 protein samples were processed and analyzed by the Orbitrap Fusion spectrometer, and the log2 fold-change was calculated using the average of the pLL-GFP samples.

For input normalization, heavy, untransduced protein lysate was mixed with equal amounts of protein lysate from the light, virally transduced cultures (Figure 5.3a). Ten samples, including two pLL-GFP samples, two pLL-OrigSpg samples, three miR-143specific-spg samples, and three miR-Xspecific-spg samples, were prepared and analyzed using tandem mass-spectrometry (Figure 5.3a). Overall, 5500 proteins were identified and quantified, with 5215 proteins identified in all ten samples. To determine the fold-change compared to the pLL-GFP control, the samples were normalized and peptide abundances were used to calculate the expression of each protein. The expression of each protein was divided by its average expression in the two pLL-GFP samples and log₂ transformed (Figure 5.3b).

5.4 Determination of Significant Expression Changes by Differential Expression Analysis

The log₂ fold changes in the eight sponge-transduced samples were used as a measure of change in protein expression. The goal was to find the proteins with significant differences in expression between the three sponges, either those that were similar between pLL-OrigSpg and the miR-143specific-spg or similar between the original sponge and miR-Xspecific-spg, and then significantly different in the other condition. Before finding the significant changes between the sponge conditions, it was necessary to first assess the quality of the data and remove outliers, then find significant changes in protein expression between the three sponges, and lastly, find proteins undergoing change specific to miR-X or miR-143 inhibition.

To assess the quality of my dataset and remove outliers, the correlations between replicates treated with the same sponge construct were evaluated. Pearson correlation coefficients were calculated between the eight samples (Figure 5.4a). One of the highest correlations was found between one replicate of miR-Xspecific-spg (miR-Xspecific-spg2) and a replicate of miR-143specific-spg (miR-143specific-spg3), which were cultured together at a later time point using older media than the other six samples. The cells transduced with pLL-GFP or other modified sponges were prepared and collected side-by-side in biological duplicates. The correlation between the two samples collected at a separate time point showed the potential technical pitfalls due to variation in media shelf-life, and these samples were taken out of further analysis due to their high correlation with each other potentially causing a batch effect. (Figure 5.4b).

To find the significantly differentially expressed proteins, a linear regression model was to be applied to the dataset. However, before the dataset could be fit to a linear regression model, proteins with a high amount of variance between biological replicates were removed. This was done so that highly variable proteins between biological replicate samples would not obscure significant expression changes between conditions, since linear regression modeling depends on changes with linear response between conditions. Proteins with greater than 0.35 differences in log₂ fold-change between two biological replicates were taken out of the dataset for all samples. The rationale for selection of the 0.35 difference in log₂ fold-change is described below. Eliminating proteins with higher differences in fold-change between two biological replicates for each sponge (Figure 5.4b). After removal of these outliers, 3110 proteins remained in the dataset and were used in linear regression analysis (Figure 5.5).





a. The correlations between each of the eight proteomics samples are shown, with the Pearson correlation coefficient between each sample shown in the intersecting square. The pLL-Orig Spg samples have two biological replicates, while the specific sponges have three biological replicates each. b. Correlations between six proteomics samples, with the Pearson correlation coefficient between each sample shown in the intersecting square. The two proteomics samples with the pearson correlation coefficient between each sample shown in the intersecting square. The two proteomics samples with a possible batch effect have been removed, as have the proteins with the highest variance between biological replicates.

Differential expression analysis by linear regression modeling was performed on the 3110 proteins reproducible between replicates using the limma R package, and multiple hypothesis testing correction was performed using the Benjamini-Hochberg method. Proteins with a false discovery rate, or q-value, of less than 0.05 were considered significantly differentially expressed. From this analysis, 454 proteins with significant changes in protein expression between sponge versions were found (Figure 5.5).



Figure 5.5 - Linear regression analysis to find differentially expressed proteins A linear regression model was applied to the dataset to find the significantly differentially expressed proteins, and 454 proteins were considered significant with a q-value (adjusted p-value) cut-off of 0.05.

Pairwise comparisons were performed between the three sponge conditions on the 454 proteins considered to be significantly differentially expressed. Based on a q-value of 0.05, 179 proteins were changed significantly between miR-Xspecific-spg and pLL-OrigSpg, while 125 proteins were changed significantly between miR-Xspecific-spg and miR-143specific-spg conditions, and 40 proteins were changed significantly between miR-143specific spg and pLL-OrigSpg. A greater number of proteins were differentially expressed between the miR-Xspecific-spg cells and pLL-OrigSpg cells than between the miR-143specific-spg and pLL-OrigSpg cells at a q-value of 0.05 (Figure 5.5). The expression levels of 454 significantly differentially expressed proteins in different sponge conditions were summarized in a heatmap (Figures 5.6).



Figure 5.6 - Differentially expressed proteins between the three sponge conditions Expression the 454 differentially expressed proteins in each of the six samples analyzed by limma. The red gradient indicates upregulated proteins with increasing positive \log_2 fold-change. The blue gradient indicates downregulated proteins with increasing negative \log_2 fold-change.
5.5 Regulation of Protein Expression Specific to miR-X or miR-143 Knockdown

The patterns of most interest in terms of decreasing non-specific binding effects were cases where a protein was upregulated due to the inhibition of miR-X or miR-143 (Figure 5.6). In these cases a protein would have a significant change in expression in the pLL-OrigSpg and miR-Xspecific-spg, but not in the miR-143specific-spg transduced cells, or would demonstrate significant expression change in the pLL-OrigSpg and miR-143specific-sponge cells, while in the miR-X specific cells it would not. These patterns of expression differences indicated protein changes due to inhibition of miR-X or miR-143 in the original sponge, and the inhibition of one specific microRNA with either of the new, modified sponges.

The linear regression analysis evaluated which proteins had significantly different expression between the three sponges, however this encompassed absolute variation between the three conditions and did not consider upregulation or downregulation of the protein (denoted by positive fold-change or negative fold-change) caused by transduction of a sponge and inhibition of a microRNA. To look more carefully at the proteins of interest in finding non-specific binding effects, I had to look not only at the significantly changed proteins between conditions, but whether the expression of the proteins had increased or decreased significantly from 0. While there are some proteins or genes with very small degrees of changes and low variation in expression that are real, often proteins with a log₂ fold-change nearing zero have less likelihood of representing a true change in expression.

For the initial two SILAC datasets, an Empirical Bayes analysis was applied to the proteomic experiments to find proteins with higher probabilities of being upregulated or downregulated. To determine the upregulated or downregulated proteins in the three conditions of the newly collected proteomics dataset, the Empirical Bayes analysis was applied to this dataset as well. This was performed on the 5215 proteins in the same six samples as were analyzed for linear regression modeling (the two samples, miR-Xspecific-spg3 and miR-143specific-spg3 were similarly not included for this analysis) (Figure 5.7). This analysis found that the posterior probability of protein being upregulated or downregulated was between 65-75% for a 0.3/-0.3 log₂ fold-change. In the previous SILAC experiments, the posterior probability of a protein being significantly changed in expression at 0.3/-0.3 log₂ fold-change was 55-65%. Using the Empirical Bayes results, proteins with a change less than -0.3 and greater than 0.3 were

considered more likely to be upregulated or downregulated (Figure 5.7). The 454 differentially expressed (DE) proteins were divided into subsets, depending on if the protein showed upregulation, downregulation, or minimal change (with log₂ fold-change values between -0.3 to 0.3), in the pLL-Orig-Spg, miR-Xspecific-spg, or miR-143specific-spg. The overlapping proteins between the subsets were examined for patterns indicative of non-specific binding. Proteins with higher probability of upregulation in both miR-X knockdown and pLL-Orig-Spg knockdown (or downregulation in these conditions), but no significant change in the miR-143 knockdown cells, demonstrated changes due to miR-143 knockdown specifically. Proteins which showed higher probability of upregulation in both miR-143 knockdown and pLL-Orig-Spg knockdown (or downregulation in both of these conditions), but less change in miR-X knockdown cells demonstrated changes due to miR-143 knockdown specifically.



ie: A protein is significantly differentially expressed between pLL-Orig-Spg and miR-Xspecific-spg, and miR-143specific-spg (Limma).

> In which sponge(s) condition is the protein upregulated or downregulated? (Emp Bayes)

Figure 5.7 - Statistical analyses applied to proteomics dataset to find expression patterns

There were 5215 proteins identified in all six samples. Linear regression analysis (Limma) was applied to proteins filtered based on reproducibility of regulation, while the Empirical Bayes analysis was applied to all 5215 proteins. Linear regression analysis finds the differentially expressed (DE) proteins between the different sponges. The Empirical Bayes analysis gives the posterior probability of a protein being upregulated or downregulated. The limma analysis finds 454 proteins from the dataset are significantly DE and the Empirical Bayes analysis finds which modified sponges transductions show upregulation of a DE protein, or downregulation of a DE protein.



Figure 5.8 - Overlapping significantly changed proteins between pLL-Orig-Spg and miR-Xspecific-spg

The 454 DE proteins were divided into subsets based on upregulation or downregulation in miR-X-KD cells or Orig.Spg cells. There are 16 proteins that are upregulated in both miR-X-KD cells and Orig.Spg cells, and represented by the pink bar. There are 14 proteins that are downregulated in both conditions, represented by the blue bar.



Figure 5.9 - Overlapping significantly changed proteins between pLL-Orig-Spg and miR-143specific-spg

The 454 DE proteins were divided into subsets based on upregulation or downregulation in miR-143-KD cells or Orig.Spg cells. There are 18 proteins that are upregulated in both miR-143-KD cells and Orig.Spg cells, and represented by the orange bar. There are 15 proteins that are downregulated in both conditions, represented by the red bar.

There were 15 proteins upregulated in both pLL-Orig-Spg and miR-Xspecific-spg transduced cells (Figure 5.8), and 18 proteins upregulated in both pLL-Orig-Spg and miR-143specific-spg transduced cells (Figure 5.9). There were also 14 commonly downregulated proteins in the former group and 15 commonly downregulated in the latter (Figures 5.8 and 5.9). However, 6 of the 15 Orig-Spg/miR-Xspecific-spg upregulated proteins were also upregulated in the miR-143specific-spg, so there were 9 proteins of interest for the miR-Xspecific-spg and 12 proteins of interest for the miR-143specific-spg (Figures 5.8, 5.9). The expression levels of the upregulated proteins in pLL-Orig-Spg and the miR-Xspecific-spg had similar expression trends (Figure 5.10a). Most proteins show upregulation to a similar degree in both conditions, except HSPB1 and STX16, which have lower upregulation in the miR-Xspecific-spg cells compared to pLL-Orig-Spg. I would expect the proteins in the specific sponges to change to a similar degree as the original sponge, so it is possible that these two proteins are not regulated by miR-X inhibition specifically. It may also be a matter of downstream network changes in pLL-Orig-Spg cells, where inhibition of either microRNA leads to increased expression of these two proteins, indirectly. The expression levels of proteins with overlap between pLL-Orig-Spg and miR-143specific-spg compared to miR-Xspecific sponge also demonstrate similar degrees of change (Figure 5.10b).



Figure 5.10 - Overlapping expression changes in significantly differentially expressed proteins a. Proteins with overlapping upregulation between pLL-Orig-Spg and miR-Xspecific-spg transductions and significant differential expression based on linear regression analysis. b. Proteins with overlapping upregulation between pLL-Orig-Spg and miR-143specific-spg transductions and significant differential expression based on linear regression analysis.

There were few overlapping upregulated and overlapping downregulated proteins between the original sponge and either of the specific sponges. The overlapping proteins need to be validated as changes due to miR-X or miR-143 loss by orthogonal methods. Validation methods could include testing if these proteins undergo opposite changes in expression when miR-X or miR-143 microRNA mimics are transfected into this cell line, or if luciferase reporters cloned with their 3'UTR sequence demonstrate derepression when miR-143 or miR-X is inhibited. Higher numbers of proteins demonstrated significant expression changes in only the miR-Xspecific-spg or miR-143specific-spg knockdowns, with minimal (less likely) change in cells transduced with the original sponge, and many of these proteins need to be investigated further as well.

The determination of biologically relevant proteins with regard to miR-143 or miR-X knockdown also begs the question: how many proteins undergoing miR-X specific expression changes or miR-143 specific expression changes are targets of miR-X or miR-143? The dataset of 5215 proteins was searched for predicted targets of miR-143 or miR-X according to TargetScan or TargetScan Custom, respectively. There were 141 predicted targets of miR-143 in the whole dataset of 5215 proteins, and 41 predicted targets of miR-X in the whole dataset of 5215 proteins. Many of these proteins were excluded from linear regression analysis because of high variation between biological replicates or did not seem likely as targets because the proteins were not undergoing significant expression changes according to linear regressiong analysis. There were two TargetScan predicted targets for miR-143 that demonstrated significant log₂ fold changes, AFF1 and DCAKD. There were also two proteins among the predicted targets for miR-X, GINS3 and ZNF706, considered significantly differentially expressed. GINS3 and ZNF706 are not among the proteins upregulated in Figure 5.10a because the proteins do not have a \log_2 fold change over 0.3. However, GINS3 has an average log₂ fold-change of 0.26 in the pLL-Orig-Spg samples, and DCAKD has an average of 0.25. These are examples of proteins that have slightly less probability of being upregulated by the original sponge, but due to their conserved microRNA target sites and significant differential expression between sponge conditions, seem likely to be microRNA targets. Other proteins following similar expression patterns, where they are upregulated in the original sponge and miR-X or miR-143specific sponge, but slightly beneath the cut-off of the $0.3 \log_2$ fold change in one of the conditions. This demonstrates that there could be numerous proteins that are biologically relevant but have a slightly lower magnitude of change due to the inhibition of a microRNA under slightly varied biological conditions (for example, a variation in stage of cell growth or metabolic state). A

larger number of replicates for proteomic analysis would help to discern those proteins which are truly changing expression due to inhibition of miR-X or miR-143, but undergo smaller amounts of expression change.

5.6 Conclusions

I have demonstrated that regulation of luciferase reporter sponges by microRNA mimics can be inhibited by modification of nucleotides outside the seed-binding site of the microRNA of interest. Mutations in nucleotides 2-4 of the MRE decreased the ability of particular microRNA to bind and regulate the sponge. I applied this same principle to knockdown of microRNA using lentiviral sponges and observed significantly different changes in protein expression between sponges intended to knockdown different microRNA. In light of these results, I have identified a new microRNA, miR-X, and determined that the sequence of a sponge for knockdown of a microRNA should be designed so that the microRNA seed-binding site remains constant but other heptamer sequences are not repeated. Random mutations in the non-seed-binding region of the microRNA binding site give less opportunity for annotated or unknown microRNAs to bind to the sponge by ensuring there are not repetitive sequences, except for the microRNA seed binding site.

6. Discussion, Conclusions, and Future Directions

In my thesis, I began by examining the effects of knockdown of miR-143 in acute myeloid leukemia cell lines and discovered the existence of a novel microRNA through analysis of the potential for non-specific binding to the miR-143 sponge. My study began with investigating the loss of miR-143 in myelodysplastic syndromes, in order to identify miR-143 targets and dysregulation of oncogenic pathways. Knockdown of miR-143 showed a small number of proteins were identified that consistently changed expression in cells with the miR-143spg, in replicate experiments. I discovered a lower expression of miR-143 in the knockdown cell line than originally thought, though the upregulation of predicted targets and knockdown of microRNA expression levels were visible in UT-7 and in other leukemic cell lines with low expression as well. Given the low expression of the microRNA, the high expression of a sponge transcript containing repetitive elements, and microRNA seeds binding to target sites of only 6-8 nucleotides, the number of potential recognition sites for a novel or annotated microRNA to bind to was high. Therefore, I investigated what potential binding sites might be present in the sponge and whether any novel microRNAs existed that could bind to these sites. I found a number of potential candidates and eliminated them based on their likelihood of regulating protein targets and their lack of resemblance to a microRNA in structure, which left only one candidate that showed strong potential of being a real microRNA and binding non-specifically to the sponge.

When the original sponge was redesigned to inhibit the binding of the potential non-specific microRNA, miR-X, or the miR-143 binding sites were mutated to inhibit the binding of miR-143 and capture miR-X instead, many proteins underwent different expression change from the original sponge. Some protein expression changes were seen in the miR-143 specific sponge and original sponge, but not for the miR-X specific sponge. There were also many proteins that demonstrated changes specific to the original sponge and the miR-X specific sponge. A number of proteins which were thought to change expression due to the inhibition of miR-143 actually appeared to be affected by the inhibition of the novel microRNA miR-X instead. This shows that non-specific binding to the sponge is a distinct possibility in experiments using the sponge method of microRNA knockdown and that there are novel microRNAs still awaiting discovery, which needs to be taken into account when designing sponges.

6.1 Methods of Discovery of Novel microRNA

This study confirms the potential for undiscovered microRNA and novel small non-coding microRNA. Original estimates of the number of microRNAs in humans were in the low hundreds, but the catalogue of microRNAs has grown to over 2500 and the unknown significance of some portions of the human genome gives the potential for more (Hrdlickova et al., 2014).

There are two main approaches to novel microRNA discovery, the experimentally driven approach and the computationally driven approach, and it is common practice to use both methods to complement each other. Experimentally driven methods often find expression of a small RNA species, through cloning and by sequencing of size-fractionated cDNA libraries, and then use bioinformatics to establish other features of microRNA (Berezikov et al., 2006a). In computationally driven approaches, structural features that can be thermodynamically predicted based on sequence are used to predict candidate microRNAs from the genome (Berezikov et al., 2007).

My route to discovering miR-X implemented both approaches as well, though it was directed by knowledge of the potential seed sites within the novel microRNAs. Small RNA-sequencing, as performed in experimentally driven approaches, provided a set of expressed novel microRNAs to search for the potential seed sites within, and the secondary structure was assessed as in computationally driven discovery techniques. In fact, the secondary structures of multiple microRNA candidates were assessed more rigorously than some computational approaches, by comparison to ideal microRNA hairpin structure and motifs found only in human microRNAs. The expression of the novel microRNA in the small RNA-seq experiments was scrutinized and fell within the normal range of expression for an average microRNA.

There are some concerns for the structure of the microRNA, such as the length of the hairpin stem in the putative primary microRNA hairpin. In the miR-X structure, there is an 8 nt sequence between the base of the hairpin and the 3p arm DROSHA cleavage site, and a 10 nt sequence between the base and the 5p arm site (Auyeung et al., 2013). The average length for this section is 11 and13 nt between the base and DROSHA cleavage, the length of a helical turn (Auyeung et al., 2013). However, there is a bulge in this 8/10 nt span, which would likely extend the span in length, and my experiment cloning the miR-X hairpin and flanking regions into an

overexpression vector and observing increased mature microRNA expression indicates that an appropriate microRNA hairpin is made and processed by DROSHA/DICER.

Another concern of miR-X is that while the sequence is conserved in primates, dog, and elephant, it has poor conservation in some mammalian species, such as mouse and rat. Numerous candidate microRNA are presented with poor conservation in other novel microRNA discovery studies, but it is noted that conservation is not a necessary feature of a functional microRNA, so long as other criteria are found in its favour (Bentwich et al., 2005; Berezikov et al., 2006a; Chen et al., 2005). Many novel microRNA discovered more recently are not located in microRNA gene clusters and do not have high expression compared to many of the initially discovered microRNA. These newer microRNAs with low expression are also often poorly conserved (Bentwich, 2005).

One factor that helps to qualify a newly discovered microRNA is detection of the sequence found on the opposing arm of the microRNA hairpin (Ambros et al., 2003). The 5p arm of the miR-X hairpin structure was searched for within the in-house and TCGA microRNA libraries, but no opposing RNA strand was detected. This has been found for other novel microRNA with expression in the low range, and may be due to the technical limits of small RNA-sequencing or low stability of the transcript. Detection of the precursor or primary microRNA transcript sequences is also not necessary for a microRNA to be considered a candidate, but it does provide more convincing evidence. In the future, optimizing the knockdown of DROSHA by siRNA transfection could be performed to enrich the nuclear primary microRNA transcripts, followed by RT-qPCR testing of nuclear RNA to detect the primary microRNA transcript of miR-X.

