Identification of miRNA pathway genes using a novel approach for identification of trans-factors acting on cis-regulatory elements in the 3′ UTR

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jacob Andrew Merle

December 2019

Identification of miRNA pathway genes using a novel approach for identification of *trans*-factors acting on cis-regulatory elements in the 3′ UTR

Jacob Andrew Merle, Ph.D.

Cornell University 2019

The 3′ untranslated regions (UTR) of mammalian genomes are well conserved and play an important role in regulation of the mRNA, serving as a binding site for many different regulatory factors. The majority of RNA binding proteins (RBP) involved in post-transcriptional regulation target *cis*-regulatory elements found within the 3′ UTR, however hundreds of proteins have been predicted to contain RNA binding domains and only a modest subset has been studied. While many of these *trans*-factors and their corresponding *cis*-regulatory elements have yet to be investigated, microRNAs (miRNA) and their corresponding target sites represent a relatively well understood *trans-cis*-regulatory paradigm for the regulatory biology of 3′ UTRs. miRNAs are an important class of small (~22nt) noncoding RNAs that regulate gene expression post transcriptionally, targeting transcripts containing complementary target sites primarily found within the 3′ UTR, and which recruit additional factors that promote accelerated decay and translational repression. Although the core components of the miRNA biogenesis and effector pathways have been discovered, there exist multiple gaps in our understanding of miRNA biogenesis and function. For example, many questions remain regarding the mechanism by which miRNAs are able to direct translational repression and accelerated mRNA

decay, and most current models suggest the existence of additional, currently unknown, components of the silencing pathway.

To identify novel miRNA pathway genes, I developed a high throughput screening method which can be used to identify *trans*-factors responsible for the function of 3′ UTR *cis*-regulatory elements. This screening method was accomplished by developing a fluorescent reporter cell line sensitive to RNAi perturbation events that alter miRNA-mediated repression and performing a high throughput RNAi screen using a pooled lentiviral short hairpin RNA (shRNA) library containing ~55,000 shRNAs targeting ~11,000 genes. Fluorescence activated cell sorting (FACS) was used to enrich cells experiencing elevated or depressed levels of miRNA mediated repression acting upon the fluorescent reporters, and high throughput sequencing to identify the shRNAs underlying altered reporter activity. Analysis of enriched shRNAs thereby allowed the identification of candidate genes involved in the miRNA pathway. The successful enrichment of known miRNA pathway genes validated the efficacy of this approach and identified candidate genes not previously known to participate in miRNA silencing. Follow-up validation experiments provided further support for 19 candidates. Finally, I generated CRISPR knockout cell lines for a single candidate, KRBOX4, analysis of which provided additional evidence, suggesting a potential role for this gene in miRNA silencing.

BIOGRAPHICAL SKETCH

Jacob Merle was born December 9th, 1989 in Westfield, New York. He grew up in the nearby town of Portland, a small New York town on the shore of Lake Erie, with a population of less than 5,000 people. He grew up doing small town things like hunting pheasants, turkey and deer, and developed a love for the outdoors. When he turned 12, he got his first job working in the vineyards that surround his hometown tying grapes, but he also worked many unpaid jobs, some would say against his will, helping his dad fix up old houses. This gave him confidence and a do-it-yourself mentality that he would later rely on during his masters and PhD. In high school he worked at a local greenhouse after school and on weekends, which put fuel in his 1993 Chevy S-10 pickup that leaked gas if the tank was more than half full but it still gave him the freedom to go to the movies nearly every weekend.

While in school Jacob developed a love of biology and after graduating high school in 2008 with his 60 classmates (one of the larger classes), he decided to stay close to home and attended Fredonia State University just 10 miles away. In May 2012, he earned a B.S. in biology and took a position as a graduate teaching assistant working in the lab of Dr. Scott Ferguson. While there he worked on a research project to show the existence of an internal ribosomal entry site in *gurken*, a developmentally important mRNA in *Drosophila*, while also teaching lab courses in Genetics and Genomics, Principles of Biology II, Biochemistry, and Microbiology. Originally Jacob had been planning to apply to medical school, but his exposure to research under the mentorship of Dr. Scott Ferguson had ignited an interest in RNA biology which led him to pursue a PhD instead.

Shortly after obtaining his M.S. in biology in May of 2014 and after accepting admission to Cornell's PhD program, Jacob found out he was going to be a father. He had intended to share a house with 5 other students but after learning of this news he convinced his landlord to

let him break his lease and he found a small studio apartment for him and his girlfriend Wendy, who made the journey with him to Ithaca. In August of 2014 he began his PhD and six months later in February of 2015, while Jacob was rotating in the lab of Dr. Andrew Grimson, Wendy gave birth to a beautiful 9.0 lb baby boy. Later that year Jacob joined the Grimson lab, where he began developing a high throughput method to screen for *trans*-factors responsible for the function of 3′ UTR *cis*-regulatory elements which he applied to identifying novel candidates involved in the miRNA pathway.

This dissertation is dedicated to Wendy Musson

ACKNOWLEDGMENTS

Wendy Musson; Thank you joining me and supporting me on this crazy journey.

My Family; I'm lumping the rest of you together, but you each know how much your support has meant to me.

Andrew Grimson; Thank you for your patience and understanding.

Mariana Wolfner; Thank you for your advice and for pushing me.

Hojoong Kwak; Thank you for your suggestions and for all of your help.

Jen Grenier; Your advice and help was invaluable.

Elizabeth Fogarty; Thank you for listening when I needed someone to listen.

Jessica West; You made the lab a brighter place.

Former and current Grimson and Pleiss lab members; Thank you for all that you did. The practice talks, the conversations, the ideas, and the laughs. I'll miss you.

Scott Ferguson; My first mentor, thank you for inspiring me.

Others I couldn't have done this without:

The Genomics Facility staff

The RNA Sequencing Core staff

The FACS Facility staff

The MBG administrative staff

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

4E-T	eIF4E-binding protein
6mer	miRNA target sites that are complementary to nucleotides 2-7 of the microRNA
7mer-A1	6mer sites that contain an A at nucleotide 1
7mer-m8	miRNA target sites that are complementary to nucleotides 2-8 of the microRNA
8mer	miRNA target sites that are complementary to nucleotides 2-8 of the microRNA and contain an A at nucleotide 1
ADARs	Adenosine deaminases acting on RNA
Ago	Argonaute
BRCA1	BRCT repeats of breast cancer, type 1
CDS	Coding sequence
DGCR8	DiGeorge Syndrome Critical Region 8
dsRBD	double stranded RNA binding domain
dsRBP	double-stranded RNA binding protein
dsRNA	double stranded RNA
EdgeR	Empirical Analysis of Digital Gene Expression Data in R
EF1α	Elongation factor 1 α
eIF	Eukaryotic initiation factor
EXP1	Exportin 1
EXP5	Exportin 5
FACS	Fluorescence activated cell sorting
GSK3β	glycogen synthase kinase 3β
Hsc70	Heat shock cognate 70
Hsp90	Heat shock protein 90
IRES	Internal ribosomal entry site
KRAB	Krueppel-associated box

m7G	7-methylguanosine
miRISC	miRNA-induced silencing complex
miRNA*	microRNA passenger strand
miRNAs	microRNAs
NMD	Nonsense mediated decay
nt	Nucleotide
PABP	poly(A) binding protein
PAM2	PABP-interacting motif 2
PARN	Poly(A)-specific ribonuclease
P-body	Processing body
PIC	Pre-initiation complex
Pol II	RNA polymerase II
Pol III	RNA polymerase III
pre-miRNA	precursor miRNA
pri-miRNA	primary-miRNA
RBPs	RNA binding proteins
RIGER	RNAi Gene Enrichment Ranking algorithm
RLU	Relative Light Unit
RNAi	RNA interference
RNase III	Ribonuclease III
shRNA	short hairpin RNA
siRNA	small interfering RNA
snoRNA	small nucleolar RNAs
TNRC6	Tri-Nucleotide Repeat Containing 6
TRBP	TAR RNA-binding protein
TRC	The RNAi Consortium

TUT2	Terminal uridylyl transferase 2
TUT4	Terminal uridylyl transferase 4
TUT7	Terminal uridylyl transferase 7
UTR	Untranslated region

Chapter 1 : Introduction

Gene Regulation

The central dogma of molecular biology describes the flow of genetic information as follows; DNA is transcribed into RNA and RNA is translated into protein. Although most genes are protein coding, some genes produce only an RNA product (e.g., ribosomal RNA and tRNA genes). Originally the term “gene” was used only to describe an abstract unit of inheritance but has evolved over time and alternate definitions have been proposed (Portin and Wilkins, 2017). For the purpose of this thesis I am defining the gene as a hereditary unit of DNA encoding a biologically functional RNA or protein.

Proper regulation of gene expression is critical for cell differentiation, development, maintaining homeostasis and responding to environmental stresses (Nadal et al., 2011; Weake and Workman, 2010). Multicellular organisms are made of many cells with specialized functions, and although each cell contains identical genetic information, regulated expression of these genes allows distinct cell types to exist and perform unique biological roles (e.g., neurons, muscle, fat, sperm, etc.). Multiple mechanisms exist within the cell to regulate when and where gene expression occurs. Transcriptional regulation is considered to have the greatest impact on gene expression (Vihervaara et al., 2018), and post-transcriptional regulation of the mRNA adds an important secondary layer of control over protein output (Corbett, 2018). However even once translated, a protein may still be subject to regulation (Knorre et al., 2009).

Post-transcriptional Regulation

Mechanisms of post-transcriptional regulation play a major role in determining the amount of protein produced by affecting the rate of mRNA decay and/or translation efficiency. A typical mRNA contains a coding region that is flanked on either side by a 5′ untranslated region (UTR) and a 3′ UTR, respectively. Post-transcriptional regulation is encoded within *cis*-regulatory elements found within 3′ UTR of the mRNA which averages ~1300 nucleotides in humans (Zhao et al., 2011) and is often highly conserved across species, implying important biological functions (Pollard et al., 2010; Siepel et al., 2005). These elements are recognized by *trans*-acting factors such as RNA binding proteins (RBPs) or microRNAs (miRNAs), whose activity can alter protein expression. RBPs can accomplish this through several means including regulation of mRNA splicing, stability, localization, translation, and decay (Glisovic et al., 2008) while miRNAs alter expression by recruiting factors that promote accelerated mRNA decay and translational repression (Iwakawa and Tomari, 2015).

microRNAs: Discovery and Conservation

Since their discovery in 1993, microRNAs (miRNAs) have transformed the field of molecular biology and have become recognized an important class of gene regulators. miRNAs are a class of small (~22-nt) noncoding RNAs and in humans it is estimated that >60% of protein coding genes are targeted by ≥1 miRNA (Lewis et al., 2005). Dysregulation of the miRNA pathway is often associated with developmental disorders and there is continually mounting evidence that miRNAs also play a large role in the etiology and pathogenesis of cancer and disease (Ardekani and Naeini, 2010; Jiang et al., 2009). Since the initial discovery of miRNAs over 25 years ago, many of the genes required for miRNA biogenesis and effector function have been discovered, and to date, over 1,800 human miRNA entries have been

catalogued in miRbase (http://mirbase.org/), an online repository for all potential miRNA sequences, annotation, nomenclature, and target prediction information (Friedman et al., 2009). However, our understanding of the miRNA pathway is still incomplete and there is reason to believe that additional factors are involved, which I discuss in detail below (page 16).

The first miRNA, *lin-4*, was discovered in *C. elegans* (Lee et al., 1993). Its expression was discovered to be necessary for the downregulation of another protein, LIN-14, but interestingly, *lin-4* itself did not encode a protein. Instead, *lin-4* expression produced 2 small RNAs, 21 and 61-nts in length. The larger transcript represented a precursor that formed a double stranded RNA (dsRNA) stem-loop structure, which was cleaved to release the smaller 21-nt RNA responsible for repression of *lin-14*. The authors observed that the lin-14 mRNA contained several sites within its 3′ UTR that were partially complementarity to the 21-nt lin-4 sequence, and that lin-4 binding of these sites resulted in translational repression of lin-14.

Originally miRNA mediated repression was thought to be a phenomenon exclusive to *C. elegans* and miRNAs were not recognized as a distinct class of biological regulators until 2000, when 2 separate groups discovered another small non-coding RNA essential to *C. elegans* development, *let-7* (Pasquinelli et al., 2000; Reinhart et al., 2000). After discovery of *let-7* it was soon realized that this miRNA had homologues in many other organisms, including humans (Pasquinelli et al., 2000). In the period that followed many more miRNAs were discovered, many of which were evolutionarily conserved. Today miRNAs are recognized as pervasive and essential regulators, which are involved in all major cellular pathways.

It is likely that the miRNA pathway evolved from the ancestral silencing pathway known as the RNA interference (RNAi) pathway which appears to have been present in the last common ancestor of eukaryotes (Fire et al., 1998; Shabalina and Koonin, 2008). The RNAi pathway requires the activity of two proteins shared by the miRNA pathway, Dicer and

Argonaute (Ago). A key difference between the RNAi pathway and the miRNA pathway is the source of the guide RNA which directs Ago to its target. The RNAi pathway utilizes small interfering RNAs (siRNAs) which are often derived from exogenous viral and transposon dsRNAs cleaved by Dicer and loaded onto Ago, while miRNAs are cleaved from endogenously expressed hairpin structures and loaded onto Ago (Carthew and Sontheimer, 2009).

Although components of the miRNA pathway machinery such as Dicer and Ago are conserved in plants and animals, important differences exist. The miRNAs themselves appear to have evolved independently, and each has a different primary mode of action. Plant miRNAs for example, target mRNAs with near-perfect base pairing (~22 nts) which induces cleavage of the transcript (Rhoades et al., 2002). Animal miRNAs on the other hand target mRNAs containing smaller sites complementary to just the first 6-8 nts of the miRNA (Brennecke et al., 2005; Lewis et al., 2005) which does not result in cleavage. In animals, miRNA mediated repression occurs primarily through recruitment of additional proteins that promote accelerated mRNA decay and translational repression versus slicing and degradation which occurs in plants (Axtell et al., 2011). However, cases of miRNA-directed slicing have also been reported in mammals (Davis et al., 2005; Park and Shin, 2014; Yekta, 2004). For the purposes of this dissertation I will be focusing on the mammalian miRNA pathway.

microRNAs: Biological Importance

Although an individual miRNA can regulate hundreds of potential target genes, their importance as regulators was not fully appreciated at first. miRNAs usually have only a modest effect on mRNA target expression, typically reducing mRNA levels by less than 50% (Baek et al., 2008; Selbach et al., 2008) and the generation and analysis of miRNA knockouts in *C. elegans* suggested that most individual microRNAs were not essential for viability or

development (Alvarez-Saavedra and Horvitz, 2010; Miska et al., 2007). However, the idea that miRNAs have either largely redundant or insignificant biological functions does not apply to flies or mammals. In *Drosophila* an analysis of 95 miRNA knockout lines deleting 130 miRNAs revealed at least one phenotype in 80% of the mutants (Chen et al., 2014), and in mice 56 of the 90 mammalian miRNA families conserved to fish have been observed to have a knockout phenotype as well (Bartel, 2018). These loss of function studies revealed that disruption of miRNA genes can result in diverse developmental and physiological defects, many of which affected embryonic and postnatal viability (Azevedo-Pouly et al., 2017; Bartel, 2018; Labialle et al., 2014; Medeiros et al., 2011; Watanabe et al., 2008). While some miRNAs are required for proper development of various organs or tissues, others influence important cellular functions and physiological processes.

It is now clear that many miRNAs play an important role in gene regulation required for normal cell function and development, so it is not surprising that disruption of miRNA expression is associated with a variety of human diseases. Although rare, mutations that affect miRNA function can result in genetic disorders (Mencía et al., 2009; Meola et al., 2009; de Pontual et al., 2011) but abnormal miRNA expression has also been observed in cancer (Ardekani and Naeini, 2010; Deng et al., 2008; Visone and Croce, 2009), cardiovascular disease (van Rooij et al., 2006; Tatsuguchi et al., 2007), neurological disorders (Amin et al., 2015; Maes et al., 2009; Schratt, 2009) as well as many others (Ardekani and Naeini, 2010). Understanding miRNA involvement in human disease allows for the development of potential treatments (Chakraborty et al., 2017; Christopher et al., 2016), and as a biomarker miRNAs can also potentially be used as a diagnostic tool to identify various diseases and monitor their progression (Faruq and Vecchione, 2015; Martinez and Peplow, 2019; Mayeux, 2004).

The Canonical miRNA Pathway

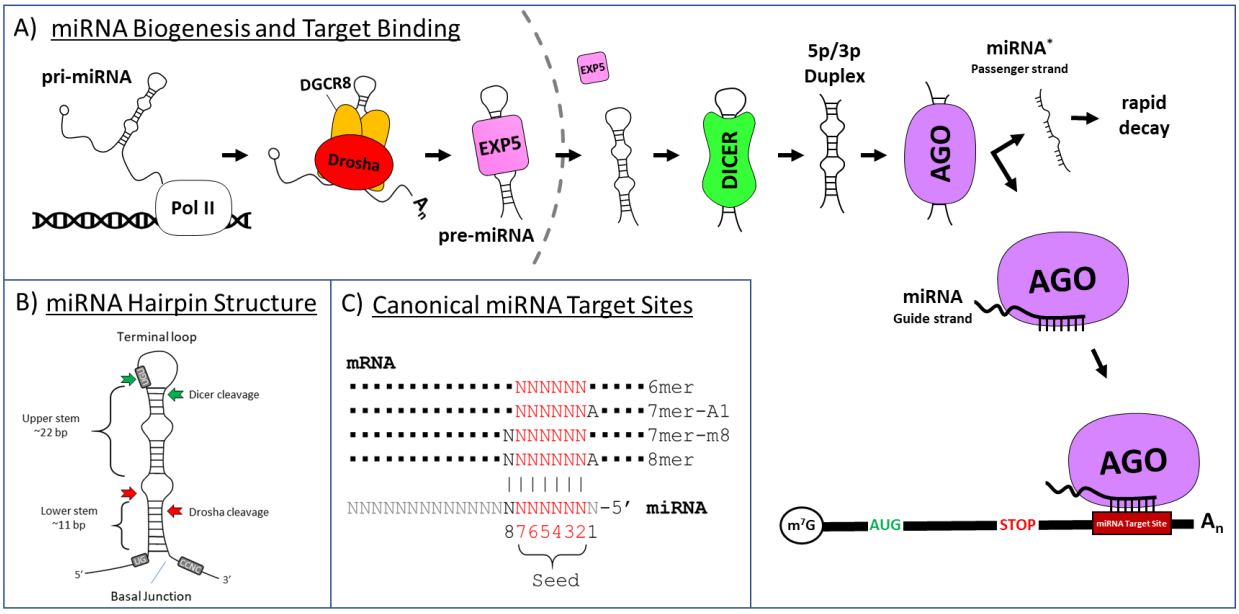

Figure 1-1: The Core Components of the miRNA Pathway

A) The primary miRNA transcript (pri-miRNA) is transcribed by RNA polymerase II. The pri-miRNA contains a hairpin structure that is recognized and bound in the nucleus by DGCR8 and Drosha whose combined action allows for Drosha to cleave the hairpin releasing the precursor miRNA (pre-miRNA). The pre-miRNA is exported from the nucleus to the cytoplasm by EXP5 where it is bound by Dicer which cleaves the terminal loop, generating a small 5p/3p RNA duplex which is loaded onto Ago. The passenger strand (miRNA*) is released and rapidly degraded, while the mature miRNA guides Ago to mRNAs containing cognate binding sites. **B)** A diagram of the functional hairpin domains and shows the location where Drosha cleavage occurs a 11bp from the basal junction (Red) and the sites of Dicer cleavage (Green). These cleavage events generate a 2 nt 3′ overhang on both sides of the duplex. Hairpins may also possess additional sequence features that can enhance processing (Grey). Figure adapted from (Bartel, 2018). **C)** Canonical miRNA target sites are complementary to the miRNA seed sequence, nucleotides 2-7. Sites that only complement the seed are known as 6mers. When the site contains an A at nucleotide 1 (7mer-A1), a complementary base pair at nucleotide 8 (7mer-

m8) or both (8mer), site efficiency is increased. Figure adapted from (Bartel, 2018).

Transcription, Nuclear Processing and Export

Biogenesis of a mature miRNA is tightly controlled requiring processing steps in both the nucleus and cytoplasm before it can be loaded onto one of the four human Argonaute (Ago1-4) proteins. Canonical miRNAs are transcribed by RNA polymerase II (Pol II) as part of a much longer capped and polyadenylated transcript (typically over 1kb), called the primary-miRNA (pri-miRNA) (Cai et al., 2004; Lee et al., 2004). The pri-miRNA contains the mature miRNA sequence within a stem-loop structure of about ~70 nts (Figure 1-1B). Some miRNAs are clustered together and are transcribed as part of a polycistronic transcript which can contain up to 6 miRNA hairpins flanked on each side by sequences allowing efficient processing (Olena and Patton, 2010). About half of all miRNA genes were found to be intragenic and most of these are found within introns (Rodriguez et al., 2004). The rest are intergenic and transcribed independently under regulation of their own promoters.

Following transcription, the recognition and cleavage of the ~70-nt hairpin structure within the pri-miRNA occurs in nucleus by the microprocessor complex, composed of one Drosha and two DiGeorge Syndrome Critical Region 8 (DGCR8) proteins. Drosha is a nuclear ribonuclease III (RNase III) enzyme that cleaves the pri-miRNA (Lee et al., 2003) and DGCR8 is a double-stranded RNA binding protein (dsRBP) which is an essential Drosha cofactor (Han et al., 2004). It is important that the Microprocessor precisely recognize and cleave the pri-miRNA. Drosha cleavage defines the pre-miRNA terminus which determines the miRNA “seed” sequence (nucleotides 2-7) responsible for specifying the target mRNAs (discussed in more detail below, page 10). Although Drosha contains a double stranded RNA binding domain (dsRBD) it is not enough for substrate interaction. Drosha is stabilized by its interaction with

DGCR8 which allows it to recognize and bind the basal ssRNA-dsRNA junction of the pri-miRNA which serves as an important reference point for specifying where cleavage occurs (Han et al., 2006). As a dimer, DGCR8 binds to the stem region of the pri-miRNA and recognizes conserved UGU motifs required for efficient and accurate cleavage (Nguyen et al., 2015; Quick-Cleveland et al., 2014). The length from the dsRNA basal junction to the cleavage site is restricted to 11 bp via a specialized helix of Drosha called the Bump helix, which contacts the basal end of the hairpin (Kwon et al., 2016). Drosha cleavage at the base of the stem-loop generates a 2-nt 3′ overhang and releases a ~65-nt hairpin, referred to as the precursor miRNA (pre-miRNA). Processing of intronic miRNAs does not affect splicing (Kim and Kim, 2007), however, when a hairpin is located within an exonic region, this can lead to destabilization of the host mRNA (Han et al., 2009).

Following Drosha processing, the pre-miRNA is recognized and exported from the nucleus to the cytoplasm in an energy dependent manner by Exportin 5 (EXP5) and Ran-GTP (Yi et al., 2003). EXP5 binds the pre-miRNA stem requiring a dsRNA region of >16 bp and a 3′ overhang which facilitates interaction, while a 5′ overhang is inhibitory (Okada et al., 2009; Zeng and Cullen, 2004). Following export, GTP is hydrolyzed and the pre-miRNA is released into the cytoplasm.

Cytoplasmic Processing and Formation of the miRNA-Induced Silencing Complex

Upon translocation to the cytoplasm the pre-miRNA is recognized via its 2-nt 3′ overhang by the RNase III endonuclease Dicer. Dicer binds the pre-miRNA and cleaves the miRNA stem near the terminal loop a specific distance (~22-nts) from the 5′ end generated by Drosha (Park et al., 2011). Like Drosha, Dicer cleavage also produces a 2-nt 3′ overhang resulting in ~22-nt

5p/3p RNA duplex. The 5p and 3p strands are named for the 5′ and 3′ arms of the pre-miRNA from which they are derived.

The 5p/3p RNA duplex is loaded onto one of the four human Argonaute (Ago1-4) proteins. All four human Ago proteins are capable of initiating miRNA mediated repression but Ago2 is unique as the only Ago to possess endonuclease ('slicer') activity (Liu et al., 2004; Meister et al., 2004). For this reason, Ago2 is the only Ago to function in both the RNAi and miRNA pathway. Ago loading of the small RNA duplex requires a conformational opening mediated by the heat shock cognate 70 (Hsc70)-heat shock protein 90 (Hsp90) chaperone complex (Iwasaki et al., 2010). Once loaded the RNA duplex is unwound in an ATP-independent manner believed to be driven by releasing structural tension imparted on Ago during the conformational opening by Hsc70-Hsp90 (Kawamata and Tomari, 2010). The guide strand is believed to be determined, in part, by the thermodynamic stability of the 5′ terminus (Khvorova et al., 2003). Although either strand can potentially function as the guide strand, there is a strong bias favoring the strand with a less stably base paired 5′ end to be retained while the miRNA* strand is released and degraded. Together the miRNA guide strand and Ago form the core miRNA-induced silencing complex (miRISC). It is important to note that strand selection is still not fully understood. The thermodynamic stability of the 5′ terminus does not explain all of strand selection as some miRNAs behave differently and for others strand preference has been observed to change in different tissues (Choo et al., 2014; Griffiths-Jones et al., 2011; Kuchenbauer et al., 2011; Ro et al., 2007). It has also been observed that Ago has a preference for guide strands with a U at the first nucleotide position (Hu et al., 2009). Therefore it is possible that additional criteria or factors are involved in strand selection.

Target Binding

After formation of the miRISC, the mature miRNA functions to guide Ago to complementary target sites which are primarily located in the 3′ untranslated region (UTR) of target mRNAs, but may potentially function in the 5′ UTR and coding sequence (CDS) as well (Lytle et al., 2007; Xu et al., 2014). Target recognition is accomplished through Watson-Crick pairing between the miRNA “seed” (nucleotides 2-7) sequence and complementary target sites within the 3′ UTR of target mRNAs (Brennecke et al., 2005; Lewis et al., 2003, 2005). Canonical target sites have 6-7 contiguous Watson-Crick base pairs (Figure 1-1C). Sites complementary to nucleotides 2-7 are referred to as 6mers and the efficacy of the site increases when there is an A across from nucleotide 1 (7mer-A1), pairing of nucleotide 8 (7mer-m8), or both (8mer).

The name seed was given to the nucleotides that define target specificity because it was believed that following the initial pairing of the seed sequence to the mRNA target, supplementary 3′ base pairing would then grow from the seed sequence in a 5′ to 3′ direction (Bartel, 2004), but structural data suggest this model is unlikely due to topological constraints that would require difficult conformational changes for the guide strand to wrap around the target mRNA (Schirle and MacRae, 2012). A more recent model proposed by Bartel, suggests pairing skips from the seed region to a second region closer to the 3′ end of the miRNA (Bartel, 2018). Structural analysis revealed that it is only nucleotides 2-5 that are accessible for the target search (Schirle and MacRae, 2012) and when the miRISC complex binds a canonical site, pairing of nucleotides 2-5 propagate to nucleotides 6-8 causing Ago to undergo a conformational change that reinforces perfect pairing of nucleotides 6 and 7 (Chandradoss et al., 2015; Klum et al., 2018; Schirle et al., 2014). The conformational change prevents pairing beyond position 8 but for many target sites nucleotides 13-16 of the miRNA can form a second RNA helix with the mRNA target (Grimson et al., 2007). At this point the additional proteins required for accelerated decay and translational repression are recruited to Ago.

miRNA Silencing

In mammals, miRNA mediated repression is accomplished primarily through accelerated mRNA decay, although most miRNAs also repress translation to some degree (Baek et al., 2008a; Eichhorn et al., 2014; Guo et al., 2010; Hendrickson et al., 2009). Many of the factors recruited to the miRISC are implicated in both processes which may be initiated simultaneously or independently. It is difficult to separate the two processes because many factors are involved in both.

After the miRISC successfully binds the target mRNA, Ago recruits Tri-Nucleotide Repeat Containing 6 protein (TNRC6, also known as GW182 in *Drosophila*), a key player in miRNA mediated repression which mediates all further downstream processes (Eulalio et al., 2009). TNRC6 acts as a hub, bridging Ago to downstream decapping, deadenylation, and degradation complexes. In humans there are three paralogs of this gene (TNRC6A-C) which share a similar structure and are believed to act redundantly. Each contains an unstructured N-terminal Ago-binding region (Yao et al., 2011), a poly(A) binding protein (PABP)-interacting motif 2 (PAM2) region (Fabian et al., 2009) and a C-terminal silencing domain which serves as a binding platform for PAN3 (Christie et al., 2013) and NOT1 (Chekulaeva et al., 2011) subunits of the PAN2/PAN3 and CCR4-NOT deadenylase complexes, respectively.

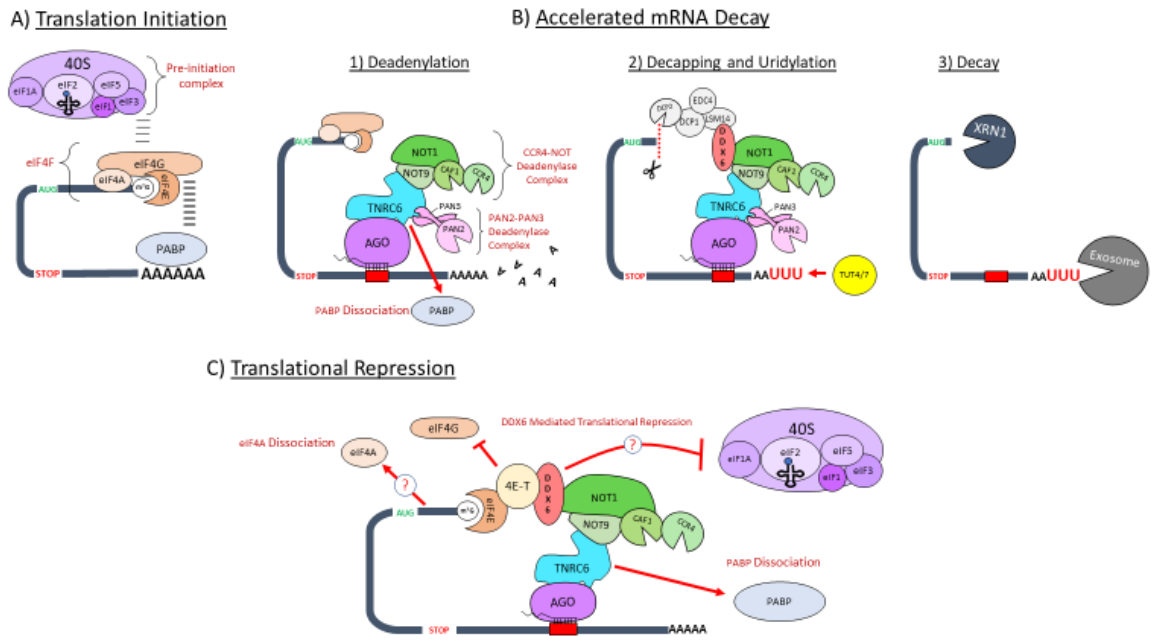

Figure 1-2: Accelerated Decay and Translational Repression

A) Translation initiation requires recognition of the 5′ 7-methylguanosine (m7G) cap by the eIF4F cap binding complex, which is comprised of eIF4G, eIF4A, and eIF4E. Once assembled, the preinitiation complex consists of the 40S ribosome subunit and several eukaryotic initiation factors. **B)** Accelerated mRNA decay proceeds in 3 steps: 1) TNRC6 promotes deadenylation by promoting PABP dissociation and recruiting the PAN2/PAN3 and CCR4-NOT deadenylase complexes. 2) CCR4-NOT recruitment of DDX6 results in recruitment of decapping proteins. The deadenylated mRNA can also be uridylated by TUT4/7 which marks the transcript for decay. 3) General decay occurs in a 3′ to 5′ direction mediated by the exosome, while XRN1 degrades the decapped transcript in a 5′ to 3′ direction. **C)** PABP dissociation, DDX6 mediated translational repression through its association with 4E-T and another unknown mechanism, and eIF4A dissociation through an unknown mechanism are all implicated in translational repression. Each is thought to prevent translation initiation.

Accelerated mRNA Decay

The first step of accelerated mRNA decay mediated by a miRNA is deadenylation by the PAN2/PAN3 and CCR4-NOT deadenylase complexes. TNRC6 dissociation of PABP from the poly(A) tail increases the accessibility of the poly(A) tail for deadenylation (Fabian et al., 2009). CCR4-NOT is composed of 10 evolutionarily conserved subunits with CNOT1 acting as a scaffold which binds the remaining subunits (Collart and Panasenko, 2012; Petit et al., 2012). CAF1 was discovered to be essential for the removal of the poly(A) tail (Fabian et al., 2011), and CNOT9 mediates CNOT1 binding to TNRC6 via tryptophan motifs (Chen et al., 2014a). Overexpression of dominant negative form of CAF1 severely inhibited deadenylation and mRNA decay (Behm-Ansmant et al., 2006; Chen et al., 2009; Zheng et al., 2008), whereas PAN3 depletion or overexpression of catalytically inactive PAN2 only had a modest effect, suggesting CCR4-NOT is the primary deadenylase complex (Chen et al., 2009).

Following deadenylation mRNAs are subjected to oligouridylation by terminal uridylyl transferases TUT4 and TUT7 which marks the mRNA for general 3′-5′ mRNA decay by the exosome (Lim et al., 2014). Although deadenylation precedes mRNA decay it is not required for decay. Transcripts lacking poly(A) tails or containing an internalized poly(A) sequence can also undergo mRNA decay (Makino et al., 2015; Nishihara et al., 2013). The CCR4-NOT complex also recruits decapping activator Dead-box RNA helicase/ATPase DDX6 which promotes removal of the 5′ 7-methylguanosine (m7G) cap via decapping enzyme DCP2 (Rouya et al., 2014). Once removed, 5′-3′ exonucleolytic decay by the XRN1 exonuclease can proceed.

Translational Repression by miRNAs

Review of Cap-Dependent Translation Initiation

In contrast to accelerated mRNA decay, the precise mechanism of translation repression remains elusive. In eukaryotes translation can be broken into 4 steps; initiation, elongation, termination and recycling. The rate limiting step is the initiation step which requires a series of reactions that result in an assembled elongation competent 80S ribosome at the AUG start codon of the mRNA. Multiple studies suggest that it is inhibition of the initiation step that is responsible for miRNA mediated translational repression (discussed below, page 15).

Canonical translation occurs in a cap-dependent manner, where the 5′ m7G cap is recognized and bound by the cap-binding complex, eIF4F, a ternary complex composed of eIF4A, eIF4E, and eIF4G (Jackson et al., 2010; Sonenberg and Hinnebusch, 2009). eIF4G is the central scaffolding protein which interacts with the eIF4E cap-binding protein and eIF4A ATP-dependent RNA helicase. eIF4G interaction with PABP associated with the 3′ poly(A) tail allows for mRNA circularization and stabilizes the eIF4F-mRNA interaction which enhances ribosome recycling and recruitment respectively (Uchida et al., 2002). Once the eIF4F complex is associated with the 5′ cap, eIF4G recruits the pre-initiation complex (PIC), comprised of the 40S small ribosomal subunit and several additional factors (eIF1, eIF1A, eIF2-GTP initiator tRNA, eIF3, and eIF5), through its interaction with eIF3 (Algire et al., 2002; Asano et al., 2001; KOLUPAEVA et al., 2005; Majumdar et al., 2003). Before the PIC can be recruited to the 5′ cap, eIF4A must unwind nearby secondary structures in the 5′ UTR. Once bound the PIC scans the 5′ UTR in a 5′ to 3′ direction until recognition of the initiation codon which is followed by recruitment of the 60S large ribosomal subunit and the elongation-competent 80S ribosome is formed (Aitken and Lorsch, 2012).

