Diversity in intracellular microRNA regulatory networks: microRNA-21 and beyond

by

John Rolf Androsavich

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in The University of Michigan
2012

Doctoral Committee:
Professor Nils G. Walter, Chair
Professor David R. Engelke
Professor Alice Telesnitsky
Associate Professor Jason Edward Gestwicki
Assistant Professor Julie Suzanne Biteen

To Rolf Hoehn

Acknowledgements

This thesis would not have been possible without the help and support of so many fantastic human beings. My gratitude extends far beyond that which can be conveyed in just a few words.

I cannot overstate my appreciation for the opportunities, resources, and guidance provided by my advisor, Dr. Nils Walter. Somehow or another, Nils has always found time to offer his scholarship to my work; and I am truly fortunate to have such a talented, dedicated and patient mentor. I am also fortunate to have received over the years encouraging and constructive feedback from the exceptional scientists on my thesis committee-- Dr. David Engelke, Dr. Alice Telesnitsky, Dr. Jason Gestwicki, and Dr. Julie Biteen. Thank you for your thoughtful comments in reviewing this thesis.

To my collaborators at Regulus Therapeutics, I am grateful for the chance to visit and be a part of the team. Thank you for welcoming me as a friend and helping to develop the miRNA-21 story. I especially like to thank Dr. B. Nelson Chau and Dr. Peter Linsley for providing me with the opportunity to research such a fascinating subject and for hours of fruitful discussion.

To my fellow Walter lab members, thank you for creating such a wonderful, fun and collaborative learning environment. I do not doubt that, with our efforts as the foundation, the Walter lab will one day succeed in surveying every last molecule of RNA! I especially wish to thank Dr. Sethu Pitchiaya for generously sharing with me from day one his talents and knowledge of science.

Certainly, this journey would have been impossible, or more likely never even attempted, if it wasn't for the support of my family and friends. I am indebted to my parents, Rolfa Jones and George Androsavich, my sisters, my grandparents, and the Zoschg family for their love, encouragement and support. I especially wish to thank my grandfather Rolf Hoehn whose company I miss dearly and whose character continues to

inspire me. To him I dedicate this thesis. I also thank all of my friends who have been a part in making so many great memories in Ann Arbor.

Last, and most importantly, I thank my wife Meredith who more than anyone else has shared in the joys of my successes and the burdens of my frustrations. I seriously could not have done this without your support; and I thank you for always reminding me, "Dreams take time."

John Androsavich
July 19, 2012

Table of Contents

List of Figures

List of Appendices

Abstract

Diversity in intracellular microRNA regulatory networks: microRNA-21 and beyond

by

John Rolf Androsavich

Chair: Nils G. Walter

microRNAs (miRNAs) comprise a broad class of short non-coding RNA that regulate gene expression by guiding the RNA induced silencing complex (RISC) to mRNAs containing complementary 'seed sites'. General principles of miRNA action in mammals have emerged over the last decade from research using cultured cancer cell models; however, it is unclear whether these principles apply *in vivo*, where few functional studies have been performed in healthy tissue. Furthermore, it is uncertain whether there are mechanistic or functional heterogeneities between the hundreds of miRNAs conserved across mammals.

The primary aim of this thesis was to compare the repressive activities of three highly abundant miRNAs— miR-21, miR-122, and let-7— in healthy liver tissue, and to contrast them with measured or previously reported activities in cancer cell lines. miRNA activities were measured based on (i) array profiling following pharmacological inhibition, and (ii) binding to polysome-associated target mRNAs. It was found that miR-21, compared to miR-122 and let-7, has surprisingly little impact on regulation of canonical seed-matched mRNAs. Moreover, miR-21 showed greatly reduced binding to polysome-associated target mRNAs, possibly due to reduced thermodynamic stability of seed pairing. Significantly, these trends are reversed in human cervical carcinoma (HeLa) cells, where miR-21 and other miRNAs showed enhanced target binding within polysomes and miR-21 triggers strong degradative activity toward target mRNAs. Taken

together, these results suggest that certain miRNA activities can be highly context dependent and miRNA pathways are overactive under pathological conditions. Additionally, bioinformatic analysis of sequence features in miR-21 and miR-122 responsive targets revealed low to moderate correspondence with previously established targeting trends derived from HeLa, exposing the complexity of *in vivo* target selection and suggesting cross-talk with other regulatory networks.

As an additional aim, a single molecule method was developed for directly observing the kinetic diversity in miRNA processes. The method combines particle tracking with step-wise photobleaching to probe dynamics and stoichiometry in complex assembly. Pilot studies indicate that fluorophore labeled let-7 miRNAs, detectable as single molecules, undergo a biphasic kinetic assembly when microinjected into live HeLa cells.

Chapter 1:
Diversity of microRNA Regulatory Networks and Implications in Disease

1.1 Introduction

The cell is a 'little box'. With this description of the black pores in the cork he observed underneath his microscope, Robert Hooke famously coined the term cell and gave it biological meaning in 1667[1]. Since then, descriptions of cells have, of course, grown more sophisticated, progressing along with philosophical and scientific understanding. More than two centuries after Hooke's publication of *Micrographia*, Theodor Schwann, in 1847, established the cell theory and described cells as "individual elementary units of all organisms[2]." By the early 1900s, following the advent of classic genetic theories of inheritance, and on the backdrop of atomic and quantum discoveries in physics, the insides of cells began to be better understood and appreciated. American zoologist Herbert Spencer Jennings commented, "A cell of a higher organism contains a thousand different substances, arranged in a complex system[3]." Soon thereafter, with the discovery in the mid-20th century of DNA as the hereditary unit of life, additional knowledge of cells came in the form of understanding its components rather than the cell itself; as Boris Ephrussi eloquently phrased it, "the cell robbed of its noblest prerogative, was no longer the ultimate unit of life. The title was now conferred on the genes, subcellular elements, of which the cell nucleus contain many thousands[4]." And so, the modern day disciplines of molecular biology, biochemistry, and genetics were born; and no longer was the cell a black box devoid of substance. Instead it was filled with DNA and protein (and RNA).

In 1958, after helping to deduce the structure of DNA, Francis Crick proposed the central dogma of molecular biology, a hypothesis that shaped the research landscape for the remainder of the century. As an extension to the 'one gene, one enzyme' theory proposed by Beadle and Tatum in 1941, the central dogma proposed that in gene

expression information passed unidirectionally between three biopolymers: hereditary information encoded in genes composed of DNA was passed down or transcribed to passive templates composed of RNA, which in turn was translated to final protein products that enacted cellular functions. In this model, DNA and protein were the active polymers, while RNA was simply the message, a temporary intermediate between the more important cellular components. In the 1970s and 80s, the passivity of RNA was challenged by several findings that certain cellular RNAs held the capacity for catalysis, much like protein, or storage of genetic information, much like DNA[5-7]. Nonetheless, RNA continued to remain largely underappreciated and in the shadows of the other biopolymers.

The Human Genome Project, launched in 1990, was an international effort to sequence the 2.85 billion nucleotides of human euchromatic DNA. The massive undertaking, pursued in parallel by the publicly funded International Human Genome Sequencing Consortium and the privately held Celera Genomics, took thirteen years to fully complete and cost several billion dollars (the private venture was shorter, less expensive). One of the primary objectives of the project was to discover novel protein coding genes and gain new insights for the development of treatments and preventions for inherited diseases[8]. Surprisingly, despite pre-sequencing estimates in the upwards of 140,000 genes[9], only 22,287 gene loci (~34,000 protein-coding transcripts) were identified[10]. Most shocking, however, is the fact that the identified protein coding regions only account for a mere ~1.2% of the human genome[10].

Why do humans have so few genes and so much DNA? Despite an initial proposal that the majority of the non-protein coding DNA was simply evolutionary remnant, so called 'junk-DNA' (a testament to the central dogma— if it does not code for protein, it's useless), with the last decade of research it has become increasingly clear that a diverse world of non-coding RNAs (ncRNAs) is richly encoded in the genome and abundantly expressed in the cell to enact a wide range of functions. High throughput efforts by the ENCODE project, tasked with interpreting the human genome sequence, have estimated that 80% of the genome is transcribed into ncRNAs[11], a phenomenon that has lead to ncRNA being called the cellular 'dark matter'[12] (in reference to physical 'dark matter', which accounts for the great majority of the matter in the universe).

So it turns out that Hooke was not too far off — the cell is a little black box, and it is filled with RNA.

1.2 microRNA: Discovery and Impact

Larval development of the *Caenorhabditis elegans* nematode worm consists of a series of temporal changes in gene expression that coordinately define the organism's body plan. In 1993, while studying early stages of larval development, Victor Ambros and coworkers discovered that, at the end of the first larval stage (L1), temporal decreases in the protein Lin-14 coincide with the appearance of a ~22 nucleotide (nt) non-coding RNA that matched the sequence of a different gene *lin-4*, which did not appear to code for protein[13]. The *lin-4* gene is essential for proper development; deletion causes the L1 stage to reiterate at later stages in development due to persisting levels of Lin-14 protein. The authors proposed that the lin-4 RNA, which is partially complementary to the 3'untranslated region (UTR) of the lin-14 transcript, was repressing Lin-14 protein expression through an antisense-like mechanism[13]. It was not until seven years later that a second short RNA, let-7, was found to be involved also in developmental changes in gene expression, this time repressing heterochronic genes *lin-14*, *lin-28*, *lin-41*, *lin-42* and *daf-12*— all with 3'UTRs containing complementary sites to let-7[14]. Soon thereafter, it was discovered that the let-7 'microRNA' is not just an idiosyncrasy of worms but, instead, it is highly conserved throughout metazoans from Drosophila to humans[15].

A boom in microRNA (miRNA) research has since ensued, with year over year of exponential growth in related publications topping 5,000 articles in 2011 (**Fig. 1.1**). Sustaining this scientific enthusiasm are the vast depths of miRNA function. To date, over 1,500 miRNAs have been identified, each believed capable of regulating several hundred genes and, taken together, >60% of all protein coding genes[16]. With this breadth, it comes as no surprise that miRNAs are involved in nearly all cellular processes from cell differentiation and growth to stress response and apoptosis[17]. Most importantly, it is evident that miRNAs are intricately linked to disease[18]. Due to their ability to alter large sets of genes, miRNAs have sparked new optimism as a means of intervention in complex multifactorial disorders such as cancer and cardiovascular disease[19].