The search for novel microRNAs with functionality in different tissues and conditions is not complete, but continually expanding. While not an ideal workflow for determining novel microRNAs, my study does start by identifying microRNAs with a functional effect by searching for microRNA capable of binding the sponge sequence. This method also incorporates some assessment of predicted target regulation, in addition to searching for small RNA sequencing reads and prediction of RNA secondary structure as performed in other novel microRNA discovery approaches.

6.2 Distinction From Other Types of Small RNA

The discovery of non-coding RNA species revised of our perception of the human genome, gene expression, and regulation. Other small non-coding RNAs have been discovered and continued to be explored in addition to microRNAs (Bartel, 2004; Bartel, 2009; Lim et al., 2003a). In a recent paper, according to information gathered in GENCODE project, there are 13,333 long non-coding RNA genes and 9078 small non-coding RNA genes, 3,086 of which are microRNA genes (Hrdlickova et al., 2014).

The novel microRNA transcripts were the output of a large-scale microRNA-seq project in the pancancer TCGA study, where short RNA transcripts were sequenced and mapped to the human genome. Short RNA transcripts that mapped to the genome but had not yet been annotated as encoding microRNA were taken as potential microRNA if they had more than one read at the genomic locus to which they mapped. The genomic sequence flanking the transcripts were assessed for RNA secondary structure by RNALfold and designated as potential novel microRNA if secondary structure was formed. My analyses of these novel microRNAs was part of a forward genetics approach, screening novel transcripts for a dozen potential microRNA seeds that had been identified by finding microRNA recognition elements in a sponge construct designed to knockdown miR-143. Most of these novel transcripts were excluded as microRNA candidates due to their hairpin structure placing the seed in an inappropriate position for DROSHA/DICER processing, or due to other criteria. There was also the possibility that novel transcripts could be other types of small non-coding RNA. Aside from canonical microRNAs, there are numerous other forms of non-coding RNA. As mentioned in the Introduction, there are piwiRNAs (piRNAs) which function similarly to microRNAs. These RNAs protect the genome from the activity of retro-transposons and are found in sperm and oocytes, as well as human cancers in germline cells (Moyano and Stefani, 2015)(Hrdlickova et al., 2014; Vagin et al., 2006). However, piRNAs do not usually have hairpin structure precursors, and are produced from long, single-stranded non-coding RNA transcripts, which makes it less likely that miR-X is part of this group (Le Thomas et al., 2014). The piRNAs have a distinct biogenesis pathway from microRNA, so to indisputably differentiate between the two groups, RNA-protein complexes found specifically in the biogenesis of the piRNA or microRNA pathways could be immunoprecipitated and tested for the presence of miR-X.

The possibility of miR-X being an ESCC microRNA is also unlikely because these microRNA are tissue-specific and developmental stage specific. Another type of well-known non-coding small RNA are transfer RNAs, which act as intermediates between transcription and translation. While tRNAs are much longer in length than a microRNA, tRNAs will produce tRNA derived small RNA fragments (tRFs). RNA fragments derived from tRNAs are Microprocessor independent and DICER dependent, highly abundant, and plentiful in many tissue types. While tRFs are found in the cytoplasm and can exhibit microRNA-like silencing (Wang et al., 2013), potential microRNAs can be distinguished from tRFs based on the RNA secondary structure of the surrounding genomic sequence. A tRF transcript would map to a genomic region coding a tRNA and the secondary structure it encoded would make the characteristic structure of a tRNA, not that of a microRNA hairpin structure. Many computational tools for novel microRNA discovery eliminate potential candidates based on similarity to known tRNA sequences. There are no tRNA encoding sequences in the genomic region of miR-X and the hairpin structure criteria that are met by miR-X demonstrated it highly unlikely that it is a tRF.

Another class of small non-coding RNAs is the small nucleolar RNAs, or snoRNAs. The snoRNAs perform modification of ribosomal RNA (rRNA) and small nuclear RNA (snRNA) involved in formation of the spliceosome (Liang et al., 2009). They modify these other RNA species by site-specific methylation and pseudouridylation, both of which are essential for rRNA processing into the cytoplasm and proper ribosome function. SnoRNAs have two main structures, C/D-boxes and H/ACA boxes, and the C/D-boxes form at single stem-loop secondary structure. Since the H/ACA boxes form two stem-loop structures, the snoRNAs forming them are not likely to be mistaken for a microRNA. However, the C/D-boxes form a single stem-loop, with the sequence motif of the C-box being RUGAUGA, where R is a purine, and the motif for the D-box is CUGA. Neither of these motifs are found in my microRNA structures of interest, however, so there is a very low likelihood of miR-X being a snoRNA.

There seems little possibility of the chr12 sequence and surrounding region of miR-X belonging to a tRF, piRNA, or snoRNA, due to its adherence to the structure of a microRNA and the tissue in which it was discovered. There is the possibility that uncharacterized tRFs or other small RNAs could bind to the sponge sequence in the cytoplasm, and a future avenue of investigation could include searching the sequences of these small RNAs to check for potential base pairing to the sponge. More importantly, numerous computational methods of novel microRNA

discovery are being refined to improve filtering of these small RNAs at earlier stages of validation.

6.3 Proteomics and Measurement of Expression Changes in microRNA Targets

The purpose of this study was to find any novel microRNAs involved in non-specific binding to the miR-143spg, and to modify the sponge strategy for knockdown of microRNA so that non-specific binding would be eliminated. This can be determined through identifying the target proteins of miR-X and miR-143 and if they undergo regulation under the original sponge versus the modified sponges for miR-X and miR-143. The model of miR-143 knockdown was created using acute myeloid leukemia cell lines and lentiviral sponges, and the effects of loss of microRNA expression were evaluated by quantitative proteomics.

In experiments to measure expression changes due to microRNA regulation, it is important to consider the correlation or differences between the expression of mRNA and of protein. This is a factor for microRNA experiments because microRNAs regulate the expression of proteins and mRNAs through multiple mechanisms. In brief, decreases in mRNA levels due to microRNA can be attributed to deadenylation, which promotes decapping and degradation of the mRNA (Guo et al., 2010). In the case of protein levels, translational inhibition is carried out by microRNAs through blockage of translation initiation. The production and stability of microRNAs, mRNA targets, and proteins, as well as the mechanism of microRNA regulation, affects the degree of changes in expression seen in the experiment.

The correlation between protein abundance and mRNA abundance in the cell, or the lack thereof, has been controversial in the past, but more recently the details of discrepancies have been investigated using high-throughput sequencing and global proteomics. In a study using parallel metabolic pulse labeling for 5,000 human genes, the mRNA and protein levels correspond better than numerous previous studies, while it was the half-lives of mRNAs and proteins that showed no correlation to each other (Schwanhausser et al., 2011). Genome scale prediction of synthesis levels for both mRNAs and proteins found that protein abundance is controlled at the level of translation (Schwanhausser et al., 2011). Proteins have a huge range of translational efficiencies, such that abundant proteins are translated 100 times more

efficiently than those of low abundance proteins, which contributes to the higher dynamic range of proteins compared to mRNAs (Maier et al., 2009; Schwanhausser et al., 2011). Genes with similar stability of mRNA and protein shared functional properties, for example, proteins involved in translational regulation such as eIF4G1, Fxr2 and tuberin had extremely low rate constants (Schwanhausser et al., 2011). It was also found that the mRNAs with longer 3'UTRs are less stable on average, and that highly structured proteins were more stable than unstructured ones. Overall translation rates are the dominant form of control for protein abundance, affected by ribosome density and occupancy, and the impact of degradation is small (Maier et al., 2009). Differences between mRNA and protein levels are attributable to various levels of regulation between mRNA and protein, and almost 40% of variance in protein levels is due to mRNA transcription (Koussounadis et al., 2015).

The impact of inhibiting a microRNA can be observed at the mRNA and protein level. In my experiment, I chose to measure changes in protein expression due to the stability of protein and due to the proteins undergoing translational inhibition that may be missed if only changes in mRNA abundances were measured. However, there may be drawbacks in the consistency of changes in expression, as I have seen in the SILAC data, and the degree of changes in expression, especially when performing microRNA inhibition as opposed to overexpression. Using protein expression changes to evaluate the effects of losing expression of the microRNA of interest is advantageous because the presence or expression of an mRNA may not directly reflect the expression of its protein (Maier et al., 2009). Many proteins are only functional in the presence of certain cofactors or only active in a complex with other proteins, which affects their stability and detection in proteomics. Thus an RNA-Seq experiment may detect mRNA transcripts leading to expression of these proteins, but the translated product may be extremely short-lived. There are a variety of modifications to a protein after translation from a ribosome, affecting activity, localization, and stability, which may also contribute to differences in the levels of mRNA transcripts and the proteome of the cell.

Discrepancies between changes in protein and changes in mRNA may be due to the collection of material, since mRNA when it is collected for RNA-seq or array experiments measure differences at one moment in time (steady state), while SILAC proteomics measures differences integrated over an extended period of protein synthesis (Guo et al., 2010). If either mRNA levels or microRNA activity change during the period of protein synthesis, correspondence between the mRNA destabilization and protein expression inhibition could become distorted (Guo et al.,

2010). An alternative method of judging expression differences between two conditions may be ribosome profiling, which determines the positions of ribosomes on mRNAs with sub-codon resolution by deep-sequencing of ribosome-protected mRNA fragments (RPFs) (Guo et al., 2010). In this study to determine the dominant mode of microRNA regulation, sequencing of mRNA, ribosome profiling, and proteomics were performed for cells transfected with miR-155 or miR-1. In the mRNA-seq and RPF data, genes with a log₂ fold-change of ≤ -0.3 in their corresponding proteins were examined and greater decreases in mRNAs or RPFs were observed for genes that had 7mer or 8mer seed target sites in their 3'UTRs. The study determined from the range of mRNA, RPF, and protein changes that the expression changes detected by proteomics "accurately represented the response of all mRNAs" (Guo et al., 2010). It was also found that microRNAs did not repress targets with low expression more potently than targets with higher expression levels (Guo et al., 2010). These latter two observations are important to note in the context of my study, to realize that they are not prominent concerns.

The ribosome profiling data was also used to determine if microRNAs could affect translation without decreasing mRNA levels. Translational repression may occur because of reduced translation initiation (Chendrimada et al., 2005) or increased ribosome drop-off (Petersen et al., 2006). The level of RPFs for a particular gene are decreased if there are fewer mRNA transcripts or fewer ribosomes occupying the mRNA and reduced initiation or increased ribosome drop-off would lead to fewer ribosomes occupying the mRNAs of a particular gene (Ingolia, 2014). It was shown that translation efficiency was decreased by a modest amount, and by coordinating the proteomic, RPF, and mRNA data, the authors deduced that 11-16% of repression in RPFs is from decreased translational efficiency and 84% of repression was due to decreased mRNA levels (Guo et al., 2010).

Often, because of their availability and ease in processing compared to proteins, the levels of mRNA between different treatments or conditions are used in differential expression analyses and taken as a proxy for the corresponding proteins (Vogel and Marcotte, 2012). The differences are taken to imply that there is a functional difference with biological relevance in the cell (Koussounadis et al., 2015). Since there is not always correlation between the levels of an mRNA and corresponding protein, are the mRNA changes signifying a true biological change? A study by Koussounadis A *et al* (2015) verified this assumption, finding that differentially expressed mRNA and their proteins have higher correlation than non-differentially expressed genes, within experimental conditions. For observing changes relevant to microRNA

knockdown, finding correlations in differentially expressed genes at the mRNA level and protein level would lend credibility to the miR-X and miR-143 specific changes, and ribosome profiling data would also be desirable to link the changes in expression between datasets. Since recent studies have shown that changes seen at the protein level are also seen at the mRNA and RPF level if the corresponding mRNA has a 3'UTR binding site. These studies have shown that significant expression changes seen at the protein level are relevant to microRNA knockdown, but additional observations of the mRNA and RPF levels would help to discriminate direct effects of microRNA regulation. I have recently acquired RNA-Seq data for cells transduced with the pLL-GFP, pLL-Orig-Spg, miR-Xspecific-spg and miR-143specific-spg lentiviruses and are beginning to correlate and compare changes.

Proteomics provides a quantitative measure of changes in protein expression, and has improved in reproducibility and detection in recent years. Advances in next-generation sequencing and proteomics have provided greater resolution and range in measuring the abundances of mRNA and protein, and for comparing expression between the two. When measuring changes in protein expression in a large-scale, discovery-based method, there are a number of challenges that interfere with the measurement of the proteome of a cell in global proteomic experiments. Firstly, is the issue of reproducibility between biological replicates. There is variation between all quantitative proteomics samples, since biological systems are highly dynamic and the amount of protein produced from one gene can vary(Mardis et al., 2009; Nagaraj et al., 2011; Viney and Reece, 2013). In a study performed in cancer cell lines, there is variation from culture dish to culture dish. There is also large cell-to-cell variability in expression of mRNAs and proteins (Paszek, 2014). Gene expression can be affected by intrinsic noise in biochemical processes related to gene expression, or by extrinsic noise from other cellular process such as cell growth, which is a major source of noise (Lei et al., 2015).

Technical variation between proteomics experiments is also a factor in differential expression experiments. The steps with the most variability are the extraction where tissue is collected, the protein denaturation and digestion by trypsin/SPE clean-up, the fluctuation of the instrument response, and the long-term instrument stability, since there is sometimes drift of quantitative response of the LC-MS/MS platform over the 2-week period (Piehowski et al., 2013). The tissue collection can be a huge source of variation depending on how the samples are processed and from where they are obtained. Digestion of proteins can also introduce considerable variation,

but some studies will minimize this contribution by automating the digestion and preparation of numerous samples (Piehowski et al., 2013).

Another issue in proteomics is the dynamic range for detecting different proteins. Protein expression can vary from one copy per cell to ten million copies per cell (Zubarev A et al 2013). To detect greater than 5,000 proteins, or half of the expressed cellular proteome, much longer periods of chromatographic separation are need. Coverage of the proteome is improving but not complete. One study reported nearly 90% coverage of the expressed proteome by using liquid chromatography (LC) and high resolution mass spectrometry to identify 10,255 proteins encoded by 9,207 genes and comparing the abundances to deep sequencing transcriptome data of the same samples (Nagaraj et al., 2011). More recently, a study identified proteins encoded by 17,294 genes in a variety of tissues, almost 84% of the protein coding genes in the human genome (Kim et al., 2014b).

Overall, increasing coverage of the proteome, consistent coverage, and automation, or other techniques to lessen variation at the protein digestion and sample preparation stage, help to alleviate the biological and technical noise in proteomics samples.

6.4 Statistical Analysis of Proteomic Datasets

Analysis of my proteomics datasets involved two major types of statistical analysis. In the first dataset, I implemented an Empirical Bayes method developed by Margolin *et al* (2009) to find the proteins with higher probability of being upregulated or downregulated. This method reassessed microRNA regulation data from previous experiments by a new statistical method to identify biologically relevant proteins based on SILAC ratio values. Their study examined previous statistical methods used for quantitative proteomics, including a Gaussian mixture model and density estimation methods for expression analysis, and found they were not robust enough or tended to overfit regions of data sparsity. The empirical Bayesian method distinguished proteins that differed from background based on the probability distributions inferred from two replicates for my experiment. However, one criticism of the method was that it did not take into account proteins with a high probability of being upregulated or downregulated in one sample, but did not share the same expression trend in another replicate. I implemented a second-step, where proteins were filtered according to their overlap in upregulation or

downregulation. However, better coverage of the proteome and a larger number of replicates would have improved this approach, since many proteins with large expression changes were eliminated due to lack of presence in the second replicate.

Another major approach, which was taken with the second proteomic dataset, was to find differential gene expression between different conditions using linear regression modeling. Linear regression modeling was performed using the "Limma" R package in Bioconductor. Limma stands for "Linear Models for Microarray Data" and is an analytical standard in differential gene expression, however is only beginning to be used in the context of proteomics experiments (Kammers et al., 2015). It is appropriate because it allows for a distribution of biological variances (different proteins to change expression by different amounts), by using the full data to shrink the observed sample variances toward a pooled estimate (Kammers et al., 2015). The linear regression analysis was used to find what constituted a significant change in expression by comparison of the whole dataset. One of the advantages of using Limma for proteomics analysis it is appropriate for smaller sample sizes like those found in proteomics data (Kammers et al., 2015). For my data, it was a useful and robust way of finding the differentially expressed proteins between my sponge "conditions." However, I could adjust the proteins evaluated by the linear regression analysis. I filtered numerous proteins from the dataset because their expression changes were not as reproducible between repilicates. However, many showed similar trends in upregulation or downregulation, and proteins that are highly variable may still be biologically relevant. The concern that prompted us to remove them was that the highly variable and noisy proteins would obscure the differences between other proteins, which followed a clear linear response. One strategy that may be useful for comparison with my current results is to apply the linear regression modeling to our entire proteomics dataset of 5215 proteins instead of the selected 3110 proteins, or simply a larger dataset with less stringent filtering.

The statistical approaches taken to find the differentially expressed and upregulated or downregulated proteins may have limited the results in other ways as well. Based on the idea that miR-X and miR-143 could both bind to the original sponge and that modification of the sponge would inhibit specifically miR-X or miR-143, I expected that numerous proteins with expression changes in cells transduced with the original sponge would share changes in either the miR-X or miR-143 specific sponge as well. However, I observed small groups of proteins undergoing overlapping changes between the original sponge and modified sponge

transductions, and large groups of proteins undergoing expression changes in only the miR-143 or miR-X specific knockdowns. As discussed in the case of the predicted target of miR-X, GINS3, many proteins undergoing upregulation in the miR-Xspecific-spg or miR-143specific-spg condition only may have proteins that are upregulated in pLL-Orig-Spg as well, but at a level slightly below the log₂ fold change cut-off of 0.3. Variation in the biological states of the cells has already been mentioned as a contributor in this situation, and a higher number of replicate samples offered as a solution. However, there are a number of other reasons for the higher number of targets occuring in only the miR-X or miR-143 specific sponge subsets.

Firstly, this may have been due to knockdown of both miR-X and miR-143 by the original sponge at the same time, and regulation of targets by one microRNA leading to downstream signaling effects that interfered with or obscured the regulation of targets by the other microRNA. It is difficult to know whether the upregulated or downregulated proteins that do not overlap between subsets are undergoing changes in expression directly due to the knockdown of a microRNA, indirectly, or due to biological noise. Another possibility is that there is another novel microRNA binding to different repeated sequences in the original sponge, and that inhibition of this microRNA leads to regulation of a set of proteins with no overlap in the miR-X or miR-143 affected proteins.

Proteins with a high degree of upregulation or downregulation in cells transduced with one type of sponge while not changing in the other two sponge treatments may still demonstrate differences in the proteome caused by the specific knockdown of one microRNA. Interestingly, the proteins undergoing significant expression changes in only miR-Xspecific-spg transduced cells form the largest group. This is expected because it is the condition with the most significantly differentially expressed proteins from the other two conditions. The reason for the greater number of upregulated miR-X proteins and the greater number of proteins undergoing significant target of miR-Xspecific-spg cells may be due to a cascade effect from one particular target of miR-X. The targets of miR-X could include a transcription factor or a protein which similarly affects the expression of a large number of other proteins, and the greater number of upregulated not similarly affects the expression of a large number of other proteins, and the greater number of upregulated number of upregulated proteins in miR-X knockdown cells may be the downstream effects of this regulation.

6.5 Target Prediction for microRNAs

One issue in determining whether distinct sets of proteins have been changed due to the inhibition of miR-143 or miR-X in my dataset is the low number of predicted targets that show appropriate patterns of significant expression changes. However, this may be due to the low specificity found for many of the available target prediction tools.