Proposed Mechanisms of miRNA Mediated Translational Repression

Three major models of how miRNAs mediate translational repression have been proposed which include: 1) PABP displacement, 2) recruitment of translational repressors and 3) dissociation of eIF4A from the cap-binding complex. These mechanisms are not mutually exclusive and may function in concert to amplify the overall silencing effect.

TNRC6 is capable of recruiting PABP via its PAM2 motif. This TNRC6-PABP interaction competes with the interaction of PABP and the eIF4G scaffolding protein needed for recruitment of the PIC. Interaction with TNRC6 leads to PABP displacement and this is believed to contribute to translational repression by disrupting mRNA circularization (Derry et al., 2006), but this is not sufficient to explain all translational repression, as other data demonstrate mRNAs lacking poly(a) tails are still capable of being translationally repressed (Zekri et al., 2013). Nonetheless, the fact that the TNRC6-PABPC interaction has been maintained throughout evolution suggests that the interaction is important for silencing (Jinek et al., 2010).

TNRC6 also recruits translational repressors through its interaction with the CCR4-NOT deadenylase complex. It was observed that tethering of CAF1 and other CCR4-NOT components to mRNAs with or without poly(A) tails induced translational repression, which showed that the deadenylase complex also plays an important role in translational repression (Chekulaeva et al., 2011; Cooke et al., 2010; Mathys et al., 2014). In yeast and fly, the CCR4-NOT complex was known to interact with translational repressor Dhh1/Me31b (Coller et al., 2001), whose orthologue in human, DDX6 (RCK/p54), was known to be required for miRNA mediated repression (Chu and Rana, 2006). This lead to the discovery that CNOT1 recruits DDX6 (Chen et al., 2014), a decapping activator which binds to the eIF4E-binding protein (4E-T) (Ozgur et al., 2015). 4E-T is known to represses translation through its binding of the eIF4E cap binding protein, preventing the interaction of eIF4E with eIF4G required for ribosome assembly

(Kamenska et al., 2014). Interestingly, translational silencing occurs even in the absence of a functional eIF4E binding domain (Kamenska et al., 2014). This suggests an alternate mechanism that does not require 4E-T binding of eIF4E is used and that other factors are involved.

The most compelling evidence that miRNAs interfere with translation initiation comes from the observation that mRNAs lacking a 5′ m7G-cap and instead contain a non-physiological A-cap structure are resistant to miRNA mediated translational repression (Mathonnet et al., 2007; Wakiyama et al., 2007). The same is true for mRNAs containing an internal ribosomal entry site (IRES), which bypasses the requirement of one or more eIFs (Humphreys et al., 2005; Pillai et al., 2005; Ricci et al., 2013). These data suggest that an eIF4F component is potentially involved in translational repression. A pull-down assay of the mRNA-protein complex performed in a cell-free system demonstrated both eIF4AI and eIF4AII were released from the eIF4F cap binding complex in the presence of miRNAs (Fukao et al., 2014). This suggested that miRNAs inhibit translation through dissociation eIF4AI and eIF4AII which are required for ribosome binding and scanning of the 5′ UTR, but it is unclear what factor mediates dissociation. Further investigation is required to determine what additional factors are involved and whether the proposed mechanisms of translational repression function redundantly to mediate translational repression or can be engaged individually to adjust the level of repression.

Potential for Novel Genes Involved in the miRNA Pathway

Although the core components of the miRNA pathway (DGCR8, Drosha, Dicer, Ago and TNRC6) are known, the full complement of factors involved in regulating miRNA mediated repression likely are not. Numerous auxiliary proteins have already been identified as regulators of miRNA mediated repression and more are expected to exist (Ha and Kim, 2014; Treiber et

al., 2017a). Broadly speaking there are 3 potential classes of novel factors that may play a role in the miRNA pathway: 1) regulators of core components, 2) miRNA specific regulators and 3) components of incompletely understood mechanisms.

Regulation of Core Components

miRNA mediated repression requires the coordinated activity of multiple protein complexes that are involved in biogenesis and effector functions to successfully mediate silencing. Therefore, each of these factors represent a potential target for regulating the miRNA pathway. One way this can be accomplished is through posttranslational modifications of the "core components" which can alter localization, stability, and processing efficiency which in turn alters the level of miRNA mediated repression experienced by target genes. For example, it was discovered that glycogen synthase kinase 3β (GSK3β) phosphorylates Drosha which allows for localization to the nucleus where processing of the pri-miRNA occurs (Tang et al., 2010, 2011). Without this posttranslational modification nuclear processing and consequently miRNA biogenesis and silencing would be negatively affected. In other cases, the association of an accessory protein alone is enough to alter the activity of a core component. In the cytoplasm processing of the pre-miRNA by Dicer can be regulated through binding of dsRBD cofactor TAR RNA-binding protein (TRBP). Although it is not essential for Dicer processing of the pre-miRNA, TRBP association with Dicer affects processing efficiency and can alter the length of some miRNAs, which can affect target specificity (Lee et al., 2013).

Regulation of the core components is not limited to post translational modification or direct interaction with cofactors. Post-transcriptional regulation of the core components has also been observed. The Dicer mRNA was discovered to be the target of the let-7 miRNA. This is a particularly interesting example because Dicer is required to produce let-7 and this creates a

negative feedback loop to control Dicer expression levels which has a broad impact on miRNA biogenesis. Another interesting example of autoregulation occurs between Drosha and DGCR8. DGCR8 interaction with Drosha stabilizes the Drosha protein, whereas Drosha destabilizes DGCR8 mRNA by cleaving a hairpin present in the second exon (Han et al., 2009). Cleavage of an exon by Drosha also demonstrates a mechanism separate from the miRNA pathway in which Drosha and DGCR8 can participate in gene regulation.

Each of the examples listed above alter miRNA biogenesis, but factors that regulate the activity of effector proteins, whose involvement is required after generation of the mature miRNA, have also been identified. A recent CRISPR screen identified CSNK1A1 kinase and ANKRD52-PP6C phosphatase complex as factors that promote continual transient phosphorylation of Ago (Golden et al., 2017). Phosphorylation of Ago by CSNK1A1 kinase impairs target association but despite this, it was observed that the phosphorylation-dephosphorylation cycle is required to maintain efficient miRNA mediated repression. The model proposed to explain this observation is that the Ago-target interaction persists longer than is necessary to promote silencing and that phosphorylation by CSNK1A1 kinase provides a mechanism of active disassembly. The subsequent phosphatase activity by ANKRD52-PP6C then allows Ago to efficiently bind to new targets. This offers an explanation for how the combined activity of a phosphatase and kinase work together to increase the effective amount of Ago available for silencing (Golden et al., 2017).

All factors involved in biogenesis and effector functions are potential targets for regulation by multiple means. The examples listed above demonstrate several ways regulation of the known core components may be accomplished. However, the activity of the accessory proteins may also be targets of regulation. For instance, phosphorylation of Dicer cofactor TRBP has been shown to upregulate Dicer processing of specific growth-promoting miRNAs, although the mechanism for the underlying specificity remains unknown (Paroo et al., 2009).

miRNA Specific Regulation

Some miRNAs are regulated through direct interactions that are miRNA specific. Many pri-miRNAs and pre-miRNAs contain *cis*-acting sequence elements that function as binding sites for proteins that can facilitate or inhibit processing. One of the best studied examples of this is LIN28 repression of let-7. LIN28 binds the terminal loop of both pri-let-7 and pre-let-7 miRNAs preventing Drosha and Dicer processing, respectively (Newman et al., 2008; Rybak et al., 2008). LIN28 also recruits terminal uridylyl transferases TUT-4 and TUT-7, which induces oligouridylation and miRNA decay by DIS3L2, a 3′-5′ exonuclease (Chang et al., 2013; Heo et al., 2008). It is therefore possible that there are additional miRNAs containing elements which are recognized by RNA binding proteins that prevent or enhance processing.

In addition to RNA binding proteins that prevent processing and/or induce decay, there are several types of RNA modification that can alter the RNA sequence and/or structure which affect maturation and turnover. Adenosine deaminases acting on RNA (ADARs) act primarily on dsRNA and convert adenosine to inosine, which is recognized by most biological machinery as guanosine (Bass and Weintraub, 1988; Wagner et al., 1989). This can inhibit Drosha or Dicer processing as is the case for pri-mir-142 or pre-mir-151, respectively (Kawahara et al., 2007; Yang et al., 2006). A-to-I editing can also enhance processing as well as alter target specificity (Bahn et al., 2015). Editing has also been observed to occur more frequently in different tissues (Paul et al., 2017). These observations illustrate that miRNA editing adds an important, but complex layer of regulation that is still not fully understood and may involve the activity of additional proteins.

Incompletely Understood Mechanisms

Strand Selection

When the miRNA duplex is loaded onto Ago, the guide strand identity is believed to be determined largely by the thermodynamic stability of the 5′ terminus, where the strand with the less stably base-paired 5′ terminus is preferentially retained by Ago becoming the guide strand (Suzuki et al., 2015). However, this only partly explains how strand selection occurs, as many miRNAs violate this rule (Meijer et al., 2014). For some miRNAs, strand preference is even observed to change in different tissues. Currently, it is not understood how the thermodynamic stability rules are potentially being overturned in certain cellular contexts or whether an additional unknown mechanism is at play in these cases. Interestingly, miRNA association with target mRNAs has been shown to protect them from degradation and may also play a role in strand selection (Tsai et al., 2016). It is possible that when target abundance is low for targets of one strand that there is increased turnover, and the other arm is preferentially retained. It may also be the case that additional factors are involved in this process.

miRNA Mediated Translational Repression

Three major mechanisms of translational repression have been proposed but questions remain regarding the precise mechanism. TNRC6 dissociation of PABP may contribute to translational repression but there is data showing mRNAs lacking poly(A) tails can still be translationally repressed which demonstrates additional factors or mechanisms must be involved (Zekri et al., 2013). DDX6 and 4E-T have been proposed as mediators of translational repression acting downstream of TNRC6, but their mode of action remains unknown (Chen et al., 2014a; Ozgur et al., 2015). Similarly, eIF4A dissociation from the cap-binding complex

provides a mechanism that allows for inhibition of translation initiation, but the factor that promotes this dissociation is still a mystery (Fukao et al., 2014). Further investigation is required to understand exactly how these processes occur and what other factors may be involved.

miRNA Turnover

Another poorly understood aspect of miRNA biology is how miRNA turnover is regulated. After loading of the 5p/3p duplex onto Ago, the miRNA* strand is released and rapidly degraded while the mature miRNA typically has a much longer half-life (Guo et al., 2015; Marzi et al., 2016; Zhang et al., 2012). Not all miRNAs are equally stable, and the turnover rate has also been observed to be isomer-specific, where related miRNAs with identical seed sequences but a different 3′ sequence (nucleotides 9-22) experience different rates of turnover (Guo et al., 2015). This indicates that turnover is linked to the miRNA sequence, but the sequence of the mRNA target appears to also play a role in turnover by modulating miRNA stability, but it is not known how this takes place or which factors are involved (Ameres et al., 2010; Xie et al., 2012). Sequence features and motifs within the miRNA have been associated with fast turnover including a 5′ guanine or cytosine (Guo et al., 2015). Additionally, turnover rates can also depend on cellular context; several neural miRNAs have been observed to have much faster turnover rates although the nuclease involved is unknown (Krol et al., 2010).

Additional Considerations for Understanding miRNA Function

Cellular Context

Much progress has been made in understanding the molecular mechanism by which the

miRISC mediates gene silencing, but until recently it remained unclear where exactly this process occurs in the cell. The presence of all four AGO proteins, TNRC6, miRNAs and targeted mRNAs in the processing bodies (P-bodies), a distinct cytoplasmic foci containing translational repressors and mRNA decay factors ,suggested that localization to the P-body may play a role in miRNA silencing (Chan and Slack, 2006). However, it was later shown that repression was independent of P-body formation (Eulalio et al., 2007). Endogenous AGO proteins were shown to be diffusely distributed in the cytoplasm with only ~1% of the total protein localized to P-body, further suggesting the major actions of the miRNAs occur elsewhere in the cell (Jain and Parker, 2013).

More recent evidence shows the miRISC and miRNAs localize to all main cellular compartments and are also present outside the cell (Makarova et al., 2016). This raised new questions about whether miRNA mediated silencing occurs in all parts of the cell, and whether there are specialized functions associated with miRISC localization. The miRISC was found to be highly associated with polysomes, which indicates that the miRISC is binding actively translated mRNAs (Maroney et al., 2006). Initially ER-bound ribosomes were thought to preferentially translate secretory and membrane proteins, but more recent findings demonstrate that they also translate many cytoplasmic proteins (Reid and Nicchitta, 2015). In fact, almost all mRNAs encoding cytoplasmic proteins are found to localize to both the cytoplasmic and the ER associated ribosomes. It was discovered that miRNA target messages are enriched on the ER membrane along with Ago2 and corresponding miRNAs, which implicates the ER membrane as an important site of target assembly (Barman and Bhattacharyya, 2015).

An unexpected discovery was the presence of the miRISC in the nucleus (Catalanotto et al., 2016; Ohrt et al., 2008; Rüdel et al., 2008). All current evidence indicates that Ago loading of the miRNA must first occur in the cytoplasm and has not been observed to occur in nuclear extracts where known loading factors are absent (Gagnon et al., 2014). Unlike cytoplasmic

miRNAs which are a part of a high molecular weight complex containing many factors involved in silencing, nuclear miRNAs are found within low molecular weight complexes and believed to be repression incompetent (Pitchiaya et al., 2017). mir-29b is an example of a miRNA that was found to be predominantly localized to the nucleus. It was discovered that mir-29b contains a 3′ hexanucleotide localization signal, a distinctive 6-nt terminal motif, at the 3′ end of the miRNA that acts as a nuclear localization element (Hwang et al., 2007). Despite having an identical seed region and the same predicted targets, mir-29a does not localize to the nucleus and a recent study showed that only a nuclear blockade of mir-29b, not mir-29a, affected nuclear morphology during mitosis (Kriegel et al., 2018), which suggests an important biological role for mir-29b in the nucleus. Other nuclear miRNAs have been identified as well, mir-21 and let-7 have been shown to cleave perfectly complementary RNA targets via Ago2 slicer activity (Meister et al., 2004; Robb et al., 2005) and although various functions ranging from transcriptional repression/activation to alternative splicing have been proposed, the function of nuclear miRNAs is still not fully understood (Schraivogel and Meister, 2014).

The miRISC has also been detected outside the cell. Circulating miRNAs are present in human plasma in a remarkably stable form that is protected from endogenous RNase activity, but it is unclear what fraction travel within exosomes, an extracellular vesicle that is released from the cell upon fusion of an endocytic compartment with the plasma membrane. Exosomes were reported to carry miRNA biogenesis components (Melo et al., 2014), Ago2–miRNA complexes (McKenzie et al., 2016) and Ago-free miRNAs (Shurtleff et al., 2016). It was observed that miRNAs delivered to another cell were functional (Valadi et al., 2007), which led to the hypothesis that they may also contribute to intracellular signaling. A major concern with these studies is the challenge of separating the miRNA-mediated effects from the changes caused by delivery of other components. Additionally, an important mechanistic aspect that remains unclear is determining if and how Ago-free miRNAs are loaded onto Ago after delivery

if at all.

The cellular localization of miRNAs and the mechanism responsible for the distribution of the miRISC throughout the cell is still under intense investigation (Akgül and Erdoğan, 2018; Leung, 2015; O'Brien et al., 2018). The physiological relevance of miRISC localization on gene regulation is still unknown, but it raises the possibility that the activity of the miRISC may have specialized functions at distinct subcellular locations. It may also provide a mechanism for regulating certain transcripts. Further research is necessary to answer these questions.

Target Prediction

On average each miRNA has hundreds of potential mRNA targets based on its seed sequence (Friedman et al., 2009). However, identification of true targets sensitive to miRNA mediated repression remains a topic of intense research and there is evidence that some miRNAs may only function to regulate certain targets in specific biological contexts. Numerous algorithms have been developed to predict miRNA targets and the most accurate quantitative models are roughly as informative as the best high throughput *in vivo* crosslinking approaches (Agarwal et al., 2015; Bartel, 2009). The first step of the prediction algorithms is to search for 7-8nt canonical sites, however not all of these sites will be sensitive to miRNA silencing (Baek et al., 2008b; Selbach et al., 2008b). Therefore, additional information is used to improve predication accuracy. For example, conservation data can reveal which sites have been under selection to be maintained which adds value, as mRNAs with conserved target sites are more likely to be genuine targets of miRNA mediated repression (Bartel, 2009). Site context within the 3′ UTR can also be used to predict target efficacy. More effective sites tend to be accessible and near the ends of the 3′ UTR, but >15nt downstream of the stop codon in regions that are not associated with secondary structure (Bartel, 2009). Additional features that have been

observed to correlate with target sensitivity to miRNA mediated repression include the number of miRNA target sites, and the potential for 3′-supplementary pairing of the miRNA (Grimson et al., 2007). mRNA features which include the length of the 3′ UTR and ORF, as well as the presence of additional marginal sites (6mers) in either the 3′ UTR or ORF also correlate with the mRNA target susceptibility to miRNA mediated repression (Garcia et al., 2011a). Even when taking all of these considerations into account, target prediction is not perfect and continues to remain an area of intense research focus. Understanding precisely which mRNAs are genuine targets of miRNA mediated repression is critical for understanding the complex regulatory networks they play a role in.

Non-canonical Biogenesis of miRNAs

The canonical biogenesis pathway described above is not the only way mature miRNAs can be produced. Some miRNAs use an alternate mechanism, usually bypassing either Drosha or Dicer processing (Abdelfattah et al., 2014). The first non-canonical miRNA described was a mirtron, which are a type of microRNA located within the introns of an mRNA host gene and use splicing to generate a small lariat precursor that bypasses the requirement of nuclear processing (Ruby et al., 2007). After splicing the lariat is debranched and refolds into a hairpin structure which can be recognized by Dicer. Drosha-mediated processing can also be bypassed by small RNAs derived from short hairpin RNAs. pre-mir-320 is a conserved miRNA transcribed by Pol II as a precursor miRNA hairpin which contains a 5′ m7G cap (Xie et al., 2013). In addition to bypassing Drosha processing pre-mir-320 is even also unique because it is exported to the cytoplasm by exportin 1 (EXP1) rather than EXP5 (Xie et al., 2013). Other noncanonical miRNAs derived from non-coding RNAs, such as tRNAs (Babiarz et al., 2008) or small nucleolar RNAs (snoRNAs) (Ender et al., 2008), have also been observed to bypass Drosha processing.

Dicer independent biogenesis has also been observed. miR-451 for example, relies on the catalytic activity of Ago2 (Cheloufi et al., 2010). This is because Drosha cleavage of the pri-miRNA generates pre-mir-451 which has an unusually short hairpin stem (18 bp) and is too short for Dicer processing. Instead, the pre-mir-451 hairpin is loaded onto Ago2 which slices the middle of the 3′ strand via its endonuclease domain generating the Ago-cleaved-pre-mir-451 (ac-pre-mir-451), a 30 nt long intermediate species which is subsequently trimmed down to ~23 nucleotides by Poly(A)-specific ribonuclease (PARN), although this trimming is not essential for miRNA silencing.

There is another class of miRNAs that do not bypass Drosha or Dicer processing but instead require monouridylation before they can undergo Dicer processing. A typical pre-miRNA has a 2 nt 3′ overhang following Drosha processing and is classified as a group I pre-miRNA. Group II pre-miRNAs on the other hand contain a shorter 1 nt overhang following Drosha cleavage, which is due to their unusual pri-miRNA structure (Heo et al., 2012). Notably, most members of the let-7 family are group II pre-miRNAs and must be extended by 1 nucleotide through monouridylation by terminal uridylyl transferases (TUT2, TUT4 and TUT7) for efficient Dicer processing. Interestingly, TUT4 and TUT7 which are required for monouridylation to generate the mature miRNA, also promote degradation of the pre-let-7 when recruited by LIN28.

The vast majority of miRNAs follow the canonical biogenesis pathway. Very few conserved miRNAs are produced without the requirement for Drosha or Dicer processing, and the other poorly conserved miRNAs are lowly expressed, and may not be functionally relevant (Abdelfattah et al., 2014; Ha and Kim, 2014). Regardless, the existence of these alternative pathways reflects the flexibility of miRNA biogenesis and for those miRNAs that bypass Drosha or Dicer processing this may offer them the ability to avoid regulation mediated by Drosha or Dicer activity. As for group II pre-miRNAs, the extra step requiring monouridylation may add an additional layer of regulation.

Conclusions

miRNAs add an important layer to the posttranscriptional regulation of gene expression in the cell and play an important role in every major biological pathway. Although the core components have been identified, there are expected to be additional auxiliary components that regulate the miRNA pathway through their interaction with other miRNA pathway components, or the miRNAs themselves. Additionally, there are several longstanding questions, which include the precise mechanisms of miRNA strand selection, miRNA mediated translational repression, and miRNA turnover. The identification of additional factors involved in this pathway will potentially answer some of these outstanding questions and provide more clarity as to how miRNA mediated repression is regulated.

In this dissertation, I will present work from my project aimed at discovering novel components of the miRNA pathway, and the assay I developed which has broader applications which includes identifying other *trans*-factors that alter gene expression. To accomplish this, I developed a fluorescent reporter cell line sensitive to RNAi events that alter miRNA-mediated repression. I then performed a high throughput RNAi screen of a pooled lentiviral short hairpin RNA (shRNA) library containing ~55,000 shRNAs targeting ~11,000 genes. Fluorescence activated cell sorting (FACS) was used to sort cells with increased or decreased miRNA mediated repression and high throughput sequencing was used to identify gene targets with enriched hairpins compared to an unsorted background. I successfully identified enrichment of known miRNA pathway genes and established a list of potential novel candidates. Subsequent validation experiments suggest I have identified one or more genes involved in the miRNA pathway.

Chapter 2 : Identification of miRNA Pathway Genes

[1]

Abstract

MicroRNAs (miRNA) are small (~22-nt), endogenous, noncoding RNAs which bind to complementary sites found within the 3′ UTR of target mRNAs. This interaction promotes accelerated mRNA decay and/or translational repression of the transcript. To exert their regulatory function, miRNAs require the activity of additional factors. Although the core components required for miRNA biogenesis and gene silencing have been established, many secondary regulatory factors have also been identified, and more are expected to exist. To identify novel miRNA pathway genes, I performed a high throughput RNAi screen of a pooled lentiviral short hairpin RNA (shRNA) library containing ~55,000 shRNAs targeting ~11,000 human genes. I developed a fluorescent reporter cell line sensitive to RNAi events that alter miRNA-mediated repression to perform the screen. The reporter cell line contains a GFP reporter targeted by let-7 and a DsRed reporter that is not, serving as an internal control for events that alter expression unrelated to miRNA mediated repression. Fluorescence activated cell sorting (FACS) was used to isolate cells with increased or decreased miRNA mediated repression and high throughput sequencing was used to identify what gene targets were enriched compared to an unsorted background. Successful enrichment of known miRNA pathway genes validated the efficacy of this approach while the enrichment of genes not previously known to participate in miRNA silencing provided a list of novel candidates to validate. To reduce the number of false positives associated with RNAi screens, individual

[1] Though the experimental design was done by the author, Jacob Merle, several experiments were done in collaboration with Elizabeth Fogarty (Figure 2-8, Figure 2-9, and Figure 2-11). Cloning of several reporter constructs was done by Andrew Grimson and Raenna Wilson. Analysis of RNA-seq data was performed by Rebekah Miller (Figure 2-10).

shRNAs targeting the candidate gene were first validated using a FACS based assay to confirm that they altered miRNA activity. A single candidate, KRBOX4, was selected for further validation. Knockdowns of KRBOX4 by two independent shRNAs were assessed by RNA-seq and small RNA-seq experiments, which showed that a single shRNA significantly reduced let-7 expression and increased expression of let-7 targets while the other did not significantly alter either. Using an orthogonal approach, KRBOX4 knockout cell lines were generated using CRISPR and a luciferase-based assay was used to measure the effect on miRNA activity, which yielded conflicting results. Data from two of the KRBOX4 knockout cell lines relieved miRNA mediated repression while the third did not.

Introduction

There are many mechanisms that the cell uses to regulate individual gene expression, and this regulation can occur at the transcriptional, post-transcriptional, and post-translational level. microRNAs (miRNAs) are an important class of small (~22nt) noncoding RNAs that are important posttranscriptional regulators of gene expression in the cell and play an important role in every major biological pathway. Understanding how the miRNA pathway is regulated and the full complement of factors involved in miRNA mediated repression is therefore critical to understanding how miRNA targets are regulated.

Canonical miRNAs are transcribed by RNA polymerase II (Pol II) as part of a long primary transcript called the primary-miRNA (pri-miRNA). The mature miRNA is embedded within a stem-loop structure that is recognized and cleaved in the nucleus by the microprocessor complex, comprised of the Drosha ribonuclease III (RNase III) enzyme and its cofactor DGCR8 (Han et al., 2004; Lee et al., 2003a). Drosha cleavage releases a ~65nt hairpin, referred to as the precursor miRNA (pre-miRNA), which is exported to the cytoplasm by

Exportin 5 (EXP5). There it undergoes further processing by the RNase III endonuclease Dicer, which binds the pre-miRNA and cleaves it near the terminal loop generating an RNA duplex 21–25 nucleotides long. The RNA duplex is then loaded onto one of the four human Argonaute (Ago1-4) proteins where the 'guide strand' (miRNA) is retained and the 'passenger strand' (miRNA*) is released and rapidly degraded.

Once associated with Ago, the miRNA acts as a guide directing the effector activity of Ago and targeting mRNAs through incomplete base pairing of its seed sequence (nucleotides 2-7) with complementary sites located primarily within the mRNAs 3′ untranslated region (3′ UTR) (Bartel, 2004). Ago functions to regulate expression through its recruitment of additional factors, including tri-Nucleotide Repeat Containing 6 (TNRC6) protein, which is implicated in both accelerated mRNA decay and translational repression (Iwakawa and Tomari, 2015). Although most miRNAs repress translation to some degree, miRNA-mediated repression is primarily accomplished through accelerated mRNA decay. This occurs through TNRC6 recruitment of the CCR4-NOT deadenylase complex as well as decapping activators. Deadenylated mRNAs are subjected to oligouridylation which promotes enhanced general 3′-5′ mRNA decay by the exosome and the removal of the 5′ cap allows 5′-3′ exonucleolytic decay by XRN1.

Although the core components (DGCR8, Drosha, Dicer, Ago and TNRC6) have been identified, there are expected to be additional auxiliary components that regulate the miRNA pathway through their interaction with other miRNA pathway components, or the miRNAs themselves. A recent example of this type of regulation was discovered using a CRISPR screening approach which discovered a novel role for the CSNK1A1 kinase and ANKRD52-PP6C phosphatase complex in miRNA mediated repression (Golden et al., 2017). Together they promote the continual transient phosphorylation of Ago which turns out to be important for efficient target association. Similarly, factors that specifically target individual miRNAs or groups of miRNAs have been also been identified. The most notable examples of this is Lin28 which

binds the terminal loop of both pri-let-7 and pre-let-7 miRNA preventing Drosha and Dicer processing, respectively (Heo et al., 2008; Newman et al., 2008). Additionally, there are several longstanding questions that have yet to be answered, which include the precise mechanisms of miRNA strand selection, miRNA mediated translational repression, and miRNA turnover. Broadly speaking there are 3 potential classes of novel factors that may play a role in the miRNA pathway; 1) regulators of core components, 2) miRNA specific regulators and 3) components of incompletely understood mechanisms.

The identification of additional factors involved in the miRNA pathway will potentially answer some of the outstanding questions that remain regarding miRNA mediated repression and provide more clarity as to how the miRNA pathway is regulated to control the expression of miRNA targets. Here I describe a high throughput RNAi screen of a short hairpin RNA (shRNA) library containing ~55,000 shRNAs targeting ~11,000 human genes to identify novel miRNA pathway components. I developed a fluorescent reporter cell line sensitive to RNAi events that alter miRNA-mediated repression to perform this screen and used fluorescence activated cell sorting (FACS) to sort cells experiencing increased or decreased levels of miRNA mediated repression. I successfully identified enrichment of known miRNA pathway genes as well as potential novel candidates. Identification of novel factors involved in miRNA mediated repression will help elucidate the multiple levels of regulation that exist in the miRNA pathway and potentially answer other longstanding questions.

Results

Generation of miRNA sensitive reporter cell lines

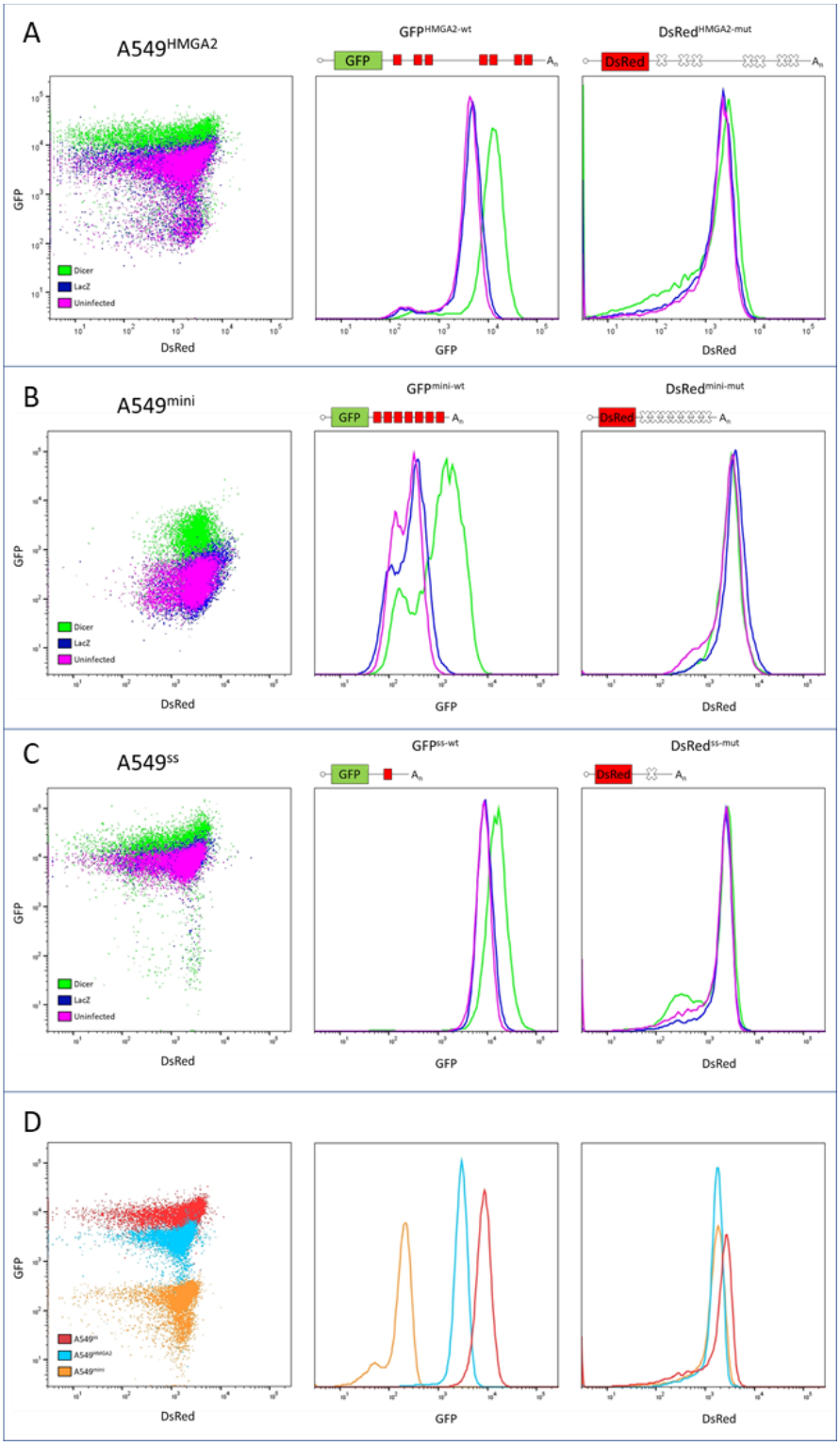

Figure 2-1: miRNA Sensitive Reporter Cell Lines

A) The A549^{HMGA2} reporter cell line expresses a GFP$^{HMGA2\text{-}wt}$ and DsRed$^{HMGA2\text{-}mut}$ reporter contain the HMGA2 3′ UTR, which possess seven let-7 target sites (red boxes), or the mutated HMGA2 3′ UTR whose let-7 seven sites have been disrupted through mutation to be nonfunctional (white crosses). **B)** The A549^{mini} contains a GFP$^{mini\text{-}wt}$ reporter with seven let-7 target sites and a DsRed$^{mini\text{-}mut}$ reporter with seven nonfunctional sites. **C)** The A549^{ss} reporter line uses a GFP$^{ss\text{-}wt}$ and DsRed$^{ss\text{-}wt}$ reporter containing a single functional and nonfunctional let-7 site, respectively. **D)** The relative GFP and DsRed expression of all 3 reporter lines is compared on a scatterplot. All have similar DsRed expression but A549^{mini} is the most repressed for GFP and A549^{ss} is the least repressed.

A549 human lung carcinoma cells obtained from American Type Culture Collection (ATCC) were selected to generate a reporter cell line sensitive to changes in miRNA activity due to RNAi *in vivo* (Giard et al., 1973). These cells were chosen for several characteristics which facilitated the genetic screen which included: expression of let-7, they are adherent cells which facilitated handling during infection and preparation for sorting, and they had the ability to survive long incubations on ice during the FACS sorting and transport required for the screen. I stably transduced A549 cells with several sets of reporter constructs to generate multiple reporter cell lines that were sensitive to the elevated and repressed levels of miRNA mediated repression.

Each reporter cell line contains a GFP reporter with a 3′ UTR containing one or more let-7 miRNA target sites, and a DsRed reporter which contains an otherwise identical 3′ UTR except that the let-7 miRNA target site(s) have been disrupted through mutation to be non-functional. The mutant sites contain a 2-nucleotide mismatch within the seed sequence which precludes let-7 binding **(Figure 2-2)** (Mayr et al., 2007). The integration of both reporters in a single cell line allows for the detection of increased or decreased levels of miRNA mediated

repression via GFP expression, while the DsRed reporter serves as an important internal control, allowing for differentiation between RNAi events that alter gene expression for reasons unrelated to miRNA biology and those that are specific to changes in miRNA mediated repression.