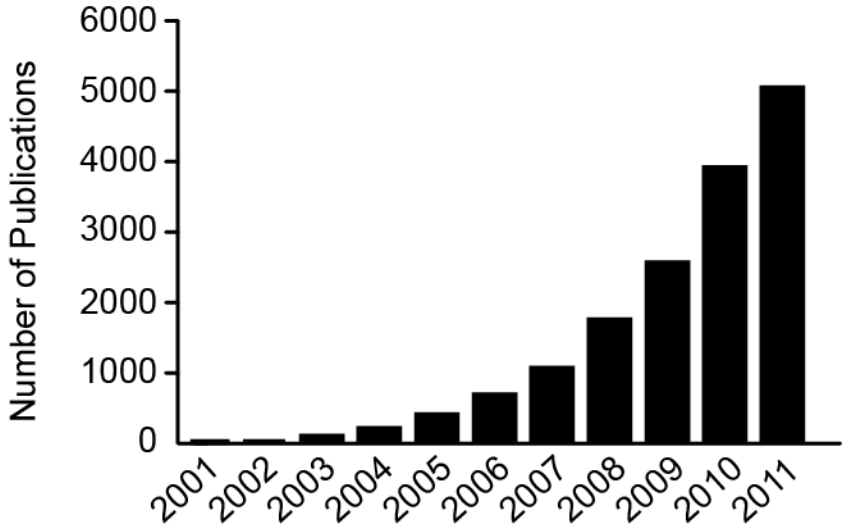


Figure 1.1 Number of miRNA-related publications by year. Based on PubMed keyword search "microRNA"

1.3 miRNA Biogenesis: From Birth to Death

miRNA expression is a highly regulated multiple-step process[20]. The main steps of miRNA biogenesis have been modestly characterized, but the timing of events and the exact factors involved are currently unsettled. In addition, there are exceptions to the rules and several instances of non-canonical miRNA biogenesis have been reported[21]. Further, while the miRNA pathway in general is highly conserved across metazoans, certain processing steps appear to diverge. The following description outlines the best-known pathway in mammals (**Fig. 1.2**).

miRNAs are encoded in both intergenic regions as independent transcriptional units as well as in the intronic regions of protein-coding genes[22]. Transcription and processing are remarkably similar in both cases. miRNAs encoded as discrete genes are transcribed by RNA polymerase II into pri-(mary)microRNA transcripts that are several kilobases in length and are capped and tailed similar to messenger RNAs[23]. Often several miRNAs can be clustered in a single pri-miRNA, such is the case for miR-23a/miR-27a/miR-24-2[24]. Pri-miRNAs are highly structured with the miRNA folded into the stem of an imperfect hairpin-loop that forms an arm in the longer transcript. This arm is then recognized by the protein factor Dgcr8 (DiGeorge Critical Region 8), which subsequently recruits the endonuclease III Drosha to cleave out the hairpin, yielding a 65-80 bp pre-(cursor)miRNA that has a 2 nt 3'overhang that serves as a handle for downstream recognition[25]. Dgcr8/Drosha also processes intronic miRNAs into pre-miRNAs. Interestingly, Drosha cleavage may actually precede the splicing step without inhibiting, and perhaps even promoting, mature message production[26,27]. It has also been reported that in *Drosophila melanogaster* and *C. elegans* certain spliced introns, termed miRtrons, can fold as pre-miRNAs and enter the miRNA pathway, bypassing Drosha/Dgcr8 cleavage[28].

Pre-miRNAs are exported from the nucleus by the exportin-5/Ran-GTP complex[29]. Once in the cytoplasm, they undergo a second cleavage step by another endonuclease III enzyme, Dicer, which generates the mature ~22 nt double stranded miRNA. Dicer contains two neighboring RNase III domains, a DExH RNA helicase/ATPase domain, a PAZ domain, and a double stranded RNA-binding domain.

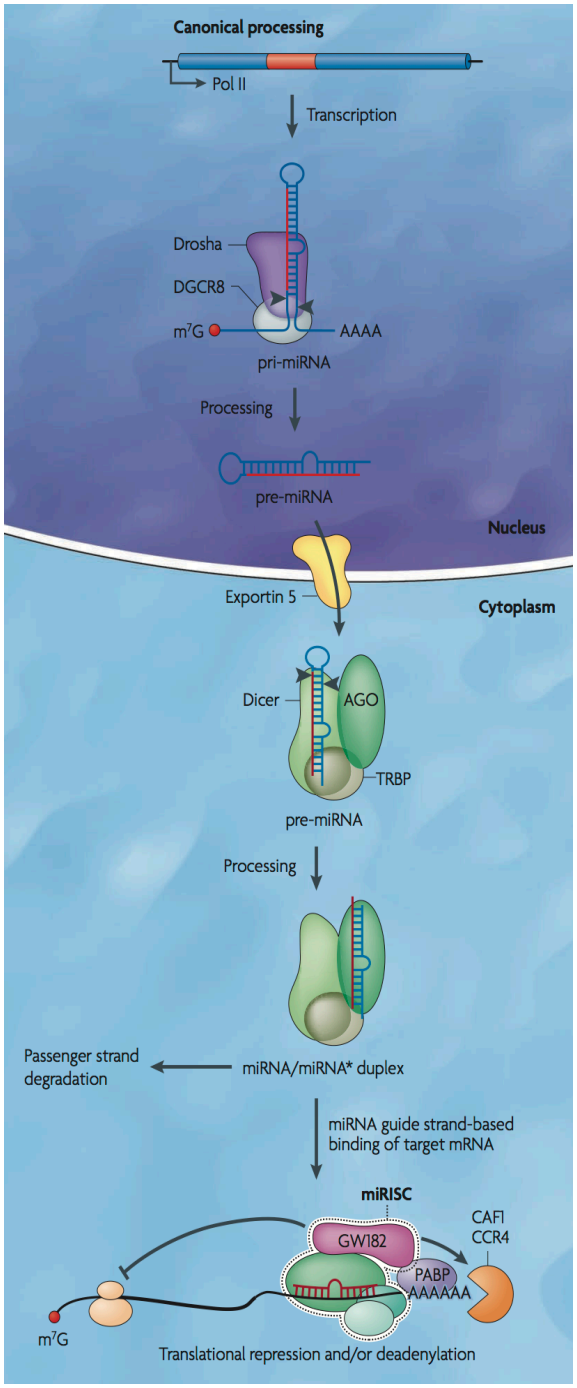


Figure 1.2 Schematic of the miRNA pathway. Reprinted with permission from Krol et al. [59].

Crystal structures of a protozoan Dicer homolog indicate that Dicer binds and spans the length of the pre-miRNA with the PAZ domain bound to the pre-miRNA 3'end overhang and the dual RNase III domains positioned at the other end of the stem just before the terminal loop[30]. The ~65 Ångstrom distance between the PAZ and RNase III domains serves as a molecular ruler, ensuring that Dicer cleaves the pre-miRNA to the proper mature length. When in position, the RNase III domains form an intramolecular dimer and cleave the pre-miRNA at staggered positions on both sides of the stem, thus excising the loop to produce the mature double-stranded ~22 nt miRNA with 3' overhangs on each end[30].

It is believed that Dicer cleavage is coupled to loading of the mature miRNA into the RNA induced silencing complex (RISC), the effector of miRNA activity. *In vitro* reconstitution experiments using purified proteins suggest that a minimum of three proteins— Dicer, Argonaute (Ago), and Trbp— are required for RISC loading[31]. Ago, as the central RISC factor, is the keystone of the pathway that is ultimately loaded with the mature miRNA. Trbp (human immunodeficiency virus trans-activating response RNA-binding protein) is required for efficient miRNA activity and is thought to recruit Ago to Dicer[32,33]. Single particle electron microscopy (EM) depicts these three factors as forming a RISC loading complex (RLC) where Dicer is in an L-shape conformation and Trbp and Ago are positioned, respectively, at the foot and along the body[34,35]. Modeling in the pre-miRNA suggests this architecture would place Ago in position for direct handoff of the cleaved mature product[34,35].

The two strands of the mature miRNA are denoted as guide and passenger strands. The guide strand goes on to guide RISC to mRNA targets, whereas the passenger strand is degraded and/or ejected. Guide strand selection is a critical process since it ultimately determines a miRNA's targets. In most cases, miRNAs are thermodynamically asymmetric and the strand with the weaker duplex at the 5' end is preferentially selected[36]. This asymmetry can be sensed by Dicer, which repositions the miRNA for directional handoff of the duplex to Ago[37]. The Ago family of proteins has a bilobal structure consisting of four conserved domains: one lobe entails an N-terminal and a PAZ domain, and the other a MID and a C-terminal PIWI domain[38]. Crystal structures of human Ago2 bound to guide-strand RNA reveal that the guide strand lies along a central

cleft between the two lobes with the 5'-phosphate positioned in a pocket of the MID domain and the 3' end anchored by the PAZ domain[39]. For miRNAs with extended duplex complementarity, unwinding is initiated by cleavage of the passenger strand by the PIWI domain, which has an RNase-H-like fold and contains the catalytic active site[40]. For most miRNAs, however, the N-terminal domain, which is positioned at the 3'end of the RNA below the PAZ domain, may serve to wedge the passenger strand from the guide strand[41].

After removal of the passenger strand, the activated miRISC is capable of binding targets. miRNAs bind targets based on complementary base pairing to the miRNA seed region, nucleotides 2-8 from the 5'end. Biochemical and structural data support a 'seed nucleation' model wherein the miRNA seed region, held in a pre-constrained A-form conformation to reduce the entropic penalty of binding targets, first binds and stabilizes the target interaction before the flexible 3'end forms additional supplementary base pairs[39,42]. Thermodynamic studies using isothermal titration calorimetry have shown that the PIWI/MID domains of an Ago archaea homolog enhance miRNA-target affinity by 300-fold mostly due to reduced entropy[43]. It is unclear which components are present in RISC at the point of target binding. Reconstitution experiments suggest Dicer and Trbp dissociate following RISC loading with miRNA[31]. Other factors deemed necessary for RISC to repress translation and/or destabilize targets include among others Tnrc6/Gw182[44-46], Rck/p54[47], Mov10[48], Rack1[49], Eif6[50], Dcp1/2[51], and Ccr4-Not[51].

Unlike the founding members let-7 and lin-4, which function as switches by turning a small number of developmental genes entirely on and off, the majority of miRNAs are believed to have a more subtle effect on gene expression, repressing protein output by 1.5 - 3 fold[17]. In this manner, miRNAs confer robustness to expression by fine-tuning protein levels, buffering against transcriptional fluctuations and stress[17,18,52]. Still, even these modest levels of repression can have a huge impact on the cell since a given miRNA can directly regulate several hundred genes[53]. Furthermore, since many miRNAs target transcription factors, perturbation of a solitary miRNA can affect thousands of downstream genes[17,54]. The significance of how this effect can shape expression patterns is exemplified by ectopic transfection of tissue specific miRNAs. For example, transfection of brain specific miR-124 into cervical cells alters the transcriptome in such

a way as to resemble that of brain cells[54]. Similarly, miR-302/367 is sufficient to induce pluripotency in transfected somatic cells without the need for other transcription factors[55].

Following target repression, miRNAs may localize to P bodies (see **Section 1.7**), where they could be recycled for multiple rounds of repression[56]. However, the number of rounds for multiple turnover may be self-limiting since target interaction may trigger non-templated addition of uridine to the 3'end of the miRNA, thereby marking the miRNA for degradation[56,57]. While uridylation has been associated with miRNA stability[57,58], miRNA degradation is the least characterized step in the pathway and a nuclease(s) responsible for miRNA decay has not been identified yet.

