A High-Throughput Screening Technology to Identify Novel MiRNA-Associating Factors

A Thesis
Presented to the Faculty of the Graduate School
Of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Masters of Science

by
RaeAnna M Wilson
May, 2014

ABSTRACT

Regulation of gene expression is critical for a multitude of processes. Perhaps the most poorly understood mode of gene regulation is the post-transcriptional regulation of mRNAs, which includes regulation by proteins and small RNA molecules. MicroRNAs (miRNAs) are small (~21nt) regulatory RNAs that repress the translation and increase the degradation rate of their target mRNAs through base-pairing interaction. Though many of the core components of the miRNA biogenesis and effector pathways have been identified, it is likely that there are many proteins yet to be identified. Herein, I describe an RNAi-based high-throughput reverse genetic screen that I have developed to identify novel miRNA interacting factors in human cell lines. The screen identified a number of candidate genes; the candidates were validated with independent RNAi-knockdowns in human cell lines using a fluorescence-based reporter that monitors perturbations in miRNA efficacy.

BIOGRAPHICAL SKETCH

The Author received a Bachelor of Science degree in Biochemistry and a Bachelor of Art in both Chemistry and Psychology from New Mexico State University in 2009. While attending New Mexico State University the Author conducted research in a biochemistry laboratory under the supervision of Dr. Shelly Lusetti and in an instrumental chemistry lab under the direction of Dr. G. A. Eiceman.

ACKNOWLEDGEMENTS

This research has been supported by an NIH training grant awarded to the Biochemistry Molecular and Cellular Biology (BMCB) program and an American Cancer Society grant awarded to the Grimson research group.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: Introduction

Proper gene expression is essential for normal development, and is often perturbed in diseased states. Regulation of gene expression can occur at any number of points: transcriptionally, post-transcriptionally, and post-translationally. One mode of post-transcriptional gene regulation involves microRNAs (miRNAs), which are small (~21nt) endogenous RNA molecules. miRNAs are integrated

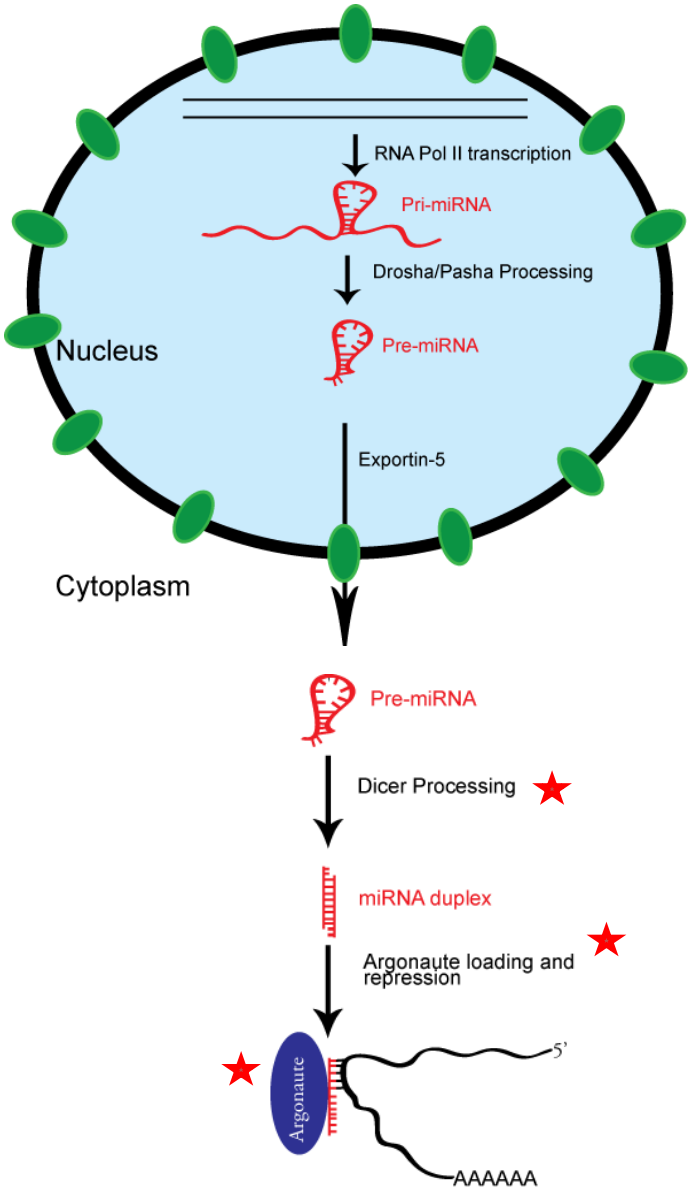

Figure 1: Diagram of Conventional miRNA pathway; stars indicate points where the mechanisms and the full complement of proteins are unknown.

into the RNA-induced silencing complex (RISC) to direct translational repression and destabilization of targeted mRNAs. The loss or dysfunction of miRNA-mediated target regulation has been implicated in cancer, heart disease, and evidence is emerging for a function for miRNAs in nearly every disease[1,2]. Many of the key components of the miRNA biogenesis and effector pathway are known, however it is likely that many factors remain to be identified. Identification of these unknown factors and elucidation of their functions will answer many of the critical questions that remain about how miRNAs are generated and function.

MiRNA Biogenesis

Mammalian miRNA genes are found in the following contexts: poly-cistronic clusters, within introns (this class of miRNAs includes mirtrons), and as autonomous genes. These loci are transcribed primarily by RNA polymerase II, and are regulated by transcription factors. In the case of poly-cistronic and autonomous miRNA genes, the product of transcription is long hairpin-containing primary-miRNA (pri-miRNA). The pri-miRNA is a substrate for the nuclear RNAseIII enzyme Drosha and its double-stranded RNA binding protein (dsRBP) cofactor Pasha (DGCR8 – Digeorge Critical Region 8 – in humans)[3,4]. The Drosha/Pasha complex cleaves the pri-miRNA at the base of the hairpin, generating an ~70nt precursor miRNA (pre-miRNA) with a 2-nt 3′ overhang. The pre-miRNA is then exported from the nucleus by the Ran-GTP-dependent Exportin-5[5]. Mirtrons bypass canonical nuclear miRNA processing and instead use the spliceosome to define the 5′ and 3′ ends of the pre-miRNA. The only known nuclear processing enzymes required for this class of miRNAs are the spliceosome and the lariat de-branching enzyme[6]. To further complicate matters, there are intronic-miRNAs that are not mirtrons. For this class of miRNAs, the mechanism that defines the 3′- and 5′-ends of the pre-miRNA is unknown[7].

In the cytoplasm, the pre-miRNA is cleaved at the base of the stem-loop by the cytoplasmic RNAseIII enzyme Dicer; this is facilitated by its dsRBP cofactor, TRBP (Transactivation responsive dsRNA Binding Protein)[8]. The result of this processing step is the generation of a ~21nt RNA duplex comprised of a miRNA guide strand (miRNA) and a passenger strand (miRNA*). The guide strand is loaded into the Argonaute (Ago) protein that is at the core of the RNA-induced silencing complex (RISC), whereas the passenger strand is rapidly degraded. For a subset of miRNAs, the guide strand is identified as the strand of the duplex with the less stably base-paired 5'end[9]; however this is not a universal convention. For those miRNAs that do not follow this convention, the strand selection mechanism is unknown. Both Dicer and TRBP participate in the loading of the miRNA into RISC and are the core components of the RISC Loading Complex (RLC), which also involves RNA helicase A, PACT, and likely many more proteins that are unknown[8,10].