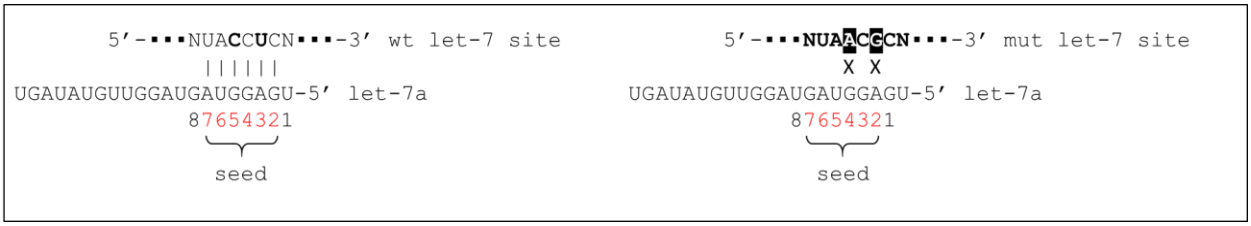

Figure 2-2: wt and mut let-7 Target Sites

The wt let-7 miRNA target sites found within the GFP reporters contain sites perfectly complementary to the nucleotide 2-7 of the miRNA (the seed sequence), while the mutant sites contain a 2-nucleotide mismatch, which prevents repression.

To generate the reporter cell lines, the reporter constructs were first cloned into a lentiviral vector and virus was generated using a 3rd generation packaging system (Dull et al., 1998) to produce a viral vector for integration of the reporter constructs (EF1α-GFP-3′ UTR or EF1α-DsRed-3′ UTR) which also contained either aminoglycoside 3′-phosphotransferase (neo) (Beck et al., 1982) or blasticidin s deaminase (bsd) (Kimura et al., 1994) antibiotic resistance genes driven by the phosphoglycerate kinase (PGK) promoter (Gerolami et al., 2000; Lizée et al., 2004). Each reporter's expression is driven by the elongation factor 1 α (EF1α) promoter, which is constitutively expressed (Kim et al., 1990; Uetsuki et al., 1989) and was selected to maintain sufficiently high expression of the reporter constructs. After transduction of A549 cells with both viral vectors carrying the wildtype and mutant reporter constructs, antibiotic selection with neomycin and blasticidin allowed for isolation of cells which had successfully integrated both reporter constructs. Using serial dilution individual cells were plated and expanded from

that single cell to generate a monoclonal cell line, and thereby reduce heterozygosity that results from the random integration of the lentivirus into the genome which may be experiencing different levels of transcriptional activity (Schröder et al., 2002).

I was able to demonstrate sensitivity to the knockdown of known components of the miRNA pathway for three separate reporter lines (**Figure 2-1A**). The A549^{HMGA2} reporter line contains the GFP$^{HGMA2\text{-}wt}$ and DsRed$^{HGMA2\text{-}mut}$ reporters which have wildtype and mutant versions of the HMGA2 3′ UTR, respectively. The HMGA2 3′ UTR was chosen because it is one of the few 3′ UTRs that has been extensively characterized and it has been shown that the seven let-7 miRNA target sites found within its 3′ UTR are the predominant *cis*-regulatory elements (Kristjánsdóttir et al., 2015). Although the HMGA2 3′ UTR does not appear to be regulated by other miRNAs, even if it was, or if it contained other *cis*-regulatory elements, this was not a major concern as knockdown of any *trans*-factor that targeted these elements should also affect the DsRed$^{HMGA2\text{-}mut}$ reporter and would therefore be distinguishable from knockdowns of interest. The HMGA2 3′ UTR was attractive because the large number of let-7 target sites in its 3′ UTR cause it to be more repressed than other endogenous targets containing fewer sites and I believed this would result in a greater fold change in reporter expression upon knockdown genes required for miRNA mediated repression which in turn would lead to a more robust enrichment of novel candidates.

The second reporter cell line A549^{mini} expresses the GFP$^{mini\text{-}wt}$ and DsRed$^{mini\text{-}mut}$ reporters which contain wildtype and mutant versions of a synthetic 3′ UTR derived from the let-7 target regions of the HMGA2 3′ UTR, respectively (**Figure 2-1B**). This 3′ UTR was chosen because it is not an endogenously expressed 3′ UTR, and less likely to contain any *cis*-regulatory elements. The lack of such elements might result in a less noisy screen and increase my ability to identify novel miRNA pathway components. It was also shown to be a much more repressive

3′ UTR than the HMGA2 3′ UTR (Figure 2.1D) which made it attractive for identifying candidates required for miRNA mediated repression.

The number of target sites and their proximity to one another can influence repression (Hon and Zhang, 2007). Therefore, the possibility also existed that the $GFP^{HMGA2\text{-}wt}$ or $GFP^{mini\text{-}wt}$ reporters would be repressed to such a degree that the knockdown of an accessory novel gene that doesn't play a central role in the miRNA pathway, or whose function is partially redundant, may not relieve repression to a detectable degree and would be missed by the screen. It was possible that a 3′ UTR with a single let-7 target site may be more sensitive. For this reason, a third reporter cell line was made. The $A549^{ss}$ cell line expresses the $GFP^{ss\text{-}wt}$ and $DsRed^{ss\text{-}mut}$ reporters which contain wildtype and mutant versions of a 3′ UTR containing a single let-7 site cloned from a 100-nucleotide region within HGMA2's 3′ UTR, respectively.

The miRNA sensitivity the wt reporter (GFP) versus the mut reporter (DsRed) was demonstrated in each cell line by robust de-repression of GFP following knockdown of Dicer while DsRed expression did not change (**Figure 2-1**). Knockdown of LacZ which served as a negative control had no effect on either GFP or DsRed in any of the reporter lines (**Figure 2-1**). The $GFP^{mini\text{-}wt}$ reporter was repressed to a greater degree than the $GFP^{HMGA2\text{-}wt}$ reporter and the $GFP^{ss\text{-}wt}$ was the least repressed (**Figure 2-1D**). Although the $A459^{mini}$ reporter showed the greatest change upon Dicer knockdown there was a sizeable portion of cells (~30%) whose GFP expression could not be differentiated from the nonfluorescent population (**Figure 2-4B**). This limited its usability for the high throughput screen of the 11,000 gene shRNA library because it would not allow for efficient sorting of the large number of cells required to perform the larger screen. However, the $A549^{mini}$ reporter was still suitable for piloting the screening approach using a smaller short hairpin RNA (shRNA) library which required less cells and I chose to use this reporter line because it was the most sensitive to the knockdown of Dicer. After validating the efficacy of the approach in the more sensitive $A549^{mini}$ reporter line, the

A549^{HMGA2} reporter could then be validated with the small shRNA library and then used to screen the larger 11,000 gene shRNA library.

Pilot Screen

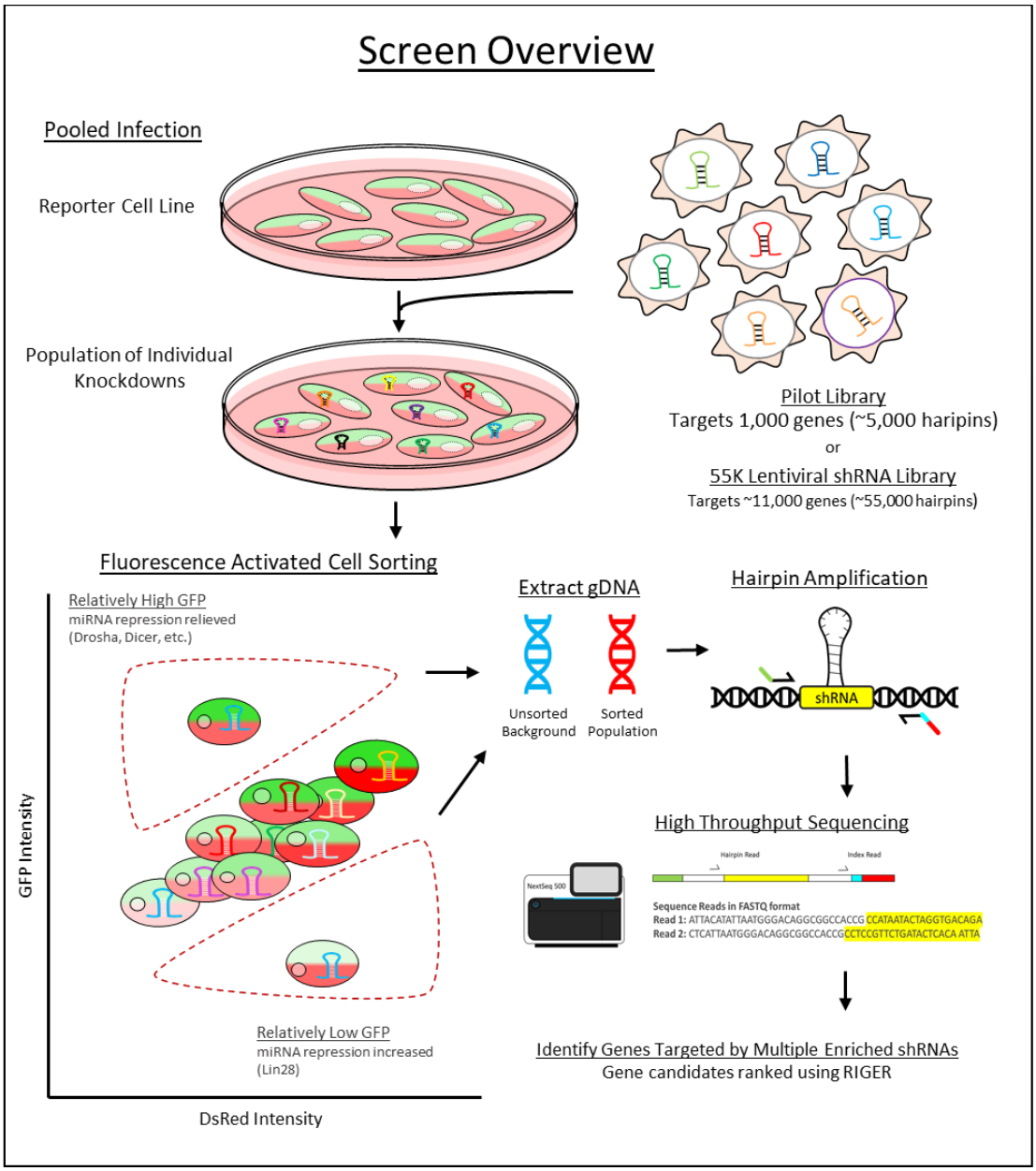

Figure 2-3: RNAi Screen Overview

The screen begins with infection of a reporter cell line that is sensitive to changes in miRNA mediated repression with a lentiviral shRNA library that targets ~11,000 genes. The cells are infected at a low MOI (~0.3) so that most cells are experiencing knockdown of a single gene. The cells are then sorted with FACS for those with low or high relative GFP expression, which indicate increased or decreased miRNA silencing, respectively. The genomic DNA (gDNA) is isolated from these sorted populations as well as an unsorted background and the hairpins are amplified by PCR and sequenced using high throughput sequencing to determine which hairpins have been enriched in the sorted population relative to the unsorted background. The gene targets of the shRNAs are identified by the hairpin sequence which is complementary to the gene and a computational approach, RIGER (Luo et al., 2008), is used to rank genes whose shRNAs are enriched in the sorted population compared to the background.

The pilot RNAi screen was performed by infecting A549^{mini} reporter cells with a small lentiviral shRNA library containing ~5,000 shRNAs targeting ~1,000 genes. Each gene is targeted on average by 5 unique hairpins increasing the probability that the knockdown will result in a phenotype. The A549^{mini} reporter cells were infected at a low multiplicity of infection (0.3), so that most cells experience knockdown of a single gene. After 7 days of growth, the cells were FACS sorted to isolate cells with the highest relative GFP$^{mini\text{-}wt}$ expression levels, as shown in **Figure 2-4A**. Gates were drawn for the top 3% and top 4-10% of cells with higher levels of GFP$^{mini\text{-}wt}$ relative to DsRed$^{mini\text{-}mut}$ which suggests they are experiencing reduced levels of miRNA mediated repression and are therefore likely to contain shRNAs targeting genes required for miRNA mediated repression (e.g. Dicer, Ago). I included a gate for the center 50% which captured half of the cells from the center of the population that did not have elevated or reduced levels of GFP$^{mini\text{-}wt}$ relative to DsRed$^{mini\text{-}mut}$ and was not expected to be enriched for hairpins targeting miRNA pathway genes. I also saved an unsorted population to be used as the background population to which the sorted populations were compared. After FACS sorting, the shRNA representation of the sorted populations and the unsorted background were determined

using high-throughput sequencing. The RNAi Gene Enrichment Ranking (RIGER) algorithm (Luo et al., 2008) was used to identify genes targeted by multiple enriched shRNAs and generate a rank ordered list of the most likely candidates.

To reduce the number of false positives due to off-target effects, which are a major problem of RNAi screens (Echeverri et al., 2006a; Mohr et al., 2014), three independent replicate screens were performed. Using this approach, I successfully detected established components of the miRNA pathway. Dicer and Ago2 were among the highest ranked gene in both the top 3% and top 4-10% populations and as expected this was not the case in the center 50% population. The enrichment of Ago2 and Dicer among the top candidates in the expected populations for each individual replicate screen validated and demonstrated the sensitivity of this assay using the A549^{mini} reporter.

The combined data from each replicate screen was used to generate the final RIGER gene score and Ago2 and Dicer were among the top ranked candidates in both the top 3% and top 4-10% populations (**Figure 2-4B**). RNAi screens like all screening approaches, are associated with false discovery. Core components of the miRNA pathway that were not among the top candidates represent false negatives which may be due to the presence of paralogs with redundant function like TNRC6A-C or may be because they are not sufficiently knocked to produce a phenotype by the shRNAs present in the library. The greater concern were false positives due to off target effects (Echeverri et al., 2006b; Sigoillot and King, 2011). For this reason, the other highly ranked genes which represent potential novel candidates involved in the miRNA pathway must be independently validated.

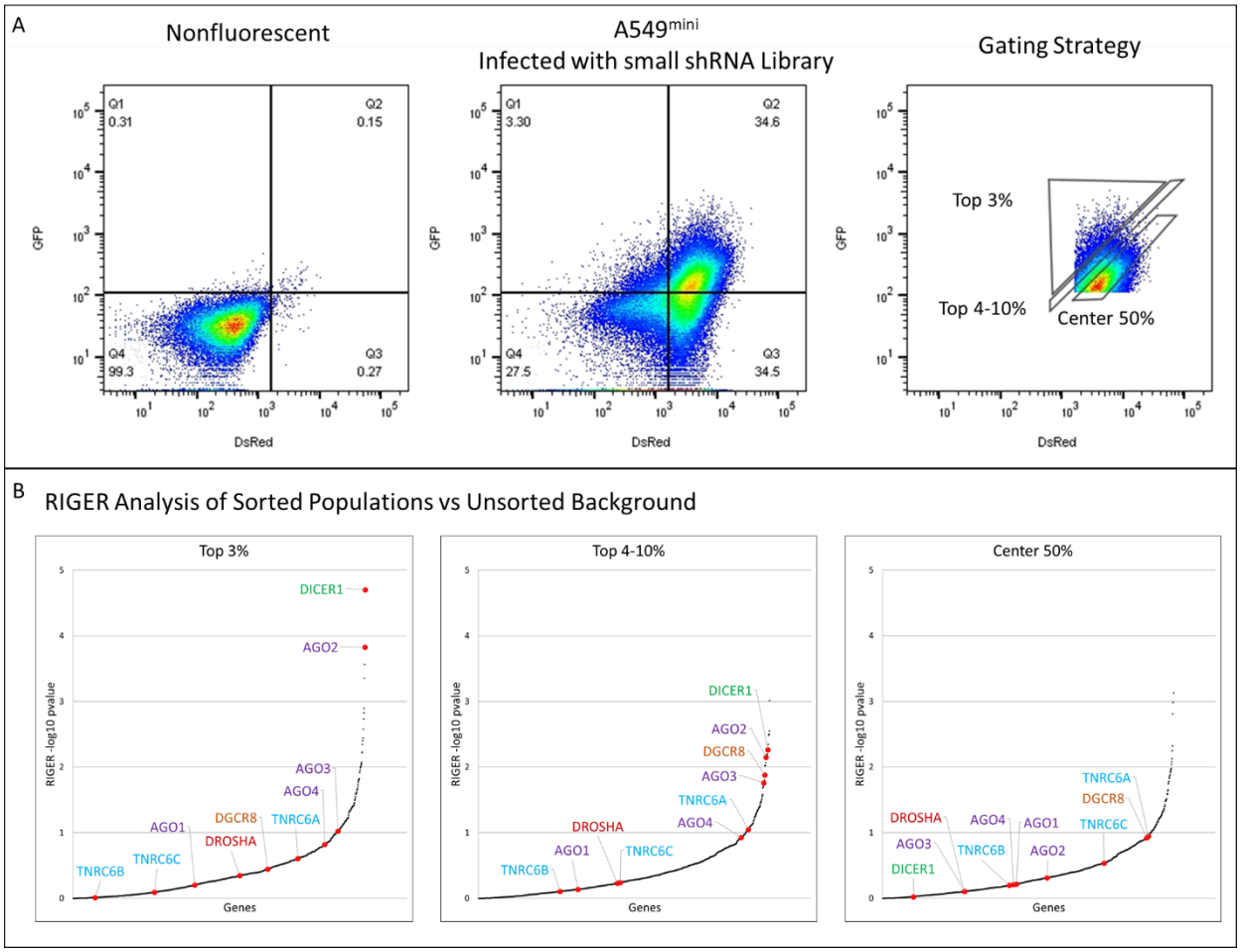

Figure 2-4: Pilot Screen of A549^{mini}

A) Gating strategy for enriching cells with high GFP expression (top 3% and top 4-10%) expected to be necessary for miRNA silencing, the center 50% population which is not expected to be enriched for any miRNA pathway components. **B)** Comparison of known miRNA pathway components and their relative enrichment (RIGER -log10 p value) in the Top3%, Top4-10% and center 50% population.

Table 2-1: Screen of the Small shRNA Library in A549^{mini}

Top 25 ranked genes in the sorted population vs unsorted background. Genes known to play a role in the miRNA pathway are highlighted in green.

Top 3% vs Unsorted

Gene	Gene rank	p-value
DICER1	1	0.00002
AGO2	2	0.000149
RPP38	3	0.000274
FAM120A	4	0.00044
PINX1	5	0.001278
SMARCAD1	6	0.001491
EIF2D	7	0.001852
ZC3H8	8	0.002646
TRMT1L	9	0.003651
PPP1R10	10	0.003973
DNA2	11	0.005906
RBMY1A1	12	0.008398
CNOT6L	13	0.004528
PRDX2	14	0.006001
INPP5A	15	0.008859
CSN3	16	0.009271
MOV10L1	17	0.009886
AICDA	18	0.0103
PRMT1	19	0.01123
SPSB1	20	0.0121
ADAD1	21	0.01525
SRPRA	22	0.0156
TERF2	23	0.009797
TARBP1	24	0.01615
RBM8A	25	0.01697

Top 4-10% vs Unsorted

Gene	Gene rank	p-value
RBMXL2	1	0.000974
DNA2	2	0.003154
RBMY1A1	3	0.00564
FAM120B	4	0.003261
PSPC1	5	0.002826
GTPBP4	6	0.004513
SECISBP2L	7	0.005072
DICER1	8	0.005483
IFIH1	9	0.006317
RPP38	10	0.006415
EIF2D	11	0.007059
AGO2	12	0.007116
PRDX2	13	0.006963
FAM120A	14	0.008612
DGCR8	15	0.01316
STAU1	16	0.009502
TRIM26	17	0.009561
METTL13	18	0.01679
MRPL44	19	0.01581
AGO3	20	0.01753
TRIM21	21	0.01934
PCMTD2	22	0.02038
OGFOD1	23	0.01728
POP7	24	0.02076
TERF2IP	25	0.00689

Center 50% vs Unsorted

Gene	Gene rank	p-value
NNMT	1	0.000742
PARN	2	0.001037
METTL9	3	0.001037
HAX1	4	0.001541
NARS	5	0.005651
PRUNE1	6	0.004773
TRNAU1AP	7	0.007491
ZGPAT	8	0.009116
SFPQ	9	0.008379
FTSJ3	10	0.01046
N4BP1	11	0.01451
PINX1	12	0.01096
OTUD4	13	0.01143
PRMT7	14	0.01206
EIF3B	15	0.01673
NEURL4	16	0.01322
APLP2	17	0.01386
SPSB1	18	0.01792
FASN	19	0.02229
EIF2S2	20	0.0229
RNF39	21	0.02522
EIF2D	22	0.02592
MKRN3	23	0.02606
TRIM10	24	0.01669
PPP1R10	25	0.0311

The screen was repeated in the A549^{HMGA2} reporter cells using a similar gating strategy to that employed in the first screens using the A549^{mini} cell line. Cells with the highest relative

GFP$^{HMGA2\text{-}wt}$ expression levels were sorted (top 3%, top 4-10%) to determine whether shRNAs targeting known miRNA pathway genes (Ago2, Dicer) were enriched in these populations. A population of the center 50% cells were also sorted which were not expected to enrich for shRNAs targeting miRNA pathway components. An additional gate (bottom 3%) was included to capture cells with decreased GFP$^{HMGA2\text{-}wt}$ expression which was expected to enrich for shRNAs targeting genes that typically prevent miRNA mediated repression. A single screen was performed in this reporter line, and the sequencing revealed that Dicer was one of the highest ranked gene targets in both the top 3% and top 4-10% populations, however Ago2 was only highly ranked in the top 4-10% population. As expected, neither Ago2 or Dicer were among the top ranked candidates in the center 50% population or the bottom 3% population. The reason for a lack of Ago2 enrichment the top3% was population was unknown. I concluded that it may have been due to an error in my sample preparation for sequencing or that the Ago2 hairpins did not produce as significant in this reporter line, but Ago2's presence in the top 4-10% population was encouraging so I proceeded to use the A549^{HMGA2} reporter for the larger screen of the 55k shRNA library targeting ~11,000 genes.

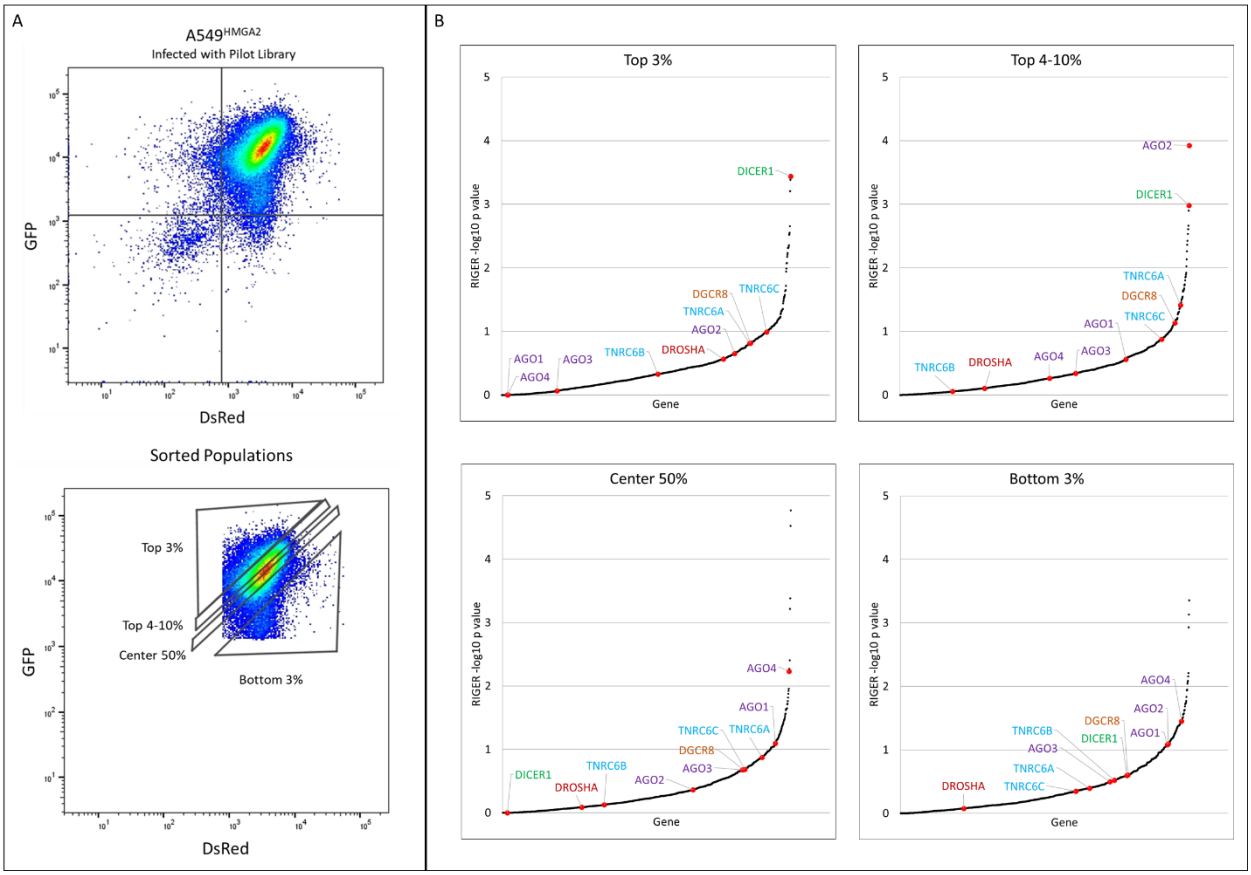

Figure 2-5: Screen of the small shRNA library with A549^{HMGA2}

A) Gating strategy for enriching cells with high GFP expression (top 3% and top 4-10%) expected to be necessary for miRNA silencing, the center 50% population which is not expected to be enriched for any miRNA pathway components, and the low GFP cells (bottom 4-10%) expected to be enriched for factors that prevent miRNA repression. **B)** Comparison of known miRNA pathway components and their relative enrichment (RIGER -log10 p value) in the top3%, top 4-10%, center 50% and bottom 4-10% populations.

		A549mini			A549HMGA2
Ranked by Geometric Mean of Gene Scores	Gene	Top 3% Replicate 1	Top 3% Replicate 2	Top 3% Replicate 3	Top 3 % Replicate 1
	DICER1	1	3	1	1
	SNRPE	77	4	18	13
	GTPBP4	102	5	6	26
	FAM120A	8	35	27	11
	SMARCAD1	6	14	121	44
	AGO2	14	21	9	205
	PRDX2	2	19	10	1631
	RPP38	7	12	8	1146
	SLIRP	70	17	32	30
	YBX3	45	27	17	138
	EIF2D	368	7	5	237
	CTIF	386	15	26	28
	MRPL44	43	50	53	47
	PAIP1	1618	16	13	17
	TARBP1	76	39	30	65
	CCSAP	1777	11	45	7
	LSM14A	94	36	51	57
	CNOT7	1600	23	16	19
	PCMTD2	93	152	65	14
	ZCCHC7	112	32	15	248
	TERF2	131	10	7	1479
	EEF2KMT	1508	2	3	1763
	LIN28A	28	123	58	84
	TRMT1L	306	9	4	1551

Table 2-2: Comparison of Rankings Across Replicate Screens

A comparison of 25 genes ranked in the top 3 % for replicate screens in 2 reporter cell lines using the small pilot shRNA library. Genes are ordered by the geometric mean of their rank in each replicate. Known components of the miRNA pathway highlighted in green are expected. Components highlighted in red were unexpected.

RNAi Screen of 55k shRNA library & identification of enriched gene targets

The 11,000 gene RNAi screen was performed in the same manner as the pilot screens except that the A549^{HMGA2} reporter cell line was infected with the 55k shRNA library targeting ~11,000 genes (Moffat et al., 2006). The A549^{HMGA2} reporter cells were infected at a low multiplicity of infection (~0.3), so that on average each cell experiences knockdown of a single gene. After 7 days of growth, the cells were FACS sorted to isolate cells with the highest and lowest relative GFP$^{HMGA2\text{-}wt}$ expression levels, as shown in **Figure 2-6A**. Gates were drawn for the top 3% and top 4-10% of the high GFP$^{HMGA2\text{-}wt}$ population which represents cells with reduced levels of miRNA mediated repression and are therefore likely to contain shRNAs that target genes required for miRNA mediated repression (e.g. Dicer, Ago2). Similarly, the gates drawn for bottom 4-10% and bottom 3% of the low GFP$^{HMGA2\text{-}wt}$ population are more likely to contain gene targets that normally prevent miRNA mediated repression (e.g. Lin28). A restrictive gate (top 3% and bottom 3%) and less restrictive gate (top 4-10% and bottom 4-10%) were drawn for the highest and lowest relative GFP$^{HMGA2\text{-}wt}$ populations because it was unknown which condition would result in better enrichment of novel candidate genes. It was possible that if the gates were drawn too narrowly, I may only see enrichment of established miRNA pathway components and if they were drawn too broadly, I may not see enrichment of any miRNA pathway genes.

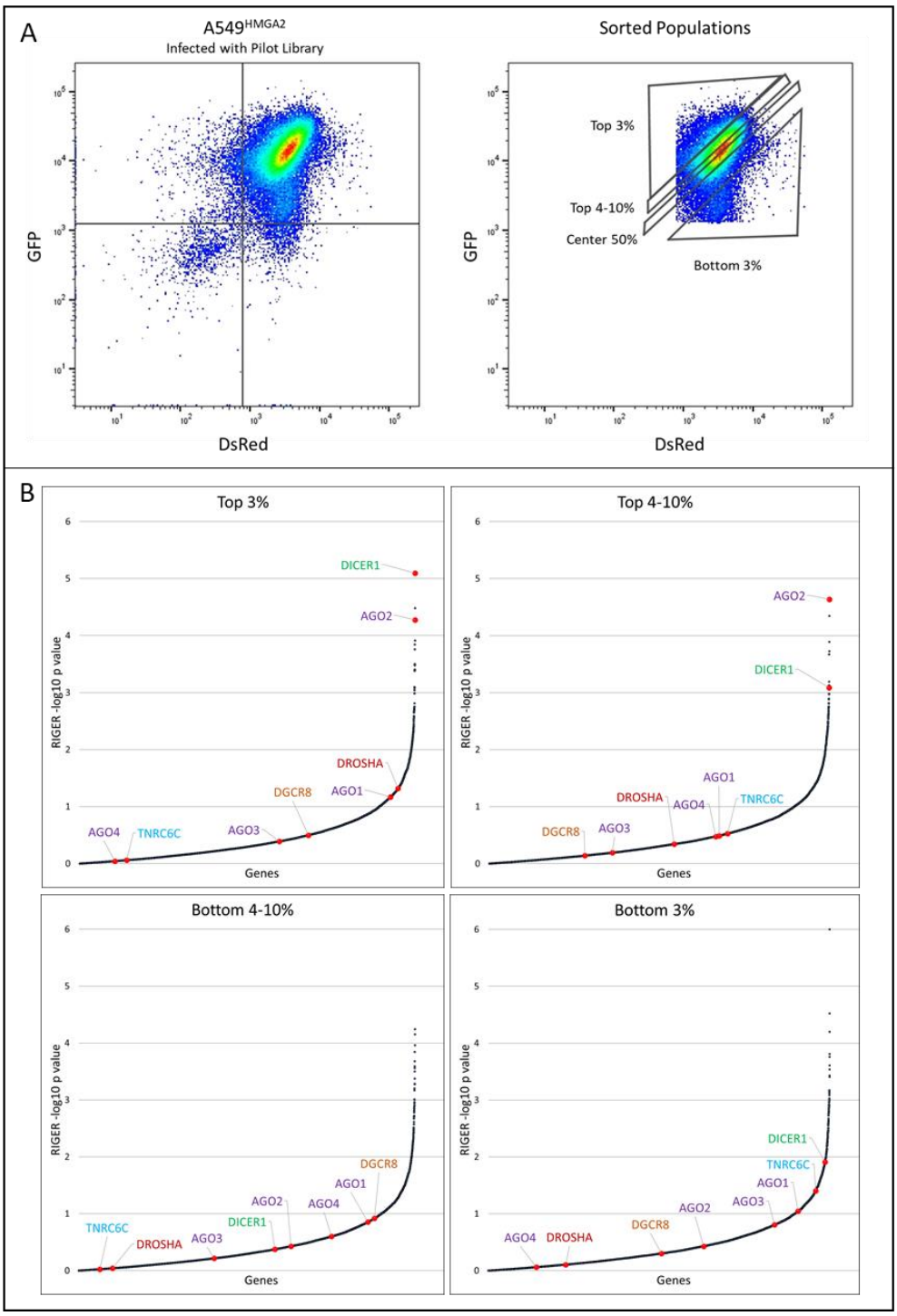

Figure 2-6: 55k Library Screen of A549^{HMGA2}

A) Gating strategy for the screen of the 55k shRNA library performed in A549^{HMGA2} reporter, enriching cells with high GFP expression (top 3% and top 4-10%) which are expected to be necessary for miRNA silencing, the center 50% population which is not expected to be enriched

for any miRNA pathway components, and the low GFP cells (bottom 4-10%) expected to be enriched for factors that prevent miRNA repression. **B)** Comparison of known miRNA pathway components and their relative enrichment (RIGER -log10 p value) in the top3%, top4-10%, center 50% and bottom 4-10% populations.

Table 2-3: Screen of the 55K shRNA Library in A549^{HMGA2}

Top 25 ranked genes in the sorted population vs unsorted background. Genes known to play a role in the miRNA pathway are highlighted in green.

Top 3% vs Unsorted

Gene	Gene Rank	p-value
DICER1	1	0.000008
AGO2	2	0.000053
PPAN	3	0.000033
GTPBP4	4	0.000143
PSPC1	5	0.000055
OR4C11	6	0.000174
GRIA1	7	0.000122
CYP7A1	8	0.000317
ZNF768	9	0.000318
TWIST1	10	0.000320
RBM45	11	0.000333
ZC3H13	12	0.000412
SLC17A5	13	0.000393
SUB1	14	0.000409
ZNF7	15	0.000828
LUC7L	16	0.000904
SECISBP2L	17	0.000910
BAG6	18	0.000804
TBPL1	19	0.001042
LIX1L	20	0.000413
AK5	21	0.001554
TARBP1	22	0.001758
IDH1	23	0.001812
PDHA1	24	0.001885
PADI1	25	0.001915

Top 4-10% vs Unsorted

Gene	Gene Rank	p-value
AGO2	1	0.000023
GTPBP4	2	0.000045
SUB1	3	0.000187
BRSK2	4	0.000215
CYP7A1	5	0.000215
GRIA1	6	0.000129
ZNF7	7	0.000854
KCNJ12	8	0.000886
MAP3K20	9	0.00074
TBPL1	10	0.001045
DLX1	11	0.000645
OR2T10	12	0.001277
ADAMTS19	13	0.00131
KRBOX4	14	0.00107
DICER1	15	0.000813
IL2RG	16	0.001535
OR4C11	17	0.001556
PDHA1	18	0.001758
CCR4	19	0.001759
SECISBP2L	20	0.001892
ZNRF1	21	0.001528
F2	22	0.002024
BFAR	23	0.00211
BCL11B	24	0.001767
FAM120A	25	0.00179

Bottom 4-10% vs Unsorted

Gene	Gene Rank	p-value
BACE1	1	0.000057
NTSR1	2	0.00007
TBC1D25	3	0.00021
MTHFD2	4	0.000109
FCGR2A	5	0.000258
HELZ	6	0.000144
ALDH1A2	7	0.000316
CFLAR	8	0.000278
TXK	9	0.00052
CCAR2	10	0.000424
MRPS16	11	0.000532
MPG	12	0.000994
ZNF140	13	0.000632
HUNK	14	0.001144
GAL3ST3	15	0.000682
EHMT2	16	0.001247
ZNF304	17	0.001254
PTPRD	18	0.001357
ING3	19	0.001097
MDM2	20	0.001416
GTF2A1L	21	0.001658
GMEB2	22	0.001275
PRDM8	23	0.001659
TMEM14B	24	0.001444
TBX15	25	0.001899

Bottom 3% vs Unsorted

Gene	Gene Rank	p-value
BACE1	1	0.000001
DAZ1	2	0.000063
HELZ	3	0.00003
TGM7	4	0.000155
HNRNPD	5	0.000289
PCDHGB1	6	0.000246
MTHFD2	7	0.000174
DDX53	8	0.000373
SENP5	9	0.000391
CDC42	10	0.000763
ZFPM2	11	0.000827
PRPF18	12	0.000727
NOSIP	13	0.000975
RBMY1A1	14	0.001636
DDX31	15	0.001077
ZNF304	16	0.001094
RFC3	17	0.000671
SMARCA1	18	0.001187
TFAP2D	19	0.001212
DYRK4	20	0.000943
CLEC4E	21	0.001227
HOXD10	22	0.001249
FSBP	23	0.001345
eGFP	24	0.001395
PRKAR2A	25	0.001425

After FACS sorting, the shRNA representation of the sorted populations and the unsorted background were determined as described previously (page 38). Again, Dicer and Ago2 were among the highest ranked genes in both the top 3% and top 4-10% populations, and as expected, were not among the highest ranked candidates in either the bottom 4-10% or bottom 3% populations (**Figure 2-6B**). These results demonstrated that the screen of the 55k library was working and the other highly ranked genes unknown to have a role in the miRNA pathway represent potential novel candidates.