1.4 Mechanisms of miRNA-Mediated Repression

When it comes to the mechanism(s) of miRNA-mediated repression, a question mark may be the most unifying point: "How do miRNAs regulate gene expression?[60,61]"; "Are the answers in sight?[62]"; "How many mechanisms?[63]" — these questions, which exemplify the current state of confusion, have been asked in the titles of recent review articles summarizing the seemingly contradictory data that exist in support of various, distinct mechanisms (**Fig. 1.3**).

Translational repression versus transcript destabilization

One topic of debate is whether miRNAs predominantly reduce protein output by repressing translation or by inducing transcript destabilization. Despite numerous reports of repression occurring only at the protein level several very recent studies using array profiling and quantitative mass spectrometry indicate that *most* repression events are a result of decreases to both protein and transcript (see **Section 1.5** for more details). The emerging picture is that translational repression occurs first, followed independently by destabilization via active recruitment of deadenylase factors[64-67].

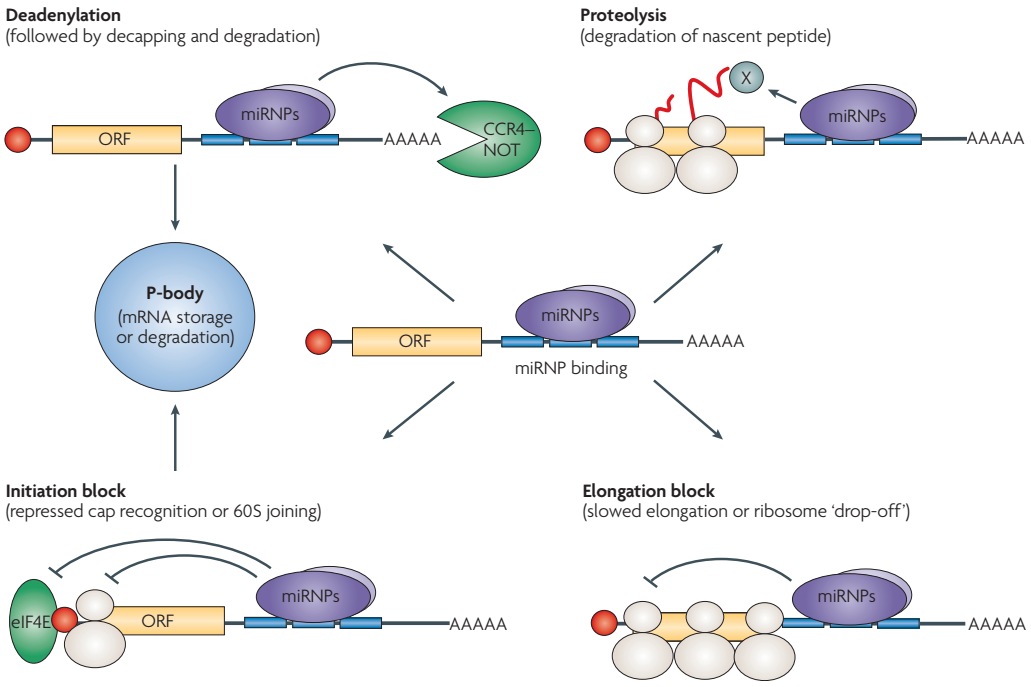


Figure 1.3 Proposed mechanisms of miRNA-mediated repression. Reprinted with permission from Filipowicz et al.[62].

Translational repression: initiation versus post-initiation

Two contrasting findings using sucrose gradients have led to proposals of multiple distinct mechanisms for how miRNAs repress translation. First, miRNAs can prevent target mRNAs from sedimenting into polysome-containing fractions[68,69]. This result is consistent with a mechanism where miRNAs block an early step in translation either at the initiation step or shortly thereafter. The details for how this might occur are unclear. Since miRNAs cannot repress reporters containing internal ribosome entry sites (IRES), RISC may interfere with initiation factor Eif4e recognizing the mRNA 5'-terminal m^7G cap, an essential step in canonical, but not IRES-mediated translation initiation[68,70,71]. The MID domain of Ago proteins contains an amino acid motif similar to that found in Eif4e, suggesting Ago may compete with Eif4e for cap binding[72]. However, structural data do not support Ago binding of cap, and pull-down assays indicate Ago only binds cap analogues non-specifically[73]. Alternatively, RISC may prevent 60S and 40S subunit joining. In cell-free extracts, targeted mRNAs are bound by 40S but not 60S subunits[74]. In agreement, RISC complexes have been identified in human cells containing 60S subunits and the negative initiation factor Eif6, which binds the 60S subunit and prevents formation of 80S assembly[50]. Eif6 is required for miRNA-mediated repression in human cells and C. elegans[50], but surprisingly not in Drosophila[75]. Therefore, multiple mechanisms may exist for translation repression at the initiation step.

The second finding using sucrose gradients, in apparent contradiction to the first, is that miRNAs are also often found in polysome fractions associated with translating target mRNAs[49,76-79]. This result is consistent with a mechanism where miRNAs block a late step in translation. Further experimentation has led to several conclusions on how this may occur, including slowing the rate of elongation[80], increased ribosome drop-off[76], and proteolysis of the nascent peptide chain[79]. Petersen et al.[76] found that ribosomes on repressed mRNAs dropped-off faster and reached the stop-codon less often than on non-target mRNAs, suggesting miRNAs reduce the production of full-length protein. In order for increased ribosome drop-off to occur without reducing the amount of steady-state mRNA in polysome fractions, initiation would have to load ribosomes faster than the drop-off rate. Nottrott et al.[79] did not observe increased drop-off rates or any other measurable effects on the translation of target mRNAs, despite reductions in target

reporter protein output. To resolve this paradox, the authors engineered reporters to express luciferase protein with eight N-terminal Myc tags that could be immunoprecipitated as soon as the nascent chain was translated. While control reporter mRNA could be immunoprecipitated by anti-Myc pull-down, reporters fused with a 3'UTR containing let-7 seed sites could not be pulled down, thus suggesting that the nascent protein was digested or bound by a proteolytic enzyme as it was being translated. Although there are no other reports of miRNA-induced nascent chain degradation, Guo et al.[81] similarly found that the ribosome densities on most target mRNAs do not change in the presence of miRNA. The authors conclude the reduced protein output is the result of destabilized mRNA, but note the possibility that the same results would be obtained if initiation and elongation rates were equally attenuated[81]. Therefore, multiple mechanisms may exist for translational repression post-initiation, and these could even work in combination with inhibition at earlier stages.

Is there a unifying mechanism to explain such contrasting results? On one hand, miRNAs prevent translation at an early step, preventing ribosomes from binding to target mRNA. On the other hand, miRNAs associate with translating mRNAs and more subtly alter translational rates. In both cases, the mechanistic details are unclear and multiple translational steps may be inhibited or other processes may have an effect. Different mechanisms may occur with specific species, cell types, time points, targets, or miRNAs.

1.5 Experimental Determination of miRNA Targets

Understanding miRNA function requires identification of target genes. As a starting point, several open source web-based tools are available to predict targets based on sequence complementarity to miRNA. Often, other parameters are also built into these tools to help improve performance. The more popular prediction methods include TargetScan[82], MiRanda[83], PicTar[84], and PITA[85]. The accuracy of these methods widely varies[86] and it is unclear in what biological contexts or with which miRNAs these tools work best.

The most common method to validate a miRNA-target interaction is a reporter assay. In this method, a luciferase reporter is fused to the 3'UTR of a gene of interest and transiently transfected in cultured cells endogenously expressing or transfected with the

miRNA in consideration. Reduced luciferase expression compared to a control reporter without the 3'UTR or, more specifically, with a mutant seed site suggests the 3'UTR is sensitive to miRNA repression. While a growing number of targets have been verified using luciferase reporter assays[87], these only account for a very small percentage of the total number of predicted targets.

Several groups have utilized higher throughput methods for identifying large groups of miRNA targets. By ectopically transfecting brain-specific miR-1 and miR-124 into HeLa cells, Lim et al.[54] used microarrays to measure downregulation in transcript abundance of 96 and 174 annotated genes, respectively. Importantly, for both miRNAs, the majority of downregulated genes contained 3'UTR seed sites, thus demonstrating for the first time that miRNA activity can be studied at the transcript level using microarray profiling. Grimson et al.[82] performed a similar analysis in HeLa cells with 11 different miRNAs and observed changes in transcript abundance of hundreds of genes, 75% of which contained 3'UTR seed matches. They also observed a correlation between gene repression and certain seed site features like A/U richness. These correlations were converted to a generalized quantitative model and incorporated into the TargetScan algorithm for target predictions[82].

How well does array profiling capture miRNA activity? Three studies (two from the Bartel group and one from the Rajewsky group) have now combined array profiling with proteomic analysis to assess how well mRNA and protein levels correlate following miRNA transfection. Using SILAC (stable isotope labeling with amino acids in cell culture) quantitative mass-spectrometry, the Bartel group first[53] found that targets whose protein levels were repressed by more than a third also showed detectable decreases in mRNA levels 48 h after transfection of miRNA into HeLa cells[53]. Subsequent analysis using ribosome profiling to assess protein output at 12 h post-transfection suggested that not only does mRNA decay reflect protein reduction, but it may also be the more significant effect of miRNA-mediated repression[81]. The Rajewsky group[88] also found a strong correlation between mRNA and protein levels using a similar mass-spectrometry approach, pulsed-SILAC (pSILAC), 12 h after transfection of HeLa cells. However, analysis at 8 h showed a much weaker correlation with more targets showing reduced protein than mRNA, consistent with conclusions of other early time point kinetic

analyses that protein inhibition precedes mRNA destabilization[64,65,89]. These studies therefore validate array profiling as a tool for studying the majority of cell-wide miRNA targeting events with the caveat that profiling should be performed ≥12 h post perturbation. Due to its convenience and cost-effectiveness, array profiling has become a common tool in the miRNA field.

As an alternative to transfection with miRNAs, miRNAs can also be inhibited to identify depressed targets. The disadvantage of transfection, on the one hand, is the requirement of ectopic expression or higher-than-physiological levels to avoid competition with endogenous miRNA; the advantage of inhibition, on the other hand, is that it can determine endogenous targets. Several different strategies have been implemented to inhibit miRNAs including morpholinos that block pre-miRNA processing[90], 'target protectors' that bind targets and prevent miRNA access [91], and plasmid expressed 'sponges'[92] and 'tough acting decoys'[93,94]. However, the most commonly used inhibitors are anti-miRs. Anti-miRs are antisense oligonucleotides (also known as ASOs) complementary to mature miRNA guide strands. They have been used both as potential therapeutics and as tools for studying miRNAs. Several different anti-miR chemistries have been developed to increase affinity, stability, and delivery. Common sugar modifications include 2'-O-methyl (2'OMe), 2'-O-methoxyethyl (2'MOE), 2'fluoro (2'F), and locked nucleic acid (LNA)[95]. LNA is a bicyclic nucleic acid analogue 'locked' into a C3-endo conformation by a bridging carbon linking the 2'O and 4'C atoms. These modifications protect from nuclease degradation and enhance affinity and increase the melting temperature (Tm) of the anti-miR:miRNA duplex. In general, LNA provides the highest affinity followed by 2'F > 2'MOE > 2'OMe > DNA[95-97]. However, often these analogues are used in combination at different base positions to fine-tune affinity and/or decrease manufacturing cost[97]. Several studies have shown that anti-miRs with higher affinity are often more potent[95,97-99], although there are exceptions to this trend[96,100]. The anti-miR backbone can also be modified. Phosphothioester (PS) interbase linkages, where a non-bridging oxygen is replaced with a sulfer atom, confer nuclease resistance and improve tissue delivery[97,101], but at the cost of minor reductions in affinity[97].