Argonaute is a small RNA binding protein that also contains an RNAseH domain[11]. There are four Argonaute proteins in humans, only one of which (Ago2) retains its nuclease function, though all are capable of repression and destabilization of target mRNAs. Ago2-catalyzed miRNA-directed cleavage is indispensable for the regulation of several mRNAs, and the loss of Ago2 is lethal[12,13]. A relative of Argonaute proteins, the PIWI proteins, participate in germ-cell maintenance through interactions with distinct class of small RNAs termed PIWI-interacting RNAs (piRNAs)[14]. These PIWI-piRNA complexes function by suppressing mobile genetic elements[15,16]. In addition to Argonaute, RISC also contains one of the three TNCR6 (Tri-Nucleotide Repeat Containing 6; GW182 in *C.elegans* and *D.melanogaster*) homologs (A-C)[17]. Once the miRNA is loaded into RISC, it serves as a guide to localize the complex to partially complementary mRNAs to direct their repression and destabilization. How the RISC complex interacts with the ribosome to repress translation is still hotly debated and

poorly understood. Similarly, the mechanism by which RISC mediated destabilization occurs is also poorly understood.

MiRNA Directed Silencing

MiRNAs direct RISC to target mRNAs through conventional Watson-Crick base-pairing interactions between the miRNA and the mRNA primarily through the 3′UTR of the target mRNA. The 5′arm of the miRNA, specifically nucleotides 2-7 (the "seed sequence"), is largely responsible for these interactions[18]. Recent studies have demonstrated that base-pairing between the miRNA central region and cognate mRNAs can also cause repression and destabilization of the target mRNA[13]. Additionally, base-pairing that extends beyond the seed sequence can result in cleavage of the target mRNA, a mode of regulation previously thought to be absent in mammalian systems[12]. Because of the small size of the seed sequence which directs miRNA-mediated repression, a single miRNA can control the expression of a large number of target mRNAs. MiRNAs functions in a distinct pathway from siRNA; siRNAs repress targets through the cleavage and consequent degradation of target mRNAs, conversely, miRNAs predominantly represses targets in a cleavage-independent manner. It has been postulated that proteins that interact with sequences on the 3′arm of the miRNA may mediate cell-type specific miRNA repression and regulation of miRNA activity[19], though this is a poorly elucidated area of research.

Interactions between the miRNA and the target mRNA usually occur within the 3′ untranslated region (3′UTR) of the target mRNA[18]. Unfortunately, sequence complementarity between the miRNA seed sequence and sequences within the 3′UTR of the target mRNA (target site) are not sufficient for repression. Functional target sites have additional context constraints. For a given target site to be optimally functional it must lie within a region where mRNA secondary structure is minimal, and where

the GC content of the surrounding area is low[18]. Additionally, when a given miRNA target site is close to another miRNA target site, the two sites act synergistically to repress translation[18].

The base-pairing between miRNAs and target mRNAs results in translational repression and increased degradation of the target mRNA[20]. Several studies have suggested that RISC interacts with translation initiation factors such as eIF4F cap recognition complex, and other ribosome loading factors to affect translation; however, these studies were done using *in vitro* viral translation experiments or *in vivo* translation of IRES containing mRNAs and therefore may not be representative of the mechanism used to inhibit cellular mRNA translation[21,22]. Recently, elements of the translational machinery have been implicated in miRNA-mediated target regulation, namely EIF4A2, which is one of three interchangeable translation initiation factors with a high degree of homology and functional redundancy[23]. Consequently, and the endogenous mechanism is poorly understood.

In contrast to the poorly understood modes of translational repression, RISC is known to increase the rate of both decapping and deadenylation of targeted mRNAs. Argonaute-TNCR6 protein complexes have been implicated in the recruitment of deadenylase factors (CAF1, and CCR4) and the initiation of poly(A)-binding protein (PABP)-dependent deadenylation of targeted mRNAs. TNCR6 can also recruit the decapping proteins Dcp1 and Dcp2 to targeted mRNAs. Both de-capping and de-adenylation initiate mRNA degradation by cytoplasmic exonucleases. However, how this recruitment occurs, and the full complement of proteins involved is not known[24,25]. In the rare case of cleavage competent miRNA:target interactions the cleavage products are degraded by Xrn1 (5′→3′ exonuclease) and the cytoplasmic exosome (3′→5′ exonuclease)[26].

Argonaute, TNCR6 and other proteins involved in miRNA-mediated mRNA repression have been found in cytoplasmic mRNA processing foci called Processing-Bodies (P-bodies)[27,17]. These repressive complexes are the locus for both transient translational repression and mRNA degradation;

however P-body localization does not necessarily lead to mRNA degradation, there is considerable evidence suggesting that mRNAs can be released from P-bodies and then be fully capable of being translated. Recently, the post-translational modification of Argonaute proteins by poly(ADP)-ribosylation has been identified as the initiator of RISC P-body localization[28]. Unfortunately, the functional significance of P-body localization of RISC and how the fate of P-body targeted mRNAs is determined is not fully understood.

Regulation of miRNAs

Some information is known about the regulation of miRNA transcription and processing. Most miRNAs are RNA polymerase II transcripts, and like all RNA polymerase II transcripts, miRNAs are subject to transcriptional regulation. miRNAs are essential regulators development and are required for maintenance of differentiated cell fates, consequently, the transcription of miRNA genes needs to be tightly regulated. Several transcription factors (cMyc, p53) are known that induce or repress miRNA genes in response to cellular stresses and developmental cues[10]. In addition, miRNAs can be regulated at each processing step. For example, several factors that alter the affinity of the Drosha/Pasha complex for subsets of miRNAs have been found, consequently increasing the processing efficacy for a subset of miRNAs while decreasing the processing of others[29]. This is exemplified by the p68 and p72 directed increase of Drosha processing of pri-miR-16-1, pri-miR-145, and pri-miR-143[10]. In the cytoplasm, phosphorylation of TRBP increases its affinity for Dicer, stabilizing the TRBP/Dicer complex and leading to a global increase in the accuracy and rate of Dicer pre-miRNA processing. The full complement of transcription factors and regulators of miRNA processing are unknown.

Argonaute proteins, the central protein component of miRNA-RISC, are regulated by post-translational modification, which modulate the localization, stability, and activity of the Argonaute proteins. One such modifications is poly(ADP)-ribosylation which results in the localization of RISC to

P-bodies and a concurrent reduction in the efficacy of miRNA-mediated repression[28]. Alternatively, the hydroxylation of the δ carbon of proline-700 in the PIWI domain of Ago2 confers stability to the protein, possibly by promoting the *cis* conformation of proline-700, and that increases its repressive capacity[11,30,31]. There are probably more mechanisms and proteins that regulate Argonaute activity that are still unknown.

Alternatively, miRNAs are directly regulated. The best characterized example of direct miRNA regulation is the regulation of let-7 by Lin-28 and terminal uridylyl transferase (TUT-4). Lin-28 regulates both the Drosha- and Dicer-mediated processing of let-7. In the nucleus Lin-28 prevents Drosha processing by preventing Drosha from accessing the pri-let-7 hairpin[10]. In the cytoplasm, Lin-28 recruits the uridylyl transferase TUT4 to the pre-let-7 structure to add a poly(U) tail, the poly-uridylation of pre-let-7 results in rapid degradation[32,33]. Given the great necessity for the proper temporal and cellular expression of miRNAs, it is likely that many miRNAs are similarly regulated, though there is little known about the regulation of the majority of miRNAs. There are many more TUTase proteins, several of which have unknown functions.