Validation of potential novel candidates

In order to reduce the number of false positives, I developed a way to quickly validate the candidates that genuinely affected the $GFP^{HMGA2\text{-}wt}$ reporter expression and were most likely to be true hits. To accomplish this, single gene knockdowns were performed in the original $A549^{HMGA2}$ reporter cell line and the effect on reporter expression was measured using flow cytometry. Although shRNAs enriched in any population are of potential interest it was expected that those enriched in the top 3% and top 4-10% populations would be more likely to contain *bona fide* miRNA pathway components than those enriched in the bottom 4-10% and bottom 3% populations. This was because there was an underlying assumption that it would be easier to disrupt the miRNA pathway than to enhance silencing.

I chose to focus on genes ranked in the top 2% (220 genes) of my sorted populations which were significantly enriched (p-value <0.05). However, due to reagent availability and time limitations, I could not validate all top 220 ranked genes from each population. I limited validation experiments to 68 genes (264 shRNAs) ranked in the top 220 of either the top 3% or top 4-10% populations, and 11 genes (39 shRNAs) ranked in the top 220 genes of either the bottom 3% or bottom 4-10% populations. I limited the validation experiments to genes that I was

confident were expressed in the reporter cell line. Using RNA-sequencing data of A549 cells, I only considered candidate genes with greater than 1 read per million (RPM). This cutoff includes ~13,000 genes, which were more likely to be expressed than transcripts with <1 RPM. Also, genes that were ranked in the top 220 candidates of both a top and bottom population were excluded from further analysis.

In addition to testing novel candidates, I also tested 8 genes (31 hairpins) that were not targets of the lentiviral library and therefore not enriched in any population but were also not expected to be involved in the miRNA pathway (Table 2-5). These served as a negative control that could be used to compare my novel candidates to. Typically, miRNAs result in less than a 1 log2 fold change in the levels of their target mRNA (Baek et al., 2008a; Selbach et al., 2008a), and although the HGMA2 3′ UTR contains more miRNA target sites than an average mRNA it was observed that knockdown of Dicer in the A549^{HMGA2} reporter produced approximately a 1 log2 fold change of GFP$^{HMGA2\text{-}wt}$/DsRed$^{HMGA2\text{-}mut}$ expression (**Figure 2-8**). Based on the knockdown data from the 8 negative control genes (31 hairpins) that were not enriched in the top or bottom populations, I chose an arbitrary cutoff of a +/- 0.5 log2 fold change for defining whether a hairpin was counted as having an effect on GFP$^{HMGA2\text{-}wt}$/DsRed$^{HMGA2\text{-}mut}$. Then only candidates that showed greater than a +/- 0.5 log2 fold change for 2 or more hairpins were considered further for validation.

Interestingly, the distribution of normalized GFP$^{HMGA2\text{-}wt}$, DsRed$^{HMGA2\text{-}mut}$, and GFP$^{HMGA2\text{-}wt}$/DsRed$^{HMGA2\text{-}mut}$ geometric means from the knockdown of individual hairpins targeting gene candidates selected from the top, bottom, or not enriched groups shows that the hairpins chosen from the top populations have significantly higher expression of GFP$^{HMGA2\text{-}wt}$ and GFP$^{HMGA2\text{-}wt}$/DsRed$^{HMGA2\text{-}mut}$ (**Figure 2-7**). This suggests that the screen did successfully enrich for genes that relieve miRNA mediated silencing upon knockdown. However, the opposite is not true for hairpins that increase repression. The candidates chosen from the bottom populations

show no statistical difference from the not enriched control populations. One possible reason there is not a significant decrease in $GFP^{HMGA2\text{-}wt}$/$DsRed^{HMGA2\text{-}mut}$ for factors enriched in the bottom populations is that because the HGMA2 3′ UTR contains seven let-7 target sites and is already strongly repressed, even if a factor that typically prevents miRNA silencing is knocked down, the dynamic range or the degree to which miRNAs can elicit additional repression of the transcript is limited.

For the 68 genes enriched in the top populations, 27 passed the cutoff with 2 or more hairpins result in a >0.5 log2 fold change of $GFP^{HMGA2\text{-}wt}$/$DsRed^{HMGA2\text{-}mut}$ (**Figure 2-8**). However, there was a concern that although the $GFP^{HMGA2\text{-}wt}$ and $DsRed^{HMGA2\text{-}mut}$ are both under control of the same promoter, they were stably integrated using separate lentiviral vectors and therefore their sites of integration in the genome were not controlled for. This meant that it was possible that nearby regulatory elements may influence reporter expression, and that knockdown of a candidate gene may increase $GFP^{HMGA2\text{-}wt}$ expression by an indirect mechanism, for example by affecting the association of a transcription factor with a nearby regulatory elements. To rule out this possibility, I repeated the individual gene knockdowns of the validated genes in the $A549^{mini}$ reporter line (**Figure 2-9**). The $A549^{mini}$ reporter cell line was generated independently of the $A549^{HMGA2}$ so it was very unlikely that the reporters would integrate to the same genomic locations or be affected by the same nearby regulatory elements. The knockdown of 22 of the 27 validated candidates in the $A549^{mini}$ reporter line allowed me to confirm 19 as likely candidates involved in the miRNA pathway and reject 3 as false positives (**Figure 2-9B**).

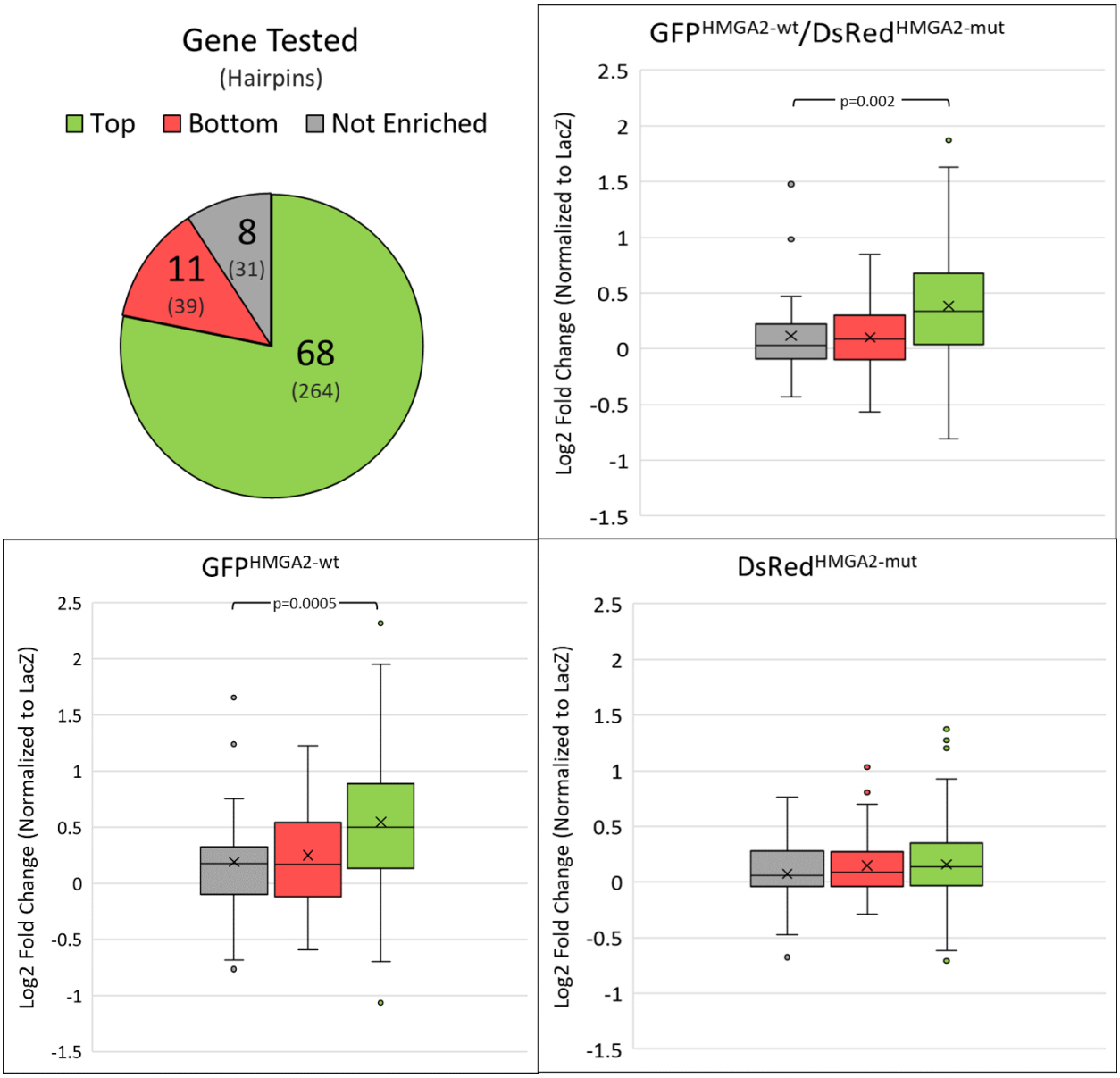

Figure 2-7: Hairpins enriched in the top and bottom populations

shRNA hairpins targeting genes enriched in top populations (top 3% or top 4-10%), bottom populations (bottom 3% or bottom 4-10%) or not enriched (not present in the 55k library) were selected to perform individual gene knockdowns. The log2 geometric means of GFP, DsRed and GFP/DsRed were normalized by the LacZ_2223 hairpin control and the distribution of values was plotted for each. Only the hairpins enriched in the top populations (68 genes, 264 hairpins) show significantly higher GFP and GFP/DsRed expression compared to the not enriched hairpins (8 genes, 31 hairpins). The hairpins in the bottom population (11 genes, 39 hairpins) were not significantly different from the not enriched genes.

Table 2-4: RIGER Ranks of All Genes Tested in Validation Experiments.

The RIGER rank of each gene is listed for each sorted population compared to the unsorted background (Top 3%, Top 4-10%, Bottom 4-10%, and Bottom 3%) and the category they belong to in **Figure 2-7**. Cells highlighted in blue are genes that had 2 or more hairpins pass the cutoff for validation in the A549^{HMGA2} reporter line, and those in bold were successfully validated in the A549^{mini} reporter line.

Gene Target	Top 3%	Top 4-10%	Bottom 4-10%	Bottom 3%	Category
DICER1	1	15	5,303	227	Positive Control
LUCIFERASE	1,230	2,506	9,026	1,932	Negative Control
lacZ	1,594	5,144	8,317	2,215	Negative Control
GTPBP4	4	2	3,631	253	Top
ZNF768	9	104	9,558	7,763	Top
TWIST1	10	205	10,689	8,896	Top
SLC17A5	13	105	10,516	5,586	Top
SUB1	14	3	6,514	1,658	Top
ZNF7	15	7	9,626	5,238	Top
LUC7L	16	172	2,368	7,924	Top
SECISBP2L	17	20	7,152	1,676	Top
TBPL1	19	10	4,983	1,569	Top
LIX1L	20	918	7,556	2,726	Top
AK5	21	32	8,945	6,121	Top
SLC30A7	27	229	7,868	10,302	Top
RPP38	30	893	5,470	1,239	Top
TPCN1	31	998	7,461	10,505	Top
PDPK1	33	806	10,098	7,005	Top
NDUFAB1	37	33	7,574	3,953	Top
CLTA	40	27	5,854	8,102	Top
INSR	41	198	9,945	9,065	Top
GIT2	44	146	2,117	532	Top
CAMKMT	47	6,080	4,215	3,364	Top
MKRN2	49	721	9,114	255	Top
GCAT	52	116	2,702	2,956	Top
PCMTD2	53	1,909	2,837	999	Top
SPTAN1	56	248	3,305	1,309	Top
CRLF3	61	424	5,772	4,606	Top
RETSAT	66	3,149	4,367	3,067	Top
RFC4	68	1,029	7,771	7,442	Top
CHD9	69	1,505	2,944	257	Top
U2AF2	70	835	7,806	3,503	Top
BLM	72	1,945	1,008	3,268	Top
FAM105A	76	41	6,894	3,176	Top
SMNDC1	81	1,491	3,473	2,174	Top
E2F8	83	2,455	7,432	10,638	Top
TTF1	84	1,091	4,875	4,941	Top
EGFL8	85	206	4,492	7,195	Top
NTRK3	86	3,489	4,751	589	Top
FDPS	88	694	3,361	6,364	Top
TTC23	92	48	8,754	5,994	Top
BRIP1	97	897	1,775	3,715	Top
NAB1	98	28	3,326	3,963	Top
NF1	99	404	10,138	8,401	Top
TCF7L2	114	99	8,178	1,699	Top
KRBOX4	138	14	3,701	9,408	Top

Gene Target	Top 3%	Top 4-10%	Bottom 4-10%	Bottom 3%	Category
SLU7	162	3,040	3,108	313	Top
ADIPOR2	173	50	3,149	8,030	Top
DHX29	202	3,276	2,939	5,000	Top
SKIV2L2	210	1,250	2,723	861	Top
DHX57	220	1,875	9,129	2,623	Top
MDH1	249	54	2,429	6,450	Top
DCTD	324	57	5,564	4,853	Top
PQBP1	399	37	7,768	8,194	Top
XK	593	100	3,322	10,548	Top
PGRMC2	663	49	2,869	4,642	Top
DPYSL3	680	63	7,915	2,402	Top
TRIM26	835	76	4,075	1,860	Top
BFAR	918	23	4,161	3,239	Top
RAB10	1,111	39	2,116	6,785	Top
PIGK	1,293	150	6,555	1,144	Top
ZNRF1	1,331	21	9,681	9,364	Top
HMOX1	1,777	55	1,534	8,566	Top
ZNF581	2,324	35	4,312	3,886	Top
TEAD3	2,330	90	9,747	2,050	Top
DCTN3	2,668	26	7,534	8,831	Top
ZNF57	2,789	66	2,550	7,241	Top
PTPN9	2,976	70	9,207	4,562	Top
ZNF185	3,310	81	8,672	7,227	Top
EDN1	3,502	60	1,707	8,793	Top
SMAD2	4,404	98	3,950	2,602	Top
B3GNT1	#N/A	#N/A	#N/A	#N/A	Not Enriched
CECR5	#N/A	#N/A	#N/A	#N/A	Not Enriched
EIF2S1	#N/A	#N/A	#N/A	#N/A	Not Enriched
KIAA1967	#N/A	#N/A	#N/A	#N/A	Not Enriched
MRE11A	#N/A	#N/A	#N/A	#N/A	Not Enriched
MTIF2	#N/A	#N/A	#N/A	#N/A	Not Enriched
RDBP	#N/A	#N/A	#N/A	#N/A	Not Enriched
SF3B14	#N/A	#N/A	#N/A	#N/A	Not Enriched
HNRNPH1	253	1,461	2,602	154	Bottom
WTAP	431	932	1,606	108	Bottom
HELZ	633	5,863	6	3	Bottom
HNRNPD	1,450	3,313	1,029	5	Bottom
TRIM27	2,265	2,243	627	29	Bottom
ZNF224	2,306	4,245	60	541	Bottom
GATAD1	4,241	4,342	188	47	Bottom
SENP5	5,228	3,455	341	9	Bottom
LARP6	5,577	3,418	5,598	93	Bottom
MTFMT	5,888	9,570	54	165	Bottom
G3BP2	7,876	1,959	1,042	96	Bottom

Of the 11 genes enriched in the Bottom populations, none tested had 2 or more hairpins pass the -0.5 log2 fold change cutoff. However, a single gene, LARP6, had 2 hairpins that resulted in a >0.5 log2 fold change which was confirmed in the A549^{mini} reporter line as well

(**Figure 2-8**, **Figure 2-9**). This demonstrates that even with this validation screening approach requiring that multiple hairpins show an effect on reporter expression in two separate cell lines, it does not necessarily eliminate false positives, and further validation is still required.

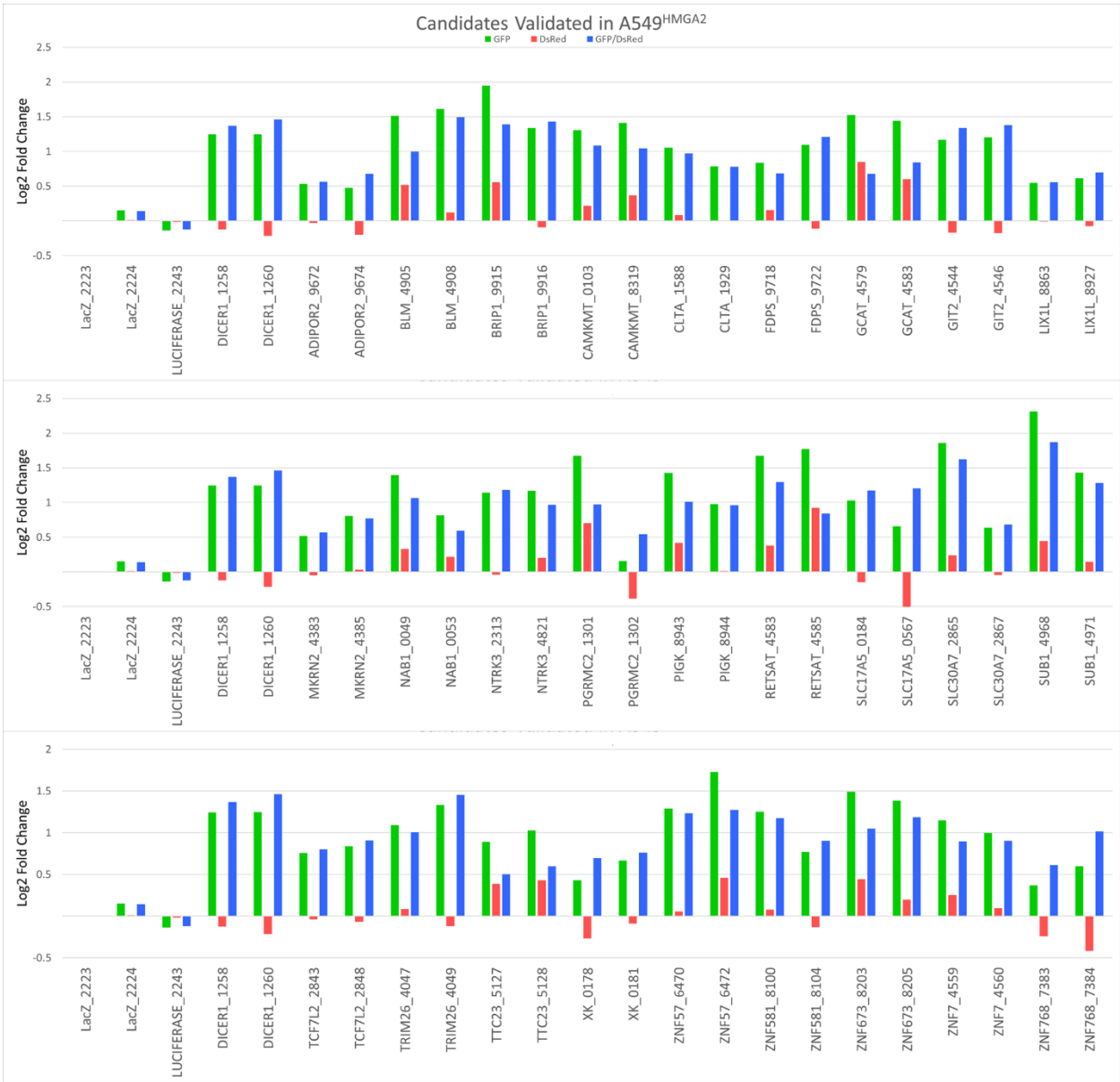

Figure 2-8: Validation of Individual Hairpins in A549^{HMGA2}

The log2 geometric mean of GFP levels normalized to LacZ_2223 control (Green), log2 geometric mean of DsRed levels normalized to LacZ_2223 control (Red), and log2 geometric mean normalized GFP/DsRed (Blue) values are shown for each hairpin. These are the 27

genes that passed the cutoff with 2 or more hairpins showing >0.5 log2 fold change in the A549^{HMGA2} reporter line.

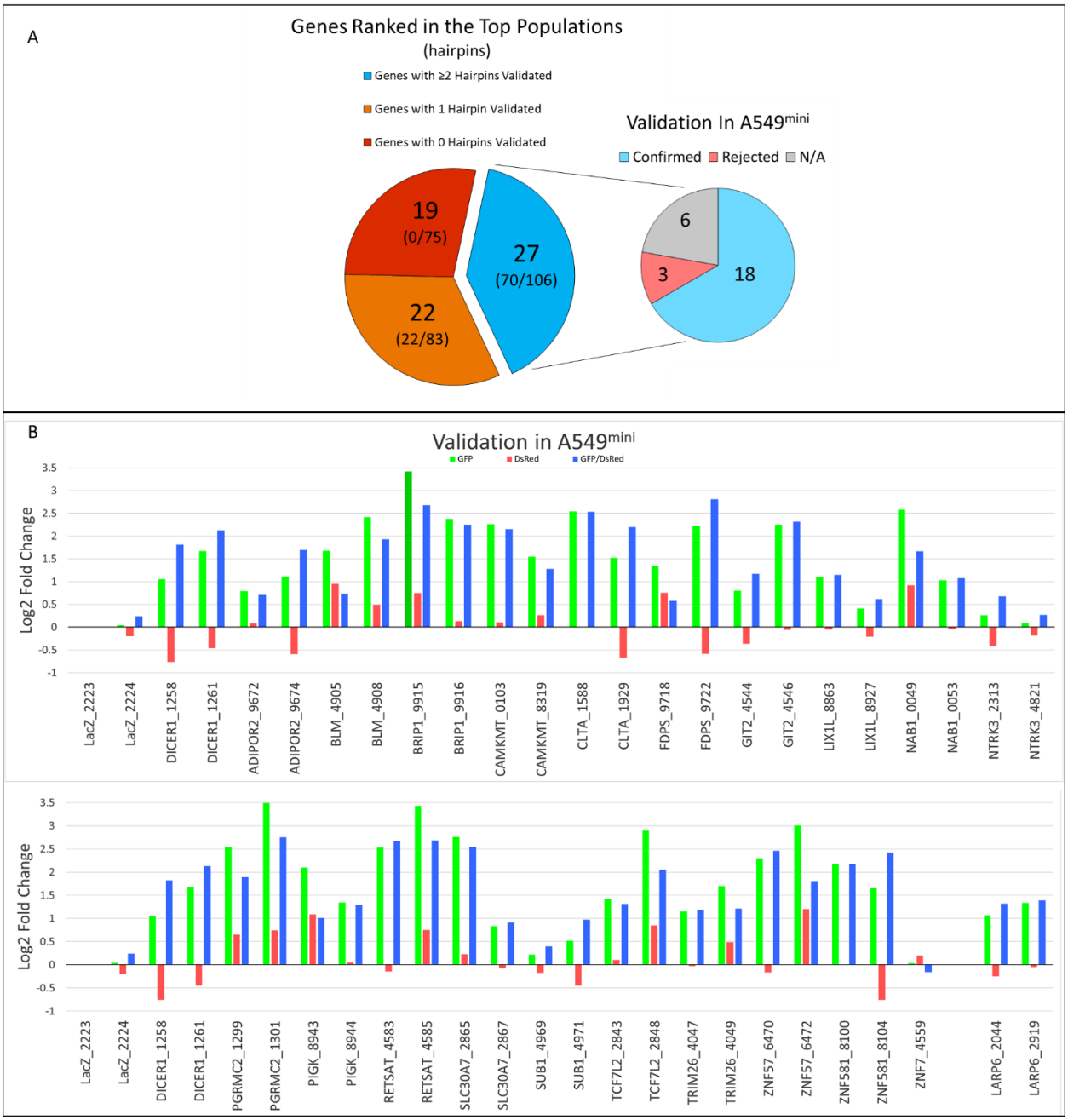

Figure 2-9: Validation of Individual Hairpins in A549^{HMGA2}

A) Of the 68 genes tested that were enriched in the top populations, 27 passed the validation cut off requiring 2 or more hairpins with a >0.5 log2 fold change of GFP/DsRed in the A549^{HMGA2}

reporter. Of these 27, 22 were validated in the A549^{mini} reporter line and 18 were confirmed as likely candidates. **B)** The log2 geometric mean of GFP (Green), DsRed, and GFP/DsRed (Blue) normalized to LacZ_2223 control are shown for each hairpin.

RNA Sequencing and Small RNA Sequencing of Candidate Knockdowns

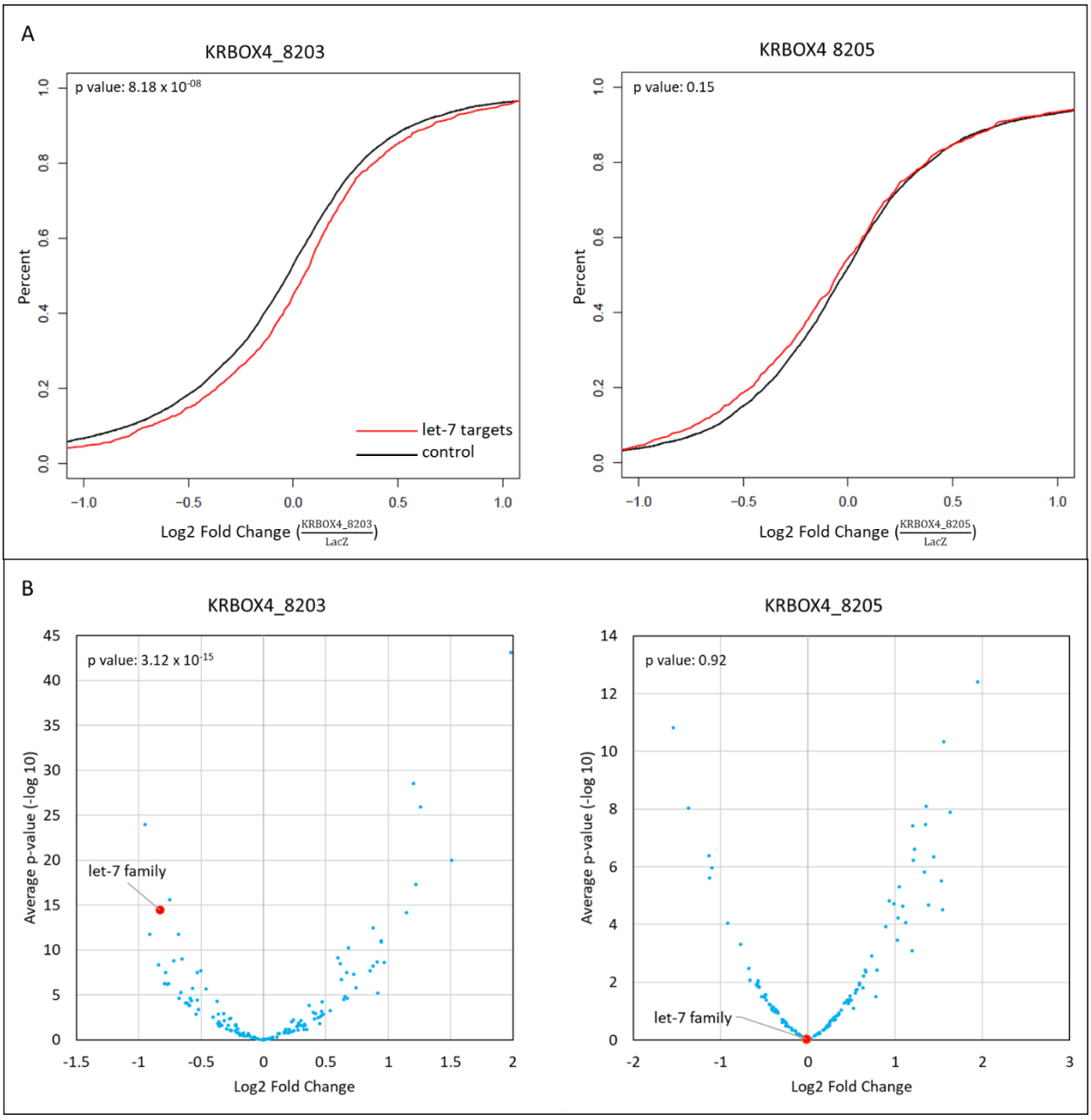

Figure 2-10: RNA sequencing and small RNA sequencing of KRBOX4 Knockdowns

A) RNA sequencing data showing targets of let-7 in knockdown of KRBOX4 using the

KRBOX4_8203 and KRBOX4_8205 hairpins. **B)** small RNA sequencing data for miRNA families with increased and decreased fold change upon knockdown with KRBOX_8203 or KRBOX_8205.

After the initial screening of the first set of hairpins targeting candidate genes in the $A549^{HMGA2}$ reporter line, KRBOX4 (KRAB Box Domain Containing 4, formerly ZNF673) was one of the only gene targets to show multiple hairpins that affected $GFP^{HMGA2\text{-}wt}$ expression. KRBOX4 contains a Krueppel-associated box (KRAB) which is a domain of approximately 75 amino acids found in the N-terminal part of about one third of eukaryotic Krueppel-type C2H2 zinc finger proteins (Bellefroid et al., 1991; Huntley et al., 2006). KRBOX4 is an interesting candidate because its function is unknown, but it appears to be a partial inverted duplication of the ZNF674 gene whose function is also unknown (Lugtenberg et al., 2006). However, unlike ZNF674, KRBOX4 contains a premature stop codon which causes a loss of the predicted zinc finger domain. To date, no known function for a KRAB domain without zinc fingers has been reported, but the functions currently associated with members of the KRAB-containing protein family include transcriptional repression, binding and splicing of RNA and control of nucleolus function (Lupo et al., 2013; Urrutia, 2003).

To investigate KRBOX4's potential involvement in the miRNA pathway further, I performed RNA sequencing and small RNA sequencing on KRBOX4 knockdowns in A549 cells. A549 cells were seeded and infected in the same manner as for the screen, with a negative control lentiviral shRNAs targeting LacZ (LacZ_2224) and 2 lentiviral shRNAs targeting KRBOX4 (KRBOX4_8203, KRBOX_8205). Two technical replicates were performed for each hairpin and after 7 days of growth the RNA was isolated for all replicates and RNA-sequencing and small RNA-sequencing libraries were prepared, sequenced and mapped to the genome.

For both the small RNA-seq and RNA-seq experiments, the technical replicates of each KRBOX4 hairpin were compared to the negative control shRNA (LacZ_2224) technical

replicates. The RNA-seq data analysis, revealed that KRBOX4_8203 resulted in a less efficient knockdown of KRBOX4 than the KRBOX_8205, although both resulted in significant repression of KRBOX4. Analysis of the RNA-seq data revealed that targets of let-7 compared to non-targets were significantly upregulated upon knockdown of KRBOX4 when using the KBROX4_8203 hairpin but not the KRBOX4_8205 hairpin which was unexpected given that the KRBOX_8205 hairpin resulted in greater repression of KRBOX4.

Reads from the small-RNA sequencing libraries were first grouped by family and then Empirical Analysis of Digital Gene Expression Data in R (EdgeR) (McCarthy et al., 2012; Robinson et al., 2010) was used to measure whether the let-7 miRNA family was significantly repressed. The data revealed that the guide strands of the let-7 family members were significantly repressed upon knockdown with the KRBOX4_8203 hairpin but not the KRBOX_8205 hairpin (**Figure 2-10B**). This may imply a role for KRBOX4 in miRNA biogenesis, but the discrepancies between the effects observed for each hairpin required additional experiments to validate this hypothesis.

Validation of Candidates Using CRISPR Knockouts

To confirm whether KRBOX4 was involved in miRNA mediated silencing I switched to using a different approach. CRISPR was used to generate knockout lines of KRBOX4 in HAP1 cells and the effect on miRNA mediated repression was confirmed using luciferase reporter constructs containing a *renilla* control reporter construct and either a firefly-wt or firefly-mut reporter containing seven functional or nonfunctional let-7 miRNA target sites, respectively. For each knockout line 3 biological replicates x 3 technical replicates were performed. Each firefly-wt and firefly-mut reporters' relative light units (RLU) were normalized to the *renilla* control RLUs to control for transfection efficiency. For each knockout line a second normalization of the

firefly-wt to the firefly-mut transfected wells was then performed to assess the effect of the knockdown on the firefly-wt miRNA target relative to the firefly-mut control. While two of the three knockouts relieved repression of the wt reporter, repression was increased in the third. (**Figure 2-11**).

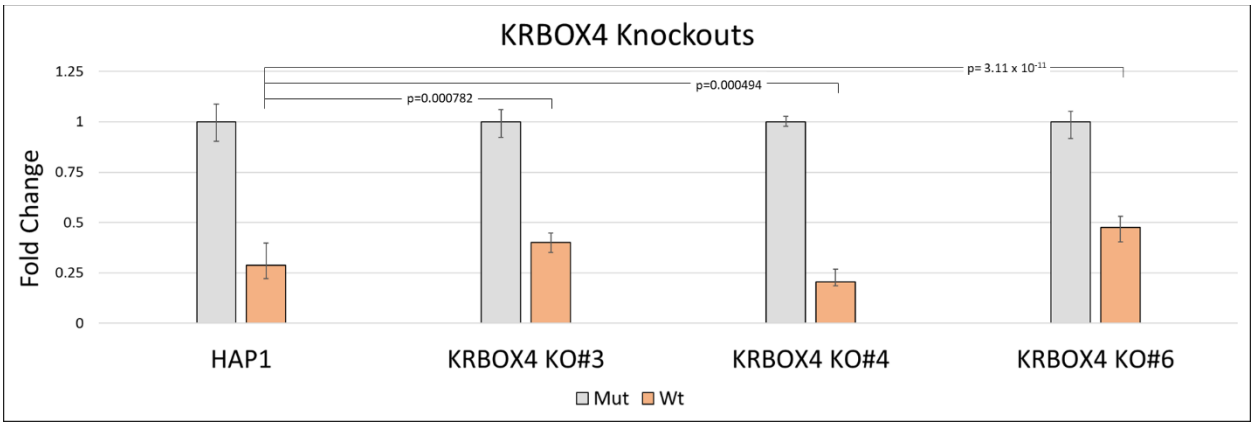

Figure 2-11: Luciferase Assay of KRBOX4 Knockouts

Luciferase assay of wildtype HAP1 cells and KRBOX4 knockout lines. Cells were transfected with a firefly-mut (Grey) reporter or a firefly-wt (Orange) reporter. Reporter expression was normalized to the firefly-mut reporter, and significance was measured using the Wilcoxon rank sum statistic.