Anti-miRs have been used as a tool for understanding miRNA function and

targeting in various cultured mammalian cell lines. For example, miR-21's targets, including Pdcd4 and Btg2, have been identified using array profiling and proteomics in HeLa and MCF-7 cells[102-104]. MCF-7 cells were also used with LNA-based anti-miRs to investigate the targets of miR-101[105]. However, it is unclear how many of the targets identified in cell culture are regulated *in vivo*.

In 2005, Krutzfeldt et al. demonstrated for the first time that miRNAs could be inhibited *in vivo* with intravenous administration of anti-miR to mice[106]. In this landmark study, liver miR-122 was inhibited using cholesterol conjugated 2'OMe-based anti-miRs, called 'antagomiRs'. miR-122 inhibition resulted in the upregulation of numerous genes with miR-122 seed sites, but it also caused several genes to be downregulated, likely as a result of secondary effects. Since these downregulated genes were functionally enriched in cholesterol biosynthesis pathways, anti-122 treatment caused reduced levels in plasma cholesterol. Moreover, other miRNAs including miR-16, miR-192, and miR-194 could also be inhibited in liver, lung, kidney, heart, intestine, fat, skin, bone marrow, muscle, ovaries, and adrenals (their targets, however, were not identified). These results, therefore, suggest that anti-122 and other anti-miRs could be of therapeutic value for the treatment of metabolic diseases and other ailments[106]. Other groups have since expanded on the link between miR-122 and cholesterol synthesis, even showing its therapeutic potential in non-human primates[99,107,108]. While anti-miRs have become a powerful tool for studying the roles of miRNAs in disease using various mouse models[109], surprisingly few inhibition studies outside of those for miR-122 discussed here have focused on the physiological roles of miRNAs in healthy tissue. Various miRNA knockout mouse models have been used for physiological target determination[110]; however, these knockouts are rare and have been studied thus far with only limited focus, mostly in the areas of cardiovascular and immunological function[111]. Moreover, there is some evidence that miRNA deletion, as opposed to short term inhibition, can be masked by compensatory mechanisms[18,112]. Therefore, *in vivo* activity and targets of most miRNAs under normal conditions are still largely unknown. This represents a major gap in knowledge; particularly considering that miRNA activity is highly dysregulated in cultured, diseased cells. Taken together, a critical current question in the field is-- Do the emerging paradigms and validated targets established in transformed cell lines apply to

miRNA activity *in vivo*?

1.6 miRNAs in Disease

miRNAs are strongly implicated in disease. Aberrant miRNA activity due to inappropriate expression has been linked to metabolic disorders such as diabetes and obesity, neurological degenerative diseases including Alzheimer's and Parkinson's, cardiovascular disease, fibrosis, and auto-inflammatory diseases[113]. Most if not all cancers show altered miRNA levels[114]. Even otherwise normal miRNA function can contribute to disease: hepatitis C virus (HCV) hijacks miR-122 to enhance replication in liver cells[115,116]. Mutations in the seed region of a miRNA, although less common, can also cause Mendelian inherited diseases such as hearing impairment[117].

Pharmacological modulation of miRNA abundance may be a powerful new strategy for treating human disease. Several biotechnology companies are pursuing miRNA-based therapeutics with a current pipeline that includes drugs for treating HCV, various cancers, and hypercholesterolemia. The drug furthest in development, an anti-miR targeting miR-122 for treatment of HCV, is currently in Phase II clinical trials. Early safety data suggests miRNA inhibition is associated with low cellular toxicity[118]. In contrast to inhibition, miRNA replacement strategies are also being developed for cases where protective miRNAs are reduced in the diseased state. Delivery of tumor suppressors let-7 and miR-34a mimics *in vivo* to Kras activated mouse models of non-small-cell lung cancer (NSCLC) caused a 60% reduction in tumor area[119,120]. Such miRNA mimics are currently only in preclinical phases of development.

miRNAs are also proving to be useful biomarkers for cancer diagnostics[114]. Profiling of miRNA expression in leukemias and solid tumors shows distinct differences in miRNA expression compared to normal tissues, with some miRNAs being upregulated and others being downregulated[121,122]. In some cases, miRNA expression signatures can be so unique and distinct that they can be used to classify cancers. Using a bead-based flow cytometric method, Lu et al.[121] were able to classify based on miRNA expression profiles the developmental lineages and differentiation states of 334 tumor samples. This level of information may one day substantially improve the diagnosis and prognosis of cancer and help in determining proper courses of treatment.

miRNA-21: a notorious oncogenic miRNA

miR-21 may be an especially effective drug target and biomarker for several diseases including cancer. In a large profiling study, miR-21 was the only miRNA found to be upregulated in all clinical samples tested including lung, breast, stomach, prostate, colon, and pancreatic tumors[122]. Increased expression of miR-21 compared to healthy tissue has also been detected in glioblastomas[123], cholangiocarcinomas[124], esophageal squamous cell cancers[125], liver cancers[126] and cervical cancers[127]. Additionally, miR-21 expression levels are a valuable prognostic marker. High levels of miR-21 have been correlated with poor patient survival and/or therapeutic outcome for several cancers including colon[128], lung[129], and prostate cancers[130], and higher levels of miR-21 are associated with aggressive forms of breast cancer[131,132].

Functional studies indicate that miR-21 has oncogenic potential in cancer cell lines. Inhibition of miR-21 caused a decrease in proliferation and invasion in MDA-MB-231 breast cancer cells[133]. Similar results were observed in Huh7 and HepG2 hepatocarcinoma cells[126]. Additionally, inhibition of miR-21 reduced invasion and metastasis of colon cancer cells [134] and suppressed growth of HeLa cells[127]. In several cell lines including glioblastoma, anti-miR-21 transfection caused cell death[135]. In contrast to inhibition, transfection of miR-21 promoted proliferation and invasion in MCF-7 and pancreatic cancer cells[130]. In the latter, increased levels of miR-21 were also found to promote chemoresistance to the cancer drug gemcitabine[130]. Several tumor suppressor genes were identified as miR-21 targets in these and similar studies including Pdcd4, Timp3, Reck, Spry1/2, Brg2, and Cdc25a[136]. Taken together, these data support that increased levels of miR-21 can promote oncogenic phenotypes, while targeted inhibition of miR-21 can reverse them.

Multiple pathways may induce miR-21 overexpression. The miR-21 gene is located on chromosome 17 in the intron of a protein-coding gene, Tmem49, with unknown function[137]. Despite this overlap, miR-21 transcription is independent and driven by its own promoter following stimulation with phorbol-12-myristate-13-acetate (PMA)[137]. Several upstream promoter elements have been computationally inferred and the transcription factor AP-1 has been shown to activate transcription of miR-21 loci[137]. Interestingly, AP-1 is repressed by Pdcd4, a target of miR-21's, indicating the existence

of a feedback loop[138]. A double negative feedback loop also exists between miR-21 and another one of its targets: Nfib, a transcriptional repressor that blocks the promoter of miR-21[137]. Both of these feedback loops could amplify changes in miR-21 expression. Indeed, miR-21 expression is stimulated by the Pdcd4/AP-1 loop in cardiac myocytes in response to oxidative stress[139]. miR-21 transcription can also be attenuated by estradiol via estrogen receptor alpha, which may play a role in miR-21's involvement in breast cancers[140].

Mature miR-21 levels can also be augmented post-transcriptionally by enhanced processing of primary miR-21 transcripts. Davis et al.[141] reported that treatment of vascular smooth muscle cells (VSMC) with bone morphogenic proteins (BMPs) and transforming growth factor beta (Tgfb) induces recruitment of Smad and RNA helicase p68 to pri-miR-21; this in turn enhances Drosha cleavage and the production of pre-miR-21[141]. The additional miR-21 levels function to transition the smooth muscle cells to a contractile phenotype by downregulating target Pdcd4[141]. Although BMPs have a stimulatory effect in vascular cells, BMP-6 inhibits miR-21 transcription in MDA-MB-231 breast cancer cells[142], suggesting miR-21 production is regulated by tissue specific controls.

Most importantly, Ras, the most frequently activated oncogene in human cancers, can induce overexpression of miR-21. Frezzetti et al. found that activated Ras induced miR-21 expression in mouse fibroblast NIH-3T3 and rat thyroid FRTL-5 cells via at least three pathways: AP-1, Raf/Mapk (mitogen activated protein kinase) and phosphatidylinositol 3-kinase (PI3K)[143].

Induction of miR-21 overexpression by Ras suggests that miR-21 contributes to cancer by supporting and reinforcing oncogenic programs initiated by other factors rather than initiating tumorigenesis itself. In agreement, using a miR-21 inducible transgenic mouse, Hatley et al. reported that miR-21 overexpression only promoted lung tumor formation when co-expressed with activated Ras[138]. Importantly, miR-21 deletion suppressed the oncogenic effects of Ras indicating that miR-21 is a primary downstream effector in the pathogenesis of Ras-induced cancers. However, in contrast to a supporting role, Medina et al.[144] found that overexpression of miR-21 alone is sufficient to cause pre-B-cell lymphoma in transgenic mice. The authors showed that tumor formation

became dependent on miR-21 overexpression, a phenomenon termed oncogene addiction; by turning off miR-21 expression from an inducible promoter, tumor growth receded due to increased apoptosis. Therefore, miR-21's contribution to pathogenesis may be dependent on cellular environment and other unknown factors. Further research is required to define the precise roles of miR-21 in cancer.

1.7 Probing miRNA Pathways with Fluorescence Microscopy

Fluorescence microscopy has been indispensable for studying the subcellular localization of miRNA and its pathway components. Both miRNA and miRISC have been found to colocalize in various granular bodies including Processing bodies.