Above, I have highlighted several features of miRNA biogenesis, regulation, and function that are poorly or incompletely understood. The results outlined below constitute an initial foray into identifying the proteins involved in the processes of miRNA biogenesis, function, and regulation and elucidate the function of these candidates. By identifying novel miRNA-associated proteins, many of the unresolved questions surrounding miRNAs can be answered.

Chapter 2: Results

Project Goals

MicroRNAs (miRNAs) are small (~21nt) post-transcriptional regulatory molecules that are loaded into RISC (RNA induced silencing complex). MiRNAs repress target mRNAs through base-pairing interactions between the miRNA and partially complementary sequences found predominantly in the 3′UTR of the target mRNA. The RISC-miRNA complex mediates the translational repression of the targeted mRNA and recruits decapping and deadenylation complexes to the mRNA through unknown mechanisms. Though the essential complement of proteins is known for the miRNA pathway, many questions remain regarding miRNA-RISC loading, regulation of the protein components of the pathway, and regulation of miRNAs themselves. I hypothesize that many proteins that interact with miRNAs and miRNA-associated factors to facilitate miRNA-mediated repression, and regulate miRNAs have yet to be identified.

To identify miRNA-associated proteins I have developed a high-throughput pooled RNAi screen. Using a fluorescent reporter system specifically sensitive to perturbations in miRNA activity, I was able to identify candidate genes that appear to be essential for the efficacy of miRNA-mediated gene repression. The reporter system is comprised of a dsRed fluorescent gene attached to the HMGA2 3′UTR, which is targeted by the let-7 miRNA; a GFP fluorescent gene attached to a version of the HMGA2 3′UTR in which the let-7 sites have been ablated lends specificity to the screen by serving as an internal reference for fluorescence intensity, excluding from consideration events that alter dsRed fluorescence independent of perturbations in miRNA activity. Because the only differences between the two construct is the presence or absence of a miRNA target site, knocking-down genes required for

miRNA efficacy results in a change in dsRed fluorescence without altering GFP fluorescence, such genes constitute candidate miRNA-associated factors. I then validated miRNA-associated candidates found in the RNAi screen by testing RNAi constructs for each candidate individually, first in the same reporter system with which they were first identified, then in a related, but independent system.

Reporter System

To monitor changes in miRNA activity in response to the RNAi gene knockdown, I have designed an integrating fluorescent reporter system that is responsive to changes in miRNA efficacy. Consequently, gene knockdowns that alter the activity of miRNAs result in a change in the fluorescence intensity levels of the reporter constructs. The reporter is comprised of a dsRed fluorescent protein gene attached to the native HMGA2 3′UTR (dsRed-HMGA2-wt), which is strongly down-regulated by the let-7 miRNA[34,35].

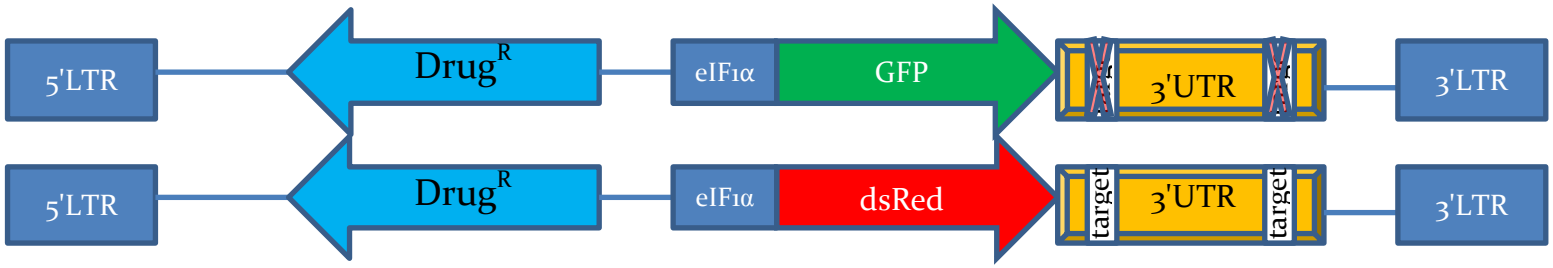

Figure 2: Reporter constructs design. LTR refers to long terminal repeat and is necessary for viral integrations. eIF1α is the eIF1α promoters. Red X's indicate ablated miRNA target sites.

The two possible phenotypic consequences of gene knockdown are an increase in dsRed and decrease in dsRed expression. Many RNAi constructs likely affect the reporter activity irrespective of involvement in miRNA metabolism; therefore I have designed a second reporter construct that functions as an internal reference of base-line fluorescence activity. This reporter is comprised of a green-fluorescent protein (GFP) gene and an HMGA2 3′UTR in which the let-7 target sites have been ablated by the mutation of 2 nucleotides, preventing let-7 mediated regulation (GFP-HMGA2-mut). Consequently, in the screen I identified RNAi constructs that increase or decrease dsRed intensity relative to that of GFP.

By identifying RNAi constructs that alter the fluorescence intensity of dsRed but not GFP, I effectively identify genes that have a function in the miRNA post-transcriptional regulation system. Increases in dsRed-to-GFP intensity are a consequence of knocking-down genes that are required for increasing miRNA activity, such as Drosha, Ago1, Ago3, and Ago4, etc. Alternatively, decreases in dsRed-to-GFP

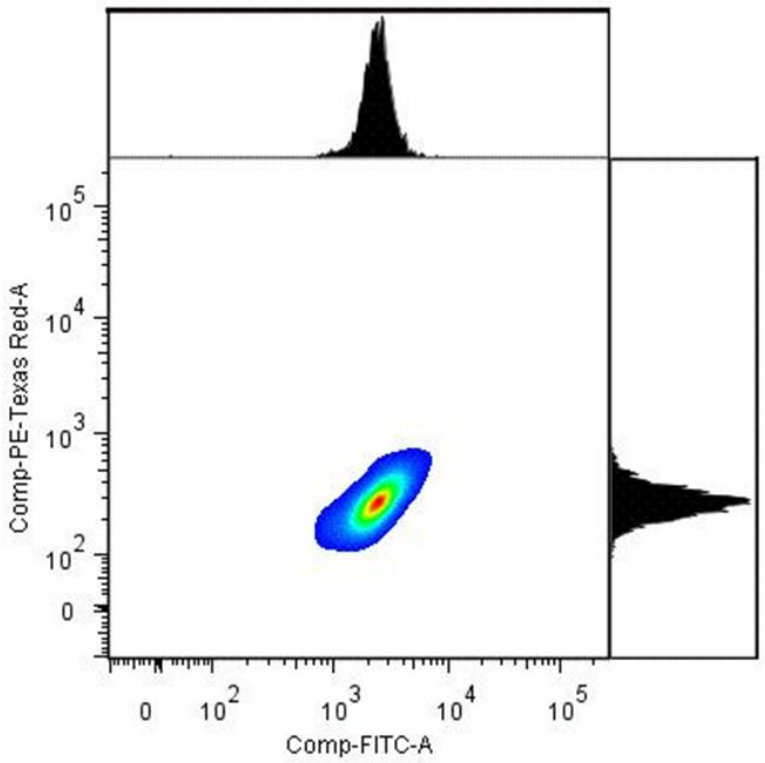

Figure 3: Two-dimensional FAC-analyzer plot. X-axis is GFP intensity (read by the FITC channel). Y-axis is dsRed intensity (read by the PE-Texas Red channel). Histograms for the dsRed and GFP intensities are shown to the right and top of the two-dimensional plot, respectively.

intensity result when genes that normally repress miRNA processing or function are knocked-down, such as Lin-28, TUT-4 etc.