Discussion

In the past RNAi has been used to identify components of the miRNA pathway in *C. elegans* (Bernstein et al., 2001; Denli et al., 2004; Ketting et al., 2001; Lee et al., 2003), and novel regulators of the miRNA pathway continue to be discovered using high throughput approaches (Golden et al., 2017; Treiber et al., 2017). The identification of additional factors involved in miRNA mediated repression will potentially answer some of the outstanding questions that remain regarding incompletely understood mechanisms such as precisely how translational repression, strand selection, and miRNA turnover occurs. Additionally, identification

of novel factors that regulate miRNA silencing through their interaction with core components of the miRNA pathway or the miRNAs themselves will provide more clarity as to how miRNA mediated repression is regulated.

In this work I demonstrate a novel approach for screening for miRNA pathway components using an RNAi screening approach in human cell lines sensitive to changes in miRNA mediated repression. I developed and tested three different reporter cell lines that all demonstrated sensitivity to the knockdown of known miRNA pathway components. In two of the reporter lines I piloted the efficacy of this approach, performing several screens using a small lentiviral library of shRNAs. Both lines successfully enriched known miRNA pathway components among their top hits while also producing a list of candidates for validation.

Individual knockdown experiments of candidate genes in two separate reporter lines produced a list of 18 candidates likely to be playing a role in miRNA mediated repression. For a single candidate, KRBOX4, RNA sequencing and small RNA-sequencing experiments were performed on knockdowns using 2 unique shRNAs targeting the gene. The results revealed one of the hairpins resulted in robust repression of let-7 family members and a corresponding increase in expression of let-7 miRNA targets, while the other hairpin did not significantly affect either. Although both hairpins significantly reduced KRBOX4 expression this did not correspond to their effect on let-7 or let-7 targets. Contrary to what I would expect, the hairpin that repressed KRBOX4 to a lesser degree (KRBOX4_8203) was the hairpin that caused both repression of the let-7 family and the increase in the expression of let-7 targets.

To determine whether KRBOX4 was a genuine novel factor I took an orthogonal approach and generated knockouts in a different cell line hoping to either rule it out as a false positive or confirm its effect on miRNA mediated repression. The level of miRNA activity was assessed using luciferase reporters containing 3′ UTRs with wildtype or mutant let-7 sites, but

this yielded conflicting results. While two of the knockout lines showed a significant increase in the wildtype reporters' expression the third showed a significant decrease. These results may suggest that there is something more complex occurring, or the contradictory result may simply be the result of an off-target effect of the CRISPR guide RNA. To better answer this question, it would be best to get a cleaner knockout, possibly by generating a larger deletion within the gene before repeating the experiment. Then the mechanism by which it is affecting miRNA mediated repression can be further investigated. Additional experiments such as immunoprecipitation coupled to mass spectrometry (IP-MS) can then be performed to identify whether KRBOX4 is interacting with any known miRNA pathway genes (Chendrimada et al., 2005a; Gregory et al., 2004a; Huang et al., 2013a).

Originally, I had selected 4 additional candidates to generate knockouts for which altered relative GFP expression in both A549^{HMGA2} and A549^{miniHMGA2} reporters (BRIP1, TRIM26, ZNF57, ZNF581), but I was only able to successfully generate and confirm knockouts for KRBOX4. These other candidates were chosen based on their known or predicted biological functions as well as other evidence which supported their potential involvement in the miRNA pathway. Although I was not able to successfully generate knockouts for the other candidates of interest, they may still play a role in the miRNA pathway and there are some good reasons to think so. BRIP1 for example is a DNA helicase that interacts with the BRCT repeats of breast cancer, type 1 (BRCA1) (Brosh and Cantor, 2014) which has been shown to regulate miRNA biogenesis through its interaction with DROSHA (Kawai and Amano, 2012). It may be that BRIP1 participates in this process. Another candidate, TRIM26, which encodes an E3 ubiquitin-protein ligase (Wang et al., 2015), was recently identified alongside another ~180 RBPs that interact specifically with distinct pre-miRNAs (Treiber et al., 2017b). The authors showed that TRIM26 interacts exclusively with miR-18b but its presence among the likely candidates identified in my screen suggests that TRIM26 may play a broader role in miRNA silencing as the

only difference between my reporters was the presence or absence of functional let-7 target sites. In the same study, the authors also identified 23 zinc-finger-containing proteins, which may represent an important class RBPs involved in miRNA biogenesis. For this reason, ZNF57 and ZNF581 were included in the list of candidates I followed up on. Very little is known about ZNF581 (HSPC189) function aside from that it contains a C2H2 zinc finger motif often associated with DNA binding, but can also bind RNA and protein (Iuchi, 2001; Zhang et al., 2000). Similarly, ZN57 (ZNF424) has not been extensively studied but has been shown to localize to the nucleus and inhibit the transcriptional activity of NFAT and p21 when overexpressed (Wang et al., 2010). ZNF57 is of particular interest because it was the only candidate tested where all 5 unique shRNAs altered the relative GFP expression in both reporter lines. Further investigation will be needed to confirm each of these, and the mechanisms by which they are acting.

Methods

Construction of miRNA sensitive reporters

To generate the plasmids for stably integrated reporter expression, pAWG014, a modified lentiviral pGFP-Neo plasmid (Clonetech) encoding the PGK promoter upstream of a GFP coding sequence followed by the HMGA2 3′ UTR (PGK-GFP-HGMA2 3′ UTR), was digested with ClaI and EcoRI overnight to remove the PGK promoter. The vector was gel-purified and isolated using Promega Wizard SV Gel and PCR Clean-up kit. The EF1α promoter was amplified from pAWG048 using primers AWG245 and AWG246 which contained ClaI and EcoRI overhangs, respectively. The amplicon was purified using the Promega Wizard PCR clean-up kit and ligated into the pAWG014 backbone using T4 DNA ligase, followed by transformation into DH5α cells, to generate pRMW027 (EF1α-GFP-HMGA2-wt).

pRMW027 was digested with SanDI overnight and gel-purified using Promega Wizard SV Gel and PCR Clean-up kit. A SanDI-NotI linker was generated by annealing oligos RMW038 and RMW039 and ligating them into the linearized pRMW027 with T4 DNA ligase, followed by transformation into DH5α cells generating pRMW032. The pRMW032 was digested with SanDI and NotI overnight in SanDI reaction buffer and gel-purified. The antibiotic resistance gene, neomycin phosphotransferase II (neo) driven by the PGK promoter (PGK-neo) was PCR amplified from pMSCV-neo (Clonetech) with NotI and SandI tailed primers (RMW035 and RMW036) so that when ligated into the pRMW032 vector, the PGK-neo sequence was oriented in the opposite direction of the EF1α-GFP-HMGA2-wt cassette, generating pRMW034 (PGK-neo, EF1α-GFP-HMGA2-wt). In the same manner, the blasticidin s deaminase (bsd) antibiotic resistance gene driven by PGK (PGK-bsd) was amplified and cloned into pRMW032 to generate pRMW053 (PGK-bsd, EF1α-GFP-HMGA2-wt).

To generate the DsRed reporter plasmid, the DsRed coding sequence and mutant HMGA2 3′ UTR (DsRed-HMGA2-mut) were amplified from pAWG017 with EcoRI and BamHI tailed primers (AWG141 and AWG126). pRMW052 was digested with EcoRI and BamHI overnight and gel purified. The PCR amplicon was ligated into the pRMW034 and transformed in the same manner as before, to create pRMW072 (PGK-bsd, EF1α-DsRed-HMGA2-mut).

The mini reporter plasmids were generated using pRMW072 and pRMW034 which were both digested with MreI and BamHI overnight. The mini-HMGA2-wt 3′ UTR and mini-HMGA2-mut 3′ UTRs were amplified from pAWG005 and pAWG007 respectively, with MreI and BamHI tailed primers (RMW119 and RMW120). Ligation of the mini-HMGA2-wt 3′ UTR into the pRMW034 vector generated pRMW118 (PGK-neo, EF1α-GFP-mini-HMGA2-wt), while ligation of the mini-HMGA2-mut 3′ UTR into pRMW072 generated pRMW125 (PGK-bsd, EF1α-DsRed-mini-HMGA2-mut). The single site reporter plasmids pJAM1 (PGK-neo, EF1α-DsRed-ss-mut) and pJAM4 (PGK-bsd, EF1α-GFP-ss-wt) were generated in a similar manner. A 100 nt region surrounding the 6th let-7 target site in the HGMA2 3′ UTR was PCR amplified from wt and mut HMGA2 3′ UTRs in pRMW34 and pRMW72 with pEAF40-100mer-3′ and pEAF40-100mer-5′ and cloned into the appropriate vector.

Stable Integration of Reporter Constructs

Cell lines stably expressing GFP and DsRed reporters were generated according to TRC protocols (https://portals.broadinstitute.org/gpp/public/). Viral supernatants were harvested 2 days post transfection and the viral titer was determined using a Resazurin (Life Technologies) cell viability assay. A549 cells were seeded in 6-well plates, 4x10^5 cells/well and infected after 24 hours with virus. Before infection, the standard growth media (DMEM 10% FBS and 1% Pen/Strep) was replaced with standard growth media containing 8 µg/ml polybrene, which

increases transduction efficiency. Virus was then added to the wells to obtain an MOI of 0.3. After 48 hours, cells were selected in growth media containing either 10 µg/ml Blasticidin or 1 mg/ml neomycin for reporter constructs. The media supplemented with antibiotics was replaced every 2 days until all uninfected cells were dead (~10 days). Cells infected with a single reporter constructs were passaged and seeded in 6-well plates at $4x10^5$ cells/well so that the infection protocol could be repeated with the complementary reporter. After the second round of antibiotic selection the reporter cells lines, now containing a GFP and DsRed reporter, were serially diluted in a 96-well plate and grown up from a single cell to generate a homogenous population.

RNAi Screens

Infection and Selection

Reporter cells were thawed and passaged in media containing selection drugs (10µg/ml Blasticidin and 1 mg/ml Geneticin) ~2 weeks prior to infection. I checked on my cells regularly the week before infection to make sure they are in log growth and that there will be enough cells for the day of infection. It is important to maintain a high representation of the library to minimize false positives and obtain consistent reproducible results. To ensure adequate representation of the small shRNA library I aimed to infect 500x more cells than the complexity of the library, but for the larger libraries, like the 55k shRNA library, I aimed for at least 200x coverage.

A library of 55,000 shRNAs requires infecting $1.1x10^7$ cells. I infected at an MOI of 0.3 which results in ~26% of cells becoming infected. Low MOI infections reduce the number of cells infected with more than one shRNA, but they require many more cells up front. To successfully infect $1.1x10^7$ cells at an MOI of 0.3 requires infecting $3.13x10^7$ cells. For A549 cells doubling time is approximately 24 hours, so $1.57x10^7$ cells need plated.

Day 0, I plated 6 T-150 (150 cm^2) tissue culture flasks with $3x10^6$ reporter cells/flask in standard growth media (DMEM, 10% FBS, 1% P/S). Day 1, after 24 hours the growth the media is replaced with media containing 8 µg/ml polybrene, a molecule that enhances transduction efficiency. The lentiviral library is then thawed and pipetted into each flask to achieve an MOI of 0.3. Day 3, 48 hours post infection, the growth media is replaced with growth media containing 3 µg/ml puromycin to kill the uninfected cells. The minimal concentration of puromycin required needs to be tested beforehand. Day 5, (a lot of cells should be dying) flasks are rinsed with PBS and each T-150 flask is split and passage into 2 T-150 flasks in standard growth media. Day 7, growth media is replaced with standard growth media. Day 8, cells were prepared for FACS.

FACS Preparation

To prepare cells for FACS sorting, three T-150 (150 cm^2) tissue culture flasks were prepared at a time. Each T-150 is rinsed with 10 ml's of PBS twice to remove any dead cells. Then 3 mls of Trypsin was added to each flask and incubated at 37 °C for 3 minutes, until cells began slough off and detach from the flask when tapped. Using a 10ml pipette pool the trypsinized cells into a single flask, and then add 6 mls of standard growth media to rinse the empty flasks before adding it to the third, inactivating the trypsin (15 ml total). The reason for this is to get as many cells to sort as possible while keeping the volumes of media used low. Once pooled together, the cells are pipetted against the flask ~8 times to ensure they are in single cell suspension. Cell clumping can ruin a screen, so it is important not to over or under trypsinize the cells. To remove any cells that may be clumping, the cells are slowly pipetted through a 40-micron strainer into a 50 ml falcon tube. 10 mls of standard growth media is used to rinse all three T-150s and passed through the 40-micron strainer maximizing the number of cells recovered (25 mls total). If there is any concern about whether the cells are in single cell

suspension, 10 µl of the pooled cells are pipetted onto a hemocytometer and verified visually using a microscope. The total number of cells is calculated by measuring the concentration with a Moxi z cell counter (Orflo).

Before proceeding these steps were repeated for 3 more flasks, yielding two 50 ml falcon tubes containing 25 mls of cells with a known number of cells in each. The 50 ml tubes are balanced and centrifuged at 180g for 3 minutes. After spinning the supernatant is aspirated, the pellets are resuspended in 6 mls of room temperature PBS and then transferred to a 15 ml falcon tube. To prevent loss of cells, each 50 ml falcon tube is washed with an additional 4 mls of PBS and added to the corresponding 15 ml tube (10 mls in each). The 15ml tubes are centrifuged again at 180g for 3 min to pellet the cells. The supernatant is aspirated, and then the cells are resuspended in PBS to achieve a final concentration of ~30 million cells/ml.

The tubes are then placed on ice and then the process is repeated for the remaining 6 flasks. After all of the T-150 flasks are prepared, the cells are pooled into a single 50 ml falcon tube, mixed and then aliquoted into three 15 ml falcon tubes and placed back on ice. It is important to maintain proper background representation to which you can compare the sorted populations. I set aside ~20 million cells (~400x) to be used as the unsorted background population. These cells are kept on ice until the genomic DNA isolation can be performed after sorting. It is also important to prepare some other controls, especially if the fluorescent proteins you have chosen requires compensation. I prepare an A549 (nonfluorescent) cell line, an A549 cell line expressing DsRed only, an A549 cell line expressing GFP only, and an uninfected reporter line from smaller T-25 flasks.

FACS Sorting and Defining Gates

Sorting was performed on a BD FACSAria II using a 100-micron nozzle, a 488 nm laser with a 505 long pass filter and 525/50 band pass filter for GFP detection and a 532 nm laser with a 575/25 bandpass filter for DsRed detection. Nonfluorescent A549 cells were used to define the GFP and DsRed positive population. At least 100,000 nonfluorescent cells were recorded and the quadrant gate was used to define what is GFP and DsRed positive, so that less than 0.25% of the nonfluorescent cells fall in the GFP or DsRed positive quadrants. Then 1 million reporter cells infected with the lentiviral library are recorded so that the sort gates could be drawn based on the shape of the population. The percentage of cells sorted by a particular gate may change slightly as the sort is performed, especially if the sorting gates are not drawn on a sufficiently large number of cells. Recording at least a million cells improves the accuracy of the gates drawn. For the screens, all sorting gates were drawn as sub-gates of the GFP and DsRed positive population. The top 3% gate was drawn based on the contour plot of GFP and DsRed expression of the $A549^{HGMA2}$ reporter cells on a log scale. The first line was drawn so that it ran through the center of the population following the diagonal line that the majority of population fell on. The gate was drawn like a triangle to capture any cells with higher GFP relative to DsRed and then the gate was adjusted upward until only ~3.0% of the parent population (GFP and DsRed positive cells) fell within the gate. The top 4-10% gate was drawn as a trapezoid maintaining the angle of the line drawn through the center of the contour plot in top 3% gate. The gate was placed as close as possible to the top 3% gate and the other side was adjusted until only ~7% of the parent population (GFP and DsRed positive cells) fell within it. Cells were sorted into 200 µl of PBS in 1.5 ml Eppendorf tubes and kept on ice until the gDNA isolation could be performed. After FACS sorting, cells were incubated on ice for 0.5-3 hours before isolating the genomic DNA (gDNA).

High Throughput Sequencing

The gDNA was extracted from the sorted and unsorted populations of cells using the Gentra Puregene Tissue Kit (Qiagen), following the protocol for DNA purification from cultured cells. After gDNA isolation the concentration of gDNA was measured by qubit assay of dsDNA. The TRC protocol for shRNA PCR for Illumina sequencing was performed on the gDNA using the TaKaRa Ex Taq polymerase (TaKaRa).

To reduce amplification bias, I determined the number of cycles for amplification of gDNA by first performing a test amplification of gDNA. I performed a 1:4 serial dilution of gDNA and performed a test amplification of 28 cycles. Then 25 µl of the 100 µl PCR reaction was run on an 8% native acrylamide gel and stained with SYBR® Gold (Invitrogen) which revealed that 0.0625 µg of gDNA produced a visible amplicon after 28 cycles, and that the PCR reaction was not yet exhausted. I was able to recover the PCR product following an overnight gel extraction in 0.4 ml of 300mM NaCl followed by ethanol precipitation. For amplification of shRNAs from my sorted cells gDNA, I adjusted the cycle number accordingly. For example, if I recovered 1 µg of gDNA I performed 24 cycles of amplification.

Each sample was amplified using PAGE purified oligos containing flow cell binding sites P5 and P7. All samples were amplified using a mix of oligos containing the P5 binding site as well as a staggered region necessary to maintain sequence diversity across the flow cell. The P5 primers (JM27-34) were mixed equimolar and diluted to 10 µM and used in all PCR reactions. For each sample a P7 primer (JM41-60) containing a unique index sequence was added to individual reactions to barcode each sample. After amplification and gel extraction, libraries were sequenced on a NextSeq500 to obtain >1,000x coverage of each sample.

Identification of Enriched Gene Targets

The software and reference files discussed below were obtained from the The RNAi Consortium (TRC) (https://portals.broadinstitute.org/gpp/public/). PoolQ is a tool that was used to quantify the results from the high throughput sequencing data. PoolQ produces a scores file containing a matrix of read counts for the shRNAs present in each sample. Reads were mapped to the TRC 90K Human Pool reference file, CP0003_reference_20150109.csv, using exact match settings. There are always some unexpected shRNAs present in the scores file due to errors caused by PCR and Illumina sequencing. To reduce the presence of these sequences for the subsequent RIGER analysis, the scores file was used to produce a log normalized score file for only shRNAs with >1 read in the background population. Otherwise unexpected shRNAs with high scores in the sorted populations that are absent in the background population will appear to be highly enriched, leading to an artificially high RIGER gene score which will produce a larger number of false positives. The log normalized score for each shRNA is calculated by taking the raw read count of an shRNA in a population and dividing it by the total number of reads for all shRNAs in that population, then multiplied by a constant factor of 1 million, adding 1, and taking the log2 base. Before the log normalized scores file can be used for RIGER analysis, an additional column is added which contains the gene target of the shRNA construct. This can be found in the CP0003_lax_gene_20170705.chip file.

RIGER, an extension of the GENE-E tool, was used to generate an enrichment score for each gene based on the enrichment of shRNAs in sorted populations compared to the unsorted background. RIGER was run using the following parameters: log fold change metric for ranking hairpins, 1×10^6 permutations, weighted sum rank scoring algorithm, and gene score adjusted to accommodate variation in harpin set size. The RIGER output generates a ranked list of genes.

For the screens of the 55k shRNA library the top 220 ranked genes (top 2%) were focused on for further validation. Due to reagent availability and time limitations, I did not validate all genes ranked in the top 2% of each population. I limited validation experiments to 68 genes ranked in the top 220 genes of either the top 3% or top 4-10% populations, and 11 genes ranked in the top 220 genes of either the bottom 3% or bottom 4-10% populations. To reduce the number of false positives I only considered candidate genes with greater than 1 read per million (RPM) in RNA-seq data of A549 cells. This included the top ~13,000 most abundant transcripts in the cell, which are more likely to be playing a functional role than very lowly expressed transcripts. I also excluded genes that showed enrichment in both a top (3% or 4-10%) and bottom (4-10% or 3%) population.

shRNA Knockdowns for Validation Experiments

shRNA hairpin plasmids designed by the TRC were used to generate shRNA virus according to TRC protocols for low throughput viral production in 6-well tissue culture plates (https://portals.broadinstitute.org/gpp/public/). Viral supernatants were harvested 2 days post transfection and stored at -80 °C.

Day 0, A549 reporter lines ($A549^{HMGA2}$ or $A549^{mini}$) were seeded in 0.5 ml of standard growth media at 40,000 cells/well in 24-well plates. Day 1, 24 hours after seeding cells were ~30% confluent and the media was replaced with media containing 8 µg/ml polybrene (Sigma) to increase transduction efficiency. Wells were then infected with virus (30 µL) to obtain an MOI of >1. Day 3, 48 hours post infection, the media was replaced with selection media containing 3 µg/ml puromycin (Sigma) and grown for 2 days. Day 5, after 48 hours in selection media, 1/4 cells were passaged and seeded in new 24-well plates in standard growth media. Day 8 cells were prepared for FACS, analyzed and recorded.

CRISPR

Guide RNAs were designed using IDT's Predesigned Alt-R® CRISPR-Cas9 guide RNA Tool (https://www.idtdna.com/site/order/designtool/index/CRISPR_PREDESIGN) and NEB base changer was used to design oligos (JM358 and JM359) for cloning sequences encoding the gRNA into the pX459V2.0-SpCas9-HF1 vector, a gift from Yuichiro Miyaoka (Addgene, catalog# 108293) (Kato-Inui et al., 2018).

HAP1 cells (Horizon) were seeded in 24-well plates in standard antibiotic free growth media (IMDM, 10% FBS) at 70,000 cells/well. After 24 hours cells were ~60-70% confluent and the transfection mixture containing, 2 µg of pX459V2.0-SpCas9-HF1, 250 ng pJAM4-GFP, and 6 µL of Turbofectin reagent (Origene) was brought to 50 µL with OPTI-MEM (Thermofisher) was added directly to each well. 48 hours after infection cells were single cell sorted on a BD FACSAria II into 96-well plates containing 100 µl of high growth media (IMDM, 1% P/S, 20% FBS) and grown from a single cell. After ~10 days cells were passaged into 24-well plates. The cells were grown to ~70% confluence, ½ were passaged and gDNA was isolated from the remaining ½ using the Gentra Puregene Tissue Kit (Qiagen). The cells were then genotyped by PCR amplification of the region targeted by the guide RNA with JM382 and JM383 which were expected to produce an amplicon of 299bp. The PCR products were run on a 2% TBE agarose gel to identify cell line whose amplicon contained 10-20bp deletions. The region targeted by the guide RNA in knockout lines KRBOX4#3, KRBOX4#4, and KRBOX4#6 was then amplified with JM443 and JM444 and cloned into pRMW34, digested overnight with KpnI, using Gibson assembly. Sanger sequencing of the knockouts was then performed with AWG147s or JM383 to confirm deletion or mutation of the KRBOX4 coding sequence.

Luciferase Assays

HAP1 cells and HAP1 knockout lines were seeded in 500 µL of antibiotic free media (IMDM with 10% FBS) at 70,000 cells/well in 24-well plates. After 24 hours, cells were transfected with 0.4 µL Lipofectamine 2000 (Invitrogen) following the manufacturer's protocols and using 20 ng of pmirGLO-derived reporter plasmids pKK512 or pJAM9 which contain either the wildtype or mutant mini 3′ UTR, respectively. Carrier DNA (pUC19) was also included to increase transfection efficiency (20 ng/well). Cells were harvested 30 hours post-transfection by removing media and washing once with PBS, and frozen at -80°C. Luciferase assays were performed using the Promega Dual-Luciferase Reporter Assay kit and Veritas Microplate Luminometer (Turner Biosystems) according to manufacturer's protocols. The firefly relative light units (RLU) were normalized to the *renilla* RLUs for each well to control for transfection efficiency. A second normalization of pKK512 to pJAM9 transfected wells was then performed to assess the effect of the knockdown on the target of miRNA mediated repression.

RNA-sequencing and small RNA-sequencing of Knockdowns

Knockdowns

Knockdowns were performed using two unique shRNAs targeting KRBOX4 (KRBOX4_8203 and KRBOX4_8205), and a negative control shRNA targeting LacZ (LacZ_2224). For each knockdown 2 technical replicates were performed. A549 cells were seeded at 40,000 cells/well in 0.5 ml of standard growth media (DMEM, 10% FBS, 1% P/S) in 24-well plates. After 24 hours wells were ~30% confluent and the media was replaced with media containing 8 µg/ml polybrene and infected with 40 µL of virus (MOI >1). After 48 hours the media was replaced with media containing 3 µg/ml puromycin. After 48 hours of puromycin selection, cells were passaged into 6 well plates in standard growth media. The following day

the media was replaced with standard growth media. At 7 days post infection, cells were washed twice with PBS and 1 ml of Trizol was added to each well and pippetted up and down to lyse the cells, before transfer to a 1.5ml Eppendorf and stored at -20 overnight.

RNA Isolation

The following day, 0.2 ml of chloroform was added to the 1.5 ml Eppendorf tube and shaken vigorously by hand for 15 seconds. The tubes were incubated for 2–3 minutes @RT and then centrifuged at 12,000 × g for 15 minutes at 4°C. The aqueous phase was removed and placed into a new tube and 0.5 ml of 100% isopropanol. Tubes were then incubated @RT for 10 minutes and centrifuged at 12,000 × g for 10 minutes at 4°C. The supernatant was removed, leaving only the RNA pellet and washed with 1 ml of 75% ethanol, vortexed briefly, and centrifuged at 7500 × g for 5 minutes at 4°C. The supernatant was poured off and washed with 75% ethanol once more. The RNA pellet was then allowed to air dry for 5–10 minutes and then resuspend in RNase-free water (25 µL) by passing the solution up and down several times and stored at -80 °C.

Library Preparation

Preparation of the mRNA-Seq libraries were generated using the TruSeq RNA Sample Preparation Kit v2 (Illumina) from 100ng of total RNA. Libraries were sequenced on the Illumina NextSeq500. Reads were aligned to the genome with Tophat ((Trapnell et al., 2009) and differential expression was found with CuffDiff ((Trapnell et al., 2013).

Preparation of small RNA libraries were generated using the TruSeq Small RNA Prep Kit (Illumina) from 100ng of total RNA. Libraries were sequenced on the Illumina NextSeq500. miRNA expression was found with MirDeep2 (Friedländer et al., 2012), and differential expression was found with EdgeR (Robinson et al., 2010).

Analysis of let-7 targets

Targets of let-7 were predicted as genes with a context + score <−0.2 from TargetScan (Garcia et al., 2011b; Grimson et al., 2007a). The background set was comprised of genes not predicted to be targets of let-7. Two-sided Kolmogorov–Smirnov tests were used to determine significant differences in targeting.

Supplemental Figures and Tables

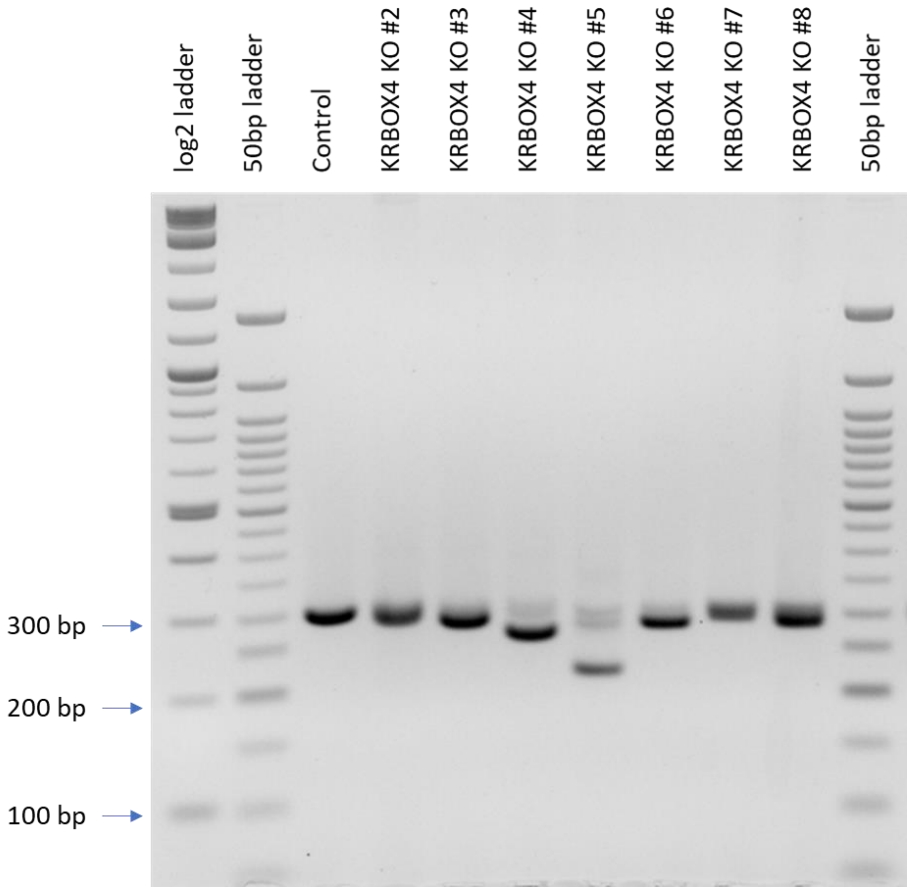

Figure 2-12: Genotyping of KRBOX4 knockout lines

PCR amplification of KRBOX4 with JM382 and JM383 was expected to produce an amplicon of 299bp. KRBOX4 KO #3, #4, #5, #6 and #8 each appear to contain a deletion. Although KRBOX4 #5 looked the most promising it did not survive and only KRBOX4 KO #3, #4, and #6 were successfully sequenced.

Figure 2-13: KRBOX4 KO Lines

Sanger sequencing and alignments of KRBOX4 knockout lines to exon 3 of KRBOX4 show that KRBOX4 KO #3 and #4 contain deletions of 13 and 30 bp respectively while KRBOX4 KO #6 contains several point mutations.

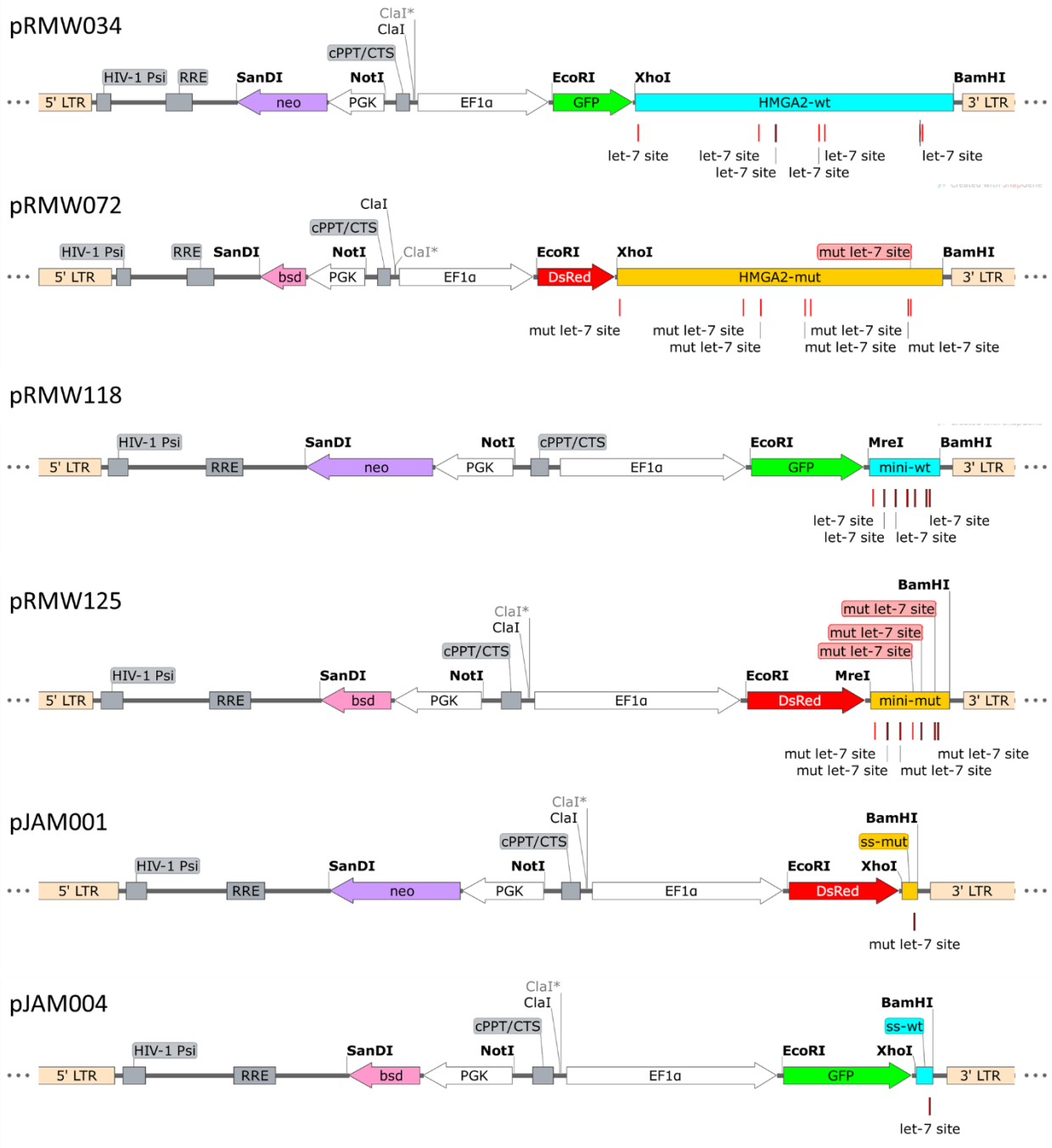

Figure 2-14: Reporter Plasmids

The plasmid maps show the region flanked by the 5′ and 3′ long terminal repeats (LTR). This is the region that is integrated into the cell upon infection. Each reporter construct contains a fluorescent protein with a different 3′ UTR driven by the EF1α promoter and an antibiotic resistance gene driven by the PGK promoter.

Table 2-5 Plasmids

Plasmid	Description		Reporter Cell Line
pRMW034	PGK-neo, eIF1α-GFP-HMGA2-wt	>	$A549^{HMGA2}$
pRMW072	PGK-bsd, eIF1α-DsRed-HMGA2-mut		
pRMW118	PGK-neo, eIF1α-GFP-mini-wt	>	$A549^{mini}$
pRMW125	PGK-bsd, eIF1α-DsRed-mini-mut		
pJAM-001	PGK-bsd, eIF1α-GFP-ss-wt	>	$A549^{ss}$
pJAM-004	PGK-neo, eIF1α-DsRed-ss-mut		
pJAM-009	pmiRGlo; firefly-mini-mut 3' UTR		
pKK512	pmiRGlo; firefly-mini-wt 3' UTR		
pmiRGlo	Dual luciferase plasmid (*renilla* and firefly)		
pX459V2.0-SpCas9-HF1	Encodes a guide RNA and SpCas9 for CRISPR		

Table 2-6 Oligos

For primers JM27-60 the prime region is color coded to indicate the flowcell attachment sequence (green), the Illumina sequencing primer (blue), vector binding sequence (red) and the stagger regions and index sequences are black.