Processing Bodies

Processing bodies or P bodies (PBs) are cytoplasmic sites of RNA degradation and storage that contain over sixty different proteins including decapping factors, nucleases and helicases[145]. Their essence is rather elusive, being defined simply as cellular foci that are microscopically visible when a PB component, or 'marker', is detected with immunofluorescence or expressed as a fluorescent fusion. They were first identified in 1997 when Bashkirov et al. found the 5'-3' exonuclease Xrn1 to localize in cytoplasmic foci[146]. Later, the mRNA decapping factors Dcp1/2 and Lsm1-7 were found to colocalize in Xrn1 foci, suggesting that cap removal and exonucleolytic decay, the last two steps in mRNA degradation preceded by deadenylation, is compartmentalized in these assembles[147,148]. Sheth and Parker reported that in *Saccharomyces cerevisiae* inhibiting mRNA turnover either before or after the decapping step altered the abundance and size of Xrn1 foci[149]. This combined with evidence that RNA decay intermediates also localized in the foci, demonstrated that mRNA decay occurs within these granules in equilibrium with the cellular RNA pool translating in polysomes[149]. Similar results were found in human cells[150], indicating that the roles of PBs in the mRNA decay process are conserved among eukaryotes.

Evidence has emerged indicating that PBs play a role in miRNA-mediated repression. First, both miRNAs and Ago2 have been found to localize to PBs[68,151-153].

Second, reporter mRNAs containing miRNA seed sites localize to PBs in a miRNA-dependent manner[68,152]. Third, several PB components, including Tnrc6/Gw182 and Rck/p54, are required for both miRNA-mediated repression and PB formation[152,154]. Knockdown of these genes depletes PBs and relieves repression of target reporters. Additionally, Ago mutants incapable of being recruited to PBs by Tnrc6/Gw182 cannot effect repression, even when directly tethered to target mRNA[155]. Fourth, PBs do not contain core components of the translational machinery, suggesting PBs could sequester repressed mRNAs away from translating polysomes[156]. Importantly, however, visible PBs can be disrupted by knockdown of Lsm1 without affecting target repression, thus suggesting miRISC and targets localize to PBs not to cause repression but as a consequence of it [47,157].

miRNA-mediated silencing can persist in the absence of visible PBs, but this independence is not reciprocal: PB formation requires active miRNA pathways to deposit mRNA. Disruption of miRNA biogenesis by depletion of Drosha/Dgcr8 in human cells or either Drosha, Dicer, or Ago in Drosophila cells results in loss of PB formation[157,158]. Interestingly, block of translation, and thus removal of mRNAs from polysomes, by puromycin treatment could not rescue PB formation in Drosophila cells depleted of essential miRNA factors[157]. This indicates that translationally inactive mRNA has to proceed through miRNA-related channels in order to enter PBs. Therefore it is likely that PBs are heavily populated with repressed miRNA targets awaiting degradation, thus making PB localization a key late step in the miRNA repression pathway.

PBs could also serve as storage facilities for temporally repressed mRNA. In yeast, mRNA localizes to PBs during stress or arrest in growth, and then reenters polysomes for translation after recovery[159]. Bhattacharyya et al.[69] have observed a similar phenomenon involving the cationic amino acid transporter (Cat-1) in liver cells. In the fed state, Cat-1 is repressed by miR-122 and localized to PBs; however, in response to amino acid starvation, the RNA binding protein (RBP) HuR reverses miR-122 repression and Cat-1 is transported from PBs to cytoplasmic polysomes[160]. Stored mRNA may be physically separated from the RNA degradation enzymes. Electron microscopy has revealed PBs are comprised of multiple compartments[161], and let-7 has previously been found to localize as foci closely adjacent to but not overlapping with Dcp1a foci[68].

Further research will be needed to determine what factors are involved in determining the fates of mRNAs transported to PBs. These mechanisms may help explain why certain mRNA-target interactions may result primarily in translational reductions. Single particle tracking experiments may be especially useful for characterizing complexes transported to PBs.

Single Particle Tracking

Single particle tracking (SPT) is a biophysical microscopy method for analyzing a molecule's change in position over time, i.e., its spatiotemporal organization in the cell. The method can be used in live cells to measure the dynamics of fluorescently labeled cellular components. From these data, properties of the tracked component can be inferred including modes of transport, relative molecular weights, and interactions with cellular substructures such as cytoskeletal filaments. Because each particle is tracked individually, the complete distribution of particle behaviors can be measured, thereby enabling subpopulations, if present, to be uncovered from the ensemble.

At the outset of this thesis work, SPT had not yet been used to study miRNAs. However, it had been used to probe the dynamics of mRNA ribonuclear protein complexes (mRNPs) and PBs, which represent two key steps in the miRNA repression pathway. Fusco et al.[162] were the first to use SPT to analyze the transport of MS2-labeled mRNPs in the cytoplasm of living cells. The authors reported that mRNPs stochastically transitioned between several different types of motion, with the majority of particles undergoing corralled motion indicating hindrance by the cytoskeleton, organelles or possibly even PBs. Consistently, disruption of microtubules caused some, but not all, of these particles to be released to more diffusive motions. Most importantly, mRNPs with a beta-actin 'zipcode' sequence, which cause mRNPs to localize to the cell periphery, displayed increased directional movement, thus suggesting that long-range transport, but not necessarily short-range, is facilitated by motor-directed, active processes[162]. Comparable experiments have also been used to investigate nuclear transport[163] and export [164]. Similar to mRNPs, Aizer et al.[165] found that PBs (marked by Dcp proteins) displayed a variety of movements; however, PB diffusion was slower than most mRNPs. Further, PBs had extensive interaction with microtubules — PBs were observed swaying

back-and-forth and, less often, processing along on the fibers. By indirectly comparing PB and mRNP dynamics, the authors speculated that mRNPs reach PBs largely by chance via diffusive motions, but the microtubule network may serve as a highway for trafficking certain complexes[165]. Two-color tracking experiments, simultaneously imaging mRNPs and PBs as they come together, are necessary to test this hypothesis and investigate how miRNAs modulate these events.

1.8 Thesis Objectives

Primary Objective — Chapters 2 & 3

It is evident that miRNA activity is highly dysregulated in cancer, yet our basic understanding of mammalian miRNA targeting and function has been derived mostly from cultured cancer cells. The primary objective of this thesis is to investigate *in vivo* the repressive activities of miRNAs under healthy cellular conditions and contrast these with activities measured or previously reported for cultured cells. I focus on the oncogenic miR-21 with the hope of gaining insights into its contributions to disease; I seek to better understand the well-documented 'abnormal' activity by better defining exactly what 'normal' activity is. In **Chapter 2**, I use anti-miR inhibition and array profiling data to measure the repression signature of miR-21 and two other similarly abundant miRNAs in healthy mouse liver. Additionally, I use sucrose gradients to investigate biochemically how these miRNAs are interacting with targets. These results are contrasted with similar measurements in HeLa cells. In **Chapter 3**, I use bioinformatics to further dissect the array profiling data from **Chapter 2** in order to better understand the underlying principles behind miR-21's target selection.

Secondary Objective — Chapter 4

Another outstanding question in the field is the mechanism(s) by which miRNAs repress targets. The seemingly contradictory mechanistic data available suggest that miRNAs may utilize different modes of repression. The secondary objective of this thesis is to develop an incisive tool for visualizing single miRNAs as they proceed through the

repression pathway(s). This tool could be used to assess heterogeneities among miRNAs in regard to their spatiotemporal assembly into repressive complexes and interactions with target mRNAs, which would be expected if certain miRNAs enact distinct mechanisms of repression. In **Chapter 4**, I discuss the development of such a method that combines intracellular single particle tracking with step-wise photobleaching.

Chapter 2:

Disease-linked microRNA-21 exhibits drastically reduced mRNA binding and silencing activity in healthy mouse liver[1]

2.1 Introduction

MicroRNAs (miRNAs) are an expansive class of evolutionarily conserved post-transcriptional gene regulators effecting RNA silencing. At ~21 nucleotides (nt), these endogenous non-coding RNAs are partially complementary to the 3' untranslated regions (3'UTRs) of mRNAs; and thus control and typically silence gene expression by recruiting proteins of the RNA induced silencing complex (RISC) to their mRNA targets[61,166]. A growing body of evidence has demonstrated that miRNAs are involved in nearly all processes of the cellular life cycle, from growth and differentiation to maintenance and cell death[167-169]. This broad capacity is achieved by the actions of nearly 1,500 miRNAs that are bioinformatically predicted to regulate expression of >60% of the human proteome based on base pairing rules involving 7- to 8-nt short 'seed sequences'[16]. Perturbing the levels of a single miRNA has been shown to affect the mRNA and protein levels of hundreds of downstream genes[53,88]. Although the precise mechanism of miRNA-mediated regulation remains debated[61], global evaluation of expression changes following transfection with or deletion of miRNAs has revealed that, at least at later time points, the majority of changes observed at the protein level are mirrored at the transcript level, indicating that mRNA destabilization is an important component of miRNA function[53,81,88,170]. Apparently, miRNA-mediated degradation can occur while the mRNA target is still actively translated within polyribosome (polysome) complexes[49,77-79,171]

[1] Androsavich, J. R., Chau, B. N., Bhat, B., Linsley, P. S., & Walter, N. G. (2012). Disease-linked microRNA-21 exhibits drastically reduced mRNA binding and silencing activity in healthy mouse liver. *RNA*. Nelson Chau performed the array profiling and Sylamer analyses. B. Bhat synthesized the anti-miRs. John Androsavich performed all other experiments and bioinformatic analyses.

Recently, the role of miRNAs in disease has become a strong focal point of research. To date, more than 4,100 miRNA-disease associations have been reported[172]. While the interactions between a miRNA and its target mRNAs are believed to mostly result in mild 'fine-tuning' of expression output, imbalances in this interplay can have deleterious effects on the cell. In some cases, dysregulation of a single miRNA-target interaction induces disease. For example, chromosomal truncations in the 3'UTR of the mRNA encoding the human proto-oncogene high mobility group A2 (Hmga2), a transcriptional modulator expressed at high levels during embryogenesis but silenced in adulthood by let-7 miRNA, can lead to cancerous transformation due to loss of let-7 miRNA-mediated repression[173]. Alternatively, disease can arise from aberrant miRNA expression, which can potentially have much broader impacts on the transcriptome. For example, miR-21 is found to be overexpressed in several types of human tumors[122] as well as heart fibroblasts following cardiac stress[174]. While this correlation is suggestive of the pathological effects of miR-21, the miRNA's precise role in disease progression remains unresolved. Overexpression of miR-21 alone in the hematopoietic system can induce tumorigenesis in transgenic mice[144]; however, in more solid tissues such as brain[144] and lung[138], activation of another oncogene, such as Kras[138], appears necessary for miR-21 to contribute to and reinforce cancerous phenotypes.

Our knowledge of miRNA function and of the mechanistic links between miRNA and disease remains incomplete. In large part, this is due to poor understanding of the multitude of targets regulated by different miRNAs. While considerable effort has been expended towards global prediction of miRNA targets[17], experimental support for these predictions is limited. Indeed, most validated targets have been confirmed using gene reporter assays in cultured cancer cell lines. Little is known about miRNAs and their targets under healthy cellular conditions.