The prediction of microRNA targets was a computationally challenging task in the time directly following their discovery. The pairing of the microRNA to the target did not seem to involve the entirety of the microRNA sequence, and many of the binding sites were discovered to contain only 7 nt matches to the 5' region of the microRNA. The requirement of conserved Watson-Crick base pairing, using the aligned regions of vertebrate 3'UTRs, to nucleotides 2-7 in the 5' region of the microRNA found a greater number of true targets (Brennecke et al., 2005b; Krek et al., 2005; Lewis et al., 2005; Lewis et al., 2003). The 5' region of the microRNA is the most conserved portion of metazoan microRNAs (Lim et al., 2003b), and it was found that mutation in this region led to disruption of target regulation (Doench and Sharp, 2004; Kloosterman et al., 2004). Initially, in D. melanogaster, the majority of sites in target mRNAs had no more 3' supplementary pairing than would be expected by random chance (Brennecke et al., 2005a) and in mammals, a 3' supplementary region in the target sites was not a strict requirement (Lewis et al., 2005). However, it was later found that some targets that lack perfect seed matches have compensatory sites in the 3' region of the microRNA, preferentially from the 13th to the 16th nucleotides (Grimson et al., 2007). With the inhibition of miR-X and miR-143 by the original sponge, there were four sites with perfect matches to miR-143 seed sites, and two sites with perfect matches to the miR-X seed site, which would have allowed binding by both microRNAs. The miR-143 specific modified sponge had strategic mutations in the region of the miR-143 binding site, which bound the 3' end of miR-143 and the 5' seed of miR-X, but had complete WC base pairing to the miR-143 seed to maintain binding ability. The miR-X specific modified sponge had the number of seed matches increased from two to four, and the miR-143 seed sites mutated, which should increase miR-X binding to the sponge.

There are three types of computer prediction tools - *ab initio* which uses sequence and structural features, machine learning which uses experimental data to train a classifier, and hybrid methods which combine the previous two (Reyes-Herrera and Ficarra, 2012). Most *ab initio* target prediction algorithms use four major factors to predict potential targets, those being:

seed match, conservation, free energy, and site accessibility (Peterson et al., 2014). The problem with *ab initio* methods is that they have a high number of false positives, and to combat this restrictions are used to filter candidate targets, which can remove true positives as well. Machine learning methods have been making advancements in recent years due to the proliferation of experimental data, and these tools learn based on positive interactions between microRNA and target, and negative interactions between microRNA and predicted target. The problem with this approach is that fewer negative interactions are validated and reported in the literature.

The target prediction tools used for evaluating potential seeds in my first SILAC dataset and for finding miR-X or miR-143 specific targets in the second dataset were TargetScan or TargetScan Custom. In evaluating the SILAC1 dataset to screen the potential seeds, numerous predicted targets were given for each seed because there was only one sample that was searched in, and no other samples were used to remove targets with conflicting expression changes. There was the potential for many false positives in this search, but the nature of the screen was simply to see if any of the potential seeds had more probability of target regulation in the dataset than miR-143 or might be expected due to random chance. In my proteomics dataset for the modified sponges, there were 20 predicted targets of miR-X found, and 120 predicted targets of miR-143, but only two proteins for each microRNA that had the appropriate expression pattern and were found to be significantly differentially expressed between conditions. Many predicted targets were eliminated from being true positives because of their variability in expression between replicates. Others were eliminated because the predicted targets did not show significant upregulation or downregulation. My study confirms the high number of false positives expected from an *ab initio* method such as TargetScan, but the number of predicted targets behaving as potential true positives, especially for miR-143, is small. Validation of these targets as true direct targets of miR-143 or miR-X needs to be performed by orthogonal methods. Other target prediction tools could be applied to the dataset to find targets that TargetScan has missed. Similarly, performing the transductions of a variety of different cell lines with the modified sponges would be able to more convincingly confirm some of the miR-X or miR-143 specific effects, and potentially find more direct targets. There are different estimates on how many proteins are affected by overexpression or inhibition of a microRNA, and in the case of inhibition of a microRNA the changes in protein expression may be small, as microRNAs often act to fine tune the expression of many proteins at once. The magnitude of change and number of targets depends on the tissue, the microenvironment, the metabolic state of the cell, and the expression

of both the microRNA and its targets. Proteins that change expression due to miR-X or miR-143 inhibition across different tissues are likely to be rare, but every microRNA has a broad range of targets, and some may be common to numerous tissue types. Repeating the sponge knockdown and proteomic differential expression analysis in other cell lines with expression of miR-X or miR-143 could potentially confirm whether the changes in expression are reproducible in a small subset of targets.

As well, it may be useful to utilize a prediction tool that takes into account non-canonical microRNA binding to the mRNA target. A number of publications have noted that microRNA can act by binding sites corresponding to the middle of the microRNA (Cloonan, 2015; Martin et al., 2014), and imply that for a number of targets the canonical seed pairing may not be as important. There are also microRNA binding sites outside of the 3'UTR of the mRNA, but multiple studies have described these sites as less efficacious (Baek et al., 2008; Selbach et al., 2008).

Other methods which could potentially be used to determine the targets of miR-143 or miR-X are cross-linking immunoprecipitation-sequencing techniques. In the high-throughput-sequencing of RNA from cross-linking immunoprecipitation (HITS-CLIP) technique, microRNA and mRNA are cross-linked to Argonaute proteins, separated, and sequenced (Chi et al., 2009). To increase the number of targets bound by Argonaute for a particular microRNA, the microRNA is usually transfected into the cells or expressed by a retro- or lentiviral vector, leading to enrichment of transcripts with the microRNA binding site in sequencing data. Another technique is the cross-linking, ligation, and sequencing of hybrids (CLASH) method (Helwak et al., 2013). This method uses cross-linking and immunoprecipitation of AGO1, followed by ligation of the cross-linked microRNA and mRNA to make hybrids, and then sequencing of the microRNA-target pairs as chimeric reads (Helwak et al., 2013). Currently efforts are ongoing to optimize a CLIP-seq method in our lab to study a number of microRNAs of interest, which would provide us with a method to cross-validate the predicted targets in our proteomics datasets.

6.6 Advantages and Disadvantages of the Sponge Method

There are a variety of methods for disrupting or inhibiting the expression of microRNAs in models of microRNA loss. As discussed in the introduction, there are genetic knockouts, which have the advantage of completely knocking out expression of the microRNA, but the disadvantage of suffering from their own variation of off-target effects. The CRISPR/Cas9 gene editing system, which uses 20 bp guide strands to target the sequence of interest, has been verified by three different groups as inducing off-target mutations at sites that differ by 5 nt from on-target sites (Fu Y et al 2013, Hsu PD et al 2013, Pattanayak V et al 2013). As well, many microRNAs are part of families with redundant function in regulation of targets, and knockout of one member of the family is compensated by regulation by other members.

Other methodology includes anti-sense oligonucleotides (ASOs) against the microRNA of interest. There are a wide variety of chemical modifications that are used to protect the ASO from exonuclease digestion and enhance potency, or to add stability to the dsRNA duplex when the ASO binds to the intended microRNA. I employed an anti-sense oligonucleotide design from Integrated DNA Technologies, which has 2'-O-methyl bases for greater stability and a novel compound, N,N-diethyl-4-(4-nitronaphthalen-1-ylazo)-phenylamine, attached at the end to of the oligonucleotide to block exonuclease degradation (Lennox KA et al 2013). This type of compound was chosen for a balance of potency, stability, and cost-effectiveness. It provided inhibition of miR-X that was observable by derepression of miR-X luciferase targets.

The use of anti-sense oligonucleotides for knockdown of microRNA expression is advancing and offers increasing options to researchers (Lennox K et al 2011, Pauli A et al 2015). In light of my research into the non-specific binding to the sponge, there is some concern over specificity of binding for the ASOs, and whether they can have off-target effects (Lennox KA et al 2011). Another downside of this method is that many types of cells can be difficult to transfect, particularly primary cells or primitive cells from the bone marrow.

A similar approach to anti-sense oligonucleotides is using microRNA target site blockers, which have recently been developed by Exiqon and implement their LNA technology. Instead of binding to the microRNA, these small LNAs bind to the microRNA target sites on mRNAs and effectively compete with RISC (Dajas-Bailador et al., 2012). The target site blockers can be specific to a particular target, and increase protein translation of that target. This is helpful for

confirming a single microRNA target, instead of inhibiting a microRNA and affecting multiple targets, then attempting to assess which are the direct and indirect effects.

Recently, a new variety of lentiviral microRNA inhibitors has emerged, similar to sponges but with modifications to improve specificity. These are known as Tough Decoy inhibitors, which come in the form of long hairpins about 60 bp long, with an internal loop containing two microRNA binding sites (Bak et al., 2013). These can be transfected as small hairpin species more similar to ASOs, or be transcribed by RNA Pol-II promoters as a transcript with a cap, poly-adenylation and a fused protein reporter like our lentiviral sponges (Haraguchi et al., 2009; Mockenhaupt et al., 2015). The hairpin shape of the Tough Decoy exposes two ~20 bp binding sites for the microRNA, which contain bulges in the reverse complement in the same manner as sponges (Mockenhaupt et al., 2015). However, the hairpin structure exposing a binding site of limited length would potentially be less thermodynamically favourable for a non-specific microRNA binding by seed site.

My study highlighted strengths and weakness of using the lentiviral sponge method of competitive inhibition of microRNA to simulate loss of a microRNA in a disease. Since microRNA are often found in families that share the same seed site, genetic knockout of a microRNA may not be sufficient, and anti-sense oligonucleotides are capable of inhibiting micro expression, but can be difficult to transfect into some cell lines or may not fully inhibit the microRNA of interest. Using sponges as a method of knockdown improves upon the other two techniques, but my study emphasizes the importance of awareness for potential non-specific binding and offers small but significant improvements to the method for eliminating the non-specific effects.

6.7 Implications for Gene Therapy and microRNA for Therapeutics

The use of microRNA for therapeutic use is promising because of the ability to control gene regulation. There are two categories of gene therapy that would involve microRNAs, either decreasing microRNA expression through microRNA antagonists, or increasing and mimicking microRNA activity through synthesized hairpins.

MicroRNA knockdown in gene therapy could potentially use LNAs, ASOs, etc, but toxicity and delivery issues will still need to be considered in this approach (Brown and Naldini, 2009). MicroRNA sponges have been viewed as a viable option because they behave more like ceRNAs, natural endogenous targets for microRNA that keep the balance of expression of certain genes, like PTENP1 (Almeida et al., 2012). In proposed use of ceRNAs as a type of gene therapy, the actual concentration of microRNAs does not change, but the ratio between the ceRNA and its corresponding mRNAs will be adjusted (Almeida et al., 2012).

Expression of microRNA sponges does offer more stability and long-term effectiveness than a LNA or ASO agonist, but the dosage would be similarly challenging to control. As well, use of retroviral or lentivral vectors in gene therapy is still fraught with difficulties due to the uncertainty of viral integration into the patient's genome.

In light of my study, yet more caution is advised for microRNA sponges beyond what is already proposed because of the potential for non-specific binding. Non-specific binding would be a risk, not only because of the challenge of controlling the dosage level of the sponge transcript, but also because, as demonstrated, there may be expression of unknown microRNAs that interferes with the intended effect of regulation.

6.8 Future Directions

One approach to validating the protein expression changes specific to miR-X or miR-143 knockdown as actual miR-X or miR-143 targets will be to complement the knockdown experiments with overexpression experiments. I have both lentiviral overexpression vectors and the microRNA hairpin mimics which were used in the luciferase sponge repression assays. Obviously, this method will lead to changes in direct and indirect targets in the same manner that knockdown of miR-143 and miR-X did. However, proteins that are direct targets should display an opposite change in expression from the knockdown datasets, and narrow the pool of proteins with which to conduct further validation assays.

I have demonstrated the concerns and the utility of using the sponge method of microRNA knockdown. One advantage of the sponge is the ability to be transduced into non-dividing cells and using this tool, there are a number of ways I could improve upon our model of miR-143 loss

in MDS. One method would be through use of primary HSPC cells from human cord blood rather than human cell lines. Using material from this source for proteomics would have been optimal but was more difficult in the initial outset of my study. Since that time, advances in mass spectrometer sensitivity in proteomics have enabled global proteomics with smaller amounts of material than before, and the ability to investigate the loss of miR-143 in a primary cell environment is more feasible.

This approach would also be potentially interesting because cancer cell lines undergo proliferation and are capable of differentiation with certain stimuli, but the myeloid leukemia cell lines are differentiated to some extent. It is possible that miR-143 targets certain genes during the transition from one type of blood cell to another, which could therefore be missed in the cell line model. Various microRNA have a gradient of expression that increases or decreases throughout lineage differentiation to a terminated cell, as discussed in the introduction, and it is possible that miR-143 has targets with important roles in differentiation. It may be possible to see expression changes in proteins that relate to or interact with these targets, but a model with differentiation would be more capable of observing such changes.

References

- Abdel-Wahab, O., T. Manshouri, J. Patel, K. Harris, J. Yao, C. Hedvat, A. Heguy, C. Bueso-Ramos, H. Kantarjian, R.L. Levine, and S. Verstovsek. 2010. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer research* 70:447-452.
- Abdelmohsen, K., and M. Gorospe. 2012. RNA-binding protein nucleolin in disease. *RNA biology* 9:799-808.
- Adolfsson, J., R. Mansson, N. Buza-Vidas, A. Hultquist, K. Liuba, C.T. Jensen, D. Bryder, L. Yang, O.J. Borge, L.A. Thoren, K. Anderson, E. Sitnicka, Y. Sasaki, M. Sigvardsson, and S.E. Jacobsen. 2005. Identification of Flt3+ lympho-myeloid stem cells lacking erythromegakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* 121:295-306.
- Akao, Y., Y. Nakagawa, Y. Kitade, T. Kinoshita, and T. Naoe. 2007a. Downregulation of microRNAs-143 and -145 in B-cell malignancies. *Cancer science* 98:1914-1920.
- Akao, Y., Y. Nakagawa, and T. Naoe. 2007b. MicroRNA-143 and -145 in colon cancer. DNA and cell biology 26:311-320.
- Allison, A.C. 1972. Immunity against viruses. *The Scientific basis of medicine annual reviews* 49-73.
- Almeida, M.I., R.M. Reis, and G.A. Calin. 2012. Decoy activity through microRNAs: the therapeutic implications. *Expert opinion on biological therapy* 12:1153-1159.
- Ambros, V., B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S.R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, and T. Tuschl. 2003. A uniform system for microRNA annotation. *Rna* 9:277-279.
- Aravin, A.A., G.J. Hannon, and J. Brennecke. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318:761-764.
- Artzi, S., A. Kiezun, and N. Shomron. 2008. miRNAminer: a tool for homologous microRNA gene search. *BMC bioinformatics* 9:39.
- Asangani, I.A., S.A. Rasheed, D.A. Nikolova, J.H. Leupold, N.H. Colburn, S. Post, and H. Allgayer. 2008. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* 27:2128-2136.

- Asikainen, S., L. Heikkinen, J. Juhila, F. Holm, J. Weltner, R. Trokovic, M. Mikkola, S. Toivonen,
 D. Balboa, R. Lampela, K. Icay, T. Tuuri, T. Otonkoski, G. Wong, and O. Hovatta. 2015.
 Selective microRNA-Offset RNA expression in human embryonic stem cells. *PloS one* 10:e0116668.
- Asirvatham, A.J., C.J. Gregorie, Z. Hu, W.J. Magner, and T.B. Tomasi. 2008. MicroRNA targets in immune genes and the Dicer/Argonaute and ARE machinery components. *Molecular immunology* 45:1995-2006.
- Auyeung, V.C., I. Ulitsky, S.E. McGeary, and D.P. Bartel. 2013. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* 152:844-858.
- Bacher, U., C. Haferlach, W. Kern, T. Haferlach, and S. Schnittger. 2008. Prognostic relevance of FLT3-TKD mutations in AML: the combination matters--an analysis of 3082 patients. *Blood* 111:2527-2537.
- Bacher, U., T. Haferlach, W. Kern, C. Haferlach, and S. Schnittger. 2007. A comparative study of molecular mutations in 381 patients with myelodysplastic syndrome and in 4130 patients with acute myeloid leukemia. *Haematologica* 92:744-752.
- Baek, D., J. Villen, C. Shin, F.D. Camargo, S.P. Gygi, and D.P. Bartel. 2008. The impact of microRNAs on protein output. *Nature* 455:64-71.
- Bak, R.O., A.K. Hollensen, M.N. Primo, C.D. Sorensen, and J.G. Mikkelsen. 2013. Potent microRNA suppression by RNA Pol II-transcribed 'Tough Decoy' inhibitors. *Rna* 19:280-293.
- Baldi, P., and A.D. Long. 2001. A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. *Bioinformatics* 17:509-519.
- Bar, N., and R. Dikstein. 2010. miR-22 forms a regulatory loop in PTEN/AKT pathway and modulates signaling kinetics. *PloS one* 5:e10859.
- Barlow, J.L., L.F. Drynan, D.R. Hewett, L.R. Holmes, S. Lorenzo-Abalde, A.L. Lane, H.E. Jolin,
 R. Pannell, A.J. Middleton, S.H. Wong, A.J. Warren, J.S. Wainscoat, J. Boultwood, and
 A.N. McKenzie. 2010a. A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nature medicine* 16:59-66.
- Barlow, J.L., L.F. Drynan, N.L. Trim, W.N. Erber, A.J. Warren, and A.N. McKenzie. 2010b. New insights into 5q- syndrome as a ribosomopathy. *Cell cycle* 9:4286-4293.
- Baron, M.H. 2013. Concise Review: early embryonic erythropoiesis: not so primitive after all. *Stem cells* 31:849-856.

- Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281-297.
- Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. Cell 136:215-233.
- Bartel, D.P., and C.Z. Chen. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature reviews. Genetics* 5:396-400.
- Behm-Ansmant, I., J. Rehwinkel, T. Doerks, A. Stark, P. Bork, and E. Izaurralde. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & development* 20:1885-1898.
- Bejar, R., R. Levine, and B.L. Ebert. 2011. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 29:504-515.
- Bench, A.J., E.P. Nacheva, T.L. Hood, J.L. Holden, L. French, S. Swanton, K.M. Champion, J. Li, P. Whittaker, G. Stavrides, A.R. Hunt, B.J. Huntly, L.J. Campbell, D.R. Bentley, P. Deloukas, and A.R. Green. 2000. Chromosome 20 deletions in myeloid malignancies: reduction of the common deleted region, generation of a PAC/BAC contig and identification of candidate genes. UK Cancer Cytogenetics Group (UKCCG). *Oncogene* 19:3902-3913.
- Bennett, J.M., D. Catovsky, M.T. Daniel, G. Flandrin, D.A. Galton, H.R. Gralnick, and C. Sultan.
 1976. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *British journal of haematology* 33:451-458.
- Bennett, J.M., D. Catovsky, M.T. Daniel, G. Flandrin, D.A. Galton, H.R. Gralnick, and C. Sultan. 1982. Proposals for the classification of the myelodysplastic syndromes. *British journal of haematology* 51:189-199.
- Bentwich, I. 2005. Prediction and validation of microRNAs and their targets. *FEBS letters* 579:5904-5910.
- Bentwich, I., A. Avniel, Y. Karov, R. Aharonov, S. Gilad, O. Barad, A. Barzilai, P. Einat, U. Einav, E. Meiri, E. Sharon, Y. Spector, and Z. Bentwich. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nature genetics* 37:766-770.
- Berezikov, E., W.J. Chung, J. Willis, E. Cuppen, and E.C. Lai. 2007. Mammalian mirtron genes. *Molecular cell* 28:328-336.
- Berezikov, E., E. Cuppen, and R.H. Plasterk. 2006a. Approaches to microRNA discovery. *Nature genetics* 38 Suppl:S2-7.
- Berezikov, E., G. van Tetering, M. Verheul, J. van de Belt, L. van Laake, J. Vos, R. Verloop, M. van de Wetering, V. Guryev, S. Takada, A.J. van Zonneveld, H. Mano, R. Plasterk, and E.

Cuppen. 2006b. Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome research* 16:1289-1298.