Central to the screen is the development of a dual-fluorescence reporter cell line. Initial efforts to this end were complicated by poor dsRed expression. After pursuing several avenues to resolve this problem, including alternate forms of red fluorescent protein (mCherry, dsRed2, and dsRedExpress) and different promoters (CMV, eIF-1α and PGK), a working, dual-fluorescence cell line was generated in

HEK293 cells. This reporter cell line was funneled through a single cell, generating a genetically homogenous clonal cell population. The genetic homogeneity of this population is indispensable to the success of the screen; this ensures that any changes in the GFP-to-dsRed intensity are due to the function of the RNAi and not stochastic events arising from a heterogeneous population of reporter cells. To confirm that the clonalized cell lines do originate from a single cell and have homogenous dsRed and GFP fluorescence, the cells were taken to the fluorescence activated cell analyzer (FAC analyzer) to evaluate the dsRed and GFP intensities. The clonal cell line displays a narrow fluorescence intensity profile, confirming that the reporter cell population is genetically homogenous. The two dimensional plotting of the fluorescence spectra displaying GFP intensity (FITC) and dsRed intensity (PE-Texas Red) with adjunct histograms displaying the spread in each channel is the simplest way to confirm that cell lines are indeed clonal. Since cell lines are derived from a diverse collection of tissues, there is much variation between cell line in terms of their mRNA and miRNA expression profiles. Consequently, it is unlikely that a single 3′UTR reporter construct, targeted by a single miRNA, in a single cell line will be adequate to identify all novel miRNA components. In addition to the development of the reporter cell line for the initial screen in HEK293 cells, I have also generated clonal reporter A549 cells, and cell lines expressing the alternate 3′UTR reporters.

Variations of the depicted constructs have been created for the purposes of validation and additional screens. Namely, a construct made with a miniaturized version of the HMGA2 3′UTR, in which the majority of the intervening sequence between the let-7 sites have been removed. This reporter construct is susceptible to the same down-regulation by let-7 that the full-length reporter experiences, without additional regulation by other miRNA target sites and unknown elements contained within the HMGA2 3′UTR. Additionally, this truncation removes a number of the alternative cleavage and poly-adenylation sites that may account for some stochastic variations within the reporter system.

Though the dual-fluorescent system likely reduces the false-discovery rate, there are still a considerable number of false-positives, which were removed through subsequent validation experiments using the same reporter construct described above and variations thereon. Specifically, by identifying candidate RNAi constructs that have a similar effect across different 3′UTR reporters, or across replicates using the same 3′UTR reporter, many of the false-positives have been weeded out. Before new genes in the miRNA pathway can be identified, it is necessary to demonstrate that the reporter cell lines respond as expected when known miRNA components are knocked-down with RNAi.

Infection with RNAi to known miRNA components

After clonalized cell lines were chosen, cells were infected at a high multiplicity of infection (MOI) with individual RNAi viruses to known miRNA components, including Drosha, DGCR8, Ago3 and Ago4; the high MOI ensure that each cell received at least one viral particle carrying the RNAi

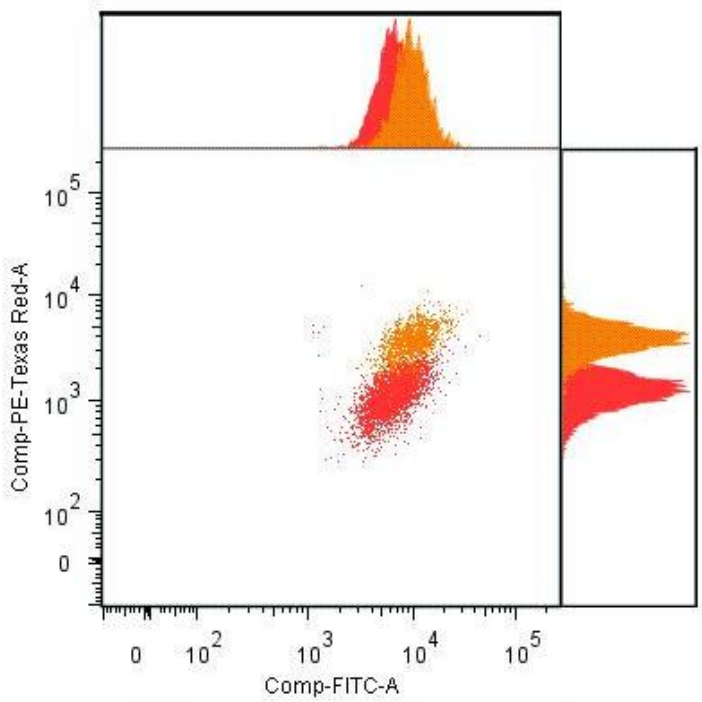

Figure 4: Two-Dimensional FAC-analyzer scatter plot of clonal cell populations treated with RNAi targeting Ago3 (orange) and clonal cell population not treated with RNAi (red).

construct to ensure maximal target gene knockdown. Following infection, cells were incubated for 6 day in DMEM with 10% FBS, Pen/Strep, and Puromycin (hereafter referred to as selective media) to select for viral integrants. Cells were taken to the FAC analyzer with control cells that were not exposed to any RNAi viruses or those that knockdown LacZ. Since the knockdown efficacy of the RNAi construct is variable, not all knockdowns created the expected result, with some having no measurable effect on fluorescence activity whatsoever. The knock-down efficacies for many, but not all, of the RNAi constructs generated by the Broad Institute have been determined. For the single RNAi virus infections that produced an observable change in fluorescence intensities of the wild-type reporter there was also a noticeable change in the mutant reporter, though not as pronounced. Several RNAi constructs to known miRNA factors displayed the expected change in the fluorescent reporter activity and the changes are consistent with the knockdown efficacies; example displayed in Figure 4.

Pilot Sort

While single RNAi virus infections produce an observable change in fluorescence intensities of the wild-type reporter without drastically altering the mutant reporter, the screen itself is infinitely more complex. Because thousands of RNAi viruses are present and exerting a multitude of effects on the infected cells, including the desired, targeted perturbation of miRNA activity, but also the disruption of essential cellular functions: metabolic activity, mitosis, chromosome repair, etc., all of which may alter

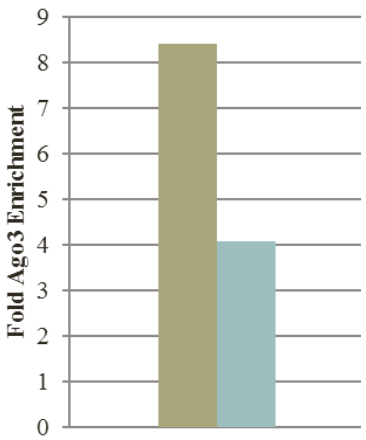

Figure 5: Fold enrichment of Ago3 relative to Dbr1 in unsorted population (Blue) and the collected top 20% of dsRed expressing cells (grey)

reporter activity adversely. To determine if RNAi virus infected cells that genuinely have altered reporter system activity can be effectively separated from those that do not have alter the reporter activity for meaningful reasons, a trial sort was undertaken. To preclude the possibly confounding situation in which a given cell receives RNAi construct to two or more different targets, I used a MOI of 0.1, such that only one out of every 10 cells is infected with a single RNAi virus, and the possibility that a cell receives two viruses is 0.01, making double infection less common and less likely to cause problems in later data analysis. In this pilot sort, cells were infected at a low MOI concurrently with RNAi viruses to Ago3 and Dbr1. Both of the RNAi constructs chosen target factors involved in RNA metabolism; however they are involved in disparate processes. These two genes were chosen because in a screen with a full RNAi library many aspects of nucleic acid metabolism will be targeted, and the reporter system must be specifically sensitive to perturbations in miRNA processes, without reacting to broader changes in RNA abundance, turnover, localization, et cetera. Cells were maintained for a week in selective media then taken to the fluorescence activated cell sorter (FACs), where the top 20% of dsRed expressing cells were collected and re-cultured.