We here have investigated the function of three highly abundant miRNAs in healthy mouse liver. We show that, with the exception of a restricted set of genes enriched in stress-response pathways, inhibition of miR-21 with a specific 'anti-miR' oligonucleotide drug or abrogation by genetic deletion has surprisingly little effect on mRNA levels of predicted and previously validated targets, in contrast to our observations for let-7 and miR-122. Consistent with a diminished repression capacity,

miR-21 displays an unusual sub-cellular distribution with reduced binding to actively translating mRNAs. In stark contrast with healthy tissue, miR-21 displays a significantly augmented repression signature and strengthened binding to mRNA targets in the cancerous HeLa cell line, as previously suggested. Our analysis gives a biochemical perspective on the disparities of miRNA-mediated repression between healthy and diseased states and suggests that miR-21 passes a threshold in its transition from a very specific regulator under normal homeostasis to an overactive, broadly repressive oncomiR in cancer.

2.2 Materials and Methods

Animal care and treatments. All animal experiments were conducted according to the Institutional AAALAC Guidelines. Male C57BL/6 mice were housed four to five animals per cage with a 12 h light/dark cycle. Oligonucleotides were dissolved in saline and administered to mice by intraperitoneal (i.p.) injection. For mRNA profiling studies, mice were treated with 20 mg/kg twice weekly for 2 weeks. miR-21 knockout mice were obtained from Dr. Stuart Orkin (Harvard University,[175]).

Cell culture and transfection. HeLa cells were cultured in DMEM with 10% FBS. Anti-miR-21 was transfected with Lipofectamine RNAiMax (Life Technology) according to manufacturer instructions. For sucrose gradients, HeLa cells were treated with 0.1 mg/ml cycloheximide 20 min before harvesting. Lysates were prepared by swelling cells on ice in hypotonic buffer (10 mM NaCl, 1.5 mM magnesium acetate, 50 ug/ml dextran sulfate, 10 mM Tris-HCl, pH 7.4) before Dounce homogenization. The homogenized lysate was collected, centrifuged at 1,200 x g for 5 min at 4 °C, and loaded on sucrose gradients (see Liver Lysates and Fractionation).

Comparison with Computationally Predicted Targets. Predicted targets for each miRNA were downloaded from the following databases: miRanda-mirSVR - mouse, good mirSVR score, conserved miRNA, updated August 2010; TargetScan(Conservation) - mouse, conserved targets Pct scores, v5.2; TargetScan(Context) - mouse, conserved targets context scores, v5.2; TargetScan(Context+) - mouse, conserved targets context+ scores, v6.0; microCosm -

mouse, v3.0; and PITA - mouse, top predictions, 3/15 flank, v6.0. Starting with the highest score, each set of top predictions were cross-referenced with genes measured by array profiling until the first 30 matches were found. All matches came from at most the top 50 predictions from any prediction dataset.

Gene Ontology Classifications and Bioinformatic Calculations. For gene ontology assignments, seed-matched or non-matched gene sets were analyzed with the Panther gene expression data analysis tool[176,177]. Enrichment statistics were calculated relative to a custom-loaded background dataset consisting of all genes measured in our array profiling. For thermodynamic calculations, seed-matched genes for each miRNA were cross-referenced with hybridization energy calculations derived from the Vienna RNA package[178] available in the PITA database[85] (Fig. 2B, top left, and Supplemental Fig. S2A) or calculated with RNAhybrid[179] (Supplemental Fig. S2B). For calculation of target site numbers, all RefSeq (NCBI) transcript sequences associated with genes measured by array profiling that mapped to a unique Ensembl ID were downloaded and searched for 3'UTR seed-matched sequences using MatLab software (Math Works) with code written in-house.

Liver Lysates and Fractionation. Prior to harvest, blood was flushed from hepatic tissue by perfusion with 1x PBS via the portal vein. Lysates were prepared in lysis buffer (20 mM Tris, pH 7.4, 100 mM NaCl, and 2.5 mM $MgCl_2$) supplemented with EDTA-free protease inhibitor cocktail tablets (Roche). Intact livers (~1.0-1.5 g) were homogenized with a glass Dounce in 3 ml ice-cold lysis buffer by twenty strokes with each of the loose and tight plungers. The homogenate was centrifuged at 1,000 x g for 10 min at 4 °C. The resulting S1 supernatant was then centrifuged twice at 16,000 x g for 10 min at 4 °C. The final S16 supernatant was immediately layered on top of a 20-60% linear sucrose gradient prepared by gently overlaying decreasing percentages (w/v) of sucrose solutions prepared in lysis buffer. Equal OD_{260} units were loaded for each sample. Sucrose gradients were ultracentrifuged at 35,000 r.p.m., ~ 218,000 × g, for 3 h at 4 °C in a Beckman SW41 rotor. After centrifugation, 0.5 ml aliquots were collected by hand by puncturing the bottom of the ultracentrifuge tube. The A_{260} of each fraction was measured with a spectrophotometer. For samples prepared with puromycin, 1 mM puromycin and

10 mM vanadyl ribonucleoside complex (NEB) were added to the homogenized lysate, the salt concentration was adjusted to 300 mM NaCl, and the lysate was incubated at 37 °C for 10 min. For sample prepared with EDTA, 0.5 M EDTA, pH 8.0, was added to a final concentration of 30 mM.

miRNA pull-down assay. miRNA pull-down assays were performed as previously described[180]. Briefly, 2'O-methyl 'capture' oligonucleotides (IDT) modified with a 5'biotin were titrated into S16 liver lysates and incubated at 25 C for 1 h. Lysates were then incubated with M280 streptavidin-coated magnetic beads (Invitrogen) to precipitate the biotin-conjugated capture strand with its complementary miRNA. Supernatants were removed and total RNA was purified using Trizol reagent (Invitrogen) and isopropanol precipitation. miRNA was detected using northern blot with ^{32}P-end-labeled complementary DNA probes and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) cross-linking, as previously described[181].

mRNA array profiling and analysis. Total RNA from tissues or cell culture was extracted as per the manufacturer's instructions (miRNeasy kit, Qiagen). mRNA expression profiles were measured for mouse liver and HeLa cells using Mouse Genome 430 2.0 arrays (Affymetrix) and U133A arrays (Affymetrix), respectively. mRNA microarrays were run in triplicate for anti-miR or saline treated and transgenic mice. For analysis by cumulative distribution frequency, genes containing one or more particular miRNA seed-matched sequences in their RefSeq curated 3'UTRs were compared to those without seeds, and the p-value significance of the difference between these curves was determined by one-sided Kolmogorov-Smirnov test. Web-based software was used to analyze the same datasets using the Sylamer algorithm[182]. Gene expression changes were compared by student t-test and ranked from highest to lowest fold-change across the x-axis. Hypergeometric enrichment scores were computed for all 4^7 (16,384) possible heptamer sequences or for all 4^8 (65,536) possible octamer sequences and were plotted on the y-axis.

Oligonucleotide synthesis and purification. Standard procedure was used to synthesize and purify anti-miRs for miR-21, miR-122, and let-7 used in current investigation. Briefly, 2'-ribofluoro (2'-F) and 2'methoxyethyl (MOE) modified oligonucleotides were

synthesized on AKTA Oligopilot 100 (GE Healthcare) synthesizer. The 2'MOE amidites were acquired from Isis pharmaceuticals (Carlsbad CA). The F-modified nucleoside phosphoramidites were purchased from ST Pharm, former Samchully Pharma (South Korea) and Innovassynth Technologies (Mumbai, India). The key steps in the solid phase synthesis of F/MOE anti-miRs 21 and 122 were (1) 3.0 min detritylation step with 15% DCA/toluene. (2) A 0.15 M solution of the phosphoramidites in anhydrous acetonitrile in each coupling step. (3) For sulfurization, 0.2 M Phenyl acetyl disulfide (PADS) in 1:1 anhydrous pyridine / CH3CN was used with 6 min contact time. After completion of the synthesis, solid support was treated with triethyl amine (TEA): acetonitrile (1:1) at room temperature for 30 min. It was followed treatment with a mixture of aqueous ammonium hydroxide (33 wt. %): ethanol (200 proof, Koptec, 3:1) and heated at 55 °C for 9 h which resulted in complete removal of all the protecting groups. The anti-miRs were purified by ion exchange chromatography on an AKTA Explorer (GE healthcare) HPLC system on a strong anion exchange column (source 30 Q, GE Healthcare). The fractions with high purity (by LCMS) of the desired full-length anti-miR were pooled together and concentrated under high vacuum to a smaller volume. The oligonucleotides were desalted by reverse phase HPLC to furnish desired F/MOE modified anti-miRs in about 40% isolated yield based on the loading of the solid support. The anti-miRs were lyophilized to a dry powder. The final purity of the anti-miRs was assessed by ion-pair-HPLC–MS analysis with Agilent 1200 HPLC 6130 MSD system from Agilent Technologies.

RNA Analysis. Following fractionation, total RNA from individual fractions was purified using Trizol (Invitrogen) extraction and miRNeasy-96 kits (Qiagen). RNA was quantified with real-time quantitative PCR (RT-qPCR), using High Capacity cDNA Reverse Transcription reagents and TaqMan MicroRNA or mRNA assays (Applied Biosystems). RT-qPCR reactions were performed in triplicate in 384-well format using an Applied Biosystems 7900HT fast real-time PCR instrument. For absolute quantification, standard curves for each amplified miRNA were run in parallel using serial dilutions of synthetic miRNA oligos (IDT) designed from guide-strand sequences listed in the mirBase database (mirbase.org)[183]. Note that the TaqMan microRNA assays are selective for only mature miRNA and cannot detect precursers or edited variants (Applied Biosystems).

Statistical Analysis. Statistical significance was tested as described in the text. ANOVA and non-parametric Kruskal-Wallis tests were calculated using Prism 5.0 (GraphPad Software). One-sided Kolmogorov-Smirnov tests were calculated using R software package (http://www. R-project.org). Linear regression analysis was calculated using OriginPro 7.0 (OriginLab). For mouse array profiling, genes that showed ≥1.10 fold-change compared to saline-treatment with $p < 0.05$ from three biological replicates per treatment group were considered significant.

3' Rapid Amplification of cDNA ends (RACE). Pdcd4 3'UTR lengths were determined using a RLM-RACE kit (Ambion) and custom nested PCR primers for Pdcd4:

inner Primer: GCACAGCAACTCTTACAGTCTTAGGTGTTAC

outer primer: GAGCTACTGAGCACAGCAACTC

The final PCR products were analyzed with agarose gel electrophoresis.

Western Blotting. Western Blot analysis was performed by first pooling together every other fraction and precipitating the protein with the addition of 3 volumes of pre-chilled neat ethanol and overnight incubation at -20 °C. Following centrifugation, the pelleted protein was washed with 70% (v/v) cold ethanol and then resuspended in 1X LDS sample loading buffer (Invitrogen). Due to the high concentration of protein in the top fractions of the gradient (low sucrose density), the first two pooled fractions were resuspended to their original volume prior to precipitation. The remaining fractions were resuspended in 1/10 of their original volume and were thus concentrated 10×. Equal volumes of each pooled fraction were resolved with SDS-PAGE and immunoblotted following a standard protocol. Rabbit polyclonal antibody to Ago2/Eif2c2 (ab5072) was purchased from Abcam. Rabbit polyclonal antibody to Trbp was developed in-house at Isis Pharmaceuticals (Carlsbad, CA). Primary antibodies were detected using IR-dye labeled secondary antibodies and membranes were scanned using an Odyssey Imaging System (Li-Cor).