- Bernasconi, P., C. Klersy, M. Boni, P.M. Cavigliano, S. Calatroni, I. Giardini, B. Rocca, R. Zappatore, M. Caresana, J. Quarna, M. Lazzarino, and C. Bernasconi. 2005. Incidence and prognostic significance of karyotype abnormalities in de novo primary myelodysplastic syndromes: a study on 331 patients from a single institution. *Leukemia* 19:1424-1431.
- Bernstein, E., A.A. Caudy, S.M. Hammond, and G.J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409:363-366.
- Bertrand, J.Y., N.C. Chi, B. Santoso, S. Teng, D.Y. Stainier, and D. Traver. 2010. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* 464:108-111.
- Boeck, R., S. Tarun, Jr., M. Rieger, J.A. Deardorff, S. Muller-Auer, and A.B. Sachs. 1996. The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *The Journal of biological chemistry* 271:432-438.
- Bohnsack, M.T., K. Czaplinski, and D. Gorlich. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* 10:185-191.
- Boland, A., E. Huntzinger, S. Schmidt, E. Izaurralde, and O. Weichenrieder. 2011. Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proceedings of the National Academy of Sciences of the United States of America* 108:10466-10471.
- Bommer, G.T., I. Gerin, Y. Feng, A.J. Kaczorowski, R. Kuick, R.E. Love, Y. Zhai, T.J. Giordano, Z.S. Qin, B.B. Moore, O.A. MacDougald, K.R. Cho, and E.R. Fearon. 2007. p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Current biology : CB* 17:1298-1307.
- Borchert, G.M., W. Lanier, and B.L. Davidson. 2006. RNA polymerase III transcribes human microRNAs. *Nature structural & molecular biology* 13:1097-1101.
- Boultwood, J., C. Fidler, S. Lewis, S. Kelly, H. Sheridan, T.J. Littlewood, V.J. Buckle, and J.S.
 Wainscoat. 1994a. Molecular mapping of uncharacteristically small 5q deletions in two patients with the 5q- syndrome: delineation of the critical region on 5q and identification of a 5q- breakpoint. *Genomics* 19:425-432.
- Boultwood, J., C. Fidler, A.J. Strickson, F. Watkins, S. Gama, L. Kearney, S. Tosi, A. Kasprzyk, J.F. Cheng, R.J. Jaju, and J.S. Wainscoat. 2002. Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. *Blood* 99:4638-4641.

Boultwood, J., S. Lewis, and J.S. Wainscoat. 1994b. The 5q-syndrome. *Blood* 84:3253-3260.

- Boultwood, J., A. Pellagatti, A.N. McKenzie, and J.S. Wainscoat. 2010. Advances in the 5qsyndrome. *Blood* 116:5803-5811.
- Bousquet, M., M.H. Harris, B. Zhou, and H.F. Lodish. 2010. MicroRNA miR-125b causes leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 107:21558-21563.
- Bousquet, M., D. Nguyen, C. Chen, L. Shields, and H.F. Lodish. 2012. MicroRNA-125b transforms myeloid cell lines by repressing multiple mRNA. *Haematologica* 97:1713-1721.
- Braun, J.E., E. Huntzinger, M. Fauser, and E. Izaurralde. 2011. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Molecular cell* 44:120-133.
- Brennecke, J., A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam, and G.J. Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* 128:1089-1103.
- Brennecke, J., D.R. Hipfner, A. Stark, R.B. Russell, and S.M. Cohen. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* 113:25-36.
- Brennecke, J., A. Stark, and S.M. Cohen. 2005a. Not miR-ly muscular: microRNAs and muscle development. *Genes & development* 19:2261-2264.
- Brennecke, J., A. Stark, R.B. Russell, and S.M. Cohen. 2005b. Principles of microRNA-target recognition. *PLoS biology* 3:e85.
- Brown, B.D., and L. Naldini. 2009. Exploiting and antagonizing microRNA regulation for therapeutic and experimental applications. *Nature reviews. Genetics* 10:578-585.
- Brown, B.D., M.A. Venneri, A. Zingale, L. Sergi Sergi, and L. Naldini. 2006. Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. *Nature medicine* 12:585-591.
- Brown, C.E., and A.B. Sachs. 1998. Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. *Molecular and cellular biology* 18:6548-6559.
- Cabrero, M., Y. Wei, H. Yang, I. Ganan-Gomez, Z. Bohannan, S. Colla, M. Marchesini, G.M.
 Bravo, K. Takahashi, C. Bueso-Ramos, and G. Garcia-Manero. 2016. Down-regulation of
 EZH2 expression in myelodysplastic syndromes. *Leukemia research* 44:1-7.
- Cai, X., C.H. Hagedorn, and B.R. Cullen. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* 10:1957-1966.
- Calin, G.A., C.D. Dumitru, M. Shimizu, R. Bichi, S. Zupo, E. Noch, H. Aldler, S. Rattan, M. Keating, K. Rai, L. Rassenti, T. Kipps, M. Negrini, F. Bullrich, and C.M. Croce. 2002.Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14

in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* 99:15524-15529.

- Calin, G.A., C. Sevignani, C.D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini, and C.M. Croce. 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America* 101:2999-3004.
- Cevec, M., and J. Plavec. 2010. Solution structure of miRNA:mRNA complex. *Methods in molecular biology* 667:251-265.
- Cevec, M., C. Thibaudeau, and J. Plavec. 2008. Solution structure of a let-7 miRNA:lin-41 mRNA complex from C. elegans. *Nucleic acids research* 36:2330-2337.
- Cevec, M., C. Thibaudeau, and J. Plavec. 2010. NMR structure of the let-7 miRNA interacting with the site LCS1 of lin-41 mRNA from Caenorhabditis elegans. *Nucleic acids research* 38:7814-7821.
- Chagraoui, J., A. Lepage-Noll, A. Anjo, G. Uzan, and P. Charbord. 2003. Fetal liver stroma consists of cells in epithelial-to-mesenchymal transition. *Blood* 101:2973-2982.
- Chang, T.C., E.A. Wentzel, O.A. Kent, K. Ramachandran, M. Mullendore, K.H. Lee, G. Feldmann, M. Yamakuchi, M. Ferlito, C.J. Lowenstein, D.E. Arking, M.A. Beer, A. Maitra, and J.T. Mendell. 2007. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Molecular cell* 26:745-752.
- Chekulaeva, M., H. Mathys, J.T. Zipprich, J. Attig, M. Colic, R. Parker, and W. Filipowicz. 2011. miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nature structural & molecular biology* 18:1218-1226.
- Chen, C.Y., L.I. Lin, J.L. Tang, B.S. Ko, W. Tsay, W.C. Chou, M. Yao, S.J. Wu, M.H. Tseng, and H.F. Tien. 2007. RUNX1 gene mutation in primary myelodysplastic syndrome--the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. *British journal of haematology* 139:405-414.
- Chen, C.Y., D. Zheng, Z. Xia, and A.B. Shyu. 2009. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nature structural & molecular biology* 16:1160-1166.
- Chen, C.Z., L. Li, H.F. Lodish, and D.P. Bartel. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303:83-86.
- Chen, P.Y., H. Manninga, K. Slanchev, M. Chien, J.J. Russo, J. Ju, R. Sheridan, B. John, D.S. Marks, D. Gaidatzis, C. Sander, M. Zavolan, and T. Tuschl. 2005. The developmental

miRNA profiles of zebrafish as determined by small RNA cloning. *Genes & development* 19:1288-1293.

- Chen, P.Y., L. Weinmann, D. Gaidatzis, Y. Pei, M. Zavolan, T. Tuschl, and G. Meister. 2008. Strand-specific 5'-O-methylation of siRNA duplexes controls guide strand selection and targeting specificity. *Rna* 14:263-274.
- Chen, Y.W., S. Song, R. Weng, P. Verma, J.M. Kugler, M. Buescher, S. Rouam, and S.M. Cohen. 2014. Systematic study of Drosophila microRNA functions using a collection of targeted knockout mutations. *Developmental cell* 31:784-800.
- Chendrimada, T.P., R.I. Gregory, E. Kumaraswamy, J. Norman, N. Cooch, K. Nishikura, and R. Shiekhattar. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436:740-744.
- Chi, S.W., J.B. Zang, A. Mele, and R.B. Darnell. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460:479-486.
- Chiang, H.R., L.W. Schoenfeld, J.G. Ruby, V.C. Auyeung, N. Spies, D. Baek, W.K. Johnston, C. Russ, S. Luo, J.E. Babiarz, R. Blelloch, G.P. Schroth, C. Nusbaum, and D.P. Bartel. 2010.
 Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes & development* 24:992-1009.
- Christiansen, D.H., M.K. Andersen, and J. Pedersen-Bjergaard. 2004. Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation. *Blood* 104:1474-1481.
- Christie, M., A. Boland, E. Huntzinger, O. Weichenrieder, and E. Izaurralde. 2013. Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. *Molecular cell* 51:360-373.
- Clapp, D.W., B. Freie, W.H. Lee, and Y.Y. Zhang. 1995. Molecular evidence that in situtransduced fetal liver hematopoietic stem/progenitor cells give rise to medullary hematopoiesis in adult rats. *Blood* 86:2113-2122.
- Claveria, C., G. Giovinazzo, R. Sierra, and M. Torres. 2013. Myc-driven endogenous cell competition in the early mammalian embryo. *Nature* 500:39-44.
- Cloonan, N. 2015. Re-thinking miRNA-mRNA interactions: intertwining issues confound target discovery. *BioEssays : news and reviews in molecular, cellular and developmental biology* 37:379-388.
- Coleman, J., P.J. Green, and M. Inouye. 1984. The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Cell* 37:429-436.

- Colgan, D.F., and J.L. Manley. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes & development* 11:2755-2766.
- Coller, J., and R. Parker. 2004. Eukaryotic mRNA decapping. *Annual review of biochemistry* 73:861-890.
- Cooper, G.M., and R.E. Hausman. 2007. The cell : a molecular approach. ASM Press ; Sinauer Associates, Washington, DC.
- Cowling, V.H. 2010. Regulation of mRNA cap methylation. *The Biochemical journal* 425:295-302.
- Creighton, C.J., J.G. Reid, and P.H. Gunaratne. 2009. Expression profiling of microRNAs by deep sequencing. *Briefings in bioinformatics* 10:490-497.
- Culjkovic, B., I. Topisirovic, and K.L. Borden. 2007. Controlling gene expression through RNA regulons: the role of the eukaryotic translation initiation factor eIF4E. *Cell cycle* 6:65-69.
- Dajas-Bailador, F., B. Bonev, P. Garcez, P. Stanley, F. Guillemot, and N. Papalopulu. 2012. microRNA-9 regulates axon extension and branching by targeting Map1b in mouse cortical neurons. *Nature neuroscience*
- Damen, J.E., L. Liu, P. Rosten, R.K. Humphries, A.B. Jefferson, P.W. Majerus, and G. Krystal. 1996. The 145-kDa protein induced to associate with Shc by multiple cytokines is an inositol tetraphosphate and phosphatidylinositol 3,4,5-triphosphate 5-phosphatase. *Proceedings of the National Academy of Sciences of the United States of America* 93:1689-1693.
- Dang, L., D.W. White, S. Gross, B.D. Bennett, M.A. Bittinger, E.M. Driggers, V.R. Fantin, H.G. Jang, S. Jin, M.C. Keenan, K.M. Marks, R.M. Prins, P.S. Ward, K.E. Yen, L.M. Liau, J.D. Rabinowitz, L.C. Cantley, C.B. Thompson, M.G. Vander Heiden, and S.M. Su. 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462:739-744.
- Davis, B.N., A.C. Hilyard, G. Lagna, and A. Hata. 2008. SMAD proteins control DROSHAmediated microRNA maturation. *Nature* 454:56-61.
- Davis, S., B. Lollo, S. Freier, and C. Esau. 2006. Improved targeting of miRNA with antisense oligonucleotides. *Nucleic acids research* 34:2294-2304.
- DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature genetics* 14:457-460.
- Dews, M., A. Homayouni, D. Yu, D. Murphy, C. Sevignani, E. Wentzel, E.E. Furth, W.M. Lee, G.H. Enders, J.T. Mendell, and A. Thomas-Tikhonenko. 2006. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nature genetics* 38:1060-1065.

- Di Tommaso, P., S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J.M. Chang, J.F. Taly, and C. Notredame. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic acids research* 39:W13-17.
- Doench, J.G., and P.A. Sharp. 2004. Specificity of microRNA target selection in translational repression. *Genes & development* 18:504-511.
- Dore, L.C., J.D. Amigo, C.O. Dos Santos, Z. Zhang, X. Gai, J.W. Tobias, D. Yu, A.M. Klein, C. Dorman, W. Wu, R.C. Hardison, B.H. Paw, and M.J. Weiss. 2008. A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proceedings of the National Academy of Sciences of the United States of America* 105:3333-3338.
- Ebert, B.L., J. Pretz, J. Bosco, C.Y. Chang, P. Tamayo, N. Galili, A. Raza, D.E. Root, E. Attar, S.R. Ellis, and T.R. Golub. 2008. Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* 451:335-339.
- Ebert, M.S., J.R. Neilson, and P.A. Sharp. 2007. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature methods* 4:721-726.
- Edery, I., M. Humbelin, A. Darveau, K.A. Lee, S. Milburn, J.W. Hershey, H. Trachsel, and N. Sonenberg. 1983. Involvement of eukaryotic initiation factor 4A in the cap recognition process. *The Journal of biological chemistry* 258:11398-11403.
- Efron, B., and R. Tibshirani. 2002. Empirical bayes methods and false discovery rates for microarrays. *Genetic epidemiology* 23:70-86.
- Eichhorn, S.W., H. Guo, S.E. McGeary, R.A. Rodriguez-Mias, C. Shin, D. Baek, S.H. Hsu, K. Ghoshal, J. Villen, and D.P. Bartel. 2014. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular cell* 56:104-115.
- Eis, P.S., W. Tam, L. Sun, A. Chadburn, Z. Li, M.F. Gomez, E. Lund, and J.E. Dahlberg. 2005. Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *Proceedings of the National Academy of Sciences of the United States of America* 102:3627-3632.
- Elbashir, S.M., J. Martinez, A. Patkaniowska, W. Lendeckel, and T. Tuschl. 2001. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *The EMBO journal* 20:6877-6888.
- Ernst, T., A.J. Chase, J. Score, C.E. Hidalgo-Curtis, C. Bryant, A.V. Jones, K. Waghorn, K. Zoi,
 F.M. Ross, A. Reiter, A. Hochhaus, H.G. Drexler, A. Duncombe, F. Cervantes, D. Oscier,
 J. Boultwood, F.H. Grand, and N.C. Cross. 2010. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nature genetics* 42:722-726.
- Eulalio, A., S. Helms, C. Fritzsch, M. Fauser, and E. Izaurralde. 2009a. A C-terminal silencing domain in GW182 is essential for miRNA function. *Rna* 15:1067-1077.
- Eulalio, A., E. Huntzinger, and E. Izaurralde. 2008. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nature structural* & *molecular biology* 15:346-353.
- Eulalio, A., E. Huntzinger, T. Nishihara, J. Rehwinkel, M. Fauser, and E. Izaurralde. 2009b. Deadenylation is a widespread effect of miRNA regulation. *Rna* 15:21-32.
- Eulalio, A., F. Tritschler, R. Buttner, O. Weichenrieder, E. Izaurralde, and V. Truffault. 2009c.The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing.*Nucleic acids research* 37:2974-2983.
- Eulalio, A., F. Tritschler, and E. Izaurralde. 2009d. The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. *Rna* 15:1433-1442.
- Fabian, M.R., M.K. Cieplak, F. Frank, M. Morita, J. Green, T. Srikumar, B. Nagar, T. Yamamoto,
 B. Raught, T.F. Duchaine, and N. Sonenberg. 2011. miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. *Nature structural & molecular biology* 18:1211-1217.
- Fabian, M.R., and N. Sonenberg. 2012. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nature structural & molecular biology* 19:586-593.
- Faraoni, I., F.R. Antonetti, J. Cardone, and E. Bonmassar. 2009. miR-155 gene: a typical multifunctional microRNA. *Biochimica et biophysica acta* 1792:497-505.
- Fazi, F., S. Racanicchi, G. Zardo, L.M. Starnes, M. Mancini, L. Travaglini, D. Diverio, E. Ammatuna, G. Cimino, F. Lo-Coco, F. Grignani, and C. Nervi. 2007. Epigenetic silencing of the myelopoiesis regulator microRNA-223 by the AML1/ETO oncoprotein. *Cancer cell* 12:457-466.
- Fazi, F., A. Rosa, A. Fatica, V. Gelmetti, M.L. De Marchis, C. Nervi, and I. Bozzoni. 2005. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. *Cell* 123:819-831.
- Felli, N., L. Fontana, E. Pelosi, R. Botta, D. Bonci, F. Facchiano, F. Liuzzi, V. Lulli, O. Morsilli, S. Santoro, M. Valtieri, G.A. Calin, C.G. Liu, A. Sorrentino, C.M. Croce, and C. Peschle. 2005. MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proceedings of the National Academy of Sciences of the United States of America* 102:18081-18086.

- Filipowicz, W., P. Pelczar, V. Pogacic, and F. Dragon. 1999. Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification. *Acta biochimica Polonica* 46:377-389.
- Fire, A., and S.Q. Xu. 1995. Rolling replication of short DNA circles. *Proceedings of the National Academy of Sciences of the United States of America* 92:4641-4645.
- Forman, J.J., A. Legesse-Miller, and H.A. Coller. 2008. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proceedings of the National Academy of Sciences of the United States of America* 105:14879-14884.
- Frank, F., N. Sonenberg, and B. Nagar. 2010. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465:818-822.
- Friedman, R.C., K.K. Farh, C.B. Burge, and D.P. Bartel. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research* 19:92-105.
- Fukao, A., Y. Mishima, N. Takizawa, S. Oka, H. Imataka, J. Pelletier, N. Sonenberg, C. Thoma, and T. Fujiwara. 2014. MicroRNAs trigger dissociation of eIF4AI and eIF4AII from target mRNAs in humans. *Molecular cell* 56:79-89.
- Fukao, T., Y. Fukuda, K. Kiga, J. Sharif, K. Hino, Y. Enomoto, A. Kawamura, K. Nakamura, T. Takeuchi, and M. Tanabe. 2007. An evolutionarily conserved mechanism for microRNA-223 expression revealed by microRNA gene profiling. *Cell* 129:617-631.
- Fukaya, T., H.O. Iwakawa, and Y. Tomari. 2014. MicroRNAs block assembly of eIF4F translation initiation complex in Drosophila. *Molecular cell* 56:67-78.
- Fukuda, T., K. Yamagata, S. Fujiyama, T. Matsumoto, I. Koshida, K. Yoshimura, M. Mihara, M. Naitou, H. Endoh, T. Nakamura, C. Akimoto, Y. Yamamoto, T. Katagiri, C. Foulds, S. Takezawa, H. Kitagawa, K. Takeyama, B.W. O'Malley, and S. Kato. 2007. DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nature cell biology* 9:604-611.
- Fuller-Pace, F.V., and H.C. Moore. 2011. RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development. *Future oncology* 7:239-251.
- Gallie, D.R. 2014. The role of the poly(A) binding protein in the assembly of the Cap-binding complex during translation initiation in plants. *Translation* 2:e959378.
- Gan, H.H., and K.C. Gunsalus. 2013. Tertiary structure-based analysis of microRNA-target interactions. *Rna* 19:539-551.