The genomic DNA from the collected cell population and the parent population (unsorted cells) were collected for further analysis. The genomic DNA was then used as a template for quantitative PCR to determine the relative abundance of the RNAi hairpin integrants for Dbr1 and Ago3 in both populations sampled. In the top 20% cell population, Ago3 hairpin was slightly over 8-fold more abundant than the Dbr1 hairpin; while in the parent population Ago3 was 4-fold more abundant (likely an artifact of titer difference or qPCR primer efficiency). Thus, this result indicates that true miRNA-associating factors can reliably be separated from those that perturb unrelated RNA processes.

While much simpler than the planned screen, this experiment demonstrates that cell populations with knockdowns of miRNA-associated genes can be effectively separated from cells that do not have

gene knockdowns associated with the miRNA pathway. Additionally, this experiment roughly establishes cutoffs for what populations should be collected, and that a smaller percentage of the parent population should be collected to give a better enrichment of miRNA-associated candidates.

RNAi Libraries

To identify putative miRNA-associated genes, I need to find genes that alter the efficacy of miRNA-mediated target repression. To identify such candidate genes I used virally delivered short-hairpin RNAs (shRNAs or RNAi) to knockdown genes across the genome and measure the effect of the knockdown on miRNA activity. This wa carried out using pooled RNAi constructs (Broad Institute RNAi Consortium). In this system, each cell had no more than one RNAi integrant, and within the population of cells every gene knockdown should have been represented several times, representing a broad number of genome integration sites. This was achieved by using an MOI of 0.1, meaning that only 1 out of every 10 cells receives a virus. Two pooled RNAi libraries were used for the experiments described below. The "Achilles" library contains an average of 5 RNAi constructs to each annotated gene in the genome (as of 2007) with a total number of 55,000 RNAi viruses; the other, is enriched for RNAi constructs targeting gene thought to be involved in RNA biology, with an average of 5 RNAi constructs to each of approximately 500 genes (total of c.a. 10,000 unique RNAi viruses). Multiple RNAi constructs to each gene will reduce false-positives attributable to RNAi off-target effects. The preliminary screens described below were carried out with both libraries, though the latter library was used more extensively.

Early Screens

The first screens were conducted in a clonal cell line with dsRed as the wild-type reporter, in a system using full-length HMGA2 3′UTR. Approximately 10 million cells were infected with the RNA Biology at an MOI of 0.1 yielding 1 million infected cells. The low multiplicity of infection and large number of singly infected cells means that a number of viral integration sites will be sampled for each RNAi hairpin. The cells were maintained in selective media for a week, with frequent media changes to reduce cellular stress. An additional screen was completed using the larger Achilles Library; 100 million cells were infected at an MOI of 0.1, maintained and treated as the cells used in the smaller RNA Biology Library. This experiment was conducted in both GFP–wild-type and dsRed–wild-type cells, however, the dsRed-wild type cells were contaminated by bacteria from the FACs facility, and therefore discarded.

Cells were prepared for FACs, leaving approximately 10% of the cells behind to re-grow and serve as a control group. During the sorting procedure, cells were sorted into 3 populations that were

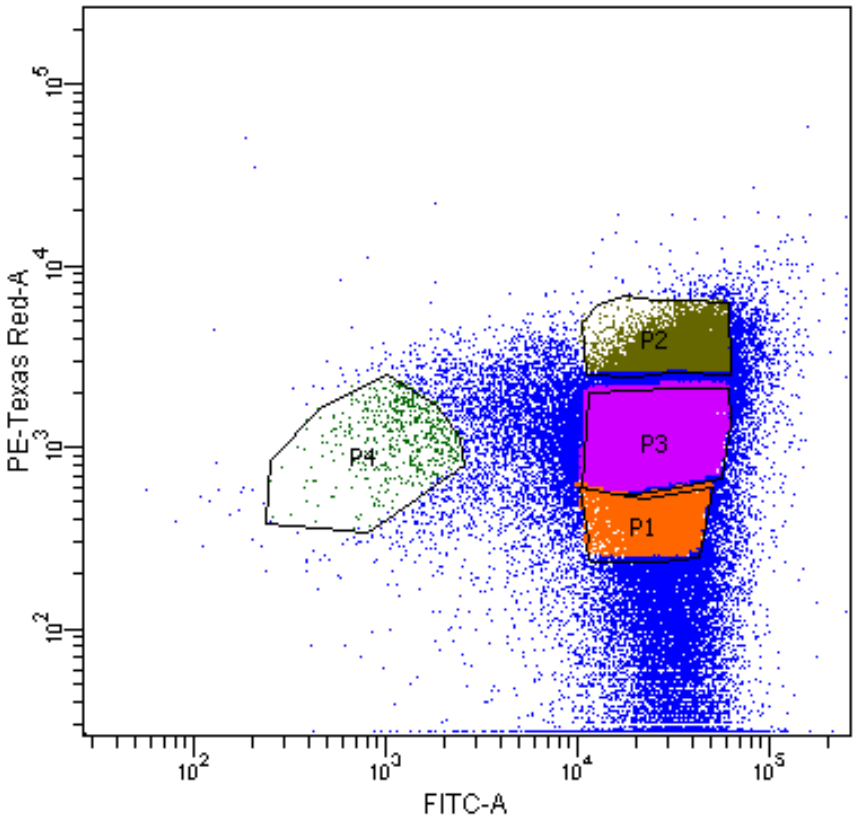

Figure 6: FAC-sorter generated image depicting populations collected (P_1, P_2, P_3, and P_4) and the overall distribution of cell fluorescent intensity on a two-dimensional scatter plot (dsRed wild-type intensity on y-axis)

determined by the negative control taken to the sorter. A population (gate) was drawn around the majority of the negative control population on the dsRed-by-GFP FACs two-way spectra; this population constitutes the "No Effect" or "Unaffected" population. The reporter cells with the RNA Biology RNAi library were then run through an initial analysis; the population that lay above the unaffected population was collected and constituted approximately 2% of the parent population. Additionally, the population that lay below the unaffected population was collected and constituted nearly 5% of the population, a number that was highly variable during the sorting. In the upper population, about 100,000 cells were collected, in the unaffected cell population there were 1 million, and in the lower population 250,000 cells were collected.

The average survival rate of sorted cells was 75%; therefore the numbers collected do not match the number of cells retrieved. Cells were returned to media and re-cultured; after 2 weeks the cells of the lower and upper populations were re-sorted. For the upper population the top 10% and the bottom 90% were collected; conversely, for the lower population the bottom 10% and 90% of the population was collected.