2.3 Results

miR-21 silences only a narrow subset of all mRNAs with matching seed sequence

To study miR-21's intrinsic silencing activity *in vivo* in healthy animals, we used array profiling to measure changes in mouse liver mRNA populations resulting from inhibition of the miRNA. Anti-miR oligonucleotides were synthetically tailored to be complementary to the full-length miRNA and featured a number of chemical modifications to enhance specificity, affinity, and stability. We chose to focus on liver tissue, for the following reasons: (a) miR-21 is among the 10 most highly abundant miRNAs in both mouse and human liver samples[184-187] (**Fig. 2.1**); (b) anti-miR delivery to this tissue by systemic administration is made highly efficient by the hepatic portal system; and (c) miR-21 is found to be elevated in human hepatocellular carcinoma clinical samples and has been found to contribute to tumorigenesis and enhanced invasiveness by silencing the expression of the tumor suppressors Pdcd4, Pten, and Reck in hepatocyte derived carcinoma cell lines[126]. In addition to miR-21, we performed similar inhibition analyses on two other miRNAs that similarly are among the most highly expressed liver miRNAs (**Fig. 2.1**): miR-122, formerly believed to be a liver specific miRNA, where it has been implicated in cancer and hepatitis C[188], but recently discovered to be expressed in other tissues as well[189]; and let-7, a conserved family of 11 miRNA isoforms with high sequence similarity that is broadly involved in cell development and tumor suppression[190].

All systemic anti-miR treatments had broad effects on the transcriptome with between ~12% and 23% of the >17,500 genes measured being statistically significantly ($p < 0.05$ by student t-test with three biological replicates per group) up- or down-regulated compared to a saline treatment as negative control (**Fig. 2.2.A, left**). The majority of these changes were mild with median changes in the range of 14 − 20% for down-regulated genes and 21 − 31% for up-regulated genes (**Fig. 2.2.A, right**). Such changes may reflect shifts in the expression levels of miRNA targets, downstream effects caused by these shifts, and/or non-specific effects of the anti-miR[106]. To determine the primary effects of the anti-miR, that is, to elucidate the genes directly regulated by a miRNA, we binned genes based on the presence of 7-nt (7m1A or 7m8) or 8-nt (8m) -

long seed sequence matches in each mRNA's 3'UTR that are complementary to the 5'-end of the miRNA (**Fig. 2.2.B**). It is believed that 8m-seed-matches are the most effective sites, with mRNAs containing these sequences having the highest likelihood for being a miRNA target[17]. In the case of miR-122 and let-7, miRNA inhibition caused a significantly greater upregulation of all seed-matched mRNA compared to non-matched mRNA for the same cumulative fraction (one-sided Kolmogorov-Smirnov p-value (KS p) = 1.44×10^{-25} for 8m-miR-122; 1.06×10^{-8} for 8m-let-7, **Fig. 2.2.C, top and mid panels**). These results are consistent with both canonical miRNA-target predictions[16] and previous results for miR-122 using similar experimental conditions[99]. Unexpectedly, this preferential effect on seed-matched targets was not strongly observed upon miR-21 inhibition (KS p = 3.74×10^{-2} for 8m-miR-21, **Fig. 2.2.C, bottom**). To verify this result, we reanalyzed the same profiling data using the Sylamer algorithm, which computes the hypergeometric significance for finding enrichment of short sequences on one end or the other of a profiling dataset ranked from most up-regulated to most down-regulated[182]. The main advantage of the Sylamer algorithm is that it does not require *a priori* assumptions regarding sequence; it instead computes and plots enrichment signatures for all possible sequences of a certain input length, thus enabling non-specific effects or other trends to be uncovered. Sylamer analysis showed that treatment with anti-miR-122 (**Fig. 2.2.D, top left**, $p < 1 \times 10^{-25}$) or let-7 (**Fig. 2.2.D, top right**, $p < 1 \times 10^{-12}$) induced upregulation of genes enriched for seed-matches unique to the inhibited miRNA, thus demonstrating the specificity of these treatments. Still, we did not find enrichment on either end of the anti-miR-21 ranked gene list for miR-21 seed matches nor any other 7 or 8-nt sequence (**Fig. 2.2.D, bottom left**), indicating that on a transcriptomic scale, the primary effects of miR-21 inhibition cannot be clearly discerned above background.

The three anti-miR oligonucleotides used were identical in all chemical modifications but sequence, thus it is unlikely that the diminished effect of anti-miR-21 was due to drug inefficacy. To further rule out this possibility, we compared livers of wild-type and *miR-21* knockout mice (for validation see [175]) and found a lack of de-repression similar to inhibition by anti-miR-21 (**Fig. 2.2.D, bottom right**).

We also noticed that, in contrast to the majority of targets containing 7 or 8-nt matches, only a narrow subset of miR-21 seed-matched transcripts were significantly up-regulated upon miR-21 inhibition, restricted to just 24 8m-seed-matches (**Table 2.1**). This observation suggests that either miR-21 has a higher selectivity for these targets or it has fewer high-potential targets compared to miR-122 and let-7. We conclude that in healthy liver tissue miR-21 demonstrates very limited mRNA target engagement, silencing fewer targets to a lesser extent than other miRNAs of similarly high abundance.

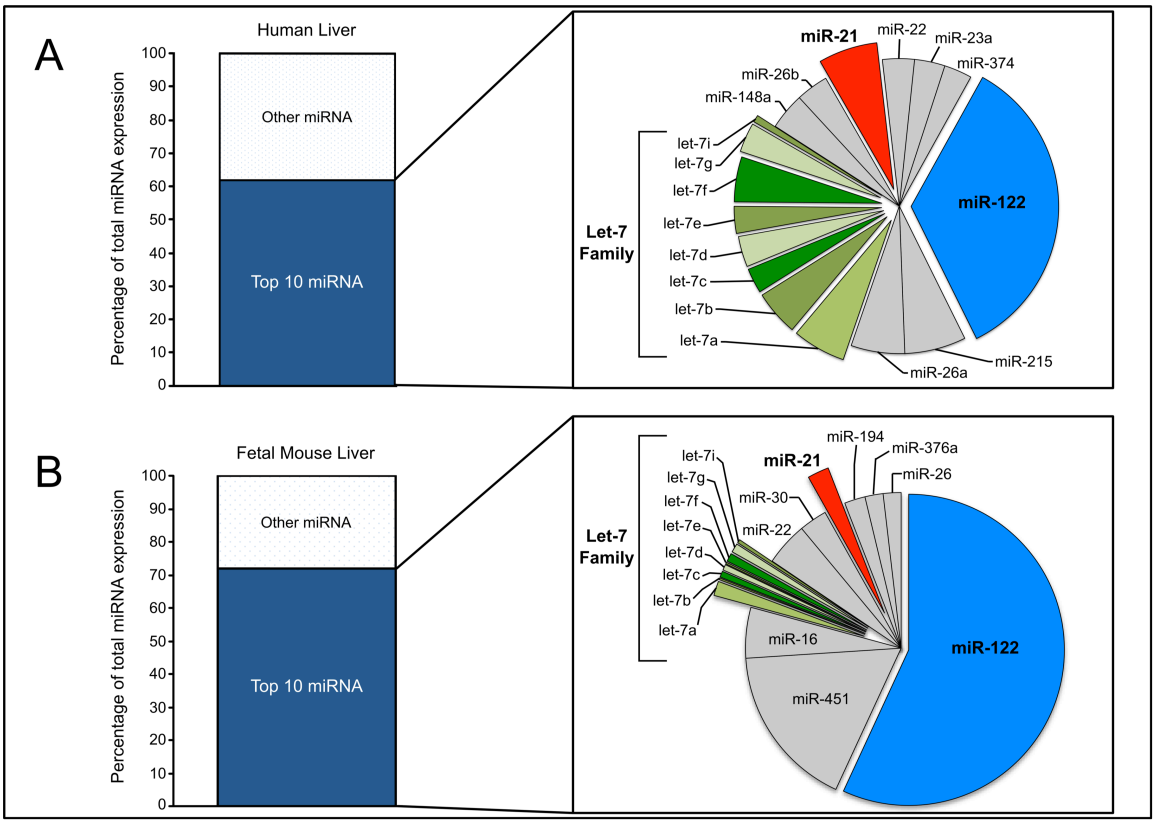


Figure 2.1 miR-21, miR-122, and let-7 family isoforms account for a large proportion of total miRNA expression in liver. Shown are data for human and mouse samples curated in the gene expression omnibus[186]. The top-10 most abundant miRNAs combined as a fraction of total miRNA expression (column graph) and their abundance relative to one another (pie graph) are shown for (**A**) qPCR profiling of human liver samples (GSE22058)[185] and (**B**) deep sequencing data from fetal mouse liver tissue (GSE21370)[184].