- Gao, X.N., J. Lin, Y.H. Li, L. Gao, X.R. Wang, W. Wang, H.Y. Kang, G.T. Yan, L.L. Wang, and
 L. Yu. 2011. MicroRNA-193a represses c-kit expression and functions as a methylationsilenced tumor suppressor in acute myeloid leukemia. *Oncogene* 30:3416-3428.
- Garzon, R., and C.M. Croce. 2008. MicroRNAs in normal and malignant hematopoiesis. *Current opinion in hematology* 15:352-358.
- Garzon, R., C.E. Heaphy, V. Havelange, M. Fabbri, S. Volinia, T. Tsao, N. Zanesi, S.M. Kornblau, G. Marcucci, G.A. Calin, M. Andreeff, and C.M. Croce. 2009a. MicroRNA 29b functions in acute myeloid leukemia. *Blood* 114:5331-5341.
- Garzon, R., S. Liu, M. Fabbri, Z. Liu, C.E. Heaphy, E. Callegari, S. Schwind, J. Pang, J. Yu, N. Muthusamy, V. Havelange, S. Volinia, W. Blum, L.J. Rush, D. Perrotti, M. Andreeff, C.D. Bloomfield, J.C. Byrd, K. Chan, L.C. Wu, C.M. Croce, and G. Marcucci. 2009b. MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. *Blood* 113:6411-6418.
- Garzon, R., F. Pichiorri, T. Palumbo, R. Iuliano, A. Cimmino, R. Aqeilan, S. Volinia, D. Bhatt, H.
 Alder, G. Marcucci, G.A. Calin, C.G. Liu, C.D. Bloomfield, M. Andreeff, and C.M. Croce.
 2006. MicroRNA fingerprints during human megakaryocytopoiesis. *Proceedings of the National Academy of Sciences of the United States of America* 103:5078-5083.
- Garzon, R., F. Pichiorri, T. Palumbo, M. Visentini, R. Aqeilan, A. Cimmino, H. Wang, H. Sun, S. Volinia, H. Alder, G.A. Calin, C.G. Liu, M. Andreeff, and C.M. Croce. 2007. MicroRNA gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia. *Oncogene* 26:4148-4157.
- Gekas, C., F. Dieterlen-Lievre, S.H. Orkin, and H.K. Mikkola. 2005. The placenta is a niche for hematopoietic stem cells. *Developmental cell* 8:365-375.
- Gentner, B., G. Schira, A. Giustacchini, M. Amendola, B.D. Brown, M. Ponzoni, and L. Naldini.
 2009. Stable knockdown of microRNA in vivo by lentiviral vectors. *Nature methods* 6:63-66.
- Gentner, B., I. Visigalli, H. Hiramatsu, E. Lechman, S. Ungari, A. Giustacchini, G. Schira, M. Amendola, A. Quattrini, S. Martino, A. Orlacchio, J.E. Dick, A. Biffi, and L. Naldini. 2010.
 Identification of hematopoietic stem cell-specific miRNAs enables gene therapy of globoid cell leukodystrophy. *Science translational medicine* 2:58ra84.
- Georgantas, R.W., 3rd, R. Hildreth, S. Morisot, J. Alder, C.G. Liu, S. Heimfeld, G.A. Calin, C.M. Croce, and C.I. Civin. 2007. CD34+ hematopoietic stem-progenitor cell microRNA

expression and function: a circuit diagram of differentiation control. *Proceedings of the National Academy of Sciences of the United States of America* 104:2750-2755.

- Gibb, E.A., R.L. Warren, G.W. Wilson, S.D. Brown, G.A. Robertson, G.B. Morin, and R.A. Holt. 2015. Activation of an endogenous retrovirus-associated long non-coding RNA in human adenocarcinoma. *Genome medicine* 7:22.
- Golomb, L., S. Volarevic, and M. Oren. 2014. p53 and ribosome biogenesis stress: the essentials. *FEBS letters* 588:2571-2579.
- Golub, R., and A. Cumano. 2013. Embryonic hematopoiesis. *Blood cells, molecules & diseases* 51:226-231.
- Gondek, L.P., R. Tiu, C.L. O'Keefe, M.A. Sekeres, K.S. Theil, and J.P. Maciejewski. 2008. Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML. *Blood* 111:1534-1542.
- Gong, L., A. Kakrana, S. Arikit, B.C. Meyers, and J.F. Wendel. 2013. Composition and expression of conserved microRNA genes in diploid cotton (Gossypium) species. *Genome biology and evolution* 5:2449-2459.
- Gorgoni, B., D. Maritano, P. Marthyn, M. Righi, and V. Poli. 2002. C/EBP beta gene inactivation causes both impaired and enhanced gene expression and inverse regulation of IL-12 p40 and p35 mRNAs in macrophages. *Journal of immunology* 168:4055-4062.
- Gorodkin, J., I.L. Hofacker, and W.L. Ruzzo. 2014. Concepts and introduction to RNA bioinformatics. *Methods in molecular biology* 1097:1-31.
- Grabow, W.W., Z. Zhuang, J.E. Shea, and L. Jaeger. 2013. The GA-minor submotif as a case study of RNA modularity, prediction, and design. *Wiley interdisciplinary reviews. RNA* 4:181-203.
- Graubert, T.A., M.A. Payton, J. Shao, R.A. Walgren, R.S. Monahan, J.L. Frater, M.A. Walshauser, M.G. Martin, Y. Kasai, and M.J. Walter. 2009. Integrated genomic analysis implicates haploinsufficiency of multiple chromosome 5q31.2 genes in de novo myelodysplastic syndromes pathogenesis. *PloS one* 4:e4583.
- Greenberg, P., C. Cox, M.M. LeBeau, P. Fenaux, P. Morel, G. Sanz, M. Sanz, T. Vallespi, T. Hamblin, D. Oscier, K. Ohyashiki, K. Toyama, C. Aul, G. Mufti, and J. Bennett. 1997.
 International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* 89:2079-2088.
- Greenberg, P.L., H. Tuechler, J. Schanz, G. Sanz, G. Garcia-Manero, F. Sole, J.M. Bennett, D.Bowen, P. Fenaux, F. Dreyfus, H. Kantarjian, A. Kuendgen, A. Levis, L. Malcovati, M.Cazzola, J. Cermak, C. Fonatsch, M.M. Le Beau, M.L. Slovak, O. Krieger, M. Luebbert, J.

Maciejewski, S.M. Magalhaes, Y. Miyazaki, M. Pfeilstocker, M. Sekeres, W.R. Sperr, R. Stauder, S. Tauro, P. Valent, T. Vallespi, A.A. van de Loosdrecht, U. Germing, and D. Haase. 2012. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* 120:2454-2465.

- Gregory, R.I., K.P. Yan, G. Amuthan, T. Chendrimada, B. Doratotaj, N. Cooch, and R. Shiekhattar. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432:235-240.
- Grifo, J.A., S.M. Tahara, M.A. Morgan, A.J. Shatkin, and W.C. Merrick. 1983. New initiation factor activity required for globin mRNA translation. *The Journal of biological chemistry* 258:5804-5810.
- Grimson, A., K.K. Farh, W.K. Johnston, P. Garrett-Engele, L.P. Lim, and D.P. Bartel. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell* 27:91-105.
- Grishok, A., A.E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D.L. Baillie, A. Fire, G. Ruvkun, and C.C. Mello. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* 106:23-34.
- Grosset, C., C.Y. Chen, N. Xu, N. Sonenberg, H. Jacquemin-Sablon, and A.B. Shyu. 2000. A mechanism for translationally coupled mRNA turnover: interaction between the poly(A) tail and a c-fos RNA coding determinant via a protein complex. *Cell* 103:29-40.
- Growney, J.D., H. Shigematsu, Z. Li, B.H. Lee, J. Adelsperger, R. Rowan, D.P. Curley, J.L. Kutok, K. Akashi, I.R. Williams, N.A. Speck, and D.G. Gilliland. 2005. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood* 106:494-504.
- Gu, W., Y. Xu, X. Xie, T. Wang, J.H. Ko, and T. Zhou. 2014. The role of RNA structure at 5' untranslated region in microRNA-mediated gene regulation. *Rna* 20:1369-1375.
- Guil, S., and J.F. Caceres. 2007. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature structural & molecular biology* 14:591-596.
- Guo, H., N.T. Ingolia, J.S. Weissman, and D.P. Bartel. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466:835-840.
- Haase, A.D., L. Jaskiewicz, H. Zhang, S. Laine, R. Sack, A. Gatignol, and W. Filipowicz. 2005. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO reports* 6:961-967.

- Haase, D., U. Germing, J. Schanz, M. Pfeilstocker, T. Nosslinger, B. Hildebrandt, A. Kundgen, M. Lubbert, R. Kunzmann, A.A. Giagounidis, C. Aul, L. Trumper, O. Krieger, R. Stauder, T.H. Muller, F. Wimazal, P. Valent, C. Fonatsch, and C. Steidl. 2007. New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* 110:4385-4395.
- Haferlach, T., Y. Nagata, V. Grossmann, Y. Okuno, U. Bacher, G. Nagae, S. Schnittger, M. Sanada, A. Kon, T. Alpermann, K. Yoshida, A. Roller, N. Nadarajah, Y. Shiraishi, Y. Shiozawa, K. Chiba, H. Tanaka, H.P. Koeffler, H.U. Klein, M. Dugas, H. Aburatani, A. Kohlmann, S. Miyano, C. Haferlach, W. Kern, and S. Ogawa. 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28:241-247.
- Hammond, S.M. 2006. MicroRNAs as oncogenes. *Current opinion in genetics & development* 16:4-9.
- Han, J., Y. Lee, K.H. Yeom, Y.K. Kim, H. Jin, and V.N. Kim. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & development* 18:3016-3027.
- Han, J., Y. Lee, K.H. Yeom, J.W. Nam, I. Heo, J.K. Rhee, S.Y. Sohn, Y. Cho, B.T. Zhang, and V.N. Kim. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125:887-901.
- Han, J., J.S. Pedersen, S.C. Kwon, C.D. Belair, Y.K. Kim, K.H. Yeom, W.Y. Yang, D. Haussler,
 R. Blelloch, and V.N. Kim. 2009. Posttranscriptional crossregulation between Drosha and
 DGCR8. *Cell* 136:75-84.
- Han, L., P.D. Witmer, E. Casey, D. Valle, and S. Sukumar. 2007. DNA methylation regulates MicroRNA expression. *Cancer biology & therapy* 6:1284-1288.
- Harada, Y., and H. Harada. 2009. Molecular pathways mediating MDS/AML with focus on AML1/RUNX1 point mutations. *Journal of cellular physiology* 220:16-20.
- Haraguchi, T., Y. Ozaki, and H. Iba. 2009. Vectors expressing efficient RNA decoys achieve the long-term suppression of specific microRNA activity in mammalian cells. *Nucleic acids research* 37:e43.
- Hausser, J., A.P. Syed, B. Bilen, and M. Zavolan. 2013. Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome research* 23:604-615.
- Havelange, V., and R. Garzon. 2010. MicroRNAs: emerging key regulators of hematopoiesis. *American journal of hematology* 85:935-942.
- He, L., X. He, L.P. Lim, E. de Stanchina, Z. Xuan, Y. Liang, W. Xue, L. Zender, J. Magnus, D. Ridzon, A.L. Jackson, P.S. Linsley, C. Chen, S.W. Lowe, M.A. Cleary, and G.J. Hannon.

2007. A microRNA component of the p53 tumour suppressor network. *Nature* 447:1130-1134.

- He, L., J.M. Thomson, M.T. Hemann, E. Hernando-Monge, D. Mu, S. Goodson, S. Powers, C. Cordon-Cardo, S.W. Lowe, G.J. Hannon, and S.M. Hammond. 2005. A microRNA polycistron as a potential human oncogene. *Nature* 435:828-833.
- Heinrichs, S., R.V. Kulkarni, C.E. Bueso-Ramos, R.L. Levine, M.L. Loh, C. Li, D. Neuberg, S.M. Kornblau, J.P. Issa, D.G. Gilliland, G. Garcia-Manero, H.M. Kantarjian, E.H. Estey, and A.T. Look. 2009. Accurate detection of uniparental disomy and microdeletions by SNP array analysis in myelodysplastic syndromes with normal cytogenetics. *Leukemia* 23:1605-1613.
- Helwak, A., G. Kudla, T. Dudnakova, and D. Tollervey. 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 153:654-665.
- Hendrickson, D.G., D.J. Hogan, H.L. McCullough, J.W. Myers, D. Herschlag, J.E. Ferrell, and P.O. Brown. 2009. Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS biology* 7:e1000238.
- Heng, Y.W., and C.G. Koh. 2010. Actin cytoskeleton dynamics and the cell division cycle. *The international journal of biochemistry & cell biology* 42:1622-1633.
- Heo, I., C. Joo, J. Cho, M. Ha, J. Han, and V.N. Kim. 2008. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Molecular cell* 32:276-284.
- Heravi-Moussavi, A., M.S. Anglesio, S.W. Cheng, J. Senz, W. Yang, L. Prentice, A.P. Fejes, C. Chow, A. Tone, S.E. Kalloger, N. Hamel, A. Roth, G. Ha, A.N. Wan, S. Maines-Bandiera, C. Salamanca, B. Pasini, B.A. Clarke, A.F. Lee, C.H. Lee, C. Zhao, R.H. Young, S.A. Aparicio, P.H. Sorensen, M.M. Woo, N. Boyd, S.J. Jones, M. Hirst, M.A. Marra, B. Gilks, S.P. Shah, W.D. Foulkes, G.B. Morin, and D.G. Huntsman. 2012. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. *The New England journal of medicine* 366:234-242.
- Herling, M., K.A. Patel, J. Khalili, E. Schlette, R. Kobayashi, L.J. Medeiros, and D. Jones. 2006.
 TCL1 shows a regulated expression pattern in chronic lymphocytic leukemia that correlates with molecular subtypes and proliferative state. *Leukemia* 20:280-285.
- Holley, C.L., and V.K. Topkara. 2011. An introduction to small non-coding RNAs: miRNA and snoRNA. *Cardiovascular drugs and therapy / sponsored by the International Society of Cardiovascular Pharmacotherapy* 25:151-159.
- Horos, R., H. Ijspeert, D. Pospisilova, R. Sendtner, C. Andrieu-Soler, E. Taskesen, A. Nieradka, R. Cmejla, M. Sendtner, I.P. Touw, and M. von Lindern. 2012. Ribosomal deficiencies in

Diamond-Blackfan anemia impair translation of transcripts essential for differentiation of murine and human erythroblasts. *Blood* 119:262-272.

- Horrigan, S.K., Z.H. Arbieva, H.Y. Xie, J. Kravarusic, N.C. Fulton, H. Naik, T.T. Le, and C.A.
 Westbrook. 2000. Delineation of a minimal interval and identification of 9 candidates for a tumor suppressor gene in malignant myeloid disorders on 5q31. *Blood* 95:2372-2377.
- Hrdlickova, B., V. Kumar, K. Kanduri, D.V. Zhernakova, S. Tripathi, J. Karjalainen, R.J. Lund, Y.
 Li, U. Ullah, R. Modderman, W. Abdulahad, H. Lahdesmaki, L. Franke, R. Lahesmaa, C.
 Wijmenga, and S. Withoff. 2014. Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. *Genome medicine* 6:88.
- Hsu, P.W., H.D. Huang, S.D. Hsu, L.Z. Lin, A.P. Tsou, C.P. Tseng, P.F. Stadler, S. Washietl, and I.L. Hofacker. 2006. miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic acids research* 34:D135-139.
- Huggins, C.E., A.A. Domenighetti, M.E. Ritchie, N. Khalil, J.M. Favaloro, J. Proietto, G.K. Smyth, S. Pepe, and L.M. Delbridge. 2008. Functional and metabolic remodelling in GLUT4-deficient hearts confers hyper-responsiveness to substrate intervention. *Journal of molecular and cellular cardiology* 44:270-280.
- Hughes, C.S., S. Foehr, D.A. Garfield, E.E. Furlong, L.M. Steinmetz, and J. Krijgsveld. 2014. Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology* 10:757.
- Huntly, B.J., and D.G. Gilliland. 2005. Leukaemia stem cells and the evolution of cancer-stemcell research. *Nature reviews. Cancer* 5:311-321.
- Hussein, K., K. Theophile, G. Busche, B. Schlegelberger, G. Gohring, H. Kreipe, and O. Bock.
 2010a. Aberrant microRNA expression pattern in myelodysplastic bone marrow cells. *Leukemia research* 34:1169-1174.
- Hussein, K., K. Theophile, G. Busche, B. Schlegelberger, G. Gohring, H. Kreipe, and O. Bock.
 2010b. Significant inverse correlation of microRNA-150/MYB and microRNA-222/p27 in myelodysplastic syndrome. *Leukemia research* 34:328-334.
- Hutvagner, G., J. McLachlan, A.E. Pasquinelli, E. Balint, T. Tuschl, and P.D. Zamore. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293:834-838.
- Hutvagner, G., and P.D. Zamore. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297:2056-2060.

- Iki, T., M. Yoshikawa, M. Nishikiori, M.C. Jaudal, E. Matsumoto-Yokoyama, I. Mitsuhara, T. Meshi, and M. Ishikawa. 2010. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Molecular cell* 39:282-291.
- Ingolia, N.T. 2014. Ribosome profiling: new views of translation, from single codons to genome scale. *Nature reviews. Genetics* 15:205-213.
- Inlay, M.A., D. Bhattacharya, D. Sahoo, T. Serwold, J. Seita, H. Karsunky, S.K. Plevritis, D.L. Dill, and I.L. Weissman. 2009. Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes & development* 23:2376-2381.
- Issler, O., and A. Chen. 2015. Determining the role of microRNAs in psychiatric disorders. *Nature reviews. Neuroscience* 16:201-212.
- Ivanovs, A., S. Rybtsov, L. Welch, R.A. Anderson, M.L. Turner, and A. Medvinsky. 2011. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. *The Journal of experimental medicine* 208:2417-2427.
- Jackson, R.J., and N. Standart. 2007. How do microRNAs regulate gene expression? *Science's STKE : signal transduction knowledge environment* 2007:re1.
- Jerez, A., Y. Sugimoto, H. Makishima, A. Verma, A.M. Jankowska, B. Przychodzen, V. Visconte, R.V. Tiu, C.L. O'Keefe, A.M. Mohamedali, A.G. Kulasekararaj, A. Pellagatti, K. McGraw, H. Muramatsu, A.R. Moliterno, M.A. Sekeres, M.A. McDevitt, S. Kojima, A. List, J. Boultwood, G.J. Mufti, and J.P. Maciejewski. 2012. Loss of heterozygosity in 7q myeloid disorders: clinical associations and genomic pathogenesis. *Blood* 119:6109-6117.
- Jinek, M., M.R. Fabian, S.M. Coyle, N. Sonenberg, and J.A. Doudna. 2010. Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. *Nature structural & molecular biology* 17:238-240.
- Johnnidis, J.B., M.H. Harris, R.T. Wheeler, S. Stehling-Sun, M.H. Lam, O. Kirak, T.R. Brummelkamp, M.D. Fleming, and F.D. Camargo. 2008. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature* 451:1125-1129.
- Johnson, S.M., H. Grosshans, J. Shingara, M. Byrom, R. Jarvis, A. Cheng, E. Labourier, K.L. Reinert, D. Brown, and F.J. Slack. 2005. RAS is regulated by the let-7 microRNA family. *Cell* 120:635-647.
- Jonas, S., and E. Izaurralde. 2015. Towards a molecular understanding of microRNA-mediated gene silencing. *Nature reviews. Genetics* 16:421-433.

- Jongen-Lavrencic, M., S.M. Sun, M.K. Dijkstra, P.J. Valk, and B. Lowenberg. 2008. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. *Blood* 111:5078-5085.
- Kadener, S., J. Rodriguez, K.C. Abruzzi, Y.L. Khodor, K. Sugino, M.T. Marr, 2nd, S. Nelson, and
 M. Rosbash. 2009. Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. *Rna* 15:537-545.
- Kahvejian, A., Y.V. Svitkin, R. Sukarieh, M.N. M'Boutchou, and N. Sonenberg. 2005.
 Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes & development* 19:104-113.
- Kammers, K., R.N. Cole, C. Tiengwe, and I. Ruczinski. 2015. Detecting Significant Changes in Protein Abundance. *EuPA open proteomics* 7:11-19.
- Kanellopoulou, C., S.A. Muljo, A.L. Kung, S. Ganesan, R. Drapkin, T. Jenuwein, D.M. Livingston, and K. Rajewsky. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes & development* 19:489-501.
- Karube, Y., H. Tanaka, H. Osada, S. Tomida, Y. Tatematsu, K. Yanagisawa, Y. Yatabe, J. Takamizawa, S. Miyoshi, T. Mitsudomi, and T. Takahashi. 2005. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. *Cancer science* 96:111-115.
- Kawamata, T., H. Seitz, and Y. Tomari. 2009. Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nature structural & molecular biology* 16:953-960.
- Kawamata, T., M. Yoda, and Y. Tomari. 2011. Multilayer checkpoints for microRNA authenticity during RISC assembly. *EMBO reports* 12:944-949.
- Kelemen, E., and M. Janossa. 1980. Macrophages are the first differentiated blood cells formed in human embryonic liver. *Experimental hematology* 8:996-1000.
- Kennerdell, J.R., and R.W. Carthew. 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. *Cell* 95:1017-1026.
- Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler. 2002. The human genome browser at UCSC. *Genome research* 12:996-1006.
- Ketting, R.F., S.E. Fischer, E. Bernstein, T. Sijen, G.J. Hannon, and R.H. Plasterk. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & development* 15:2654-2659.
- Khalaj, M., M. Tavakkoli, A.W. Stranahan, and C.Y. Park. 2014. Pathogenic microRNA's in myeloid malignancies. *Frontiers in genetics* 5:361.