Primer	Sequence 5'->3'	Purpose
JM27	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTTCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 1
JM28	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTATCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 2
JM29	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTGATCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 3
JM30	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTCGATCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 4
JM31	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTACGATCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 5
JM32	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAGAATCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 6
JM33	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTGACGACATCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 7
JM34	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTTGGACACATCTTGTGGAAAGGACGA	shRNA Amplification P5 Primer Stagger 8
JM41	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 1
JM42	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 2
JM43	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 3
JM44	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 4
JM45	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 5
JM46	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 6
JM47	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 7
JM48	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 8
JM49	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 9
JM50	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 10
JM51	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 11
JM52	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 12
JM53	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 13
JM54	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 14
JM55	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 15
JM56	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 16
JM57	CAAGCAGAAGACGGCATACGAGATCTCTACGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 17
JM58	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 18
JM59	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 19

JM60	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTATTCTTTCCCCTGCACTGT	shRNA Amplification P7 Primer Index 20
JM358_KRBOX4_g1_fwd	ggaaaggacgaaacaccGCCATCCAGGACTCTTCACCgttttagag ctagaaatagcaag	gRNA cloning into pX459V2.0-SpCas9-HF1
JM359_KRBOX4_g1_rev	cttgctatttctagctctaaaacGGTGAAGAGTCCTGGATGGCggt gtttcgtcctttcc	gRNA cloning into pX459V2.0-SpCas9-HF1
JM_382_KRBOX4_fw_2	gccatcaccttcagcctgc	genotyping KRBOX4
JM_383_KRBOX4_rev_2	catgtccacaagtccgaggc	genotyping KRBOX4
JM443_KRBOX4_g1/g2_fwd	agtgcgatccataggtacCCTTGCTTGCTATGTGAATG	genotyping KRBOX4
JM444_KRBOX4_g1/g2_rev	cattggtcttaaaggtacACAGTTCTGCTCATCAGAGC	genotyping KRBOX4
AWG125Z	atatctcgaGCCGACATTCAATTTCTACCTC	HMGA2; pIS1_Hmga2UTR-fwd, XhoI-tailed
AWG126Z	atatggatccAGTCAGAGGGCACACAAAGG	HMGA2; pIS1_Hmga2UTR-rev, BamHI-tailed
AWG141Z	atatgaattcGCCACCATGGACAACACC	pLVX_DsRed_MonoC1, EcoRI tailed
AWG142Z	atactcgagttactagggcccGGACTGGGAGCCGGAGTG	pLVX_DsRed_MonoC1, XhoI tailed; ApaI, stops added
AWG147S	TCGTTGGGAGTGAATTAGCC	pLVX sequencing primer
AWG161Z	atggatcCGTTCCATTGGACACAAGCC	BamHI tailedX
AWG245Z	attaatcgaTCTAGGTCTTGAAAGGAGTGCC	ClaI-tailed
AWG246Z	atatgaattcGAGTCGTATTAGTACCAAGCTAATTCC	EcoRI-tailed
pEAF40-100mer-3'	atatggatccTATACTGGGACTTAAGAGGTAGTAG	100nt region that includes the 6th let-7 site in HMGA2, BamHI tailed
pEAF40-100mer-5'	atatctcgaGCACACCGAGCAATTTCC	100nt region that includes the 6th let-7 site in HMGA2, XhoI tailed
RMW038	gacccatatatgcggccgcat	SanDI-NotI linker forward
RMW039	gtcatgcggccgcatatatgg	SanDI-NotI linker reverse
RMW119	atatcgccggcgGGGCGCCGACATTCAATTTC	mini-HMGA2 UTR primer-fwd, MreI tailled
RMW120	atatggatccGCGTTCCATTGGACACAAGCC	

Table 2-7: Individual Gene Knockdowns

In the table below the individual shRNAs used for the validation experiments are listed, as well as their gene targets, TRCN Clone ID, the target sequence, and the values for the log2 fold change of the GFP/DsRed geometric mean normalized to LacZ_2223 upon knockdown in the A549^{HMGA2} reporter (red) and A549^{mini} reporter (orange). Cells highlighted in blue are genes that had 2 or more hairpins pass the cutoff for validation in the A549HMGA2 reporter line, and those in bold were successfully validated in the A549^{mini} reporter line.

shRNA	Gene Target	TRCN Clone ID	RNAi Target Seq	Category	A549^{HMGA2} GFP/DsRed	A549^{mini} GFP/DsRed
DICER1_1258	DICER1	TRCN0000051258	GCTCGAAATCTTACGCAAATA	Control	1.3672	1.8158
DICER1_1260	DICER1	TRCN0000051260	CCACACATCTTCAAGACTTAA	Control	1.4609	1.2419
LacZ_2223	lacZ	TRCN0000072223	TGTTCGCATTATCCGAACCAT	Control	N/A	N/A
LacZ_2224	lacZ	TRCN0000072224	CGCGATCGTAATCACCCGAGT	Control	0.1420	0.2403
LUCIFERASE_2243	LUCIFERASE	TRCN0000072243	CTTCGAAATGTCCGTTCGGTT	Control	-0.1216	0.1396
ADIPOR2_9649	**ADIPOR2**	TRCN0000299649	CCTTGCTTCATCTACTTGATT	Top	-0.1658	
ADIPOR2_9672	**ADIPOR2**	TRCN0000299672	GCCAGATATAAGGCTCAGAAA	Top	0.5622	0.7077
ADIPOR2_9673	**ADIPOR2**	TRCN0000299673	GCTCTTCTCTAAACTGGATTA	Top	0.1377	
ADIPOR2_9674	**ADIPOR2**	TRCN0000299674	GCCTGAGTGGAATCATTCCTA	Top	0.6779	1.6976
AK5_5867	AK5	TRCN0000335867	GGCAGTTGACAACAAGTTATT	Top	0.0425	
AK5_5948	AK5	TRCN0000335948	CGATATGGATTCCAATACATT	Top	0.4691	
AK5_5949	AK5	TRCN0000335949	CCTTGATCCTTCGATGATATT	Top	0.0805	
AK5_5950	AK5	TRCN0000335950	GCACAGCTATTGACTCTATTT	Top	0.6774	
BFAR_6429	BFAR	TRCN0000296429	TACTCTCAGTTCTGGAATTAT	Top	0.8709	
BFAR_8233	BFAR	TRCN0000308233	GGACTGGTTGGAGGTCCATTA	Top	0.2051	
BLM_3475	**BLM**	TRCN0000273475	GACGCTAGACAGATAAGTTTA	Top	0.3782	
BLM_3476	**BLM**	TRCN0000273476	ACCGAATCTCAATGTACATAG	Top	0.2379	
BLM_4905	**BLM**	TRCN0000004905	CCATCAATGATTGGGATGATA	Top	0.9976	0.7300
BLM_4906	**BLM**	TRCN0000004906	CGCTTATGTGATGCTCGGAAA	Top	-0.1984	
BLM_4908	**BLM**	TRCN0000004908	CGAAGGAAGTTGTATGCACTA	Top	1.4903	1.9283
BRIP1_9913	**BRIP1**	TRCN0000049913	GCTAAGAAACAGGCATCCATA	Top	0.3650	
BRIP1_9914	**BRIP1**	TRCN0000049914	CGTCAGAACTTGGTGTTACAT	Top	0.4103	0.6023
BRIP1_9915	**BRIP1**	TRCN0000049915	GCTCTTATTCTAGTGGATGAT	Top	1.3910	2.6724
BRIP1_9916	**BRIP1**	TRCN0000049916	CCAAGGAACTTCACGTCATTT	Top	1.4285	2.2532
CAMKMT_0045	**CAMKMT**	TRCN0000150045	CTGTTAACTGATGGGAATGAA	Top	0.9740	1.1245
CAMKMT_0103	**CAMKMT**	TRCN0000130103	CTACTGCCTCAAGCACAATAA	Top	1.0863	2.1533
CAMKMT_8319	**CAMKMT**	TRCN0000148319	CCATCAGAAATGTGCAAGACA	Top	1.0419	1.2833
CAMKMT_9144	**CAMKMT**	TRCN0000129144	GCCTGTCTGTGACTTGAATTT	Top	0.6029	
CAMKMT_9944	**CAMKMT**	TRCN0000129944	GAAGAAACGTATCAAGTGCAT	Top	0.2170	
CHD9_0235	CHD9	TRCN0000230235	GATCAATCCAGGGACTATAAA	Top	0.3498	
CHD9_0236	CHD9	TRCN0000230236	TTCGTACGTGGACTGATATTA	Top	0.0615	
CHD9_8966	CHD9	TRCN0000218966	TGCATACCAGCGTACTAATAA	Top	0.0698	
CLTA_0295	**CLTA**	TRCN0000380295	AGACAGTTATGCAGCTATTTC	Top	-0.0231	
CLTA_1091	**CLTA**	TRCN0000381091	TGGAAGCCCTTGATGCCAATT	Top	-0.2749	
CLTA_1588	**CLTA**	TRCN0000291588	AGCCTGAAAGTATCCGTAAAT	Top	0.9714	2.5354
CLTA_1929	**CLTA**	TRCN0000381929	AGTCTAGCAAGCAGGCCAAAG	Top	0.7816	2.1943
CLTA_2229	**CLTA**	TRCN0000382229	ACTACATCTGCAATATCTTAA	Top	0.2324	
CRLF3_0918	CRLF3	TRCN0000300918	GCACTTCGGAACGATTCTGAA	Top	0.4884	
CRLF3_0919	CRLF3	TRCN0000300919	CCAGTACAGATAGAAGAACTA	Top	0.3245	
CRLF3_0986	CRLF3	TRCN0000300986	CCATTGAACAGGAGACCATTA	Top	0.2266	
DCTD_4022	DCTD	TRCN0000344022	TCATCATCCAGGCAGGTATAA	Top	0.5101	
DCTD_4023	DCTD	TRCN0000344023	GAATAAGCTGGACACCAAATA	Top	-0.0525	
DCTD_4083	DCTD	TRCN0000344083	GCTGCGAGGCTCCTGTTTAAT	Top	-0.1480	
DCTD_4084	DCTD	TRCN0000344084	TTGTCGGGATTGGGTACAATG	Top	-0.0627	
DCTN3_0895	DCTN3	TRCN0000300895	TGATGCCTCTAAGCTGCAATT	Top	0.4397	
DCTN3_0896	DCTN3	TRCN0000300896	GATCTGATCAAGTACCTGGAT	Top	-0.0006	
DCTN3_0956	DCTN3	TRCN0000300956	CAATGCTTCTCTCCAAGCAAT	Top	-0.0324	
DCTN3_0959	DCTN3	TRCN0000300959	GCAGGACCAGTGTGTGGAAAT	Top	0.8137	

shRNA	Gene Target	TRCN Clone ID	RNAi Target Seq	Category	A549^{HMGA2} GFP/DsRed	A549^{mini} GFP/DsRed
DHX29_1238	DHX29	TRCN0000051238	CCTCAAATACAAGCCACTATT	Top	0.1586	
DHX29_1240	DHX29	TRCN0000051240	CCTGGTAGTATGCCCTACAAT	Top	0.4507	0.4372
DHX29_1241	DHX29	TRCN0000051241	CGTGTACCTTTGGAGGAATTA	Top	1.3234	1.1376
DHX29_1242	DHX29	TRCN0000051242	CCTAAGTATCAGAAACTTCTA	Top	0.3027	
DHX57_1884	DHX57	TRCN0000241884	GACCTAGCAACAGTAACATAA	Top	-0.2633	
DHX57_1886	DHX57	TRCN0000241886	ACGGAACTGTTATCGGATATA	Top	0.3757	
DHX57_1887	DHX57	TRCN0000241887	ACTATACCAGGTCGTACATTT	Top	0.1224	
DHX57_3050	DHX57	TRCN0000073050	CCTTGACATTTGCGGAAACTT	Top	0.9902	
DHX57_3052	DHX57	TRCN0000073052	GCATCACTAGAGCATCTCCTT	Top	0.0099	
DPYSL3_6848	DPYSL3	TRCN0000046848	CGGCATAGATGGAACCCATTA	Top	-0.0532	
DPYSL3_6849	DPYSL3	TRCN0000046849	GCTGATATTTACATGGAAGAT	Top	-0.0826	
DPYSL3_6850	DPYSL3	TRCN0000046850	GCGGCAGAGTACAACATCTTT	Top	-0.1966	
DPYSL3_6851	DPYSL3	TRCN0000046851	CGACTATGTCTACAAGCGCAT	Top	-0.5351	
DPYSL3_6852	DPYSL3	TRCN0000046852	GCTCAAGTTCATGCTGAGAAT	Top	0.4705	
E2F8_7428	E2F8	TRCN0000017428	GCCGCAAAGACAAGTCTTTAA	Top	-0.2673	
E2F8_7430	E2F8	TRCN0000017430	CGCCGAGCAGATTATGATGAT	Top	0.0996	
EDN1_2481	EDN1	TRCN0000272481	CCATGAGAAACAGCGTCAAAT	Top	0.4076	
EDN1_4761	EDN1	TRCN0000284761	AGACAAACATGCAAGTAAAGA	Top	-0.3770	
EDN1_4762	EDN1	TRCN0000284762	GCAGTTAGTGAGAGGAAGAAA	Top	1.0305	
EGFL8_1891	EGFL8	TRCN0000281891	CAACCAGTGCCAGCATACTCA	Top	-0.1441	
EGFL8_2299	EGFL8	TRCN0000272299	TACAAGCCCTACCTGACCTTG	Top	-0.1120	
EGFL8_2301	EGFL8	TRCN0000272301	CTCGGCGTCAATCATCGATAA	Top	-0.1382	
EGFL8_2352	EGFL8	TRCN0000272352	CAGAAACCGGACCCACTAATC	Top	-0.2911	
EGFL8_4725	EGFL8	TRCN0000284725	CCTCTGCTCGCACCATTGTTT	Top	0.0139	
FAM105A_4204	FAM105A	TRCN0000294204	AGCAGCTGCACTGATACATTT	Top	-0.2181	
FAM105A_4205	FAM105A	TRCN0000294205	TGGACTAGAGCAGATTGATAT	Top	-0.1818	
FAM105A_4273	FAM105A	TRCN0000294273	TATCAAGTCACTGAAGTTTAT	Top	-0.0363	
FAM105A_6921	FAM105A	TRCN0000286921	CCACTACCACATTCCAGTCTT	Top	-0.1648	
FAM105A_6922	FAM105A	TRCN0000286922	GCTCTCAGATCAGTGTTATTT	Top	0.2422	
FDPS_0493	**FDPS**	TRCN0000310493	GTCCTGGAGTACAATGCCATT	Top	0.2464	
FDPS_9718	**FDPS**	TRCN0000299718	GCCATGTACATGGCAGGAATT	Top	0.6806	0.5765
FDPS_9721	**FDPS**	TRCN0000299721	CCCAGAGATAGGAGATGCTAT	Top	-0.3870	
FDPS_9722	**FDPS**	TRCN0000299722	CCAGCAGTGTTCTTGCAATAT	Top	1.2087	2.8043
GCAT_4579	GCAT	TRCN0000034579	CCTTAACTTCTGTGCCAACAA	Top	0.6754	
GCAT_4580	GCAT	TRCN0000034580	CCAGAGGTTCCGTAGTAAGAT	Top	0.4971	
GCAT_4581	GCAT	TRCN0000034581	CATGCTGAAGAGAGGCATCTT	Top	0.5002	
GCAT_4583	GCAT	TRCN0000034583	CAGAGCATCCACAAGAATCTA	Top	0.8399	
GIT2_4544	**GIT2**	TRCN0000364544	GAGTACTCATCAACACGAAAT	Top	1.3381	1.1733
GIT2_4545	**GIT2**	TRCN0000364545	ATAACGGTGCTAACTCTATAT	Top	0.8390	1.1345
GIT2_4546	**GIT2**	TRCN0000364546	TTGCTGGACCCTGCGTCTATT	Top	1.3773	2.3132
GIT2_4618	**GIT2**	TRCN0000364618	ACAACGATCAGCCCGACTATG	Top	1.1545	1.2523
GTPBP4_6790	GTPBP4	TRCN0000286790	GCTGGAGAGTATGACAGTGTA	Top	0.4232	
GTPBP4_6791	GTPBP4	TRCN0000286791	GCTCATCGAGTGGAAACCAAA	Top	0.2671	
GTPBP4_7974	GTPBP4	TRCN0000047974	GCTGGAGAGTATGACAGTGTA	Top	1.3472	
HMOX1_5248	HMOX1	TRCN0000045248	GCTGAGTTCATGAGGAACTTT	Top	0.4594	
HMOX1_5249	HMOX1	TRCN0000045249	CGGGCCAGCAACAAAGTGCAA	Top	-0.0438	
HMOX1_5250	HMOX1	TRCN0000045250	ACAGTTGCTGTAGGGCTTTAT	Top	0.3572	
HMOX1_5251	HMOX1	TRCN0000045251	CAACAAGGAGAGCCCAGTCTT	Top	-0.0279	
INSR_0379	INSR	TRCN0000000379	GCTCTGTTACTTGGCCACTAT	Top	0.4960	
INSR_1221	INSR	TRCN0000121221	AGAGACATCTATGAAACGGAT	Top	0.1433	
INSR_4847	INSR	TRCN0000194847	CCTATACATTTCTGTTCATCT	Top	0.2400	
INSR_5618	INSR	TRCN0000195618	CCTTACCAAGGCCTGTCTAAT	Top	0.0075	
INSR_9622	INSR	TRCN0000199622	GCGCATGTGCTGGCAATTCAA	Top	-0.7042	
KRBOX4_8203	KRBOX4	TRCN0000018203	GCACAGAATTTGATTCTGTAA	Top	1.0495	
KRBOX4_8205	KRBOX4	TRCN0000018205	GCATTTATCTGAGCACAGAAT	Top	1.1866	
LIX1L_7578	**LIX1L**	TRCN0000297578	CCTATTTGCTAAGAGATCCCT	Top	-0.0216	
LIX1L_7579	**LIX1L**	TRCN0000297579	CCTTCCGATTCATGCTGGAAT	Top	0.4817	0.7324
LIX1L_8863	**LIX1L**	TRCN0000278863	TGCTGCAAAGATTGCGCTAAT	Top	0.5601	1.1462
LIX1L_8927	**LIX1L**	TRCN0000278927	CTGGAATCCAACAAGGGCAAA	Top	0.6952	0.6126
LIX1L_8931	**LIX1L**	TRCN0000278931	GAACATCCTTCCCGAAGAATC	Top	0.5099	

shRNA	Gene Target	TRCN Clone ID	RNAi Target Seq	Category	A549^{HMGA2} GFP/DsRed	A549^{mini} GFP/DsRed
LUC7L_4895	LUC7L	TRCN0000314895	CGAGGTCTGTTCAGCCTACCT	Top	-0.4540	
LUC7L_4966	LUC7L	TRCN0000314966	GAGGAAATCAGTGCGGAAGTT	Top	1.0684	
LUC7L_4973	LUC7L	TRCN0000314973	GCTGAATGTGATCGGAGAACT	Top	-0.2432	
MDH1_5198	MDH1	TRCN0000275198	GCCTATAATTCTTGTGCTGTT	Top	0.2144	
MDH1_5199	MDH1	TRCN0000275199	CTTCAGTTGCTTGACTCGTTT	Top	-0.0584	
MDH1_5248	MDH1	TRCN0000275248	CCCTGTTGTAATCAAGAATAA	Top	0.0612	
MDH1_5325	MDH1	TRCN0000285325	AGAATCTAAATGTCGTCTTTG	Top	0.6263	
MKRN2_4382	MKRN2	TRCN0000004382	CCTATGGAACTCGGTGCAGAT	Top	0.1967	
MKRN2_4383	MKRN2	TRCN0000004383	CAGCAAACAAAGTGGAAGTAT	Top	0.5664	
MKRN2_4384	MKRN2	TRCN0000004384	CACGTACTGTTTGTCCTGCAT	Top	0.3858	
MKRN2_4385	MKRN2	TRCN0000004385	GTCCAGAATGCCGTGTGATAT	Top	0.7706	
NAB1_0049	**NAB1**	TRCN0000020049	CCACACAAAGAGGAGGAAATT	Top	1.0631	1.6643
NAB1_0050	**NAB1**	TRCN0000020050	CCGAATGCCTAATTTACAGAA	Top	0.0855	
NAB1_0052	**NAB1**	TRCN0000020052	GCTCGACAGATTTCTCGAGAA	Top	0.4028	
NAB1_0053	**NAB1**	TRCN0000020053	CAACTCTGTGTGAAGGATAAT	Top	0.5971	1.0789
NDUFAB1_7151	NDUFAB1	TRCN0000027151	CCGTGTTCTTTACGTATTGAA	Top	0.0776	
NDUFAB1_7184	NDUFAB1	TRCN0000027184	TGCAGATAAGAAGGATGTATA	Top	0.3041	
NDUFAB1_7196	NDUFAB1	TRCN0000027196	GTCCACAAGAAATTGTAGATT	Top	0.2833	
NDUFAB1_7202	NDUFAB1	TRCN0000027202	CGCAGGTTCCTGGTAGAGTTA	Top	0.1007	
NF1_4915	NF1	TRCN0000234915	TAAGCGGCCTCACTACTATTT	Top	0.1071	
NF1_8778	NF1	TRCN0000238778	TGCGCAGTTAGCAGTTATAAA	Top	0.3509	
NF1_9715	NF1	TRCN0000039715	CCTCACAACAACCAACACTTT	Top	0.3580	
NTRK3_2309	NTRK3	TRCN0000002309	CACTACAACAATGGCAACTAT	Top	0.2773	-0.2581
NTRK3_2313	NTRK3	TRCN0000002313	CACGGACATCTCAAGGAATAT	Top	1.1840	0.6773
NTRK3_4821	NTRK3	TRCN0000194821	CCAATCTACCTGGACATTCTT	Top	0.9673	0.2706
PCMTD2_0365	PCMTD2	TRCN0000020365	GCATGAAGAGTACATGAAGAA	Top	0.8255	
PCMTD2_5858	PCMTD2	TRCN0000275858	TGATGAGCTGATAGATAATTT	Top	0.3723	
PCMTD2_5919	PCMTD2	TRCN0000275919	AGTCTTTGCCAGTCGGATTTC	Top	0.4791	
PDPK1_0413	PDPK1	TRCN0000010413	CAACATAGAGCAGTACATTCA	Top	0.4896	
PDPK1_1479	PDPK1	TRCN0000001479	CGGAAGGGTTTATTTGCAAGA	Top	0.3386	
PDPK1_1541	PDPK1	TRCN0000221541	CAAAGTTCTGAAAGGTGAAAT	Top	0.2469	
PDPK1_4792	PDPK1	TRCN0000234792	GGATAAGCGGAAGGGTTTATT	Top	-0.0389	
PDPK1_4794	PDPK1	TRCN0000234794	ACGCCTAACAGGACGTATTAT	Top	0.7439	
PGRMC2_1298	**PGRMC2**	TRCN0000061298	GCACTTAGAGATGAATATGAT	Top	0.1536	
PGRMC2_1299	**PGRMC2**	TRCN0000061299	CAGATTTGAATGCAGTACAAA	Top	0.5039	1.8912
PGRMC2_1301	**PGRMC2**	TRCN0000061301	GCGGGTCCATATGGAATATTT	Top	0.9732	2.7529
PGRMC2_1302	**PGRMC2**	TRCN0000061302	CGCGGTCAATGGGAAAGTCTT	Top	0.5411	
PIGK_8940	**PIGK**	TRCN0000288940	GCTCAGATAATACACCAGAAA	Top	0.3861	
PIGK_8941	**PIGK**	TRCN0000288941	CGCTACAATGAGCTACTGTTT	Top	0.4353	
PIGK_8943	**PIGK**	TRCN0000288943	GCAACTGCTTAATGGCACTAA	Top	1.0102	1.0140
PIGK_8944	**PIGK**	TRCN0000288944	GCTAGTCATATCGAGGATCAA	Top	0.9613	1.2904
PQBP1_5575	PQBP1	TRCN0000285575	CCATGGACCCTAGCTCATACT	Top	-0.1516	
PQBP1_6435	PQBP1	TRCN0000276435	AGTAATGCAGATGCTGAAGAA	Top	0.0342	
PQBP1_6438	PQBP1	TRCN0000276438	TCGAGAGCGTGGCTATGACAA	Top	-0.6975	
PQBP1_6489	PQBP1	TRCN0000276489	CCCTTACTACTGGAATGCAGA	Top	0.2366	
PQBP1_6490	PQBP1	TRCN0000276490	GAGATCATTGCCGAGGACTAT	Top	0.5599	
PTPN9_1117	PTPN9	TRCN0000381117	CCTTTCCCTTTGGCTAGATAA	Top	0.0497	
PTPN9_8263	PTPN9	TRCN0000338263	CCATACTCAGACAGATTACAT	Top	0.2091	
PTPN9_8264	PTPN9	TRCN0000338264	GAGTGCCAAGCACGTTTATAA	Top	0.2286	
PTPN9_8265	PTPN9	TRCN0000338265	CGTGGAGAACATGAATCATTA	Top	0.0811	
PTPN9_8266	PTPN9	TRCN0000338266	AGTCCGGGAGAGGATTCAAAT	Top	0.0626	
RAB10_9189	RAB10	TRCN0000029189	GCCTTCAATACTACCTTTATT	Top	-0.0326	
RAB10_9190	RAB10	TRCN0000029190	CCAATGAAGATGTGGAAAGAA	Top	1.0701	
RAB10_9191	RAB10	TRCN0000029191	CCTCACGTTAGCTGAAGATAT	Top	0.3611	
RAB10_9193	RAB10	TRCN0000029193	CAAGTGTGATATGGACGACAA	Top	-0.0913	
RETSAT_4583	**RETSAT**	TRCN0000064583	GCACGTATAAAGCACTCTAAT	Top	1.2979	2.6707
RETSAT_4585	**RETSAT**	TRCN0000064585	CCGAATCCTTTCTCCGAAGAT	Top	0.8427	2.6793
RETSAT_4586	**RETSAT**	TRCN0000064586	GCCAGGAAGAAGGTTCTCAAA	Top	0.7597	1.9682
RFC4_4233	RFC4	TRCN0000074233	GCATCTGATGAACGTGGAATA	Top	1.6277	
RFC4_4234	RFC4	TRCN0000074234	CCTGACCTCTAGATGTTCAAA	Top	0.4533	

shRNA	Gene Target	TRCN Clone ID	RNAi Target Seq	Category	A549^{HMGA2} GFP/DsRed	A549^{mini} GFP/DsRed
RFC4_4235	RFC4	TRCN0000074235	CCGATTCTGTCTTATCTGTAA	Top	0.2611	
RFC4_4236	RFC4	TRCN0000074236	CAGAAGTCTATTATCACAGAA	Top	0.0423	
RPP38_9873	RPP38	TRCN0000049873	GCTGTTAGTTCTGGTGTGTAA	Top	0.0052	
RPP38_9874	RPP38	TRCN0000049874	CGGCAGGCTTCTGTAACATTA	Top	-0.0107	
RPP38_9875	RPP38	TRCN0000049875	GCAGCATTGCTGTTGATATTA	Top	0.6071	
RPP38_9876	RPP38	TRCN0000049876	CCTCACACTTGATTCAGTTAA	Top	0.0885	
SECISBP2L_1879	SECISBP2L	TRCN0000061879	CGCCTGTTTCTACAGAGTATA	Top	0.3244	
SECISBP2L_1880	SECISBP2L	TRCN0000061880	CCTGATTACTTGTTACCCATT	Top	-0.1635	
SECISBP2L_1881	SECISBP2L	TRCN0000061881	CGATCTCTGAAGTAAATGAAA	Top	0.2317	
SKIV2L2_6268	SKIV2L2	TRCN0000296268	AGCAGGACCACTTCGTCAAAT	Top	0.2717	
SKIV2L2_6300	SKIV2L2	TRCN0000296300	AGCTAAAGGAATGTGAGATTT	Top	0.0721	
SKIV2L2_6301	SKIV2L2	TRCN0000296301	AGAAGTTGCTTGGGTTATATT	Top	-0.0525	
SKIV2L2_7086	SKIV2L2	TRCN0000307086	CCCAGGATAGAAGAGTCAATA	Top	0.0830	
SKIV2L2_8175	SKIV2L2	TRCN0000308175	GATTGATAAAGGCCTTATTTG	Top	-0.1999	
SLC17A5_0184	SLC17A5	TRCN0000300184	GCTATCGTAGTTGCACACTTT	Top	1.1731	
SLC17A5_0185	SLC17A5	TRCN0000300185	GCCGTTGCTTTCCTAACTATA	Top	0.4932	
SLC17A5_0186	SLC17A5	TRCN0000300186	GCTCGTTACAACTTAGCAATT	Top	0.1119	
SLC17A5_0567	SLC17A5	TRCN0000310567	CGTTAGTGGATATGGTAGATT	Top	1.2049	
SLC30A7_2863	**SLC30A7**	TRCN0000042863	GCAGATCCTATCTGTTCAATT	Top	0.6465	
SLC30A7_2864	**SLC30A7**	TRCN0000042864	GCCATGTCGATCATTGCCATA	Top	0.4888	
SLC30A7_2865	**SLC30A7**	TRCN0000042865	GCAACTGCTTAGGCTTGATTT	Top	1.6208	2.5343
SLC30A7_2867	**SLC30A7**	TRCN0000042867	CCTGAACCTCTCTTTCGCTTT	Top	0.6811	0.9092
SLU7_1145	SLU7	TRCN0000001145	CCAAGATAGATGCAGCTGATA	Top	0.4617	
SLU7_1146	SLU7	TRCN0000001146	CGAACATTGAAAGCCCAGAAA	Top	-0.8052	
SLU7_1147	SLU7	TRCN0000001147	CACAGGAGATACCATTTCAAT	Top	0.3204	
SLU7_1149	SLU7	TRCN0000001149	GATCCAAATTCTGCCTACTAT	Top	-0.0228	
SMAD2_5815	SMAD2	TRCN0000295815	CAAGTACTCCTTGCTGGATTG	Top	0.2667	
SMAD2_5870	SMAD2	TRCN0000295870	CATGATCCAGTATCACAGTAT	Top	-0.5549	
SMAD2_8592	SMAD2	TRCN0000288592	GCGTTGCTCAAGCATGTCATA	Top	0.1112	
SMAD2_8651	SMAD2	TRCN0000288651	CGATTAGATGAGCTTGAGAAA	Top	0.4763	
SMAD2_8652	SMAD2	TRCN0000288652	CCTAAGTGATAGTGCAATCTT	Top	-0.1923	
SMNDC1_0599	SMNDC1	TRCN0000010599	CGTCATTGGTTGGCTTCAGTA	Top	-0.0189	
SMNDC1_1136	SMNDC1	TRCN0000001136	GCTTCTACTCAACCTACTCAT	Top	-0.0310	
SMNDC1_1137	SMNDC1	TRCN0000001137	CAGTGTTATGAAGCGGAGATT	Top	0.5779	
SMNDC1_1138	SMNDC1	TRCN0000001138	CTGGTAAAGTTGGAGTAGGAA	Top	0.0646	
SPTAN1_3668	SPTAN1	TRCN0000053668	GCCCATGAAGACAGCTTCAAA	Top	1.0378	
SPTAN1_3669	SPTAN1	TRCN0000053669	GCCACTGAACTGAAAGGAATA	Top	0.2123	
SPTAN1_3670	SPTAN1	TRCN0000053670	GCCATTGTTAAGCTGGATGAA	Top	0.2387	
SPTAN1_3672	SPTAN1	TRCN0000053672	GCTCTAAACACAGACAATTAT	Top	-0.1711	
SUB1_4968	SUB1	TRCN0000014968	GAACAGATTTCTGACATTGAT	Top	1.8686	
SUB1_4969	SUB1	TRCN0000014969	ACATTGATGATGCAGTAAGAA	Top	1.1779	0.3973
SUB1_4970	SUB1	TRCN0000014970	GTACGTTAGTGTTCGCGATTT	Top	0.4506	0.3353
SUB1_4971	SUB1	TRCN0000014971	AGCCAGCTGAAGGAACAGATT	Top	1.2821	0.9695
SUB1_4972	SUB1	TRCN0000014972	GAAGGAACAGATTTCTGACAT	Top	0.5212	0.1716
TBPL1_4262	TBPL1	TRCN0000274262	CCAAACCTTGCTGTAATATAA	Top	0.0757	
TBPL1_4264	TBPL1	TRCN0000274264	ACCTAGAATTACAGCTACAAT	Top	0.3849	
TBPL1_4265	TBPL1	TRCN0000274265	CTGCTGTGTGCTATCGGATAA	Top	-0.0426	
TBPL1_4329	TBPL1	TRCN0000274329	TGTGGAACAGATTTACCCATT	Top	-0.0597	
TCF7L2_1897	**TCF7L2**	TRCN0000061897	CCCACATAAAGAAACCTCTTA	Top	-0.0065	
TCF7L2_2843	**TCF7L2**	TRCN0000262843	TAGCTGAGTGCACGTTGAAAG	Top	0.7975	1.3089
TCF7L2_2847	**TCF7L2**	TRCN0000262847	CCTTTCACTTCCTCCGATTAC	Top	0.5456	0.4048
TCF7L2_2848	**TCF7L2**	TRCN0000262848	AGAGAAGAGCAAGCGAAATAC	Top	0.9053	2.0531
TEAD3_5948	TEAD3	TRCN0000015948	GCCACTGTTCTGCGCTTTAAT	Top	0.6269	
TEAD3_5949	TEAD3	TRCN0000015949	CCATGTCTACAAGCTCGTCAA	Top	-0.1123	
TEAD3_5950	TEAD3	TRCN0000015950	CCTGGAGTATTCAGCCTTCAT	Top	-0.4109	
TEAD3_5951	TEAD3	TRCN0000015951	CTCTGCTGATAGCATGACCAT	Top	0.3392	
TEAD3_5952	TEAD3	TRCN0000015952	GAGTTGATTGCACGCTATATT	Top	-0.1075	
TPCN1_5041	TPCN1	TRCN0000005041	CGCCTCTACTTCATGACCTTT	Top	0.4394	
TPCN1_5814	TPCN1	TRCN0000335814	GACGATCATTGTCGCCTTTAT	Top	-0.1468	
TPCN1_5898	TPCN1	TRCN0000335898	GCTGAGGTTGTTTAAGTTGAA	Top	-0.4192	