Screens Completes and Submitted for Sequencing	
Reporter Cell Line	***RNAi Library***
dsRed wild-type clonal cell line 1	RNA Biology Library
dsRed wild-type clonal cell line 2	RNA Biology Library
GFP wild-type clonal cell line 1	Achilles Library

Table 1: completed screens submitted for Illumina sequencing

Genomic DNA was harvested from the collected population and the retained unsorted population once the cells numbered between 10 million and 20 million cells. The genomic DNA was then subjected to a PCR protocol designed by the Broad Institute that amplifies the RNAi hairpin from the genomic DNA and appends a barcode, allowing the originating population to be identified. Using controls to

ensure that no reagent was contaminated between samples, each population from three independent sorts were prepared and submitted for sequencing (two from dsRed wild-type and one from GFP wild-type).

Data Analysis and Candidate Selection

Once sequencing runs were completed, the data was processed such that for each gene, the number of reads for each hairpin to that gene was tabulated, and the gene candidates were ranked by the second-most abundantly represented RNAi hairpin. By using the second-most abundantly represented hairpin to rank each gene, the ranking will not be subjected to additional variability on account of off-target effects of a single hairpin. Consequently, genes with widely variable enrichment values between RNAi hairpins were removed from consideration as possible candidate genes.

A further criterion was added to select candidates, the knock-down efficacies for each hairpin were considered, and I required the knock-down efficacies to correlate with the enrichment levels (i.e. the least enriched hairpin should be the one with the poorest knockdown efficacy, and the top 2 or 3 most enriched RNAi hairpins for a given gene should be the most efficacious). Additionally, correlations of rankings between screens were used as further criteria for selecting likely candidates. Ideally, true candidates should appear highly enriched across multiple screens, there is however a large amount of variability both within and between screens that artificially enrich genes that are not likely candidates.

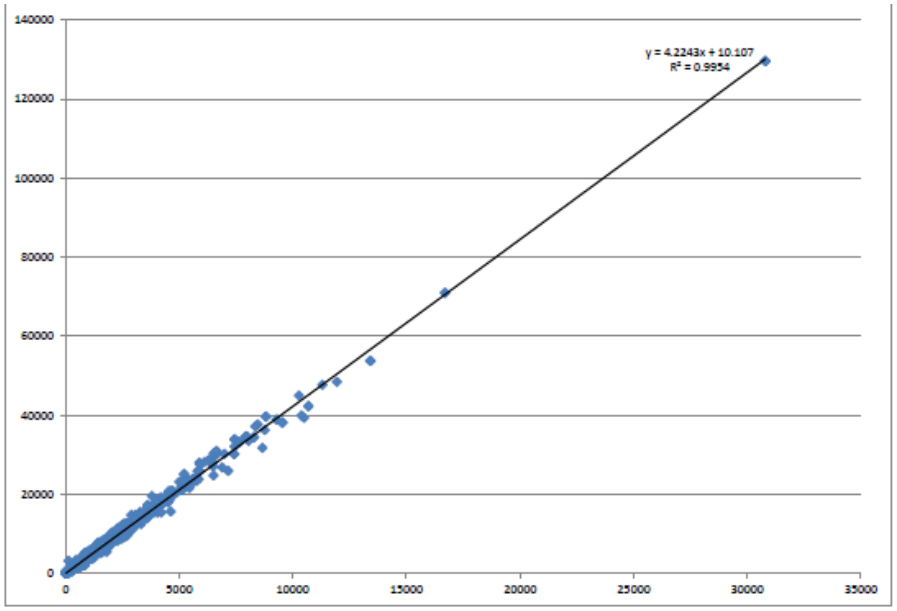

Figure 7: Scatter plot of raw sequencing read counts for technical replicate 1 (x-axis) versus technical replicate 2 (y-axis) with linear best-fit line

Within each screen, the technical replicates – independently bar-coded sequencing samples that came from the same collected cell population, but went through DNA collection and PCR preparation independently – were highly correlated, with R^2 values between 0.97 and 0.99, indicating that little variability is added to the screen by sequencing sample preparation. However, the correlation between independent screens using the same RNAi library (i.e. the two screens in the dsRed reporter cell lines) was much weaker, with R^2 values ranging from 0.4 to 0.6. Consequently, it appears that much of the variability in the screen arises from the FACs collection, and possibly from the infection and handling of cells.

To visualize the spread of enrichment values for all genes in the screen juxtaposed with known miRNA associated factors S-curves were generated, wherein for a given population comparison (i.e. hairpins enriched in the "Increased" cell population relative to the entire "Unsorted" population) the individual RNAi hairpins were ranked by enrichment values then plotted. From such graphs it is readily apparent that the RNAi constructs targeting known miRNA-associated factors are not the most enriched, but most core components do appear in the top 500 hairpins. Additionally, RNAi to factors that are not required for miRNA metabolism are depleted. The screen provided the initial data to identify possible novel miRNA-associated factors, however both false-positives and false-negatives are likely in a high-throughput technology such as this, therefore additional validation is necessary. Approximately 125

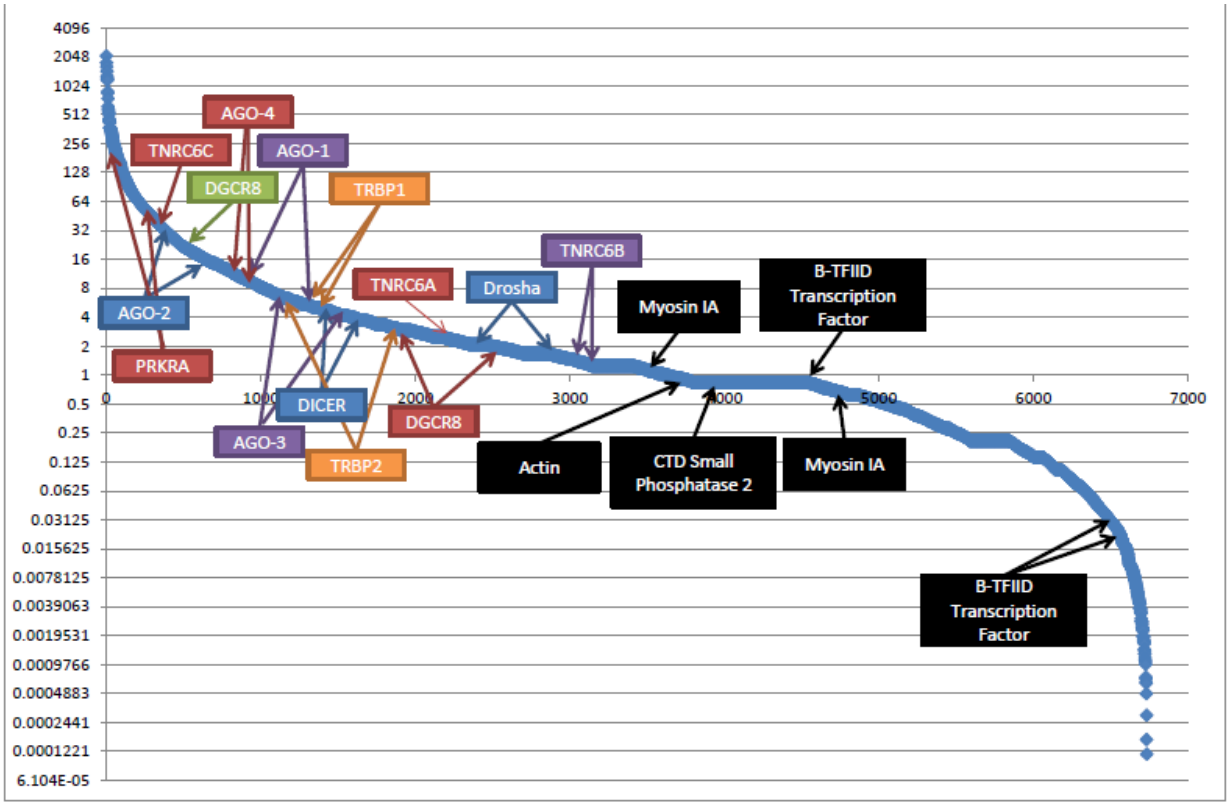

Figure 8: S-Curve of enrichment values all reads in the increased population relative to unaffected population. In colors are known miRNA-associated factors, in black are factors known to have no effect on miRNA activity.