- Khvorova, A., A. Reynolds, and S.D. Jayasena. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115:209-216.
- Kiel, M.J., T. Iwashita, O.H. Yilmaz, and S.J. Morrison. 2005. Spatial differences in hematopoiesis but not in stem cells indicate a lack of regional patterning in definitive hematopoietic stem cells. *Developmental biology* 283:29-39.
- Kikuchi, K., and M. Kondo. 2006. Developmental switch of mouse hematopoietic stem cells from fetal to adult type occurs in bone marrow after birth. *Proceedings of the National Academy of Sciences of the United States of America* 103:17852-17857.
- Kim, K.Y., Y.J. Hwang, M.K. Jung, J. Choe, Y. Kim, S. Kim, C.J. Lee, H. Ahn, J. Lee, N.W. Kowall, Y.K. Kim, J.I. Kim, S.B. Lee, and H. Ryu. 2014a. A multifunctional protein EWS regulates the expression of Drosha and microRNAs. *Cell death and differentiation* 21:136-145.
- Kim, M.S., S.M. Pinto, D. Getnet, R.S. Nirujogi, S.S. Manda, R. Chaerkady, A.K. Madugundu, D.S. Kelkar, R. Isserlin, S. Jain, J.K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N.A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L.D. Selvan, A.H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S.K. Sreenivasamurthy, A. Marimuthu, G.J. Sathe, S. Chavan, K.K. Datta, Y. Subbannayya, A. Sahu, S.D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K.R. Murthy, N. Syed, R. Goel, A.A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T.C. Huang, J. Zhong, X. Wu, P.G. Shaw, D. Freed, M.S. Zahari, K.K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C.J. Mitchell, S.K. Shankar, P. Satishchandra, J.T. Schroeder, R. Sirdeshmukh, A. Maitra, S.D. Leach, C.G. Drake, M.K. Halushka, T.S. Prasad, R.H. Hruban, C.L. Kerr, G.D. Bader, C.A. lacobuzio-Donahue, H. Gowda, and A. Pandey. 2014b. A draft map of the human proteome. *Nature* 509:575-581.
- Kim, V.N. 2004. MicroRNA precursors in motion: exportin-5 mediates their nuclear export. *Trends Cell Biol* 14:156-159.
- Kim, V.N., J. Han, and M.C. Siomi. 2009. Biogenesis of small RNAs in animals. *Nature reviews. Molecular cell biology* 10:126-139.
- Kingsley, P.D., E. Greenfest-Allen, J.M. Frame, T.P. Bushnell, J. Malik, K.E. McGrath, C.J. Stoeckert, and J. Palis. 2013. Ontogeny of erythroid gene expression. *Blood* 121:e5-e13.
- Kloosterman, W.P., E. Wienholds, R.F. Ketting, and R.H. Plasterk. 2004. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic acids research* 32:6284-6291.

- Knudsen, S. 1999. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* 15:356-361.
- Komrokji, R.S., L. Zhang, and J.M. Bennett. 2010. Myelodysplastic syndromes classification and risk stratification. *Hematology/oncology clinics of North America* 24:443-457.
- Kondo, M., I.L. Weissman, and K. Akashi. 1997. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 91:661-672.
- Kosan, C., and M. Godmann. 2016. Genetic and Epigenetic Mechanisms That Maintain Hematopoietic Stem Cell Function. *Stem cells international* 2016:5178965.
- Kosmider, O., V. Gelsi-Boyer, L. Slama, F. Dreyfus, O. Beyne-Rauzy, B. Quesnel, M. Hunault-Berger, B. Slama, N. Vey, C. Lacombe, E. Solary, D. Birnbaum, O.A. Bernard, and M. Fontenay. 2010. Mutations of IDH1 and IDH2 genes in early and accelerated phases of myelodysplastic syndromes and MDS/myeloproliferative neoplasms. *Leukemia* 24:1094-1096.
- Koussounadis, A., S.P. Langdon, I.H. Um, D.J. Harrison, and V.A. Smith. 2015. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Scientific reports* 5:10775.
- Kozlov, G., M. Menade, A. Rosenauer, L. Nguyen, and K. Gehring. 2010. Molecular determinants of PAM2 recognition by the MLLE domain of poly(A)-binding protein. *Journal of molecular biology* 397:397-407.
- Krek, A., D. Grun, M.N. Poy, R. Wolf, L. Rosenberg, E.J. Epstein, P. MacMenamin, I. da Piedade, K.C. Gunsalus, M. Stoffel, and N. Rajewsky. 2005. Combinatorial microRNA target predictions. *Nature genetics* 37:495-500.
- Krutzfeldt, J., N. Rajewsky, R. Braich, K.G. Rajeev, T. Tuschl, M. Manoharan, and M. Stoffel. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 438:685-689.
- Kumar, M.S., A. Narla, A. Nonami, A. Mullally, N. Dimitrova, B. Ball, J.R. McAuley, L. Poveromo, J.L. Kutok, N. Galili, A. Raza, E. Attar, D.G. Gilliland, T. Jacks, and B.L. Ebert. 2011.
 Coordinate loss of a microRNA and protein-coding gene cooperate in the pathogenesis of 5q- syndrome. *Blood* 118:4666-4673.
- Kwak, P.B., and Y. Tomari. 2012. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature structural & molecular biology* 19:145-151.
- Lagos-Quintana, M., R. Rauhut, W. Lendeckel, and T. Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294:853-858.
- Lagos-Quintana, M., R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel, and T. Tuschl. 2002. Identification of tissue-specific microRNAs from mouse. *Current biology : CB* 12:735-739.

- Lai, E.C. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature genetics* 30:363-364.
- Lai, V.K., M. Ashraf, S. Jiang, and K. Haider. 2012. MicroRNA-143 is a critical regulator of cell cycle activity in stem cells with co-overexpression of Akt and angiopoietin-1 via transcriptional regulation of Erk5/cyclin D1 signaling. *Cell cycle* 11:767-777.
- Laiosa, C.V., M. Stadtfeld, and T. Graf. 2006. Determinants of lymphoid-myeloid lineage diversification. *Annual review of immunology* 24:705-738.
- Lamontagne, B., S. Larose, J. Boulanger, and S.A. Elela. 2001. The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. *Current issues in molecular biology* 3:71-78.
- Landthaler, M., D. Gaidatzis, A. Rothballer, P.Y. Chen, S.J. Soll, L. Dinic, T. Ojo, M. Hafner, M. Zavolan, and T. Tuschl. 2008. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *Rna* 14:2580-2596.
- Langenberger, D., C. Bermudez-Santana, J. Hertel, S. Hoffmann, P. Khaitovich, and P.F. Stadler. 2009. Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics* 25:2298-2301.
- Lau, N.C., L.P. Lim, E.G. Weinstein, and D.P. Bartel. 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* 294:858-862.
- Le Thomas, A., K.F. Toth, and A.A. Aravin. 2014. To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome biology* 15:204.
- Lee, H.Y., K. Zhou, A.M. Smith, C.L. Noland, and J.A. Doudna. 2013. Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. *Nucleic acids research* 41:6568-6576.
- Lee, L.K., M. Ueno, B. Van Handel, and H.K. Mikkola. 2010. Placenta as a newly identified source of hematopoietic stem cells. *Current opinion in hematology* 17:313-318.
- Lee, R.C., and V. Ambros. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* 294:862-864.
- Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75:843-854.
- Lee, S.Y., A. De La Torre, D. Yan, S. Kustu, B.T. Nixon, and D.E. Wemmer. 2003a. Regulation of the transcriptional activator NtrC1: structural studies of the regulatory and AAA+ ATPase domains. *Genes & development* 17:2552-2563.

- Lee, Y., C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V.N. Kim. 2003b. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425:415-419.
- Lee, Y., I. Hur, S.Y. Park, Y.K. Kim, M.R. Suh, and V.N. Kim. 2006. The role of PACT in the RNA silencing pathway. *The EMBO journal* 25:522-532.
- Lee, Y., K. Jeon, J.T. Lee, S. Kim, and V.N. Kim. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO journal* 21:4663-4670.
- Lee, Y., M. Kim, J. Han, K.H. Yeom, S. Lee, S.H. Baek, and V.N. Kim. 2004. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* 23:4051-4060.
- Lehmann, U., B. Hasemeier, M. Christgen, M. Muller, D. Romermann, F. Langer, and H. Kreipe. 2008. Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *The Journal of pathology* 214:17-24.
- Lei, X., W. Tian, H. Zhu, T. Chen, and P. Ao. 2015. Biological Sources of Intrinsic and Extrinsic Noise in cl Expression of Lysogenic Phage Lambda. *Scientific reports* 5:13597.
- Lennox, K.A., and M.A. Behlke. 2010. A direct comparison of anti-microRNA oligonucleotide potency. *Pharmaceutical research* 27:1788-1799.
- Leontis, N.B., and E. Westhof. 2003. Analysis of RNA motifs. *Current opinion in structural biology* 13:300-308.
- Leuschner, P.J., S.L. Ameres, S. Kueng, and J. Martinez. 2006. Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO reports* 7:314-320.
- Lewis, B.P., C.B. Burge, and D.P. Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120:15-20.
- Lewis, B.P., I.H. Shih, M.W. Jones-Rhoades, D.P. Bartel, and C.B. Burge. 2003. Prediction of mammalian microRNA targets. *Cell* 115:787-798.
- Liang, X.H., Q. Liu, and M.J. Fournier. 2009. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *Rna* 15:1716-1728.
- Lim, L.P., M.E. Glasner, S. Yekta, C.B. Burge, and D.P. Bartel. 2003a. Vertebrate microRNA genes. *Science* 299:1540.
- Lim, L.P., N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, D.P. Bartel, P.S. Linsley, and J.M. Johnson. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433:769-773.

- Lim, L.P., N.C. Lau, E.G. Weinstein, A. Abdelhakim, S. Yekta, M.W. Rhoades, C.B. Burge, and D.P. Bartel. 2003b. The microRNAs of Caenorhabditis elegans. *Genes & development* 17:991-1008.
- Lim, Z.Y., L. Pearce, A.Y. Ho, L. Barber, W. Ingram, M. Usai, K. Tobal, S. Devereux, A. Pagliuca, and G.J. Mufti. 2007. Delayed attainment of full donor chimaerism following alemtuzumab-based reduced-intensity conditioning haematopoeitic stem cell transplantation for acute myeloid leukaemia and myelodysplastic syndromes is associated with improved outcomes. *British journal of haematology* 138:517-526.
- Lin, K.Y., X.J. Zhang, D.D. Feng, H. Zhang, C.W. Zeng, B.W. Han, A.D. Zhou, L.H. Qu, L. Xu, and Y.Q. Chen. 2011. miR-125b, a target of CDX2, regulates cell differentiation through repression of the core binding factor in hematopoietic malignancies. *The Journal of biological chemistry* 286:38253-38263.
- Lin, R.J., Y.C. Lin, J. Chen, H.H. Kuo, Y.Y. Chen, M.B. Diccianni, W.B. London, C.H. Chang, and A.L. Yu. 2010. microRNA signature and expression of Dicer and Drosha can predict prognosis and delineate risk groups in neuroblastoma. *Cancer research* 70:7841-7850.
- Lin, S., and R.I. Gregory. 2015. MicroRNA biogenesis pathways in cancer. *Nature reviews. Cancer* 15:321-333.
- Liu, J., M.A. Carmell, F.V. Rivas, C.G. Marsden, J.M. Thomson, J.J. Song, S.M. Hammond, L. Joshua-Tor, and G.J. Hannon. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305:1437-1441.
- Loya, C.M., C.S. Lu, D. Van Vactor, and T.A. Fulga. 2009. Transgenic microRNA inhibition with spatiotemporal specificity in intact organisms. *Nature methods* 6:897-903.
- Lu, J., G. Getz, E.A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B.L.
 Ebert, R.H. Mak, A.A. Ferrando, J.R. Downing, T. Jacks, H.R. Horvitz, and T.R. Golub.
 2005. MicroRNA expression profiles classify human cancers. *Nature* 435:834-838.
- Lu, J., S. Guo, B.L. Ebert, H. Zhang, X. Peng, J. Bosco, J. Pretz, R. Schlanger, J.Y. Wang, R.H. Mak, D.M. Dombkowski, F.I. Preffer, D.T. Scadden, and T.R. Golub. 2008. MicroRNAmediated control of cell fate in megakaryocyte-erythrocyte progenitors. *Developmental cell* 14:843-853.
- Lucitti, J.L., E.A. Jones, C. Huang, J. Chen, S.E. Fraser, and M.E. Dickinson. 2007. Vascular remodeling of the mouse yolk sac requires hemodynamic force. *Development* 134:3317-3326.
- Lund, E., and J.E. Dahlberg. 2006. Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harbor symposia on quantitative biology* 71:59-66.

- Lund, E., S. Guttinger, A. Calado, J.E. Dahlberg, and U. Kutay. 2004. Nuclear export of microRNA precursors. *Science* 303:95-98.
- Ma, J.B., Y.R. Yuan, G. Meister, Y. Pei, T. Tuschl, and D.J. Patel. 2005. Structural basis for 5'end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* 434:666-670.
- Madison, B.B., Q. Liu, X. Zhong, C.M. Hahn, N. Lin, M.J. Emmett, B.Z. Stanger, J.S. Lee, and A.K. Rustgi. 2013. LIN28B promotes growth and tumorigenesis of the intestinal epithelium via Let-7. *Genes & development* 27:2233-2245.
- Maier, T., M. Guell, and L. Serrano. 2009. Correlation of mRNA and protein in complex biological samples. *FEBS letters* 583:3966-3973.
- Makishima, H., A.M. Jankowska, R.V. Tiu, H. Szpurka, Y. Sugimoto, Z. Hu, Y. Saunthararajah,
 K. Guinta, M.A. Keddache, P. Putnam, M.A. Sekeres, A.R. Moliterno, A.F. List, M.A.
 McDevitt, and J.P. Maciejewski. 2010. Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. *Leukemia* 24:1799-1804.
- Mardis, E.R., L. Ding, D.J. Dooling, D.E. Larson, M.D. McLellan, K. Chen, D.C. Koboldt, R.S. Fulton, K.D. Delehaunty, S.D. McGrath, L.A. Fulton, D.P. Locke, V.J. Magrini, R.M. Abbott, T.L. Vickery, J.S. Reed, J.S. Robinson, T. Wylie, S.M. Smith, L. Carmichael, J.M. Eldred, C.C. Harris, J. Walker, J.B. Peck, F. Du, A.F. Dukes, G.E. Sanderson, A.M. Brummett, E. Clark, J.F. McMichael, R.J. Meyer, J.K. Schindler, C.S. Pohl, J.W. Wallis, X. Shi, L. Lin, H. Schmidt, Y. Tang, C. Haipek, M.E. Wiechert, J.V. Ivy, J. Kalicki, G. Elliott, R.E. Ries, J.E. Payton, P. Westervelt, M.H. Tomasson, M.A. Watson, J. Baty, S. Heath, W.D. Shannon, R. Nagarajan, D.C. Link, M.J. Walter, T.A. Graubert, J.F. DiPersio, R.K. Wilson, and T.J. Ley. 2009. Recurring mutations found by sequencing an acute myeloid leukemia genome. *The New England journal of medicine* 361:1058-1066.
- Margolin, A.A., S.E. Ong, M. Schenone, R. Gould, S.L. Schreiber, S.A. Carr, and T.R. Golub. 2009. Empirical Bayes analysis of quantitative proteomics experiments. *PloS one* 4:e7454.
- Markham, R., and J.D. Smith. 1951. Structure of ribonucleic acid. *Nature* 168:406-408.
- Martin, H.C., S. Wani, A.L. Steptoe, K. Krishnan, K. Nones, E. Nourbakhsh, A. Vlassov, S.M. Grimmond, and N. Cloonan. 2014. Imperfect centered miRNA binding sites are common and can mediate repression of target mRNAs. *Genome biology* 15:R51.
- Masaki, S., R. Ohtsuka, Y. Abe, K. Muta, and T. Umemura. 2007. Expression patterns of microRNAs 155 and 451 during normal human erythropoiesis. *Biochemical and biophysical research communications* 364:509-514.

- Master, A., A. Wojcicka, K. Gizewska, P. Poplawski, G.R. Williams, and A. Nauman. 2016. A Novel Method for Gene-Specific Enhancement of Protein Translation by Targeting 5'UTRs of Selected Tumor Suppressors. *PloS one* 11:e0155359.
- Matera, A.G., R.M. Terns, and M.P. Terns. 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature reviews. Molecular cell biology* 8:209-220.
- Mathonnet, G., M.R. Fabian, Y.V. Svitkin, A. Parsyan, L. Huck, T. Murata, S. Biffo, W.C. Merrick, E. Darzynkiewicz, R.S. Pillai, W. Filipowicz, T.F. Duchaine, and N. Sonenberg. 2007. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* 317:1764-1767.
- Mathys, H., J. Basquin, S. Ozgur, M. Czarnocki-Cieciura, F. Bonneau, A. Aartse, A. Dziembowski, M. Nowotny, E. Conti, and W. Filipowicz. 2014. Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Molecular cell* 54:751-765.
- Matranga, C., Y. Tomari, C. Shin, D.P. Bartel, and P.D. Zamore. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* 123:607-620.
- McCarthy, D.J., and G.K. Smyth. 2009. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25:765-771.
- McNerney, M.E., C.D. Brown, X. Wang, E.T. Bartom, S. Karmakar, C. Bandlamudi, S. Yu, J. Ko,
 B.P. Sandall, T. Stricker, J. Anastasi, R.L. Grossman, J.M. Cunningham, M.M. Le Beau,
 and K.P. White. 2013. CUX1 is a haploinsufficient tumor suppressor gene on chromosome
 7 frequently inactivated in acute myeloid leukemia. *Blood* 121:975-983.
- Medvinsky, A., and E. Dzierzak. 1996. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* 86:897-906.
- Meister, G., M. Landthaler, Y. Dorsett, and T. Tuschl. 2004. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *Rna* 10:544-550.
- Melo, S.A., C. Moutinho, S. Ropero, G.A. Calin, S. Rossi, R. Spizzo, A.F. Fernandez, V. Davalos, A. Villanueva, G. Montoya, H. Yamamoto, S. Schwartz, Jr., and M. Esteller. 2010. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. *Cancer cell* 18:303-315.
- Meng, F., R. Henson, H. Wehbe-Janek, K. Ghoshal, S.T. Jacob, and T. Patel. 2007. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* 133:647-658.

- Merritt, W.M., Y.G. Lin, L.Y. Han, A.A. Kamat, W.A. Spannuth, R. Schmandt, D. Urbauer, L.A. Pennacchio, J.F. Cheng, A.M. Nick, M.T. Deavers, A. Mourad-Zeidan, H. Wang, P. Mueller, M.E. Lenburg, J.W. Gray, S. Mok, M.J. Birrer, G. Lopez-Berestein, R.L. Coleman, M. Bar-Eli, and A.K. Sood. 2008. Dicer, Drosha, and outcomes in patients with ovarian cancer. *The New England journal of medicine* 359:2641-2650.
- Michlewski, G., S. Guil, C.A. Semple, and J.F. Caceres. 2008. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Molecular cell* 32:383-393.
- Misquitta, L., and B.M. Paterson. 1999. Targeted disruption of gene function in Drosophila by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation. *Proceedings of the National Academy of Sciences of the United States of America* 96:1451-1456.
- Miyoshi, K., H. Tsukumo, T. Nagami, H. Siomi, and M.C. Siomi. 2005. Slicer function of Drosophila Argonautes and its involvement in RISC formation. *Genes & development* 19:2837-2848.
- Mizuno, T., M.Y. Chou, and M. Inouye. 1984. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proceedings of the National Academy of Sciences of the United States of America* 81:1966-1970.
- Mockenhaupt, S., S. Grosse, D. Rupp, R. Bartenschlager, and D. Grimm. 2015. Alleviation of off-target effects from vector-encoded shRNAs via codelivered RNA decoys. *Proceedings of the National Academy of Sciences of the United States of America* 112:E4007-4016.
- Molnar, A., F. Schwach, D.J. Studholme, E.C. Thuenemann, and D.C. Baulcombe. 2007. miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. *Nature* 447:1126-1129.
- Moore, M.J., T.K. Scheel, J.M. Luna, C.Y. Park, J.J. Fak, E. Nishiuchi, C.M. Rice, and R.B. Darnell. 2015. miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nature communications* 6:8864.
- Morin, R.D., N.A. Johnson, T.M. Severson, A.J. Mungall, J. An, R. Goya, J.E. Paul, M. Boyle, B.W. Woolcock, F. Kuchenbauer, D. Yap, R.K. Humphries, O.L. Griffith, S. Shah, H. Zhu, M. Kimbara, P. Shashkin, J.F. Charlot, M. Tcherpakov, R. Corbett, A. Tam, R. Varhol, D. Smailus, M. Moksa, Y. Zhao, A. Delaney, H. Qian, I. Birol, J. Schein, R. Moore, R. Holt, D.E. Horsman, J.M. Connors, S. Jones, S. Aparicio, M. Hirst, R.D. Gascoyne, and M.A. Marra. 2010. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature genetics* 42:181-185.