shRNA	Gene Target	TRCN Clone ID	RNAi Target Seq	Category	A549^{HMGA2} GFP/DsRed	A549^{mini} GFP/DsRed
TPCN1_5961	TPCN1	TRCN0000335961	GCCTACCTCTTTGCACACAAT	Top	0.9844	
TRIM26_3005	**TRIM26**	TRCN0000433005	CAATTACGCCTAAGATCATTT	Top	0.2458	
TRIM26_4047	**TRIM26**	TRCN0000004047	ACACCGAGAGAAGCTGCACTA	Top	1.0040	1.1814
TRIM26_4049	**TRIM26**	TRCN0000004049	GCTGAGAGACTTGGAATATAA	Top	1.4533	1.2120
TTC23_5126	TTC23	TRCN0000005126	GCTGCTGTTAGCATCACTCAT	Top	0.3454	
TTC23_5127	TTC23	TRCN0000005127	GTACCCATATTGAGAGAATTA	Top	0.5040	
TTC23_5128	TTC23	TRCN0000005128	GCGTAGCACTGACAAGAATTT	Top	0.5978	
TTF1_5519	TTF1	TRCN0000275519	GCCAAGGTTAGAACCTGCAAA	Top	-0.5128	
TTF1_5520	TTF1	TRCN0000275520	GCACAAGGTGTCGCTATTAAA	Top	-0.6231	
TTF1_5586	TTF1	TRCN0000275586	GTGAAGACCTGTCACTATTAA	Top	-0.3061	
TWIST1_0539	TWIST1	TRCN0000020539	GCATTCTGATAGAAGTCTGAA	Top	0.3002	
TWIST1_0540	TWIST1	TRCN0000020540	CCTGAGCAACAGCGAGGAAGA	Top	0.7414	
U2AF2_0475	U2AF2	TRCN0000350475	GTGTTGGCTGTGCAGATTAAC	Top	0.1371	
U2AF2_4892	U2AF2	TRCN0000314892	CGACGAGGAGTATGAGGAGAT	Top	0.2674	
U2AF2_4894	U2AF2	TRCN0000314894	CCTTTGACCAGAGGCGCTAAA	Top	-0.1268	
U2AF2_4964	U2AF2	TRCN0000314964	GAAGAAGAAGGTCCGTAAATA	Top	0.2392	
XK_0178	XK	TRCN0000060178	CCATTGTACTATGCTTTCTAA	Top	0.6954	
XK_0179	XK	TRCN0000060179	GCTTGTATTCTATCAGTTCTT	Top	0.6349	
XK_0180	XK	TRCN0000060180	GCAAACTAATCACCCACCGAT	Top	0.4186	
XK_0181	XK	TRCN0000060181	GCCATTCTCTTCATGCTTGTA	Top	0.7596	
ZNF185_8021	ZNF185	TRCN0000108021	CCAAAGATTACCCTAGAACAT	Top	0.1109	
ZNF185_8022	ZNF185	TRCN0000108022	TGTCCAAAGATTACCCTAGAA	Top	0.2126	
ZNF185_8023	ZNF185	TRCN0000108023	CCCTGCTGATAGGAAGAGCAA	Top	0.2094	
ZNF57_6468	**ZNF57**	TRCN0000016468	CCAGAGCATTCCAAGGTCATT	Top	0.6531	0.5773
ZNF57_6469	**ZNF57**	TRCN0000016469	CCAGATCATTCCGAGGTCATT	Top	1.1902	2.5553
ZNF57_6470	**ZNF57**	TRCN0000016470	GCCATCCATCAAATGCCAGAT	Top	1.2322	2.4597
ZNF57_6471	**ZNF57**	TRCN0000016471	CGAATGTCAGTCATACTCAAA	Top	0.8449	1.5120
ZNF57_6472	**ZNF57**	TRCN0000016472	CAGGATATGTACGGGCAAGAA	Top	1.2707	1.8009
ZNF581_8100	**ZNF581**	TRCN0000108100	GAAGGAGGAATCCGTGAGTAA	Top	1.1736	2.1652
ZNF581_8101	**ZNF581**	TRCN0000108101	CCCAACCACTACCTGCTTATT	Top	0.1482	
ZNF581_8104	**ZNF581**	TRCN0000108104	CTCAAGGGTCTTCGAGTACAT	Top	0.9043	2.4166
ZNF7_4559	ZNF7	TRCN0000014559	CACAGCTTACAATACATCAAA	Top	0.8957	-0.1621
ZNF7_4560	ZNF7	TRCN0000014560	GCTCTGCATTGGGAAATTAAT	Top	0.9023	
ZNF7_4561	ZNF7	TRCN0000014561	TCCACCTTTGTGAGCCGTAAA	Top	0.2917	
ZNF7_4562	ZNF7	TRCN0000014562	TGGTTGTAGTTCACGGCTTAT	Top	0.0573	
ZNF768_5250	ZNF768	TRCN0000285250	GTGCAGAGTTCTGACGAAATG	Top	-0.0339	
ZNF768_7383	ZNF768	TRCN0000017383	GCCCAGAATTTGAAGCTCAAA	Top	0.6109	
ZNF768_7384	ZNF768	TRCN0000017384	GCAGTTTGAGATGCTTCAGAA	Top	1.0145	
ZNRF1_3044	ZNRF1	TRCN0000073044	ACAACGATGATGTGCTGACTA	Top	0.4712	
ZNRF1_3045	ZNRF1	TRCN0000073045	CGAGATGGAAATGCACTTTAT	Top	0.9062	
ZNRF1_3047	ZNRF1	TRCN0000073047	CAGCTCGCATAGTGGTTTCAA	Top	0.3591	
B3GNT1_5081	B3GNT1	TRCN0000235081	CTATGCCCTGGTGATCGATGT	Not Enriched	-0.4324	
B3GNT1_5083	B3GNT1	TRCN0000235083	CAGCACAATAAGATCCTATAT	Not Enriched	-0.0181	
B3GNT1_5084	B3GNT1	TRCN0000235084	TTCTGTAAGCTTCGTACAAAT	Not Enriched	0.0793	
CECR5_3150	CECR5	TRCN0000353150	CTGTGCCTGGAAACCATTTAC	Not Enriched	0.0331	
CECR5_3151	CECR5	TRCN0000353151	ACGTGGTGAATGACGTGAATG	Not Enriched	0.1122	
CECR5_3643	CECR5	TRCN0000333643	AGGTGGATGCAGACCAAGTTA	Not Enriched	-0.0890	
CECR5_3644	CECR5	TRCN0000333644	GATAACCCTATGTCTGACGTA	Not Enriched	-0.0492	
CECR5_4867	CECR5	TRCN0000344867	TTTCCCTGCTGGGTAGCATTT	Not Enriched	-0.0811	
EIF2S1_0040	EIF2S1	TRCN0000220040	TATGGTGCCTATGATGCATTT	Not Enriched	0.4726	
EIF2S1_0042	EIF2S1	TRCN0000220042	GCCCATTAAGATTAATCTAAT	Not Enriched	-0.2030	
EIF2S1_0292	EIF2S1	TRCN0000160292	CTGGAATACAACAACATTGAA	Not Enriched	0.4253	
EIF2S1_3119	EIF2S1	TRCN0000163119	GCAGTCTCAGACCCATCTATT	Not Enriched	-0.0321	
KIAA1967_0846	KIAA1967	TRCN0000290846	CGGGTCTTCACTGGTATTGTT	Not Enriched	0.2222	
KIAA1967_0847	KIAA1967	TRCN0000290847	GCCAAAGGAAAGGATCTCTTT	Not Enriched	-0.1574	
KIAA1967_0848	KIAA1967	TRCN0000290848	GCATTGATTTGAGCGGCTGTA	Not Enriched	0.9818	
KIAA1967_0849	KIAA1967	TRCN0000290849	CCTCTGAAGCAGATTAAGTTT	Not Enriched	0.4120	
KIAA1967_0850	KIAA1967	TRCN0000290850	CCCATCTGTGACTTCCTAGAA	Not Enriched	-0.0866	
MRE11A_8324	MRE11A	TRCN0000338324	GCTGATGACCTTATGAGTATA	Not Enriched	-0.1361	
MRE11A_8391	MRE11A	TRCN0000338391	ACGACTGCGAGTGGACTATAG	Not Enriched	-0.1152	

shRNA	Gene Target	TRCN Clone ID	RNAi Target Seq	Category	A549^{HMGA2} GFP/DsRed	A549^{mini} GFP/DsRed
MRE11A_8393	MRE11A	TRCN0000338393	TGTTGGTTTGCTGCGTATTAA	Not Enriched	0.1863	
MRE11A_8395	MRE11A	TRCN0000338395	ACGGGAACGTCTGGGTAATTC	Not Enriched	0.0566	
MTIF2_0857	MTIF2	TRCN0000150857	GAAGAGATCAATTCTACGGTT	Not Enriched	0.0324	
MTIF2_1530	MTIF2	TRCN0000151530	CCTATTATCCTTGCCGTAAAT	Not Enriched	1.4757	
RDBP_1033	RDBP	TRCN0000291033	CAAAGTCAACATAGCCCGAAA	Not Enriched	-0.0975	
RDBP_6857	RDBP	TRCN0000306857	ACCCAGATTGTCTACAGTGAT	Not Enriched	0.1450	
RDBP_6858	RDBP	TRCN0000306858	GTGGTGTCAAACGCTCACTAT	Not Enriched	0.4413	
RDBP_7260	RDBP	TRCN0000307260	CTGGATTCCTTGTGCCTCATA	Not Enriched	-0.0871	
SF3B14_1630	SF3B14	TRCN0000221630	GCATCAACACAGATCCACCAA	Not Enriched	-0.3352	
SF3B14_1631	SF3B14	TRCN0000221631	ACCTGAAGTAAATCGGATATT	Not Enriched	0.3311	
SF3B14_1632	SF3B14	TRCN0000221632	CGGGATTCAATGTTTGTAACA	Not Enriched	0.0423	
SF3B14_1633	SF3B14	TRCN0000221633	CTGAAACTAGAGGAACAGCTT	Not Enriched	0.0934	
G3BP2_7548	G3BP2	TRCN0000047548	CGGGAGTTTGTGAGGCAATAT	Bottom	0.1437	
G3BP2_7550	G3BP2	TRCN0000047550	CCACAAAGTATTATCTCTGAA	Bottom	-0.0255	
G3BP2_7552	G3BP2	TRCN0000047552	CCGGAATATTTACACAGGTTT	Bottom	-0.0645	
GATAD1_6008	GATAD1	TRCN0000016008	CCTGAGTCAGTTTCCACTATA	Bottom	-0.1568	
GATAD1_6009	GATAD1	TRCN0000016009	CCTGCAATAACAATTAAGGAA	Bottom	0.2055	
GATAD1_6010	GATAD1	TRCN0000016010	GCTCAGAAACACTAAATACAA	Bottom	0.0019	
GATAD1_6011	GATAD1	TRCN0000016011	CAAGGGAGTATATTACCAAAT	Bottom	0.0232	
HELZ_2269	HELZ	TRCN0000162269	CCTGAGAAAGTGCTTAGTGAA	Bottom	0.3518	
HELZ_8502	HELZ	TRCN0000158502	CCCAGAAAGAAGATATTCTTA	Bottom	0.7401	
HELZ_9972	HELZ	TRCN0000159972	CGAGGAATATACTTTGTGTAA	Bottom	-0.3209	
HNRNPD_3283	HNRNPD	TRCN0000293283	TCGAAGGAACAATATCAGCAA	Bottom	-0.0559	
HNRNPD_3352	HNRNPD	TRCN0000293352	AGTAAGAACGAGGAGGATGAA	Bottom	0.2265	
HNRNPD_3353	HNRNPD	TRCN0000293353	AGAGTGGTTATGGGAAGGTAT	Bottom	0.0227	
HNRNPH1_6085	HNRNPH1	TRCN0000236085	AGCTGAAGTTAGAACTCATTA	Bottom	0.3573	
HNRNPH1_6086	HNRNPH1	TRCN0000236086	TTGCCCTTTGCCACGTTAAAT	Bottom	0.2005	
HNRNPH1_6088	HNRNPH1	TRCN0000236088	CAAGTCAAACAACGTTGAAAT	Bottom	0.0854	
HNRNPH1_7227	HNRNPH1	TRCN0000257227	GATCCACCACGAAAGCTTATG	Bottom	0.1800	
LARP6_1898	LARP6	TRCN0000141898	GCCACCAGAATCTCTTTCTGA	Bottom	0.1697	
LARP6_2044	LARP6	TRCN0000122044	CTCTCTTCTTTGGTATGGAAT	Bottom	0.8444	1.3222
LARP6_2919	LARP6	TRCN0000142919	CGTTAAGCTACTCACATCCTT	Bottom	0.7695	1.3905
LARP6_5086	LARP6	TRCN0000145086	GCAAGTGTATGGATTATTCCT	Bottom	0.2424	
MTFMT_5076	MTFMT	TRCN0000035076	CCCTATGGCATATTGAATGTT	Bottom	0.3416	
MTFMT_5077	MTFMT	TRCN0000035077	GCTCTTATTCCAGGATCAGTA	Bottom	-0.0951	
MTFMT_5078	MTFMT	TRCN0000035078	CGATCAGTGATGCTCAAGAAA	Bottom	-0.3771	
SENP5_2312	SENP5	TRCN0000062312	GACTGCTGTTACGAAGTGTAT	Bottom	0.4190	
SENP5_4413	SENP5	TRCN0000294413	CACTATTGTTACCTCAAATTT	Bottom	-0.5695	
SENP5_7079	SENP5	TRCN0000287079	CCAACACTTGTGCATTCTGAA	Bottom	-0.3564	
SENP5_7080	SENP5	TRCN0000287080	CCTTACCAGAACATCGTTCTA	Bottom	-0.4368	
TRIM27_0261	TRIM27	TRCN0000280261	GAGAAGATTGTTTGGGAGTTT	Bottom	-0.1669	
TRIM27_0262	TRIM27	TRCN0000280262	CAGGGCTGAAAGAATCAGGAT	Bottom	-0.2355	
TRIM27_0318	TRIM27	TRCN0000280318	GCCCTACTTCAGTCTGAGTTA	Bottom	0.2974	
TRIM27_0319	TRIM27	TRCN0000280319	GAATTAAGAGAGGCTCAGTTA	Bottom	0.1280	
WTAP_1422	WTAP	TRCN0000231422	ATGGCAAGAGATGAGTTAATT	Bottom	0.0377	
WTAP_1423	WTAP	TRCN0000231423	GGCAAGTACACAGATCTTAAC	Bottom	0.4860	
WTAP_1424	WTAP	TRCN0000231424	GCAAGAGTGTACTACTCAAAT	Bottom	-0.0336	
ZNF224_3068	ZNF224	TRCN0000013068	GCAGAGAAACACCTCTCAAAT	Bottom	0.3942	
ZNF224_3069	ZNF224	TRCN0000013069	CAATGGATATAAACCATCCTT	Bottom	0.0258	
ZNF224_3070	ZNF224	TRCN0000013070	CCGATGTGATACGTGTGATAA	Bottom	0.3010	
ZNF224_3071	ZNF224	TRCN0000013071	GTGGGAAGAGATTTACTCAAA	Bottom	-0.1647	

Table 2-8: Small shRNA Screen Library Preparation

Sort Date	Cell Line	Population	# Sorted	gDNA (µg) Recovered	µg/ PCR Rxn	# of PCR Rxns	Total µg amplified	Cycle #	Primer	# Reads
10/28/16	mini	Top 3%	1.21E+05	0.61	0.13	1	0.13	29	JM41	1.66E+07
10/28/16	mini	Top 4-10%	2.64E+05	1.09	0.25	1	0.25	28	JM42	3.13E+07
10/28/16	mini	Center 50%	7.15E+05	4.08	0.25	2	0.50	28	JM44, JM45	2.17E+07
10/28/16	mini	Unsorted	1.00E+07	82.96	1.00	5	5.00	26	JM47-JM51	1.19E+08
2/27/17	mini	Top 3%	2.91E+04	0.10	0.05	2	0.10	26	JM41	3.00E+07
2/27/17	mini	Top 4-10%	1.55E+05	0.45	0.22	2	0.45	24	JM42	5.96E+07
2/27/17	mini	Center 50%	1.03E+06	4.47	1.12	4	4.47	22	JM43, JM44	2.38E+07
2/27/17	mini	Unsorted	2.00E+07	171.54	4.00	4	16.00	20	JM46, JM47	1.62E+07
3/17/17	mini	Top 3%	1.41E+04	0.35	0.17	2	0.35	24	JM48	8.63E+06
3/17/17	mini	Top 4-10%	8.31E+04	0.83	0.42	2	0.83	23	JM49	4.04E+07
3/17/17	mini	Center 50%	6.46E+05	6.28	1.05	6	6.28	22	JM50, 51	3.75E+07
3/17/17	mini	Unsorted	1.00E+07	99.71	4.00	4	16.00	20	JM54, JM55	2.49E+07
7/11/17	HMGA2	Top 3%	5.70E+04	0.04	0.04	1	0.04	27	JM55	1.66E+07
7/11/17	HMGA2	Top 4-10%	1.68E+05	0.43	0.43	1	0.43	24	JM54	1.09E+07
7/11/17	HMGA2	Center 50%	1.60E+06	13.80	8.00	3	24.00	19	JM50	2.31E+07
7/11/17	HMGA2	Bottom 3%	5.51E+04	0.10	0.10	1	0.10	26	JM46	2.04E+07
7/11/17	HMGA2	Unsorted	3.00E+06	118.70	8.00	3	24.00	19	JM60	2.01E+07

Table 2-9: 55K shRNA Screen Library Preparation

Sort Date	Cell Line	Population	# Sorted	Recovered gDNA (µg)	gDNA (µg) / PCR Rxn	# of PCR Rxns	Total µg amplified	# of Cycles	Primer	Sequencing Reads
12/19/18	HMGA2	Top 3%	2.63E+05	1.16	0.25	4	1.00	26	JM41	3.17E+07
12/19/18	HMGA2	Top 4-10%	6.67E+05	6.28	2.00	3	6.00	23	JM44	3.66E+07
12/19/18	HMGA2	Bottom 4-10%	7.97E+05	6.36	2.00	3	6.00	23	JM45	4.14E+07
12/19/18	HMGA2	Bottom 3%	2.10E+05	0.36	0.18	2	0.36	26	JM47	3.48E+07
12/19/18	HMGA2	Unsorted	2.00E+07	108.80	6.80	13	88.40	21	JM50	2.91E+07
1/30/18	HMGA2	Top 3%	3.58E+05	1.16	0.58	2	1.16	25	JM41	5.59E+06
1/30/18	HMGA2	Top 4-10%	7.74E+05	4.76	2.38	2	4.76	23	JM44	6.89E+06
1/30/18	HMGA2	Bottom 4-10%	9.37E+05	4.44	2.22	2	4.44	23	JM47	7.67E+06
1/30/18	HMGA2	Bottom 3%	5.00E+05	1.75	0.88	2	1.75	25	JM48	7.56E+06
1/30/18	HMGA2	Unsorted	2.20E+07	N/A	N/A	N/A	N/A	N/A	N/A	N/A
3/28/18	HMGA2	Top 3%	4.05E+05	1.00	1.00	1	1.00	23	JM48, JM44	4.80E+07
3/28/18	HMGA2	Top 4-10%	8.50E+05	2.00	2.00	1	2.00	23	JM44, JM46, JM51	4.98E+07
3/28/18	HMGA2	Bottom 4-10%	1.00E+06	4.50	4.50	1	4.50	21	JM47	4.18E+04
3/28/18	HMGA2	Bottom 3%	5.00E+05	1.00	1.00	1	1.00	23	JM52	3.39E+07
3/28/18	HMGA2	Unsorted	8.50E+06	93.60	5.85	11	64.35	21	JM57	6.22E+07
3/28/18	HMGA2	Unsorted	8.50E+06	86.00	5.38	11	59.13	21	JM59	

Chapter 3 : Future Directions

Validating Novel Candidates

There are many potential ways a novel candidate could be affecting miRNA mediated repression. To fully understand the biological relevance and functional role of the candidate requires a variety of follow up experiments. I first attempted to validate the results of the shRNA knockdowns by generating CRISPR knockout lines. To facilitate generation of the knockout lines, these experiments were performed in the HAP1 cell line (Horizon) which is a near haploid cell line. However, it may be the case that a knockout of the candidate gene is lethal. In these cases, multiple unique shRNAs effect on the candidate will need to be validated, to reduce the chance that the effect on the miRNA pathway is due to an off-target effect. Another important consideration is that the candidate may be acting generally or in a context-specific manor on the miRNA pathway. General factors are involved in all, or most, miRNA mediated repression across different cell types, while context-specific factors only regulate activity for specific miRNAs, 3′ UTRs or cell types. Therefore, it is important to note, a drawback of this approach is that it may incorrectly reject genuine context specific candidates that do not alter miRNA mediated repression in other cell lines.

For candidates that are successfully confirmed using an orthogonal CRISPR approach in a separate cell line, small RNA sequencing and mRNA sequencing experiments on the knockouts can be performed to give some insight into how the protein may be functioning. Factors that affect miRNA biogenesis would, upon knockdown, be expected to result in a decrease in expression of the mature miRNA while factors involved in miRNA silencing that are involved after miRNA biogenesis has taken place would not. If it is discovered that there are reduced levels of the mature miRNA, northern blots can be used to provide further insight into

whether Drosha or Dicer processing is being affected. For example, if there is an increased abundance of the pre-miRNA it would indicate the novel factor is affecting Dicer processing of the pre-miRNA. If Drosha processing is being affected on the other hand, I would expect to see a decrease in the level of pre-miRNA transcripts. The mRNA sequencing data can provide potential information on the miRNA pathway components that may be under regulation by the candidate gene. If the mRNA sequencing data reveals significantly elevated or repressed levels of known miRNA pathway genes (Dicer, Ago, etc.) upon knockdown of a candidate, it would suggest that the candidate may be involved in the regulation of an established miRNA pathway component.

Even if the candidate has no effect on transcript levels of known components of the miRNA pathway, it may still be mediating its effect post-translationally through its association with one or more known components. Known miRNA pathway proteins that associate with the candidate can be identified by cloning epitope-tagged versions of the candidates and using a tag-specific antibody to perform immunoprecipitation followed by mass spectrometry (Chendrimada et al., 2005; Gregory et al., 2004; Huang et al., 2013). The identification of interacting proteins may help elucidate a possible function for the candidate gene in the miRNA pathway. Association of the candidate with factors required for nuclear (Drosha, DGCR8) and cytoplasmic (Dicer) processing would suggest that the candidate may be involved in biogenesis of the miRNA, while association with Ago, TNRC6 or other downstream effectors would more likely suggest a role in miRNA mediated translational repression or accelerated mRNA decay.

For any candidates found to interact with known miRNA pathway components, follow up experiments can be tailored to confirm or reject how the candidate may be functioning. For example, if a novel candidate associates with a factor such as DDX6 or eIF4AII which play a role in miRNA mediated translational repression (Chu and Rana, 2006b; Fukao et al., 2014), it may be that the candidate is involved in translational repression. In such a case, this could be

tested by performing ribosomal profiling and comparing the levels of the unbound and polysomal fractions of the target mRNA in the presence and knockout or knockdown of the novel candidate (Fromm-Dornieden et al., 2012).

Additional Screens

Other miRNAs

One potential issue that may limit target discovery with the current screening approach was that the only difference between my wildtype and mutant reporter was the presence or absence of let-7 target sites. This may limit target discovery because it will not allow for the identification of factors that target other miRNAs or miRNA families. let-7 was one of the first miRNAs identified (Pasquinelli et al., 2000; Reinhart et al., 2000) and it has been very well characterized and studied (Lee et al., 2016; Powers et al., 2016) so there may not be many let-7 specific regulators left unidentified. In humans, there are currently 1917 annotated hairpin precursors in miRbase (Kozomara et al., 2019). There are many other miRNAs which have not been studied and some that behave unusually, such as miR-29b which localizes to the nucleus (Hwang et al., 2007). Non-canonical miRNAs may be uniquely regulated by factors that do not regulate let-7, and by switching which miRNAs target my reporter constructs, it may help identify these non-canonical factors.

Strand Selection

This screening approach can also be applied in a more focused manner to answer specific questions regarding some of the less well understood mechanism of miRNA biology. As

previously stated, the mechanism of strand selection is still incompletely understood and for some miRNAs, arm switching has been observed. Arm switching is defined as the phenomenon where the mature miRNA switches between one of the two arms of the precursor miRNA and has been observed in different tissues and different stages of development (Choo et al., 2014; Griffiths-Jones et al., 2011; Kuchenbauer et al., 2011; Ro et al., 2007). mir-423 for instance, has been observed to switch strand preference from favoring miR-423-5p in normal tissue to miR-423-3p in lung adenocarcinoma cells (Chen et al., 2018). In order to specifically screen for potential factors involved in strand selection, I designed a GFP and DsRed reporter which were targeted by miR-423-5p and miR-423-3p, respectively. These reporter constructs were successfully integrated into A549 cells and expression of $GFP^{mir\text{-}423\text{-}5p}$ and $DsRed^{mir\text{-}423\text{-}3p}$ was observed using flow cytometry (Figure 3.1). Although the design is similar to the previous screens, the beauty of using this approach for identifying a novel factor involved in arm-switching is that the signal should be fairly strong, because as the targeting of one reporter is reduced the targeting of the other is increased. A risk of performing this screen is that the potential factor may not be a target in our library or may not have suitable hairpins to produce an efficient knockdown capable of eliciting a phenotype and there would be no way of knowing this. However, a different library could be used as well as a different method or perturbation. This approach is equally suitable to using a CRISPR based method for perturbing gene expression and can be easily modified for such an approach.

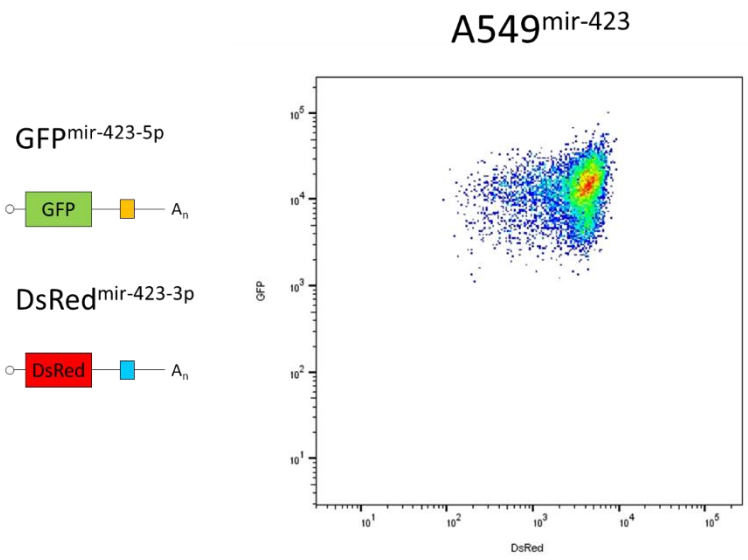

Figure 3-1: The A549^{mir423} Reporter

The A549^{mir423} reporter containing GFP$^{mir423\text{-}5p}$ and DsRed$^{mir423\text{-}3p}$.

Length Effect Screen

The work in this thesis thus far has focused specifically on the identification of novel genes involved in the miRNA pathway, and while there are still many candidates to continue to follow up on, this assay has the potential to be used more broadly. microRNAs are just one of the many *trans*-factors acting on *cis*-regulatory elements found within the 3′ UTR. Most regulatory RBPs that have been studied target *cis*-regulatory elements found within the 3′ UTR, however hundreds of proteins are predicted to contain RNA binding domains and only a subset has been carefully studied. The assay I have developed lends itself to identification of any *trans*-factor that regulates gene expression through its interaction with *cis*-regulatory elements found within the 3′ UTR.

In addition to sequence-specific elements, the length of the 3′ UTR itself may contain

regulatory information. In Kristjánsdóttir et al. (2015), analysis of a series of Hmga2 constructs with varied 3′ UTR lengths demonstrated a positive correlation between 3′ UTR length and repression (Kristjánsdóttir et al., 2015). A negative correlation between the 3′ UTR length and transcript stability has been observed previously, though it was not consistent across cell types (Yang et al., 2003). The only known mechanism that responds to 3′ UTR length is the nonsense-mediated-decay (NMD) pathway (Popp and Maquat, 2013). However, Kristjansdottir et al., (2017) showed that the knockdown of critical NMD pathway components UPF1 and SMG6 resulted in a negligible loss of 3′ UTR length-mediated repression, which suggests a novel mechanism of regulation dependent on the 3′ UTR length exists that is independent of the NMD pathway. Further supporting this result was the observation that the length effect persisted after treatment with puromycin which inhibits translation and therefore is occurring in an NMD independent manner (Kristjansdottir et al., 2017).

To identify the *trans*-acting factor responsible for this length-effect observed for longer 3′ UTRs I developed a reporter cell line which expresses a GFP and DsRed reporter containing a long (1200 nt) and short (400 nt) 3′ UTR, respectively. These reporters were screened using the small lentiviral library described previously, and then sequenced and analyzed for a single screen (**Table 2-1**) Before selecting candidates for follow up validation experiments, replicate screens should either be performed with the small library or using the 55k library in order to identify a candidate gene responsible for the length effect.

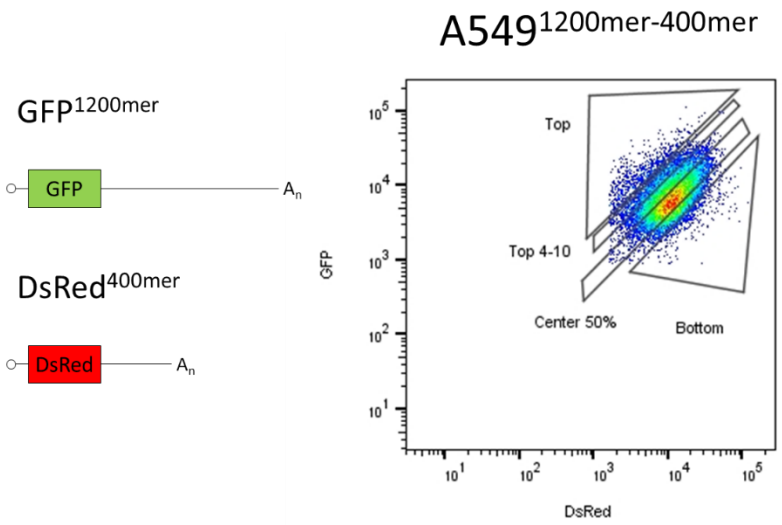

Figure 3-2: The A549$^{1200mer\text{-}400mer}$ Reporter

The A549$^{1200mer\text{-}400mer}$ reporter containing GFP^{1200mer} and DsRed^{400mer}.

Gene	Top3
MAGOHB	1
DNAJC3	2
GIPC2	3
METTL6	4
GSPT1	5
NHP2	6
PLOD3	7
DBR1	8
ZNF365	9
MTMR3	10
PRR5L	11
IRF2BPL	12
CUL9	13
RPL13A	14
RETSAT	15
MRPL42	16
WDR83	17
PRKCB	18
ZFP41	19
MMP20	20

Gene	Top 4-10
COL17A1	1
MYADML	2
APLP2	3
REXO4	4
LIPG	5
PRDX2	6
PSMC1	7
IFT46	8
PAPOLA	9
ATG7	10
NHP2	11
TRIM72	12
CHST11	13
ESRP2	14
TRIM26	15
GADD45B	16
ZNF768	17
APOBEC2	18
MEX3B	19
THUMPD1	20

Gene	Center
REXO4	1
PSMC1	2
FAM13A	3
DTWD1	4
INSR	5
UBE2A	6
DKC1	7
ENTPD6	8
GOT2	9
TDRD3	10
PRDX2	11
TDGF1	12
XPO1	13
PAF1	14
NHP2	15
TRIM26	16
SWT1	17
IQGAP1	18
SMG6	19
N6AMT1	20

Gene	Bottom 3
CWC15	1
METTL6	2
PABPC3	3
TRIM26	4
NSUN4	5
AEN	6
TUFM	7
RMI1	8
RBM43	9
CPSF2	10
PAPOLA	11
PAPOLB	12
TRIML1	13
FMR1	14
SRM	15
DRG2	16
SLC4A1AP	17
TDRD5	18
SMARCAL1	19
EXO1	20

Table 3-1: Comparison of Top 20 Candidates from A549$^{1200mer\text{-}400mer}$ Screen

The top 20 genes enriched in each condition, top 3%, top 4-10%, center 50%, and bottom 3%.

Bibliography

Abdelfattah, A.M., Park, C., and Choi, M.Y. (2014). Update on non-canonical microRNAs. Biomol. Concepts *5*, 275–287.

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. Elife *4*, e05005.

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat. Struct. Mol. Biol. *19*, 568–576.

Akgül, B., and Erdoğan, İ. (2018). Intracytoplasmic Re-localization of miRISC Complexes. Front. Genet. *9*.

Algire, M.A., Maag, D., Savio, P., Acker, M.G., Tarun, S.Z., Sachs, A.B., Asano, K., Nielsen, K.H., Olsen, D.S., Phan, L., et al. (2002). Development and characterization of a reconstituted yeast translation initiation system. RNA *8*, 382–397.

Alvarez-Saavedra, E., and Horvitz, H.R. (2010). Many families of C. elegans microRNAs are not essential for development or viability. Curr. Biol. CB *20*, 367–373.

Ameres, S.L., Horwich, M.D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA–Directed Trimming and Tailing of Small Silencing RNAs. Science *328*, 1534–1539.

Amin, N.D., Bai, G., Klug, J.R., Bonanomi, D., Pankratz, M.T., Gifford, W.D., Hinckley, C.A., Sternfeld, M.J., Driscoll, S.P., Dominguez, B., et al. (2015). Loss of motoneuron-specific microRNA-218 causes systemic neuromuscular failure. Science *350*, 1525–1529.

Ardekani, A.M., and Naeini, M.M. (2010). The Role of MicroRNAs in Human Diseases. Avicenna J. Med. Biotechnol. *2*, 161–179.

Asano, K., Shalev, A., Phan, L., Nielsen, K., Clayton, J., Valášek, L., Donahue, T.F., and Hinnebusch, A.G. (2001). Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. EMBO J. *20*, 2326–2337.

Axtell, M.J., Westholm, J.O., and Lai, E.C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. *12*, 221.

Azevedo-Pouly, A.C.P., Sutaria, D.S., Jiang, J., Elgamal, O.A., Amari, F., Allard, D., Grippo, P.J., Coppola, V., and Schmittgen, T.D. (2017). miR-216 and miR-217 expression is reduced in transgenic mouse models of pancreatic adenocarcinoma, knockout of miR-216/miR-217 host gene is embryonic lethal. Funct. Integr. Genomics *17*, 203–212.

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

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008a). The impact of microRNAs on protein output. Nature *455*, 64–71.

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008b). The impact of microRNAs on protein output. Nature *455*, 64–71.

Bahn, J.H., Ahn, J., Lin, X., Zhang, Q., Lee, J.-H., Civelek, M., and Xiao, X. (2015). Genomic Analysis of ADAR1 Binding and its Involvement in Multiple RNA Processing Pathways. Nat. Commun. *6*, 6355.

Barman, B., and Bhattacharyya, S.N. (2015). mRNA Targeting to Endoplasmic Reticulum Precedes Ago Protein Interaction and MicroRNA (miRNA)-mediated Translation Repression in Mammalian Cells. J. Biol. Chem. *290*, 24650–24656.

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. (2018). Metazoan MicroRNAs. Cell *173*, 20–51.

Bass, B.L., and Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell *55*, 1089–1098.

Beck, E., Ludwig, G., Auerswald, E.A., Reiss, B., and Schaller, H. (1982). Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. Gene *19*, 327–336.