candidates were selected based on the criteria described above combined with research into the known pathway involvements, gene function, or in the case of unknown genes, analog function and predicted functions based of homology. The candidates selected for further validation were genes that appeared within the top 200 most enriched genes in both screens that used the RNA Biology RNAi library; very few candidates were selected using the data acquired from the Achilles Library screen.

Candidate Validation

Cells expressing the miniaturized version of the HMGA2 3'UTR were infected at a high MOI with individual RNAi viruses (2 to 3 RNAi constructs that were enriched in the primary screen for each gene) to each of the candidate genes in 24-well format. After 1 week of selection for integrants, cells were taken to the FAC analyzer with a control cell samples infected with RNAi virus knocking-down LacZ and other genes not thought to be involved in miRNA biology. Two candidates showed increased dsRed activity across all RNAi constructs evaluated: EIF4A1, a core component to the translation initiation machinery, and ZCCHC8, a zinc finger protein with high homology to a number of alternative polyA polymerases. Given that only 2 of the 125 candidate genes evaluated pass the first round of validation, I must conclude that there is an extremely high false-positive rate for the screen, something that should be improved upon.

The two genes that passed primary validation: EIF4A1 and ZCCHC8 are both from gene families that are known to have a function in miRNA metabolism. Specifically, ZCCHC8 is one of many homologs of TUT-4, which is a known negative regulator of the miRNA *let-7* and associates with the protein Lin-28 to actively degrade the *let-7* miRNA[36]. ZCCHC8 was identified as being enriched in the increased cell population collected for both screens with the RNA Biology RNAi library; this precludes it from being a negative regulator of miRNA and suggests that it may function as a positive effector of

miRNA activity. Additionally, though ZCCHC8 comes from a family of alternative poly-A polymerases, there is no published data supporting a polymerase function for this protein.

The second candidate to pass primary validation, EIF4A1 is one of 3 EIF4A proteins which are a part of the EIF4F initiation component present in 40S initiation complex. Earlier this year, EIF4A2, but not EIF4A1, was shown to participate in miRNA-mediated translational repression. However the system being used to explore the function of the EIF4A proteins was based on IRES systems, which in the history of the miRNA field have been largely inconsistent. Additionally, given the high degree of homology between the two proteins (97%) it is likely that both proteins function in the miRNA pathway[23].

Chapter 3: Discussion

Initial validation evidence seems to suggest that the screen has truly identified 2 novel candidates out of the possible 125 tested with FAC analysis of reporter cell-line knock-down treatment. The screen identified two putative true-positives, which include EIF4A1 and ZCCHC8, both of which have relatives already implicated in the processing and regulation of miRNAs. However, further validation using the traditional luciferase-based reporter assays and potentially protein-based assays are necessary to confirm

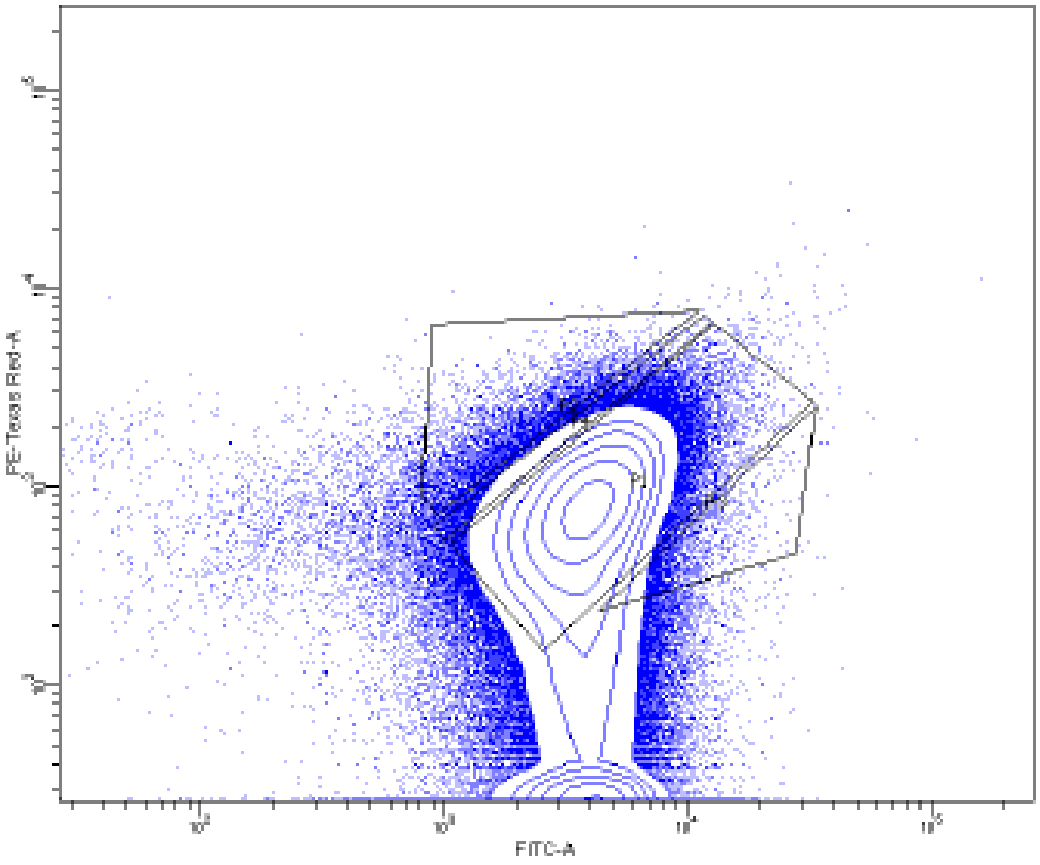

Figure 9: Two-dimensional plot generated by the FAC-sorter displaying the modified population selection criteria. P1: Top 2%, P2: Upper population, P3: Unaffected population, P4: Decreased population

these candidates.