- Morin, R.D., M.D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A.L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C.J. Eaves, and M.A. Marra. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome research* 18:610-621.
- Morrison, S.J., and I.L. Weissman. 1994. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* 1:661-673.
- Mott, J.L., S. Kobayashi, S.F. Bronk, and G.J. Gores. 2007. mir-29 regulates Mcl-1 protein expression and apoptosis. *Oncogene* 26:6133-6140.
- Moyano, M., and G. Stefani. 2015. piRNA involvement in genome stability and human cancer. *Journal of hematology & oncology* 8:38.
- Muljo, S.A., K.M. Ansel, C. Kanellopoulou, D.M. Livingston, A. Rao, and K. Rajewsky. 2005. Aberrant T cell differentiation in the absence of Dicer. *The Journal of experimental medicine* 202:261-269.
- Muller, A.M., A. Medvinsky, J. Strouboulis, F. Grosveld, and E. Dzierzak. 1994. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity* 1:291-301.
- Nagaraj, N., J.R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Paabo, and M. Mann. 2011. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular systems biology* 7:548.
- Navarro, F., and J. Lieberman. 2010. Small RNAs guide hematopoietic cell differentiation and function. *Journal of immunology* 184:5939-5947.
- Nelson, P.T., D.A. Baldwin, L.M. Scearce, J.C. Oberholtzer, J.W. Tobias, and Z. Mourelatos. 2004. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nature methods* 1:155-161.
- Newman, M.A., J.M. Thomson, and S.M. Hammond. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *Rna* 14:1539-1549.
- Nicholson, R.H., and A.W. Nicholson. 2002. Molecular characterization of a mouse cDNA encoding Dicer, a ribonuclease III ortholog involved in RNA interference. *Mammalian genome : official journal of the International Mammalian Genome Society* 13:67-73.
- Nielsen, C.B., N. Shomron, R. Sandberg, E. Hornstein, J. Kitzman, and C.B. Burge. 2007. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *Rna* 13:1894-1910.
- Nishihara, T., L. Zekri, J.E. Braun, and E. Izaurralde. 2013. miRISC recruits decapping factors to miRNA targets to enhance their degradation. *Nucleic acids research* 41:8692-8705.

- Notredame, C., D.G. Higgins, and J. Heringa. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* 302:205-217.
- O'Connell, R.M., A.A. Chaudhuri, D.S. Rao, and D. Baltimore. 2009. Inositol phosphatase SHIP1 is a primary target of miR-155. *Proceedings of the National Academy of Sciences of the United States of America* 106:7113-7118.
- O'Connell, R.M., D.S. Rao, A.A. Chaudhuri, M.P. Boldin, K.D. Taganov, J. Nicoll, R.L. Paquette, and D. Baltimore. 2008. Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder. *The Journal of experimental medicine* 205:585-594.
- O'Donnell, K.A., E.A. Wentzel, K.I. Zeller, C.V. Dang, and J.T. Mendell. 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435:839-843.
- Obernosterer, G., P.J. Leuschner, M. Alenius, and J. Martinez. 2006. Post-transcriptional regulation of microRNA expression. *Rna* 12:1161-1167.
- Okamura, K., J.W. Hagen, H. Duan, D.M. Tyler, and E.C. Lai. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell* 130:89-100.
- Olsen, P.H., and V. Ambros. 1999. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental biology* 216:671-680.
- Ono, M., S. Bolland, P. Tempst, and J.V. Ravetch. 1996. Role of the inositol phosphatase SHIP in negative regulation of the immune system by the receptor Fc(gamma)RIIB. *Nature* 383:263-266.
- Orkin, S.H., and L.I. Zon. 2008. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132:631-644.
- Ottersbach, K., and E. Dzierzak. 2005. The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. *Developmental cell* 8:377-387.
- Paddison, P.J., A.A. Caudy, E. Bernstein, G.J. Hannon, and D.S. Conklin. 2002. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes & development* 16:948-958.
- Palis, J., S. Robertson, M. Kennedy, C. Wall, and G. Keller. 1999. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development* 126:5073-5084.
- Paquette, R.L., E.M. Landaw, R.V. Pierre, J. Kahan, M. Lubbert, O. Lazcano, G. Isaac, F. McCormick, and H.P. Koeffler. 1993. N-ras mutations are associated with poor prognosis and increased risk of leukemia in myelodysplastic syndrome. *Blood* 82:590-599.

- Paroo, Z., X. Ye, S. Chen, and Q. Liu. 2009. Phosphorylation of the human microRNAgenerating complex mediates MAPK/Erk signaling. *Cell* 139:112-122.
- Paschka, P., R.F. Schlenk, V.I. Gaidzik, M. Habdank, J. Kronke, L. Bullinger, D. Spath, S. Kayser, M. Zucknick, K. Gotze, H.A. Horst, U. Germing, H. Dohner, and K. Dohner. 2010.
 IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28:3636-3643.
- Pasquinelli, A.E., A. McCoy, E. Jimenez, E. Salo, G. Ruvkun, M.Q. Martindale, and J. Baguna.
 2003. Expression of the 22 nucleotide let-7 heterochronic RNA throughout the Metazoa: a role in life history evolution? *Evolution & development* 5:372-378.
- Paszek, P. 2014. From measuring noise toward integrated single-cell biology. *Frontiers in genetics* 5:408.
- Patterson, T.A., E.K. Lobenhofer, S.B. Fulmer-Smentek, P.J. Collins, T.M. Chu, W. Bao, H. Fang, E.S. Kawasaki, J. Hager, I.R. Tikhonova, S.J. Walker, L. Zhang, P. Hurban, F. de Longueville, J.C. Fuscoe, W. Tong, L. Shi, and R.D. Wolfinger. 2006. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nature biotechnology* 24:1140-1150.
- Peculis, B.A. 2000. RNA-binding proteins: if it looks like a sn(o)RNA. *Current biology : CB* 10:R916-918.
- Pedersen-Bjergaard, J., D.H. Christiansen, F. Desta, and M.K. Andersen. 2006. Alternative genetic pathways and cooperating genetic abnormalities in the pathogenesis of therapy-related myelodysplasia and acute myeloid leukemia. *Leukemia* 20:1943-1949.
- Pekarsky, Y., N. Zanesi, R.I. Aqeilan, and C.M. Croce. 2007. Animal models for chronic lymphocytic leukemia. *Journal of cellular biochemistry* 100:1109-1118.
- Pellagatti, A., and J. Boultwood. 2015. Recent Advances in the 5q- Syndrome. *Mediterranean journal of hematology and infectious diseases* 7:e2015037.
- Pellagatti, A., E. Hellstrom-Lindberg, A. Giagounidis, J. Perry, L. Malcovati, M.G. Della Porta, M. Jadersten, S. Killick, C. Fidler, M. Cazzola, J.S. Wainscoat, and J. Boultwood. 2008.
 Haploinsufficiency of RPS14 in 5q- syndrome is associated with deregulation of ribosomal- and translation-related genes. *British journal of haematology* 142:57-64.
- Petersen, C.P., M.E. Bordeleau, J. Pelletier, and P.A. Sharp. 2006. Short RNAs repress translation after initiation in mammalian cells. *Molecular cell* 21:533-542.

- Peterson, S.M., J.A. Thompson, M.L. Ufkin, P. Sathyanarayana, L. Liaw, and C.B. Congdon. 2014. Common features of microRNA target prediction tools. *Frontiers in genetics* 5:23.
- Petriv, O.I., F. Kuchenbauer, A.D. Delaney, V. Lecault, A. White, D. Kent, L. Marmolejo, M. Heuser, T. Berg, M. Copley, J. Ruschmann, S. Sekulovic, C. Benz, E. Kuroda, V. Ho, F. Antignano, T. Halim, V. Giambra, G. Krystal, C.J. Takei, A.P. Weng, J. Piret, C. Eaves, M.A. Marra, R.K. Humphries, and C.L. Hansen. 2010. Comprehensive microRNA expression profiling of the hematopoietic hierarchy. *Proceedings of the National Academy of Sciences of the United States of America* 107:15443-15448.
- Piehowski, P.D., V.A. Petyuk, D.J. Orton, F. Xie, R.J. Moore, M. Ramirez-Restrepo, A. Engel, A.P. Lieberman, R.L. Albin, D.G. Camp, R.D. Smith, and A.J. Myers. 2013. Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *Journal of proteome research* 12:2128-2137.
- Pikielny, C.W., and M. Rosbash. 1986. Specific small nuclear RNAs are associated with yeast spliceosomes. *Cell* 45:869-877.
- Polioudakis, D., A.A. Bhinge, P.J. Killion, B.K. Lee, N.S. Abell, and V.R. Iyer. 2013. A MycmicroRNA network promotes exit from quiescence by suppressing the interferon response and cell-cycle arrest genes. *Nucleic acids research* 41:2239-2254.
- Rajkowitsch, L., D. Chen, S. Stampfl, K. Semrad, C. Waldsich, O. Mayer, M.F. Jantsch, R. Konrat, U. Blasi, and R. Schroeder. 2007. RNA chaperones, RNA annealers and RNA helicases. *RNA biology* 4:118-130.
- Rakheja, D., K.S. Chen, Y. Liu, A.A. Shukla, V. Schmid, T.C. Chang, S. Khokhar, J.E. Wickiser,
 N.J. Karandikar, J.S. Malter, J.T. Mendell, and J.F. Amatruda. 2014. Somatic mutations in
 DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in
 Wilms tumours. *Nature communications* 2:4802.
- Rand, T.A., S. Petersen, F. Du, and X. Wang. 2005. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* 123:621-629.
- Raver-Shapira, N., E. Marciano, E. Meiri, Y. Spector, N. Rosenfeld, N. Moskovits, Z. Bentwich, and M. Oren. 2007. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Molecular cell* 26:731-743.
- Reid, J.G., A.K. Nagaraja, F.C. Lynn, R.B. Drabek, D.M. Muzny, C.A. Shaw, M.K. Weiss, A.O. Naghavi, M. Khan, H. Zhu, J. Tennakoon, G.H. Gunaratne, D.B. Corry, J. Miller, M.T. McManus, M.S. German, R.A. Gibbs, M.M. Matzuk, and P.H. Gunaratne. 2008. Mouse let-7 miRNA populations exhibit RNA editing that is constrained in the 5'-seed/

cleavage/anchor regions and stabilize predicted mmu-let-7a:mRNA duplexes. *Genome research* 18:1571-1581.

- Reinhart, B.J., F.J. Slack, M. Basson, A.E. Pasquinelli, J.C. Bettinger, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 403:901-906.
- Reyes-Herrera, P.H., and E. Ficarra. 2012. One decade of development and evolution of microRNA target prediction algorithms. *Genomics, proteomics & bioinformatics* 10:254-263.
- Rosa, A., M. Ballarino, A. Sorrentino, O. Sthandier, F.G. De Angelis, M. Marchioni, B. Masella,
 A. Guarini, A. Fatica, C. Peschle, and I. Bozzoni. 2007. The interplay between the master transcription factor PU.1 and miR-424 regulates human monocyte/macrophage differentiation. *Proceedings of the National Academy of Sciences of the United States of America* 104:19849-19854.
- Rouhi, A., D.L. Mager, R.K. Humphries, and F. Kuchenbauer. 2008. MiRNAs, epigenetics, and cancer. *Mammalian genome : official journal of the International Mammalian Genome Society* 19:517-525.
- Ruby, J.G., C.H. Jan, and D.P. Bartel. 2007. Intronic microRNA precursors that bypass Drosha processing. *Nature* 448:83-86.
- Sachs, A.B., and G. Varani. 2000. Eukaryotic translation initiation: there are (at least) two sides to every story. *Nature structural biology* 7:356-361.
- Saito, Y., and P.A. Jones. 2006. Epigenetic activation of tumor suppressor microRNAs in human cancer cells. *Cell cycle* 5:2220-2222.
- Sargin, B., C. Choudhary, N. Crosetto, M.H. Schmidt, R. Grundler, M. Rensinghoff, C. Thiessen,
 L. Tickenbrock, J. Schwable, C. Brandts, B. August, S. Koschmieder, S.R. Bandi, J.
 Duyster, W.E. Berdel, C. Muller-Tidow, I. Dikic, and H. Serve. 2007. Flt3-dependent transformation by inactivating c-Cbl mutations in AML. *Blood* 110:1004-1012.
- Sashida, G., H. Harada, H. Matsui, M. Oshima, M. Yui, Y. Harada, S. Tanaka, M. Mochizuki-Kashio, C. Wang, A. Saraya, T. Muto, Y. Hayashi, K. Suzuki, H. Nakajima, T. Inaba, H. Koseki, G. Huang, T. Kitamura, and A. Iwama. 2014. Ezh2 loss promotes development of myelodysplastic syndrome but attenuates its predisposition to leukaemic transformation. *Nature communications* 5:4177.
- Schena, M., D. Shalon, R. Heller, A. Chai, P.O. Brown, and R.W. Davis. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences of the United States of America* 93:10614-10619.

- Schlenner, S.M., V. Madan, K. Busch, A. Tietz, C. Laufle, C. Costa, C. Blum, H.J. Fehling, and H.R. Rodewald. 2010. Fate mapping reveals separate origins of T cells and myeloid lineages in the thymus. *Immunity* 32:426-436.
- Schneider, R.K., V. Adema, D. Heckl, M. Jaras, M. Mallo, A.M. Lord, L.P. Chu, M.E. McConkey,
 R. Kramann, A. Mullally, R. Bejar, F. Sole, and B.L. Ebert. 2014. Role of casein kinase
 1A1 in the biology and targeted therapy of del(5q) MDS. *Cancer cell* 26:509-520.
- Schramedei, K., N. Morbt, G. Pfeifer, J. Lauter, M. Rosolowski, J.M. Tomm, M. von Bergen, F. Horn, and K. Brocke-Heidrich. 2011. MicroRNA-21 targets tumor suppressor genes ANP32A and SMARCA4. Oncogene 30:2975-2985.
- Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. 2011. Global quantification of mammalian gene expression control. *Nature* 473:337-342.
- Schwarz, D.S., G. Hutvagner, T. Du, Z. Xu, N. Aronin, and P.D. Zamore. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115:199-208.
- Seggerson, K., L. Tang, and E.G. Moss. 2002. Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. *Developmental biology* 243:215-225.
- Selbach, M., B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* 455:58-63.
- Sewer, A., N. Paul, P. Landgraf, A. Aravin, S. Pfeffer, M.J. Brownstein, T. Tuschl, E. van Nimwegen, and M. Zavolan. 2005. Identification of clustered microRNAs using an ab initio prediction method. *BMC bioinformatics* 6:267.
- Shi, W., D. Hendrix, M. Levine, and B. Haley. 2009. A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. *Nature structural & molecular biology* 16:183-189.
- Shin, C., J.W. Nam, K.K. Farh, H.R. Chiang, A. Shkumatava, and D.P. Bartel. 2010. Expanding the microRNA targeting code: functional sites with centered pairing. *Molecular cell* 38:789-802.
- Siddiqui, N., D.A. Mangus, T.C. Chang, J.M. Palermino, A.B. Shyu, and K. Gehring. 2007. Poly(A) nuclease interacts with the C-terminal domain of polyadenylate-binding protein domain from poly(A)-binding protein. *The Journal of biological chemistry* 282:25067-25075.
- Siminovitch, L., E.A. McCulloch, and J.E. Till. 1963. The Distribution of Colony-Forming Cells among Spleen Colonies. *Journal of cellular physiology* 62:327-336.

- Simmons, A. 1997. Hematology : a combined theoretical and technical approach. Butterworth-Heinemann, Boston. x, 507 p. pp.
- Slade, I., C. Bacchelli, H. Davies, A. Murray, F. Abbaszadeh, S. Hanks, R. Barfoot, A. Burke, J. Chisholm, M. Hewitt, H. Jenkinson, D. King, B. Morland, B. Pizer, K. Prescott, A. Saggar, L. Side, H. Traunecker, S. Vaidya, P. Ward, P.A. Futreal, G. Vujanic, A.G. Nicholson, N. Sebire, C. Turnbull, J.R. Priest, K. Pritchard-Jones, R. Houlston, C. Stiller, M.R. Stratton, J. Douglas, and N. Rahman. 2011. DICER1 syndrome: clarifying the diagnosis, clinical features and management implications of a pleiotropic tumour predisposition syndrome. *Journal of medical genetics* 48:273-278.
- Smalheiser, N.R. 2003. EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome biology* 4:403.
- Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology* 3:Article3.
- Sokol, L., G. Caceres, S. Volinia, H. Alder, G.J. Nuovo, C.G. Liu, K. McGraw, J.A. Clark, C.A. Sigua, D.T. Chen, L. Moscinski, C.M. Croce, and A.F. List. 2011. Identification of a risk dependent microRNA expression signature in myelodysplastic syndromes. *British journal* of haematology 153:24-32.
- Sonenberg, N., K.M. Rupprecht, S.M. Hecht, and A.J. Shatkin. 1979. Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proceedings of the National Academy of Sciences of the United States of America* 76:4345-4349.
- Song, J.J., S.K. Smith, G.J. Hannon, and L. Joshua-Tor. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305:1434-1437.
- Song, S.J., K. Ito, U. Ala, L. Kats, K. Webster, S.M. Sun, M. Jongen-Lavrencic, K. Manova-Todorova, J. Teruya-Feldstein, D.E. Avigan, R. Delwel, and P.P. Pandolfi. 2013. The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation. *Cell stem cell* 13:87-101.
- Song, S.J., and P.P. Pandolfi. 2014. MicroRNAs in the pathogenesis of myelodysplastic syndromes and myeloid leukaemia. *Current opinion in hematology* 21:276-282.
- Starczynowski, D.T., F. Kuchenbauer, B. Argiropoulos, S. Sung, R. Morin, A. Muranyi, M. Hirst, D. Hogge, M. Marra, R.A. Wells, R. Buckstein, W. Lam, R.K. Humphries, and A. Karsan.

2010. Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. *Nature medicine* 16:49-58.