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

Bellefroid, E.J., Poncelet, D.A., Lecocq, P.J., Revelant, O., and Martial, J.A. (1991). The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proc. Natl. Acad. Sci. U. S. A. *88*, 3608–3612.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol. *3*, e85.

Brosh, R.M., and Cantor, S.B. (2014). Molecular and cellular functions of the FANCJ DNA helicase defective in cancer and in Fanconi anemia. Front. Genet. *5*, 372.

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

1966.

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

Catalanotto, C., Cogoni, C., and Zardo, G. (2016). MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. Int. J. Mol. Sci. *17*.

Chakraborty, C., Sharma, A.R., Sharma, G., Doss, C.G.P., and Lee, S.-S. (2017). Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine. Mol. Ther. - Nucleic Acids *8*, 132–143.

Chan, S.-P., and Slack, F.J. (2006). microRNA-Mediated Silencing Inside P Bodies. RNA Biol. *3*, 97–100.

Chandradoss, S.D., Schirle, N.T., Szczepaniak, M., MacRae, I.J., and Joo, C. (2015). A Dynamic Search Process Underlies MicroRNA Targeting. Cell *162*, 96–107.

Chang, H.-M., Triboulet, R., Thornton, J.E., and Gregory, R.I. (2013). A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature *497*, 244–248.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4–NOT through conserved W-containing motifs. Nat. Struct. Mol. Biol. *18*, 1218–1226.

Cheloufi, S., Dos Santos, C.O., Chong, M.M.W., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584–589.

Chen, C.-Y.A., Zheng, D., Xia, Z., and Shyu, A.-B. (2009). Ago–TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. Nat. Struct. Mol. Biol. *16*, 1160–1166.

Chen, L., Sun, H., Wang, C., Yang, Y., Zhang, M., and Wong, G. (2018). miRNA arm switching identifies novel tumour biomarkers. EBioMedicine *38*, 37–46.

Chen, Y., Boland, A., Kuzuoğlu-Öztürk, D., Bawankar, P., Loh, B., Chang, C.-T., Weichenrieder, O., and Izaurralde, E. (2014a). A DDX6-CNOT1 Complex and W-Binding Pockets in CNOT9 Reveal Direct Links between miRNA Target Recognition and Silencing. Mol. Cell *54*, 737–750.

Chen, Y.-W., Song, S., Weng, R., Verma, P., Kugler, J.-M., Buescher, M., Rouam, S., and Cohen, S.M. (2014b). Systematic Study of Drosophila MicroRNA Functions Using a Collection of Targeted Knockout Mutations. Dev. Cell *31*, 784–800.

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

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and

Shiekhattar, R. (2005b). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740–744.

Choo, K.B., Soon, Y.L., Nguyen, P.N.N., Hiew, M.S.Y., and Huang, C.-J. (2014). MicroRNA-5p and -3p co-expression and cross-targeting in colon cancer cells. J. Biomed. Sci. *21*, 95.

Christie, M., Boland, A., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2013). Structure of the PAN3 Pseudokinase Reveals the Basis for Interactions with the PAN2 Deadenylase and the GW182 Proteins. Mol. Cell *51*, 360–373.

Christopher, A.F., Kaur, R.P., Kaur, G., Kaur, A., Gupta, V., and Bansal, P. (2016). MicroRNA therapeutics: Discovering novel targets and developing specific therapy. Perspect. Clin. Res. *7*, 68–74.

Chu, C., and Rana, T.M. (2006). Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. PLoS Biol. *4*, e210.

Collart, M.A., and Panasenko, O.O. (2012). The Ccr4–Not complex. Gene *492*, 42–53.

Coller, J.M., Tucker, M., Sheth, U., Valencia-Sanchez, M.A., and Parker, R. (2001). The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. RNA N. Y. N *7*, 1717–1727.

Cooke, A., Prigge, A., and Wickens, M. (2010). Translational Repression by Deadenylases. J. Biol. Chem. *285*, 28506–28513.

Corbett, A.H. (2018). Post-transcriptional regulation of gene expression and human disease. Curr. Opin. Cell Biol. *52*, 96–104.

Davis, E., Caiment, F., Tordoir, X., Cavaillé, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr. Biol. CB *15*, 743–749.

Deng, S., Calin, G.A., Croce, C.M., Coukos, G., and Zhang, L. (2008). Mechanisms of microRNA deregulation in human cancer. Cell Cycle Georget. Tex *7*, 2643–2646.

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

Derry, M.C., Yanagiya, A., Martineau, Y., and Sonenberg, N. (2006). Regulation of Poly(A)-binding Protein through PABP-interacting Proteins. Cold Spring Harb. Symp. Quant. Biol. *71*, 537–543.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J. Virol. *72*, 8463–8471.

Echeverri, C.J., Beachy, P.A., Baum, B., Boutros, M., Buchholz, F., Chanda, S.K., Downward, J., Ellenberg, J., Fraser, A.G., Hacohen, N., et al. (2006a). Minimizing the risk of reporting false

positives in large-scale RNAi screens. Nat. Methods *3*, 777–779.

Echeverri, C.J., Beachy, P.A., Baum, B., Boutros, M., Buchholz, F., Chanda, S.K., Downward, J., Ellenberg, J., Fraser, A.G., Hacohen, N., et al. (2006b). Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat. Methods *3*, 777–779.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.-H., Ghoshal, K., Villén, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol. Cell *56*, 104–115.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A human snoRNA with microRNA-like functions. Mol. Cell *32*, 519–528.

Eulalio, A., Behm-Ansmant, I., Schweizer, D., and Izaurralde, E. (2007). P-Body Formation Is a Consequence, Not the Cause, of RNA-Mediated Gene Silencing. Mol. Cell. Biol. *27*, 3970–3981.

Eulalio, A., Tritschler, F., and Izaurralde, E. (2009). The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. RNA N. Y. N *15*, 1433–1442.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A., et al. (2009). Mammalian miRNA RISC Recruits CAF1 and PABP to Affect PABP-Dependent Deadenylation. Mol. Cell *35*, 868–880.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F., et al. (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4–NOT. Nat. Struct. Mol. Biol. *18*, 1211–1217.

Faruq, O., and Vecchione, A. (2015). microRNA: Diagnostic Perspective. Front. Med. *2*.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. *40*, 37–52.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2008). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

Fromm-Dornieden, C., von der Heyde, S., Lytovchenko, O., Salinas-Riester, G., Brenig, B., Beissbarth, T., and Baumgartner, B.G. (2012). Novel polysome messages and changes in

translational activity appear after induction of adipogenesis in 3T3-L1 cells. BMC Mol. Biol. *13*, 9.

Fukao, A., Mishima, Y., Takizawa, N., Oka, S., Imataka, H., Pelletier, J., Sonenberg, N., Thoma, C., and Fujiwara, T. (2014a). MicroRNAs Trigger Dissociation of eIF4AI and eIF4AII from Target mRNAs in Humans. Mol. Cell *56*, 79–89.

Fukao, A., Mishima, Y., Takizawa, N., Oka, S., Imataka, H., Pelletier, J., Sonenberg, N., Thoma, C., and Fujiwara, T. (2014b). MicroRNAs Trigger Dissociation of eIF4AI and eIF4AII from Target mRNAs in Humans. Mol. Cell *56*, 79–89.

Gagnon, K.T., Li, L., Chu, Y., Janowski, B.A., and Corey, D.R. (2014). RNAi Factors Are Present and Active in Human Cell Nuclei. Cell Rep. *6*, 211–221.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011a). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat. Struct. Mol. Biol. *18*, 1139–1146.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. (2011b). Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nat. Struct. Mol. Biol. *18*, 1139–1146.

Gerolami, R., Uch, R., Jordier, F., Chapel, S., Bagnis, C., Bréchot, C., and Mannoni, P. (2000). Gene transfer to hepatocellular carcinoma: transduction efficacy and transgene expression kinetics by using retroviral and lentiviral vectors. Cancer Gene Ther. *7*, 1286–1292.

Giard, D.J., Aaronson, S.A., Todaro, G.J., Arnstein, P., Kersey, J.H., Dosik, H., and Parks, W.P. (1973). In Vitro Cultivation of Human Tumors: Establishment of Cell Lines Derived From a Series of Solid Tumors. JNCI J. Natl. Cancer Inst. *51*, 1417–1423.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. *582*, 1977–1986.

Golden, R.J., Chen, B., Li, T., Braun, J., Manjunath, H., Chen, X., Wu, J., Schmid, V., Chang, T.-C., Kopp, F., et al. (2017a). An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. Nature *542*, 197–202.

Golden, R.J., Chen, B., Li, T., Braun, J., Manjunath, H., Chen, X., Wu, J., Schmid, V., Chang, T.-C., Kopp, F., et al. (2017b). An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. Nature *542*, 197–202.

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

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

Griffiths-Jones, S., Hui, J.H.L., Marco, A., and Ronshaugen, M. (2011). MicroRNA evolution by arm switching. EMBO Rep. *12*, 172–177.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007a). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell *27*, 91–105.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007b). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell *27*, 91–105.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guo, Y., Liu, J., Elfenbein, S.J., Ma, Y., Zhong, M., Qiu, C., Ding, Y., and Lu, J. (2015). Characterization of the mammalian miRNA turnover landscape. Nucleic Acids Res. *43*, 2326–2341.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. *15*, 509–524.

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

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887–901.

Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.-K., Yeom, K.-H., Yang, W.-Y., Haussler, D., Blelloch, R., and Kim, V.N. (2009). Posttranscriptional Crossregulation between Drosha and DGCR8. Cell *136*, 75–84.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol. *7*, e1000238.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 Mediates the Terminal Uridylation of let-7 Precursor MicroRNA. Mol. Cell *32*, 276–284.

Heo, I., Ha, M., Lim, J., Yoon, M.-J., Park, J.-E., Kwon, S.C., Chang, H., and Kim, V.N. (2012). Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. Cell *151*, 521–532.

Hon, L.S., and Zhang, Z. (2007). The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. Genome Biol. *8*, R166.

Hu, H.Y., Yan, Z., Xu, Y., Hu, H., Menzel, C., Zhou, Y.H., Chen, W., and Khaitovich, P. (2009). Sequence features associated with microRNA strand selection in humans and flies. BMC

Genomics *10*, 413.

Huang, T.-C., Pinto, S.M., and Pandey, A. (2013a). Proteomics for understanding miRNA biology. Proteomics *13*.

Huang, T.-C., Pinto, S.M., and Pandey, A. (2013b). Proteomics for understanding miRNA biology. Proteomics *13*.

Humphreys, D.T., Westman, B.J., Martin, D.I.K., and Preiss, T. (2005). MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. Proc. Natl. Acad. Sci. U. S. A. *102*, 16961–16966.

Huntley, S., Baggott, D.M., Hamilton, A.T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E., and Stubbs, L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. Genome Res. *16*, 669–677.

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

Iuchi, S. (2001). Three classes of C2H2 zinc finger proteins. Cell. Mol. Life Sci. CMLS *58*, 625–635.

Iwakawa, H., and Tomari, Y. (2015). The Functions of MicroRNAs: mRNA Decay and Translational Repression. Trends Cell Biol. *25*, 651–665.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol. Cell *39*, 292–299.

Jackson, R.J., Hellen, C.U.T., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat. Rev. Mol. Cell Biol. *11*, 113–127.

Jain, S., and Parker, R. (2013). The Discovery and Analysis of P Bodies. In Ten Years of Progress in GW/P Body Research, E.K.L. Chan, and M.J. Fritzler, eds. (New York, NY: Springer New York), pp. 23–43.

Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G., and Liu, Y. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. *37*, D98-104.

Jinek, M., Fabian, M.R., Coyle, S.M., Sonenberg, N., and Doudna, J.A. (2010). Structural insights into the human GW182-PABC interaction in microRNA-mediated deadenylation. Nat. Struct. Mol. Biol. *17*, 238–240.

Kamenska, A., Lu, W.-T., Kubacka, D., Broomhead, H., Minshall, N., Bushell, M., and Standart, N. (2014). Human 4E-T represses translation of bound mRNAs and enhances microRNA-mediated silencing. Nucleic Acids Res. *42*, 3298–3313.

Kato-Inui, T., Takahashi, G., Hsu, S., and Miyaoka, Y. (2018). Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 with improved proof-reading enhances homology-directed repair. Nucleic Acids Res. *46*, 4677–4688.

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

Kawai, S., and Amano, A. (2012). BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. J. Cell Biol. *197*, 201–208.

Kawamata, T., and Tomari, Y. (2010). Making RISC. Trends Biochem. Sci. *35*, 368–376.

Ketting, R.F., Fischer, S.E.J., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H.A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654–2659.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209–216.

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

Kim, D.W., Uetsuki, T., Kaziro, Y., Yamaguchi, N., and Sugano, S. (1990). Use of the human elongation factor 1 alpha promoter as a versatile and efficient expression system. Gene *91*, 217–223.

Kimura, M., Kamakura, T., Zhou Tao, Q., Kaneko, I., and Yamaguchi, I. (1994). Cloning of the blasticidin S deaminase gene (BSD) from Aspergillus terreus and its use as a selectable marker for Schizosaccharomyces pombe and Pyricularia oryzae. Mol. Gen. Genet. MGG *242*, 121–129.

Klum, S.M., Chandradoss, S.D., Schirle, N.T., Joo, C., and MacRae, I.J. (2018). Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition. EMBO J. *37*, 75–88.

Knorre, D.G., Kudryashova, N.V., and Godovikova, T.S. (2009). Chemical and Functional Aspects of Posttranslational Modification of Proteins. Acta Naturae *1*, 29–51.

KOLUPAEVA, V.G., UNBEHAUN, A., LOMAKIN, I.B., HELLEN, C.U.T., and PESTOVA, T.V. (2005). Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. RNA *11*, 470–486.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162.

Kriegel, A.J., Terhune, S.S., Greene, A.S., Noon, K.R., Pereckas, M.S., and Liang, M. (2018). Isomer-specific effect of microRNA miR-29b on nuclear morphology. J. Biol. Chem. *293*, 14080–14088.

Kristjansdottir, K., and Link to external site, this link will open in a new window (2017). 3’

UTR sequences and syntax: Investigating cis-element density and interactions, and the role of 3' UTR length in gene regulation. Ph.D. Cornell University.

Kristjánsdóttir, K., Fogarty, E.A., and Grimson, A. (2015). Systematic analysis of the Hmga2 3′ UTR identifies many independent regulatory sequences and a novel interaction between distal sites. RNA *21*, 1346–1360.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schübeler, D., et al. (2010). Characterizing Light-Regulated Retinal MicroRNAs Reveals Rapid Turnover as a Common Property of Neuronal MicroRNAs. Cell *141*, 618–631.

Kuchenbauer, F., Mah, S.M., Heuser, M., McPherson, A., Rüschmann, J., Rouhi, A., Berg, T., Bullinger, L., Argiropoulos, B., Morin, R.D., et al. (2011). Comprehensive analysis of mammalian miRNA* species and their role in myeloid cells. Blood *118*, 3350–3358.

Kwon, S.C., Nguyen, T.A., Choi, Y.-G., Jo, M.H., Hohng, S., Kim, V.N., and Woo, J.-S. (2016). Structure of Human DROSHA. Cell *164*, 81–90.

Labialle, S., Marty, V., Bortolin-Cavaillé, M.-L., Hoareau-Osman, M., Pradère, J.-P., Valet, P., Martin, P.G.P., and Cavaillé, J. (2014). The miR-379/miR-410 cluster at the imprinted Dlk1-Dio3 domain controls neonatal metabolic adaptation. EMBO J. *33*, 2216–2230.

Lee, H., Han, S., Kwon, C.S., and Lee, D. (2016). Biogenesis and regulation of the let-7 miRNAs and their functional implications. Protein Cell *7*, 100–113.

Lee, H.Y., Zhou, K., Smith, A.M., Noland, C.L., and Doudna, J.A. (2013). Differential roles of human Dicer-binding proteins TRBP and PACT in small RNA processing. Nucleic Acids Res. *41*, 6568–6576.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003a). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003b). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

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

Leung, A.K.L. (2015). The Whereabouts of miRNA Actions: Cytoplasm and Beyond. Trends Cell Biol. *25*, 601–610.

Lewis, B.P., Shih, I. -hung, Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787–798.

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

Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., and Kim, V.N. (2014). Uridylation by TUT4 and TUT7 Marks mRNA for Degradation. Cell *159*, 1365–1376.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 Is the Catalytic Engine of Mammalian RNAi. Science *305*, 1437–1441.

Lizée, G., Gonzales, M.I., and Topalian, S.L. (2004). Lentivirus vector-mediated expression of tumor-associated epitopes by human antigen presenting cells. Hum. Gene Ther. *15*, 393–404.

Lugtenberg, D., Yntema, H.G., Banning, M.J.G., Oudakker, A.R., Firth, H.V., Willatt, L., Raynaud, M., Kleefstra, T., Fryns, J.-P., Ropers, H.-H., et al. (2006). ZNF674: A New Krüppel-Associated Box–Containing Zinc-Finger Gene Involved in Nonsyndromic X-Linked Mental Retardation. Am. J. Hum. Genet. *78*, 265–278.

Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J.S., Beroukhim, R., Weir, B.A., et al. (2008). Highly parallel identification of essential genes in cancer cells. Proc. Natl. Acad. Sci. U. S. A. *105*, 20380–20385.

Lupo, A., Cesaro, E., Montano, G., Zurlo, D., Izzo, P., and Costanzo, P. (2013). KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. Curr. Genomics *14*, 268–278.

Lytle, J.R., Yario, T.A., and Steitz, J.A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. Proc. Natl. Acad. Sci. *104*, 9667–9672.

Maes, O.C., Chertkow, H.M., Wang, E., and Schipper, H.M. (2009). MicroRNA: Implications for Alzheimer Disease and other Human CNS Disorders. Curr. Genomics *10*, 154–168.

Majumdar, R., Bandyopadhyay, A., and Maitra, U. (2003). Mammalian Translation Initiation Factor eIF1 Functions with eIF1A and eIF3 in the Formation of a Stable 40 S Preinitiation Complex. J. Biol. Chem. *278*, 6580–6587.

Makarova, J.A., Shkurnikov, M.U., Wicklein, D., Lange, T., Samatov, T.R., Turchinovich, A.A., and Tonevitsky, A.G. (2016). Intracellular and extracellular microRNA: An update on localization and biological role. Prog. Histochem. Cytochem. *51*, 33–49.

Makino, S., Mishima, Y., Inoue, K., and Inada, T. (2015). Roles of mRNA Fate Modulators Dhh1 and Pat1 in TNRC6-dependent Gene Silencing Recapitulated in Yeast. J. Biol. Chem. *290*, 8331–8347.

Maroney, P.A., Yu, Y., Fisher, J., and Nilsen, T.W. (2006). Evidence that microRNAs are associated with translating messenger RNAs in human cells. Nat. Struct. Mol. Biol. *13*, 1102.

Martinez, B., and Peplow, P.V. (2019). MicroRNAs as diagnostic and therapeutic tools for

Alzheimer's disease: advances and limitations. Neural Regen. Res. *14*, 242–255.

Marzi, M.J., Ghini, F., Cerruti, B., de Pretis, S., Bonetti, P., Giacomelli, C., Gorski, M.M., Kress, T., Pelizzola, M., Muller, H., et al. (2016). Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. Genome Res. *26*, 554–565.

Mathonnet, G., Fabian, M.R., Svitkin, Y.V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W.C., Darzynkiewicz, E., Pillai, R.S., et al. (2007). MicroRNA Inhibition of Translation Initiation in Vitro by Targeting the Cap-Binding Complex eIF4F. Science *317*, 1764–1767.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and Biochemical Insights to the Role of the CCR4-NOT Complex and DDX6 ATPase in MicroRNA Repression. Mol. Cell *54*, 751–765.

Mayeux, R. (2004). Biomarkers: potential uses and limitations. NeuroRx J. Am. Soc. Exp. Neurother. *1*, 182–188.

Mayr, C., Hemann, M.T., and Bartel, D.P. (2007). Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. Science *315*, 1576–1579.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

McKenzie, A.J., Hoshino, D., Hong, N.H., Cha, D.J., Franklin, J.L., Coffey, R.J., Patton, J.G., and Weaver, A.M. (2016). KRAS-MEK Signaling Controls Ago2 Sorting into Exosomes. Cell Rep. *15*, 978–987.

Medeiros, L.A., Dennis, L.M., Gill, M.E., Houbaviy, H., Markoulaki, S., Fu, D., White, A.C., Kirak, O., Sharp, P.A., Page, D.C., et al. (2011). Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. Proc. Natl. Acad. Sci. U. S. A. *108*, 14163–14168.

Meijer, H.A., Smith, E.M., and Bushell, M. (2014). Regulation of miRNA strand selection: follow the leader? Biochem. Soc. Trans. *42*, 1135–1140.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. Mol. Cell *15*, 185–197.

Melo, S.A., Sugimoto, H., O'Connell, J.T., Kato, N., Villanueva, A., Vidal, A., Qiu, L., Vitkin, E., Perelman, L.T., Melo, C.A., et al. (2014). Cancer Exosomes Perform Cell-Independent MicroRNA Biogenesis and Promote Tumorigenesis. Cancer Cell *26*, 707–721.

Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., Aguirre, L.A., del Castillo, I., Steel, K.P., Dalmay, T., et al. (2009). Mutations in the seed region

of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat. Genet. *41*, 609–613.

Meola, N., Gennarino, V.A., and Banfi, S. (2009). microRNAs and genetic diseases. PathoGenetics *2*, 7.

Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. PLoS Genet. *3*, e215.

Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepfer, A.M., Hinkle, G., Piqani, B., Eisenhaure, T.M., Luo, B., Grenier, J.K., et al. (2006). A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. Cell *124*, 1283–1298.

Mohr, S.E., Smith, J.A., Shamu, C.E., Neumüller, R.A., and Perrimon, N. (2014). RNAi screening comes of age: improved techniques and complementary approaches. Nat. Rev. Mol. Cell Biol. *15*, 591–600.

Nadal, E. de, Ammerer, G., and Posas, F. (2011). Controlling gene expression in response to stress. Nat. Rev. Genet. *12*, 833–845.

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

Nguyen, T.A., Jo, M.H., Choi, Y.-G., Park, J., Kwon, S.C., Hohng, S., Kim, V.N., and Woo, J.-S. (2015). Functional Anatomy of the Human Microprocessor. Cell *161*, 1374–1387.

Nishihara, T., Zekri, L., Braun, J.E., and Izaurralde, E. (2013). miRISC recruits decapping factors to miRNA targets to enhance their degradation. Nucleic Acids Res. *41*, 8692–8705.

O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front. Endocrinol. *9*.

Ohrt, T., Mütze, J., Staroske, W., Weinmann, L., Höck, J., Crell, K., Meister, G., and Schwille, P. (2008). Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. Nucleic Acids Res. *36*, 6439–6449.

Okada, C., Yamashita, E., Lee, S.J., Shibata, S., Katahira, J., Nakagawa, A., Yoneda, Y., and Tsukihara, T. (2009). A high-resolution structure of the pre-microRNA nuclear export machinery. Science *326*, 1275–1279.

Olena, A.F., and Patton, J.G. (2010). Genomic Organization of microRNAs. J. Cell. Physiol. *222*, 540–545.

Ozgur, S., Basquin, J., Kamenska, A., Filipowicz, W., Standart, N., and Conti, E. (2015). Structure of a Human 4E-T/DDX6/CNOT1 Complex Reveals the Different Interplay of DDX6-Binding Proteins with the CCR4-NOT Complex. Cell Rep. *13*, 703–711.

Park, J.H., and Shin, C. (2014). MicroRNA-directed cleavage of targets: mechanism and experimental approaches. BMB Rep. *47*, 417–423.

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

Paroo, Z., Ye, X., Chen, S., and Liu, Q. (2009). Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. Cell *139*, 112–122.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Müller, P., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86–89.

Paul, D., Sinha, A.N., Ray, A., Lal, M., Nayak, S., Sharma, A., Mehani, B., Mukherjee, D., Laddha, S.V., Suri, A., et al. (2017). A-to-I editing in human miRNAs is enriched in seed sequence, influenced by sequence contexts and significantly hypoedited in glioblastoma multiforme. Sci. Rep. *7*, 2466.

Petit, A.-P., Wohlbold, L., Bawankar, P., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2012). The structural basis for the interaction between the CAF1 nuclease and the NOT1 scaffold of the human CCR4-NOT deadenylase complex. Nucleic Acids Res. *40*, 11058–11072.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science *309*, 1573–1576.

Pitchiaya, S., Heinicke, L.A., Park, J.I., Cameron, E.L., and Walter, N.G. (2017). Resolving Subcellular miRNA Trafficking and Turnover at Single-Molecule Resolution. Cell Rep. *19*, 630–642.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110.

de Pontual, L., Yao, E., Callier, P., Faivre, L., Drouin, V., Cariou, S., Van Haeringen, A., Geneviève, D., Goldenberg, A., Oufadem, M., et al. (2011). Germline deletion of the miR-17∼92 cluster causes skeletal and growth defects in humans. Nat. Genet. *43*, 1026–1030.

Popp, M.W.-L., and Maquat, L.E. (2013). Organizing principles of mammalian nonsense-mediated mRNA decay. Annu. Rev. Genet. *47*, 139–165.

Portin, P., and Wilkins, A. (2017). The Evolving Definition of the Term “Gene.” Genetics *205*, 1353–1364.

Powers, J.T., Tsanov, K.M., Pearson, D.S., Roels, F., Spina, C.S., Ebright, R., Seligson, M., de Soysa, Y., Cahan, P., Theißen, J., et al. (2016). Multiple mechanisms disrupt the *let-7* microRNA family in neuroblastoma. Nature *535*, 246–251.

Quick-Cleveland, J., Jacob, J.P., Weitz, S.H., Shoffner, G., Senturia, R., and Guo, F. (2014). The DGCR8 RNA-binding heme domain recognizes primary microRNAs by clamping the hairpin. Cell Rep. *7*, 1994–2005.

Reid, D.W., and Nicchitta, C.V. (2015). Diversity and selectivity in mRNA translation on the endoplasmic reticulum. Nat. Rev. Mol. Cell Biol. *16*, 221–231.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901–906.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513–520.

Ricci, E.P., Limousin, T., Soto-Rifo, R., Rubilar, P.S., Decimo, D., and Ohlmann, T. (2013). miRNA repression of translation in vitro takes place during 43S ribosomal scanning. Nucleic Acids Res. *41*, 586–598.

Ro, S., Park, C., Young, D., Sanders, K.M., and Yan, W. (2007). Tissue-dependent paired expression of miRNAs. Nucleic Acids Res. *35*, 5944–5953.

Robb, G.B., Brown, K.M., Khurana, J., and Rana, T.M. (2005). Specific and potent RNAi in the nucleus of human cells. Nat. Struct. Mol. Biol. *12*, 133.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

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

van Rooij, E., Sutherland, L.B., Liu, N., Williams, A.H., McAnally, J., Gerard, R.D., Richardson, J.A., and Olson, E.N. (2006). A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. Proc. Natl. Acad. Sci. U. S. A. *103*, 18255–18260.

Rouya, C., Siddiqui, N., Morita, M., Duchaine, T.F., Fabian, M.R., and Sonenberg, N. (2014). Human DDX6 effects miRNA-mediated gene silencing via direct binding to CNOT1. RNA *20*, 1398–1409.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83–86.

Rüdel, S., Flatley, A., Weinmann, L., Kremmer, E., and Meister, G. (2008). A multifunctional human Argonaute2-specific monoclonal antibody. RNA *14*, 1244–1253.

Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E.E., Nitsch, R., and Wulczyn, F.G. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat. Cell Biol. *10*, 987–993.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037–1040.

Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural Basis for microRNA Targeting. Science *346*, 608–613.

Schraivogel, D., and Meister, G. (2014). Import routes and nuclear functions of Argonaute and other small RNA-silencing proteins. Trends Biochem. Sci. *39*, 420–431.

Schratt, G. (2009). microRNAs at the synapse. Nat. Rev. Neurosci. *10*, 842–849.

Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. Cell *110*, 521–529.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008a). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008b). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63.

Shabalina, S.A., and Koonin, E.V. (2008). Origins and evolution of eukaryotic RNA interference. Trends Ecol. Evol. *23*, 578–587.

Shurtleff, M.J., Temoche-Diaz, M.M., Karfilis, K.V., Ri, S., and Schekman, R. (2016). Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction. ELife *5*.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034.

Sigoillot, F.D., and King, R.W. (2011). Vigilance and Validation: Keys to Success in RNAi Screening. ACS Chem. Biol. *6*, 47–60.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. Cell *136*, 731–745.

Suzuki, H.I., Katsura, A., Yasuda, T., Ueno, T., Mano, H., Sugimoto, K., and Miyazono, K. (2015). Small-RNA asymmetry is directly driven by mammalian Argonautes. Nat. Struct. Mol. Biol. *22*, 512–521.

Tang, X., Zhang, Y., Tucker, L., and Ramratnam, B. (2010). Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization. Nucleic Acids Res. *38*, 6610–6619.

Tang, X., Li, M., Tucker, L., and Ramratnam, B. (2011). Glycogen Synthase Kinase 3 Beta (GSK3β) Phosphorylates the RNAase III Enzyme Drosha at S300 and S302. PLoS ONE *6*, e20391.

Tatsuguchi, M., Seok, H.Y., Callis, T.E., Thomson, J.M., Chen, J.-F., Newman, M., Rojas, M., Hammond, S.M., and Wang, D.-Z. (2007). Expression of microRNAs is dynamically regulated during cardiomyocyte hypertrophy. J. Mol. Cell. Cardiol. *42*, 1137–1141.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. *25*, 1105–1111.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiß, J.-L., Eichner, N., Lehmann, G., Schall, K., Urlaub, H., and Meister, G. (2017a). A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. Mol. Cell *66*, 270-284.e13.

Treiber, T., Treiber, N., Plessmann, U., Harlander, S., Daiß, J.-L., Eichner, N., Lehmann, G., Schall, K., Urlaub, H., and Meister, G. (2017b). A Compendium of RNA-Binding Proteins that Regulate MicroRNA Biogenesis. Mol. Cell *66*, 270-284.e13.

Tsai, K.-W., Leung, C.-M., Lo, Y.-H., Chen, T.-W., Chan, W.-C., Yu, S.-Y., Tu, Y.-T., Lam, H.-C., Li, S.-C., Ger, L.-P., et al. (2016). Arm Selection Preference of MicroRNA-193a Varies in Breast Cancer. Sci. Rep. *6*, 28176.

Uchida, N., Hoshino, S., Imataka, H., Sonenberg, N., and Katada, T. (2002). A Novel Role of the Mammalian GSPT/eRF3 Associating with Poly(A)-binding Protein in Cap/Poly(A)-dependent Translation. J. Biol. Chem. *277*, 50286–50292.

Uetsuki, T., Naito, A., Nagata, S., and Kaziro, Y. (1989). Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha. J. Biol. Chem. *264*, 5791–5798.

Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. Genome Biol. *4*, 231.

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J., and Lötvall, J.O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat. Cell Biol. *9*, 654–659.

Vihervaara, A., Duarte, F.M., and Lis, J.T. (2018). Molecular mechanisms driving transcriptional stress responses. Nat. Rev. Genet. *19*, 385–397.

Visone, R., and Croce, C.M. (2009). MiRNAs and cancer. Am. J. Pathol. *174*, 1131–1138.

Wagner, R.W., Smith, J.E., Cooperman, B.S., and Nishikura, K. (1989). A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. Proc. Natl. Acad. Sci. U. S. A. *86*, 2647–2651.

Wakiyama, M., Takimoto, K., Ohara, O., and Yokoyama, S. (2007). Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. Genes Dev.

21, 1857–1862.

Wang, P., Zhao, W., Zhao, K., Zhang, L., and Gao, C. (2015). TRIM26 Negatively Regulates Interferon-β Production and Antiviral Response through Polyubiquitination and Degradation of Nuclear IRF3. PLoS Pathog. *11*.

Wang, Y., Zhou, J., Ye, X., Wan, Y., Li, Y., Mo, X., Yuan, W., Yan, Y., Luo, N., Wang, Z., et al. (2010). ZNF424, a novel human KRAB/C2H2 zinc finger protein, suppresses NFAT and p21 pathway. BMB Rep. *43*, 212–218.

Watanabe, T., Sato, T., Amano, T., Kawamura, Y., Kawamura, N., Kawaguchi, H., Yamashita, N., Kurihara, H., and Nakaoka, T. (2008). Dnm3os, a non-coding RNA, is required for normal growth and skeletal development in mice. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *237*, 3738–3748.

Weake, V.M., and Workman, J.L. (2010). Inducible gene expression: diverse regulatory mechanisms. Nat. Rev. Genet. *11*, 426–437.

Xie, J., Ameres, S.L., Friedline, R., Hung, J.-H., Zhang, Y., Xie, Q., Zhong, L., Su, Q., He, R., Li, M., et al. (2012). Long-term, efficient inhibition of microRNA function in mice using rAAV vectors. Nat. Methods *9*, 403–409.

Xie, M., Li, M., Vilborg, A., Lee, N., Shu, M.-D., Yartseva, V., Šestan, N., and Steitz, J.A. (2013). Mammalian 5′-capped microRNA precursors that generate a single microRNA. Cell *155*, 1568–1580.

Xu, W., San Lucas, A., Wang, Z., and Liu, Y. (2014). Identifying microRNA targets in different gene regions. BMC Bioinformatics *15*, S4.

Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell, J.E. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res. *13*, 1863–1872.

Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., and Nishikura, K. (2006). Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat. Struct. Mol. Biol. *13*, 13–21.

Yao, B., Li, S., Lian, S.L., Fritzler, M.J., and Chan, E.K.L. (2011). Mapping of Ago2-GW182 functional interactions. Methods Mol. Biol. Clifton NJ *725*, 45–62.

Yekta, S. (2004). MicroRNA-Directed Cleavage of HOXB8 mRNA. Science *304*, 594–596.

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

Zekri, L., Kuzuoğlu-Öztürk, D., and Izaurralde, E. (2013a). GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. EMBO J. *32*, 1052–1065.

Zekri, L., Kuzuoğlu-Öztürk, D., and Izaurralde, E. (2013b). GW182 proteins cause PABP dissociation from silenced miRNA targets in the absence of deadenylation. EMBO J. *32*, 1052–1065.

Zeng, Y., and Cullen, B.R. (2004). Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res. *32*, 4776–4785.

Zhang, Q.-H., Ye, M., Wu, X.-Y., Ren, S.-X., Zhao, M., Zhao, C.-J., Fu, G., Shen, Y., Fan, H.-Y., Lu, G., et al. (2000). Cloning and Functional Analysis of cDNAs with Open Reading Frames for 300 Previously Undefined Genes Expressed in CD34+ Hematopoietic Stem/Progenitor Cells. Genome Res. *10*, 1546–1560.

Zhang, Z., Qin, Y.-W., Brewer, G., and Jing, Q. (2012). MicroRNA degradation and turnover: regulating the regulators: MicroRNA degradation and turnover. Wiley Interdiscip. Rev. RNA *3*, 593–600.

Zhao, W., Blagev, D., Pollack, J.L., and Erle, D.J. (2011). Toward a Systematic Understanding of mRNA 3′ Untranslated Regions. Proc. Am. Thorac. Soc. *8*, 163–166.

Zheng, D., Ezzeddine, N., Chen, C.-Y.A., Zhu, W., He, X., and Shyu, A.-B. (2008). Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. J. Cell Biol. *182*, 89–101.