The current screening methodology has an appreciably high false-positive rate; one path to finding additional candidates includes reducing the number of false positives and increasing the true

positives. To accomplish this goal, one must consider the FACs selection criterion to be the most likely source of noise; therefore this is the one area that can be refined. Consequently, I bring up the following

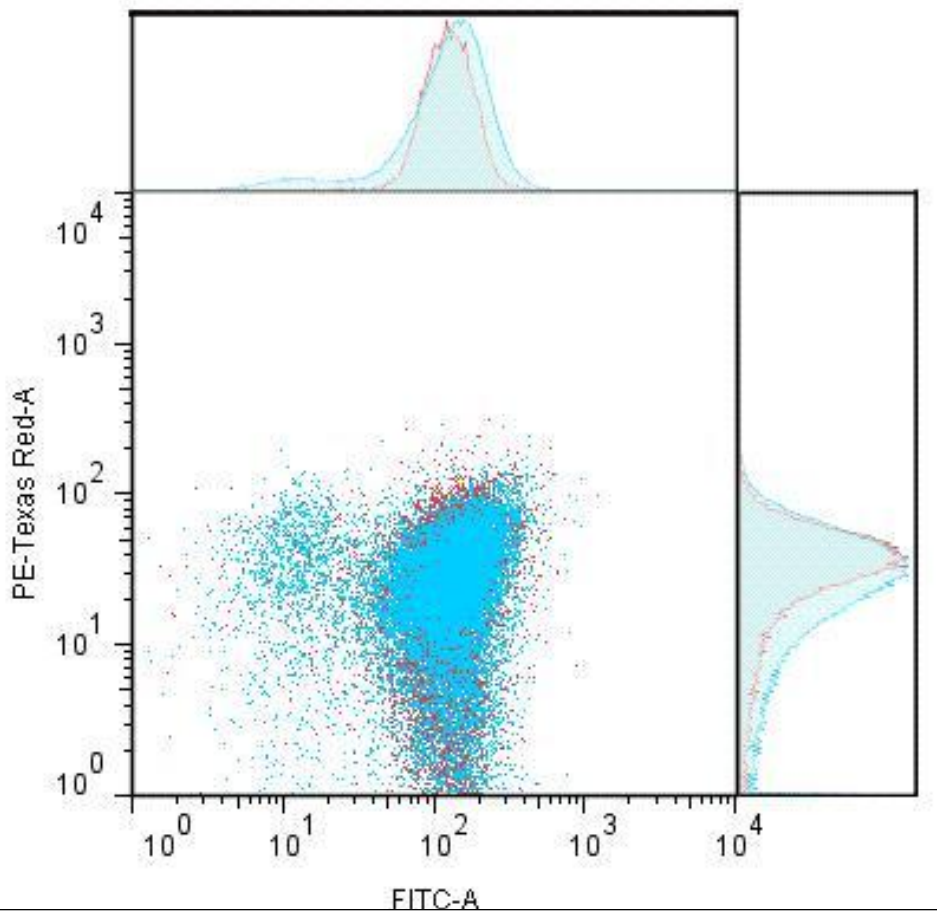

Figure 10: 2 Dimensional FACs overlay image of sorted Unaffected population (Red) and Top 2% population (Blue)

observations: first, there is a wide stochastic variation along both the wild-type and mutant reporter axes, resulting in a population spread along a diagonal expanse. Consequently, modifying the selection criteria to take this into consideration, by using a scaffold that takes that fact into account may yield more favorable results. Figure 9 depicts an example of such favorable population selection criterion. As an aside, the current iteration of the first round of population sorting does not consistently shift the sorted population relative to the unsorted population (see figure 10).

In closing, I would like to acknowledge that this screening technology identifies novel factors involved in miRNA biogenesis and action. Additionally, it can be amended to search for any *trans*-

factor required for the regulation of a *known* nucleic sequence element, regardless of whether it is an effector of DNA or RNA. To further demonstrate this point, I would like to point out a parallel study in our lab that identified and validated several proteins using this methodology that appear to regulate a novel sequence elements that are highly conserved across vertebrates'.

Works Cited

1. Liu, X. et al. A Necessary Role of miR-221 and miR-222 in Vascular Smooth Muscle Cell Proliferation and Neointimal Hyperplasia. *Circulation Research* **104**, 476-487(2009).

2. Mayr, C. & Bartel, D.P. Widespread Shortening of 32UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* **138**, 673-684(2009).

3. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419(2003).

4. Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Current Biology* **14**, 2162-2167(2004).

5. Yi, R. et al. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development* **17**, 3011-3016(2003).

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

7. Lagos-Quintana, M. New microRNAs from mouse and human. *Rna* **9**, 175-179(2003).

8. Chendrimada, T.P. et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**, 740-744(2005).

9. Schwarz, D.S. et al. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208(2003).

10. Winter, J. et al. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology* **11**, 228-234(2009).

11. Song, J.-J. et al. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science (New York, N.Y.)* **305**, 1434-7(2004).

12. Karginov, F.V. et al. Diverse Endonucleolytic Cleavage Sites in the Mammalian Transcriptome Depend upon MicroRNAs, Drosha, and Additional Nucleases. *Molecular cell* **38**, 781-788(2010).

13. Shin, C. et al. Expanding the MicroRNA Targeting Code: Functional Sites with Centered Pairing. *Molecular cell* **38**, 789-802(2010).

14. Carmell, M.A. et al. The Argonaute family : tentacles that reach into RNAi , developmental control , stem cell maintenance , and tumorigenesis. *Genes & Development* 2733-2742(2002).doi:10.1101/gad.1026102.GENES

15. Girard, A. et al. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199-202(2006).

16. Malone, C.D. et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**, 522-35(2009).

17. Liu, J. et al. A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* **7**, 1261-1266(2005).

18. Grimson, A. et al. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular cell* **27**, 91-105(2007).

19. Bartel, D.P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**, 281-297(2004).

20. Guo, H. et al. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835-840(2010).

21. Mathonnet, G. et al. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science (New York, N.Y.)* **317**, 1764-7(2007).

22. Pillai, R.S. et al. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science (New York, N.Y.)* **309**, 1573-6(2005).

23. Meijer, H.A. et al. Translational Repression and eIF4A2 Activity Are Critical for MicroRNA-Mediated Gene Regulation. *Science* **340**, 82-85(2013).

24. Fabian, M.R. et al. Mammalian miRNA RISC Recruits CAF1 and PABP to Affect PABP-Dependent Deadenylation. *Molecular cell* **35**, 868-880(2009).

25. Chen, C.-Y.A. et al. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol* **16**, 1160-1166(2009).

26. Valencia-Sanchez, M.A. et al. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & development* **20**, 515-24(2006).

27. Parker, R. & Sheth, U. P bodies and the control of mRNA translation and degradation. *Molecular Cell* **25**, 635-646(2007).

28. Leung, A.L. et al. Poly(ADP-Ribose) Regulates Stress Responses and MicroRNA Activity in the Cytoplasm. *Molecular cell* **42**, 489-499(2011).

29. Davis, B.N. et al. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**, 56-61(2008).

30. Qi, H.H. et al. Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature* **455**, 421-424(2008).

31. Vitagliano, L. et al. Preferred proline puckerings in cis and trans peptide groups: Implications for collagen stability. *Protein Science* **10**, 2627-2632(2001).

32. Heo, I. et al. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Molecular Cell* **32**, 276-284(2008).

33. Newman, M.A., Thomson, J.M. & Hammond, S.M. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *Rna New York Ny* **14**, 1539-1549(2008).

34. Ormö, M. et al. Crystal Structure of the Aequorea victoria Green Fluorescent Protein . *Science* **273** , 1392-1395(1996).

35. Prasher, D.C. et al. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **111**, 229-233(1992).

36. Heo, I. et al. TUT4 in Concert with Lin28 Suppresses MicroRNA Biogenesis through Pre-MicroRNA Uridylation. *Cell* **138**, 696-708(2009).