- Starczynowski, D.T., R. Morin, A. McPherson, J. Lam, R. Chari, J. Wegrzyn, F. Kuchenbauer,
 M. Hirst, K. Tohyama, R.K. Humphries, W.L. Lam, M. Marra, and A. Karsan. 2011.
 Genome-wide identification of human microRNAs located in leukemia-associated genomic alterations. *Blood* 117:595-607.
- Starega-Roslan, J., J. Krol, E. Koscianska, P. Kozlowski, W.J. Szlachcic, K. Sobczak, and W.J. Krzyzosiak. 2011. Structural basis of microRNA length variety. *Nucleic acids research* 39:257-268.
- Steensma, D.P., and A.F. List. 2005. Genetic testing in the myelodysplastic syndromes: molecular insights into hematologic diversity. *Mayo Clinic proceedings* 80:681-698.
- Stoffers, S.L., S.E. Meyer, and H.L. Grimes. 2012. MicroRNAs in the midst of myeloid signal transduction. *Journal of cellular physiology* 227:525-533.
- Su, J., H. Liang, W. Yao, N. Wang, S. Zhang, X. Yan, H. Feng, W. Pang, Y. Wang, X. Wang, Z. Fu, Y. Liu, C. Zhao, J. Zhang, C.Y. Zhang, K. Zen, X. Chen, and Y. Wang. 2014. MiR-143 and MiR-145 regulate IGF1R to suppress cell proliferation in colorectal cancer. *PloS one* 9:e114420.
- Su, X., D. Chakravarti, M.S. Cho, L. Liu, Y.J. Gi, Y.L. Lin, M.L. Leung, A. El-Naggar, C.J. Creighton, M.B. Suraokar, I. Wistuba, and E.R. Flores. 2010. TAp63 suppresses metastasis through coordinate regulation of Dicer and miRNAs. *Nature* 467:986-990.
- Subtelny, A.O., S.W. Eichhorn, G.R. Chen, H. Sive, and D.P. Bartel. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508:66-71.
- Tagawa, H., and M. Seto. 2005. A microRNA cluster as a target of genomic amplification in malignant lymphoma. *Leukemia* 19:2013-2016.
- Tang, J.L., H.A. Hou, C.Y. Chen, C.Y. Liu, W.C. Chou, M.H. Tseng, C.F. Huang, F.Y. Lee, M.C. Liu, M. Yao, S.Y. Huang, B.S. Ko, S.C. Hsu, S.J. Wu, W. Tsay, Y.C. Chen, L.I. Lin, and H.F. Tien. 2009. AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood* 114:5352-5361.
- Tanner, A., S.E. Taylor, W. Decottignies, and B.K. Berges. 2014. Humanized mice as a model to study human hematopoietic stem cell transplantation. *Stem cells and development* 23:76-82.
- Tarasov, V., P. Jung, B. Verdoodt, D. Lodygin, A. Epanchintsev, A. Menssen, G. Meister, and H. Hermeking. 2007. Differential regulation of microRNAs by p53 revealed by massively

parallel sequencing: miR-34a is a p53 target that induces apoptosis and G1-arrest. *Cell cycle* 6:1586-1593.

- Tavassoli, M. 1991. Embryonic and fetal hemopoiesis: an overview. *Blood cells* 17:269-281; discussion 282-266.
- Tefferi, A. 2010. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. *Leukemia* 24:1128-1138.
- Thol, F., E.M. Weissinger, J. Krauter, K. Wagner, F. Damm, M. Wichmann, G. Gohring, C. Schumann, G. Bug, O. Ottmann, W.K. Hofmann, B. Schlegelberger, A. Ganser, and M. Heuser. 2010. IDH1 mutations in patients with myelodysplastic syndromes are associated with an unfavorable prognosis. *Haematologica* 95:1668-1674.
- Thomson, J.M., M. Newman, J.S. Parker, E.M. Morin-Kensicki, T. Wright, and S.M. Hammond. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes & development* 20:2202-2207.
- Till, J.E., and C.E. Mc. 1961. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation research* 14:213-222.
- Tokumaru, S., M. Suzuki, H. Yamada, M. Nagino, and T. Takahashi. 2008. let-7 regulates Dicer expression and constitutes a negative feedback loop. *Carcinogenesis* 29:2073-2077.
- Torrezan, G.T., E.N. Ferreira, A.M. Nakahata, B.D. Barros, M.T. Castro, B.R. Correa, A.C. Krepischi, E.H. Olivieri, I.W. Cunha, U. Tabori, P.E. Grundy, C.M. Costa, B. de Camargo, P.A. Galante, and D.M. Carraro. 2014. Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. *Nature communications* 5:4039.
- Toth, K.F., D. Pezic, E. Stuwe, and A. Webster. 2016. The piRNA Pathway Guards the Germline Genome Against Transposable Elements. *Advances in experimental medicine and biology* 886:51-77.
- Tucker, M., M.A. Valencia-Sanchez, R.R. Staples, J. Chen, C.L. Denis, and R. Parker. 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell* 104:377-386.
- Tusher, V.G., R. Tibshirani, and G. Chu. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* 98:5116-5121.
- Urbach, A., A. Yermalovich, J. Zhang, C.S. Spina, H. Zhu, A.R. Perez-Atayde, R. Shukrun, J. Charlton, N. Sebire, W. Mifsud, B. Dekel, K. Pritchard-Jones, and G.Q. Daley. 2014. Lin28 sustains early renal progenitors and induces Wilms tumor. *Genes & development* 28:971-982.

- Vagin, V.V., A. Sigova, C. Li, H. Seitz, V. Gvozdev, and P.D. Zamore. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313:320-324.
- Van den Berghe, H., J.J. Cassiman, G. David, J.P. Fryns, J.L. Michaux, and G. Sokal. 1974. Distinct haematological disorder with deletion of long arm of no. 5 chromosome. *Nature* 251:437-438.
- Vanhee-Brossollet, C., H. Thoreau, N. Serpente, L. D'Auriol, J.P. Levy, and C. Vaquero. 1995. A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals. *Virology* 206:196-202.
- Vardiman, J.W., J. Thiele, D.A. Arber, R.D. Brunning, M.J. Borowitz, A. Porwit, N.L. Harris, M.M. Le Beau, E. Hellstrom-Lindberg, A. Tefferi, and C.D. Bloomfield. 2009. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 114:937-951.
- Vasilatou, D., S.G. Papageorgiou, G. Dimitriadis, and V. Pappa. 2013. Epigenetic alterations and microRNAs: new players in the pathogenesis of myelodysplastic syndromes. *Epigenetics* 8:561-570.
- Vella, M.C., K. Reinert, and F.J. Slack. 2004. Architecture of a validated microRNA::target interaction. *Chemistry & biology* 11:1619-1623.
- Ventura, A., A.G. Young, M.M. Winslow, L. Lintault, A. Meissner, S.J. Erkeland, J. Newman, R.T. Bronson, D. Crowley, J.R. Stone, R. Jaenisch, P.A. Sharp, and T. Jacks. 2008.
 Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* 132:875-886.
- Viney, M., and S.E. Reece. 2013. Adaptive noise. *Proceedings. Biological sciences / The Royal Society* 280:20131104.
- Viswanathan, S.R., G.Q. Daley, and R.I. Gregory. 2008. Selective blockade of microRNA processing by Lin28. *Science* 320:97-100.
- Viswanathan, S.R., J.T. Powers, W. Einhorn, Y. Hoshida, T.L. Ng, S. Toffanin, M. O'Sullivan, J. Lu, L.A. Phillips, V.L. Lockhart, S.P. Shah, P.S. Tanwar, C.H. Mermel, R. Beroukhim, M. Azam, J. Teixeira, M. Meyerson, T.P. Hughes, J.M. Llovet, J. Radich, C.G. Mullighan, T.R. Golub, P.H. Sorensen, and G.Q. Daley. 2009. Lin28 promotes transformation and is associated with advanced human malignancies. *Nature genetics* 41:843-848.
- Vogel, C., and E.M. Marcotte. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics* 13:227-232.

- Votavova, H., M. Grmanova, M. Dostalova Merkerova, M. Belickova, A. Vasikova, R. Neuwirtova, and J. Cermak. 2011. Differential expression of microRNAs in CD34+ cells of 5q- syndrome. *Journal of hematology & oncology* 4:1.
- Wada, T., J. Kikuchi, N. Nishimura, R. Shimizu, T. Kitamura, and Y. Furukawa. 2009. Expression levels of histone deacetylases determine the cell fate of hematopoietic progenitors. *The Journal of biological chemistry* 284:30673-30683.
- Walz, A.L., A. Ooms, S. Gadd, D.S. Gerhard, M.A. Smith, J.M. Guidry Auvil, D. Meerzaman, Q.R. Chen, C.H. Hsu, C. Yan, C. Nguyen, Y. Hu, R. Bowlby, D. Brooks, Y. Ma, A.J. Mungall, R.A. Moore, J. Schein, M.A. Marra, V. Huff, J.S. Dome, Y.Y. Chi, C.G. Mullighan, J. Ma, D.A. Wheeler, O.A. Hampton, N. Jafari, N. Ross, J.M. Gastier-Foster, and E.J. Perlman. 2015. Recurrent DGCR8, DROSHA, and SIX homeodomain mutations in favorable histology Wilms tumors. *Cancer cell* 27:286-297.
- Wang, P.W., J.D. Eisenbart, R. Espinosa, 3rd, E.M. Davis, R.A. Larson, and M.M. Le Beau. 2000. Refinement of the smallest commonly deleted segment of chromosome 20 in malignant myeloid diseases and development of a PAC-based physical and transcription map. *Genomics* 67:28-39.
- Wang, Q., I. Lee, J. Ren, S.S. Ajay, Y.S. Lee, and X. Bao. 2013. Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. *Molecular therapy : the journal of the American Society of Gene Therapy* 21:368-379.
- Wang, T., G. Wang, D. Hao, X. Liu, D. Wang, N. Ning, and X. Li. 2015a. Aberrant regulation of the LIN28A/LIN28B and let-7 loop in human malignant tumors and its effects on the hallmarks of cancer. *Molecular cancer* 14:125.
- Wang, Y., J. Chen, W. Yang, F. Mo, J. Senz, D. Yap, M.S. Anglesio, B. Gilks, G.B. Morin, and D.G. Huntsman. 2015b. The oncogenic roles of DICER1 RNase IIIb domain mutations in ovarian Sertoli-Leydig cell tumors. *Neoplasia* 17:650-660.
- Wang, Y., S. Juranek, H. Li, G. Sheng, T. Tuschl, and D.J. Patel. 2008a. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 456:921-926.
- Wang, Y., G. Sheng, S. Juranek, T. Tuschl, and D.J. Patel. 2008b. Structure of the guidestrand-containing argonaute silencing complex. *Nature* 456:209-213.
- Ward, P.S., J. Patel, D.R. Wise, O. Abdel-Wahab, B.D. Bennett, H.A. Coller, J.R. Cross, V.R.Fantin, C.V. Hedvat, A.E. Perl, J.D. Rabinowitz, M. Carroll, S.M. Su, K.A. Sharp, R.L.Levine, and C.B. Thompson. 2010. The common feature of leukemia-associated IDH1 and

IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2hydroxyglutarate. *Cancer cell* 17:225-234.

- Watanabe, T., and H. Lin. 2014. Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs. *Molecular cell* 56:18-27.
- Waterhouse, P.M., M.W. Graham, and M.B. Wang. 1998. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proceedings of the National Academy of Sciences of the United States of America* 95:13959-13964.
- Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 75:855-862.
- Wilson, R.C., A. Tambe, M.A. Kidwell, C.L. Noland, C.P. Schneider, and J.A. Doudna. 2015. Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Molecular cell* 57:397-407.
- Wolf, J., and L.A. Passmore. 2014. mRNA deadenylation by Pan2-Pan3. *Biochemical Society transactions* 42:184-187.
- Wright, G.W., and R.M. Simon. 2003. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics* 19:2448-2455.
- Xia, Z., D.R. Bell, Y. Shi, and P. Ren. 2013a. RNA 3D structure prediction by using a coarsegrained model and experimental data. *The journal of physical chemistry. B* 117:3135-3144.
- Xia, Z., T. Huynh, P. Ren, and R. Zhou. 2013b. Large domain motions in Ago protein controlled by the guide DNA-strand seed region determine the Ago-DNA-mRNA complex recognition process. *PloS one* 8:e54620.
- Yamashita, A., T.C. Chang, Y. Yamashita, W. Zhu, Z. Zhong, C.Y. Chen, and A.B. Shyu. 2005. Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nature structural & molecular biology* 12:1054-1063.
- Ye, Y., M.A. McDevitt, M. Guo, W. Zhang, O. Galm, S.D. Gore, J.E. Karp, J.P. Maciejewski, J. Kowalski, H.L. Tsai, L.P. Gondek, H.C. Tsai, X. Wang, C. Hooker, B.D. Smith, H.E. Carraway, and J.G. Herman. 2009. Progressive chromatin repression and promoter methylation of CTNNA1 associated with advanced myeloid malignancies. *Cancer research* 69:8482-8490.
- Yekta, S., I.H. Shih, and D.P. Bartel. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304:594-596.

- Yelamanchili, S.V., A.D. Chaudhuri, L.N. Chen, H. Xiong, and H.S. Fox. 2010. MicroRNA-21 dysregulates the expression of MEF2C in neurons in monkey and human SIV/HIV neurological disease. *Cell death & disease* 1:e77.
- Yi, R., Y. Qin, I.G. Macara, and B.R. Cullen. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & development* 17:3011-3016.
- Yoda, M., T. Kawamata, Z. Paroo, X. Ye, S. Iwasaki, Q. Liu, and Y. Tomari. 2010. ATPdependent human RISC assembly pathways. *Nature structural & molecular biology* 17:17-23.
- Yu, D., C.O. dos Santos, G. Zhao, J. Jiang, J.D. Amigo, E. Khandros, L.C. Dore, Y. Yao, J. D'Souza, Z. Zhang, S. Ghaffari, J. Choi, S. Friend, W. Tong, J.S. Orange, B.H. Paw, and M.J. Weiss. 2010. miR-451 protects against erythroid oxidant stress by repressing 14-3-3zeta. *Genes & development* 24:1620-1633.
- Yuan, Y.R., Y. Pei, J.B. Ma, V. Kuryavyi, M. Zhadina, G. Meister, H.Y. Chen, Z. Dauter, T. Tuschl, and D.J. Patel. 2005. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Molecular cell* 19:405-419.
- Zeng, Y., and B.R. Cullen. 2004. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic acids research* 32:4776-4785.
- Zeng, Y., and B.R. Cullen. 2005. Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *The Journal of biological chemistry* 280:27595-27603.
- Zeng, Y., R. Yi, and B.R. Cullen. 2005. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *The EMBO journal* 24:138-148.
- Zhang, B., E.J. Stellwag, and X. Pan. 2009. Large-scale genome analysis reveals unique features of microRNAs. *Gene* 443:100-109.
- Zhao, T., G. Li, S. Mi, S. Li, G.J. Hannon, X.J. Wang, and Y. Qi. 2007a. A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. *Genes & development* 21:1190-1203.
- Zhao, Y., J.F. Ransom, A. Li, V. Vedantham, M. von Drehle, A.N. Muth, T. Tsuchihashi, M.T. McManus, R.J. Schwartz, and D. Srivastava. 2007b. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129:303-317.
- Zhou, H., M.L. Arcila, Z. Li, E.J. Lee, C. Henzler, J. Liu, T.M. Rana, and K.S. Kosik. 2012. Deep annotation of mouse iso-miR and iso-moR variation. *Nucleic acids research* 40:5864-5875.

- Zhou, H., and I. Rigoutsos. 2014. MiR-103a-3p targets the 5' UTR of GPRC5A in pancreatic cells. *Rna* 20:1431-1439.
- Zwieb, C. 2014. The principles of RNA structure architecture. *Methods in molecular biology* 1097:33-43.

Appendix

For Chapter 4, Molecular Analysis of Potential microRNA, Supplemental Data

In Chapter 4, the pLL-miR-143spg was analyzed for heptameric repeats. Any heptamer that was repeated more than once was considered a potential seed binding site. Considering the number of transcripts produced from the CMV promoter of the pLL-miR-143spg vector, it is possible that heptamers only repeated once may act as seed binding sites as well. Searching for these sites in the output of new novel microRNA analyses could be a future avenue of investigation.

Table A.0.1 - Discovery of Repetitive Sequences within Sponge

A scanning window of 7 nucleotides (see box overlaid on text) was used to find repetitive sequences within the four tandem repeats of the sponge transcript. The underlined text is the miR-143 seed site, while the central bolded letters CGA or CAA are the opposite of the reverse complement sequence designed to create a bulge, the 4-5 letters at the end of each repeat are the random nucleotides inserted for spacing and to prevent self-annealing of the transcript. Repeat 1

Repeat 2

Repeat 3

Repeat 4

GAGCTACAGCGATCATCTCA GCTA GgGCTACAGCGATCATCTCA TTAAG GAGCTgCAGCAATCATCTCA AACCT GAGCTACgGCGATCATCTCA atCA

Seed Binding Site	Repeats	Difference from Major Repeat	Reverse Complement
GAGCTAC	2		GUAGCUC
G g GCTAC	1	1	GUAGCCC
GAGCT g C	1	1	GCAGCUC
AGCTACA	1		UGUAGCU
g GCTACA	1	1	UGUAGCC
AGCT g CA	1	1	UGCAGCU
AGCTAC g	1	1	CGUAGCU
GCTACAG	2		CUGUAGC
GCT g CAG	1	1	CUGCAGC
GCTAC g G	1	1	CCGUAGC
CTACAG C	2		GCUGUAG
CTgCAG C	1	1	GCUGCAG
CTACgG C	1	1	GCCGUAG
TACAG CG	2		CGCUGUA
TgCAG CA	1	2	UGCUGCA
TACgG CG	1	1	CGCCGUA
ACAG CGA	2		UCGCUGU
gCAG CAA	1	2	UUGCUGC
ACgG CGA	1	1	UCGCCGU
CAG CGA T	2		AUCGCUG
CAG CAA T	1	1	AUUGCUG
CgG CGA T	1	1	AUCGCCG
AG CGA TC	2		GAUCGCU
AG CAA TC	1	1	GAUUGCU
gGCGATC	1	1	GAUCGCC
GCGATCA	3		UGAUCGC

	ipiement
GCAATCA 1 UGAUUGC	
CGATCAT 3 AUGAUCG	
CAATCAT 1 AUGAUUG	
GATCATC 3 GAUGAUC	
AATCATC 1 1 GAUGAUU	
ATCATCT 4 AGAUGAU	
ATCTCAA/a 2 UUGAGAU	
ATCTCAG 1 CUGAGAU	
ATCTCAT 1 AUGAGAU	
TCTCAAA 1 UUUGAGA	
TCTCAat 1 AUUGAGA	
TCTCATT 1 AAUGAGA	
TCTCAGC 1 GCUGAGA	
CTCAAAC 1 GUUUGAG	
CTCATTA 1 UAAUGAG	
CTCAGCT 1 AGCUGAG	
CTCAatC 1 GAUUGAG	
TCAAACC 1 GGUUUGA	
TCATTAA 1 UUAAUGA	
TCAGCTA 1 UAGCUGA	
TCAatCA 1 UGAUUGA	
CAAACCT 1 AGGUUUG	
CATTAAG 1 CUUAAUG	
CAGCTAG 1 CUAGCUG	
CAatCAG 1 CUGAUUG	
AAACCTG 1 CAGGUUU	
ATTAAGG 1 CCUUAAU	
AGCTAGg 1 CCUAGCU	
AatCAGA 1 UCUGAUU	
AACCTGA 1 UCAGGUU	
TTAAGGA 1 UCCUUAA	
GCTAGqG 1 CCCUAGC	
atCAGAG 1 CUCUGAU	
ACCTGAG 1 CUCAGGU	
TAAGGAG 1 CUCCUUA	
CTAG g GC 1 GCCCUAG	
tCAGAGC 1 GCUCUGA	
CCTGAGC 1 GCUCAGG	
AAGGAGC 1 GCUCCUU	
TAGaGCT 1 AGCCCUA	
CAGAGCT 1 1 AGCUCUG	
CTGAGCT 1 2 AGCUCAG	
AGAGCTA 1 UAGCUCU	

For Chapter 5, Empirical Bayes Analysis, Supplemental Data

The Empirical Bayes analysis was performed for the modified sponges dataset, and the posterior probability of upregulation and downregulation for each protein in each of the samples was determined. If the protein on average had a log₂ fold-change of greater than 0.3, it had a 65-75% posterior probability of upregulation. If the change was less than -0.3, there was a similar posterior probability of downregulation.





Empirical bayes analysis of proteins found in the two original sponge replicates used in limma analysis. The proteins in each sample are evaluated to find the posterior probability of them being upregulated (green dots) or downregulated (pink dots).



Figure A.2 - Empirical Bayes analysis of modified sponge samples

Empirical bayes analysis of the proteins found in the two miR-143specific-spg replicates used in limma analysis. The proteins in each sample are evaluated to find the posterior probabilities for upregulation (green dots) and downregulation (pink dots).



Figure A.3 - Empirical Bayes analysis of modified sponge samples

Empirical bayes analysis of the proteins found in the two miR-Xspecific-spg replicates used in limma analysis. The proteins in each sample are evaluated to find the posterior probabilities for upregulation (green dots) and downregulation (pink dots).