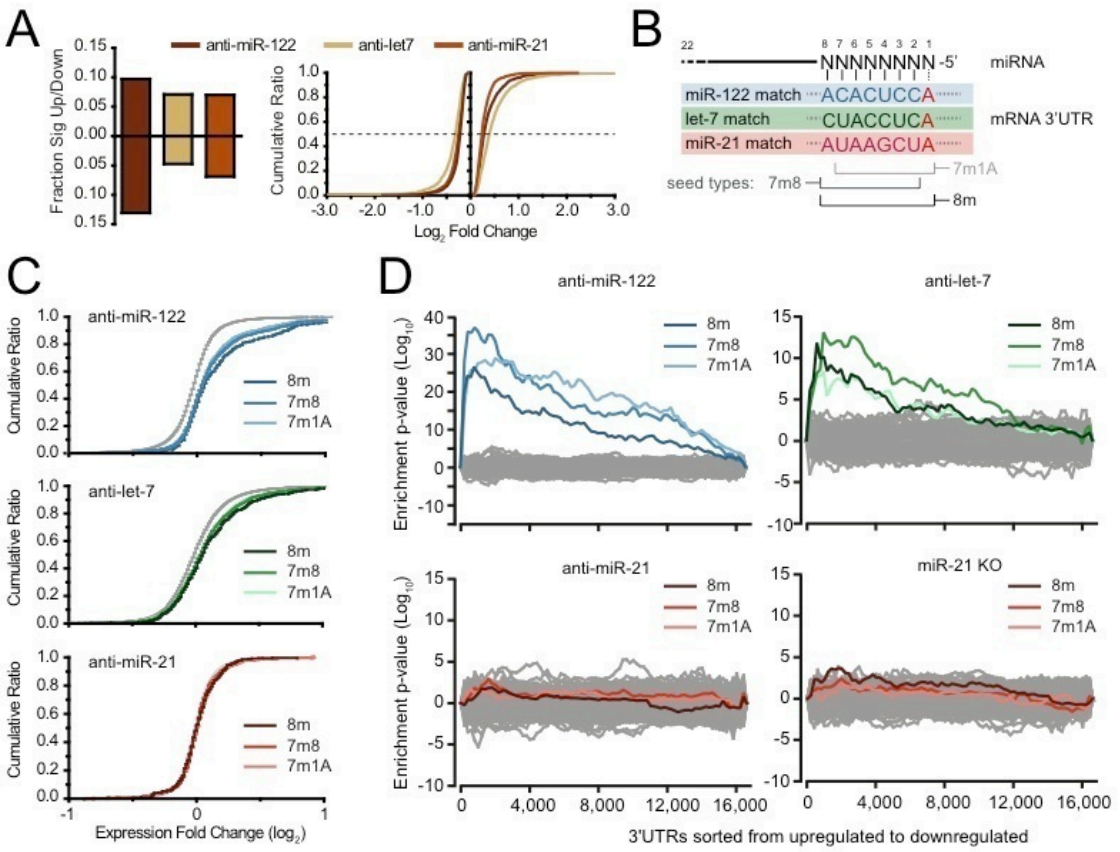


Figure 2.2 Array profiling of liver mRNA following miRNA inhibition or knockout. (**A**) Total gene expression changes in response to anti-miR treatment. Left, fraction of total genes measured which were significantly ($p < 0.05$, student t-test) upregulated (positive y-axis) or down-regulated (negative y-axis). Right, the magnitudes of change in expression for the same fractions plotted as cumulative distribution frequencies. The dotted line represents the median. (**B**) Schematic of canonical miRNA seed:mRNA seed-matched pairing for miR-122, let-7 and miR-21. The miRNA seed region, nucleotides 2-8 in the 5' end of a miRNA, binds to 7- or 8-nt seed-matched sequences in the 3'UTRs of mRNA transcripts. Three variations of seed-matches have been identified as being the most critical for miRNA binding: heptanucleotide sequences complementary to either position 2-8 of the miRNA (7m8) or position 2-7 with an adenosine at position 1 (7m1A), or octanucleotide sequences complementary to position 2-8 plus an adenosine at position 1 (8m). Note that while Watson-Crick base pairing is not required across from position 1, it frequently occurs since a majority of miRNAs contain uracils at this position. (**C**) Cumulative distribution frequencies for all profiled genes plotted as a function of fold-change in expression following anti-miR treatment against miR-122 (top), let-7 (middle), or miR-21 (bottom). Genes were binned based on the presence (colored lines) or absence (grey line) of seed-matched sequences complementary to the inhibited miRNA. (**D**) Heptanucleotide Sylamer analysis of the same datasets shown in C for each indicated treatment. The seed matches for each miRNA are highlighted. All other possible 7-nt sequences not related to the seed match are shown as grey lines and thus represent statistical background noise. The 8m seed match from the octanucleotide analysis is

shown super-imposed. Heptanucleotide and octanucleotide analyses had similar backgrounds. The peaking of enrichment, calculated as a hypergeometric p-value, for seed-matched sequences on the left hand side of the plot for miR-122 (top left) and let-7 (top right) indicates that the genes most up-regulated upon miRNA inhibition are enriched for the corresponding seed-matched sequence. Enrichment for miR-21 seed-matched genes is not observed when miR-21 is inhibited (bottom left) or knocked-out (bottom right).

Table 2.1 All miR-21 8m-seed-matched genes showing significant upregulation upon anti-miR-21 treatment in mouse liver.

Gene	Fold-Change	p-value	Seed-Type	RefSeq ID
Dnaja1	1.73	0.0033	8m	NM_001164671///NM_001164672///NM_00829
Taf7	1.51	0.0001	8m	NM_175770///NM_17577
Skp2	1.33	0.00001	8m	NM_013787///NM_14546
Timp3	1.28	0.001	8m	NM_011595
Pex10	1.26	0.0164	8m	NM_001042407
Tbcel	1.24	0.0307	8m	NM_173038
5730403B10Rik	1.24	0.0043	8m	NM_025670
E2f2	1.23	0.0129	8m	NM_177733
Cdc25a	1.23	0.0019	8m	NM_007658
Yap1	1.21	0.0004	8m	NM_001171147///NM_00953
Fmo2	1.2	0.0095	8m	NM_018881
Zranb2	1.18	0.0018	8m	NM_017381
Nkiras1	1.18	0.0015	8m	NM_023526
Cand1	1.18	0.0169	8m	NM_027994///NM_027994///NM_02799
Ubl4	1.17	0.0068	8m	NM_14505
Ogfod1	1.17	0.0128	8m	NM_001093757///NM_17776
Lipk	1.17	0.0312	8m	NM_172837
Hpgds	1.17	0.0384	8m	NM_019455

Sbno1	1.15	0.0442	8m	NM_001081203
Ehd1	1.14	0.0051	8m	NM_010119
A030009H04Rik	1.12	0.0111	8m	NM_020591///NR_02782
2610507B11Rik	1.12	0.0487	8m	NM_001002004
Agps	1.10	0.031	8m	NM_172666
4632428N05Rik	1.10	0.0226	8m	NM_001159572///NM_02873

miR-21 repression corresponds poorly with targeting predictions and known targets

We next sought to compare our experimental results with the most probable targets derived from the available target prediction databases. Not all mRNAs containing a seed match are expected to be repressed by a co-expressed miRNA, as the likelihood for an interaction can be dependent on several factors such as binding site accessibility and/or hybridization free energy. Several target prediction algorithms have been developed to take these factors into account, with the goal of decreasing the frequency of false-positives[16,82,84,85,183,191,192]. Scores are assigned based on heavy weighting of local mRNA context features, evolutionary conservation, and/or thermodynamics in an effort to rank the most likely or most potent targets of a miRNA. Indeed, for miR-122 and let-7 we found that the experimental changes in mRNA level for the majority of top predicted targets across seven algorithms (see Materials and Methods) were on average greater than the changes for all 7- or 8-nt seed matches together (**Fig. 2.3.A, top**), indicating that they were more predictive than just seed-matching alone. Our understanding of target prediction remains incomplete, as the mean changes of the top predicted targets were below those of the experimentally observed top-30 seed-matched targets, and several top predictions particularly for let-7 were not significantly up-regulated upon anti-miR treatment (**Fig. 2.3.A, bottom**). These observations, however, are in agreement with a similar comparison between top prediction sets and proteomic data following miR-223 genetic deletion in mouse neutrophils[53], indicating a similar, if somewhat limited accuracy of the predicted responses for target proteins and mRNAs upon impairment of miR-223, miR-122, and let-7. By comparison, we observed that far fewer predicted targets of miR-21 were regulated. Although a few targets were predicted correctly (**Fig. 2.3.A, bottom**), these showed only very mild increases upon anti-miR treatment, while the majority of predicted targets were unresponsive or slightly down-regulated (**Fig. 2.3.A, top**). Surprisingly, none of the TargetScan predictions, among the most accurate for miR-122 and let-7, were correct for miR-21; the Context+ algorithm, which was designed to account for miRNAs with low 5' G/C content, a category into which miR-21 falls, did not significantly remedy this lack of accuracy (**Fig. 2.3.A**).

A miRNA's base composition can have a significant influence on mRNA target binding capacity. Sequence analysis revealed full-length miR-21 as slightly more A/U-rich (64%) than miR-122 (54%) and let-7 (59%). These differences are enhanced in the seed region: miR-21 (71%), miR-122 (43%), let-7d (43%) (**Fig. 2.2.B**). Since the seed region often accounts for the majority of miRNA:mRNA base-pairing interactions, we hypothesized that the release of free energy upon target binding would be much lower for miR-21 than miR-122 or let-7. Indeed, computational analysis estimated the free energy of miR-21 binding to targets detected in our array profiling to be significantly less than for the other miRNAs (**Fig. 2.3.B**), indicating that 3'end pairing does not sufficiently compensate for weak seed pairing for a majority of targets. This reduced binding energy may partially explain miR-21's inactivity; however, we did not find much of a correlation between calculated binding energies and changes in mRNA levels (**Fig. 2.4.A**). Thermodynamics also does not satisfactorily explain why miR-21 does not repress Pdcd4, a validated target with extended complementary pairing and binding energies similar to those of miR-122 target AldoA (**Fig. 2.4.B**). Furthermore, the 5' miRNA seed region, critical for target recognition, may disproportionally contribute to the thermodynamics of target binding by serving to nucleate the miRNA:mRNA interaction. In line with a seed-nucleation binding model, structural probing of guide-strand loaded RISC suggests that the 5' segment is preorganized by protein-RNA interactions in a conformation favorable for Watson-Crick base pairing, while the 3' end remains unconstrained, free to adopt a more flexible, less-primed conformation[42]. Recent data from Bartel and colleagues emphasized the relationship between seed base composition and repression by demonstrating that the low repression proficiencies of the *C. elegans* miRNA lsy-6 and the human miRNA miR-23 were due, in part, to weak seed pairing; the other decisive factor was concluded to be high target abundance, which can dilute the effects a miRNA exerts on any given specific mRNA[191]. The influence of these factors was weighted in a modified version of the TargetScan context prediction scores, termed Context+. Interestingly, miR-21 and miR-23 seeds are similarly A/U-rich (2/8 bases are G/C); however, miR-21 is not believed to have high target abundance by our calculation (**Fig. 2.3.C**) or the Bartel group's (39^{th} percentile compared to the 90^{th} percentile for miR-23). As a result, Context+ consistently ranked miR-21 targets higher than those of miR-

23, but often lower than those of miR-122 or let-7 (**Fig. 2.3.D**). In relation to our comparison between array profiling results and target predictions, the mean scores for the top 30 Context+ predictions were not significantly different for miR-21 and miR-122 targets (**Fig. 2.3.E**), yet predictions for miR-122 were far more accurate (**Fig. 2.3.A**). Therefore, we conclude that miR-21, like miR-23 and lsy-6, binds targets with unusually low thermodynamic stability, possibly explaining its substandard repression proficiency. Still, the exhibited activity of miR-21 is not entirely explained by the most recent predictive models nor is it completely analogous to miR-23 or lsy-6 as it has comparably few target sites.

We also compared our experimental results with known targets of each miRNA, supported by experimental evidence in mice or human samples and curated in the miRecords database of validated miRNA:target associations[193]. We found only 12.5% of all validated miR-21 targets to be significantly upregulated upon anti-miR treatment, with the prominent miR-21 targets Pdcd4, Pten, and Reck among the majority of targets that were not up-regulated (**Fig. 2.3.F**). In comparison, miR-122 and let-7 inhibition derepressed 90% and 15.8% of all previously reported targets, respectively. The discrepancy between miR-122 and let-7 was unexpected considering the overall strong repression signatures and agreement with prediction algorithms found for these two miRNAs. We note, however, that many of the curated targets for miR-122 are derived from a single study that used anti-miRs to inhibit liver miR-122[108], similar to our own, which explains the consistency. By contrast, the curated targets for let-7 are individually referenced from a number of different studies from a wide range of biological samples. These observations emphasize the importance of biological context when assessing specific miRNA-target pairings.

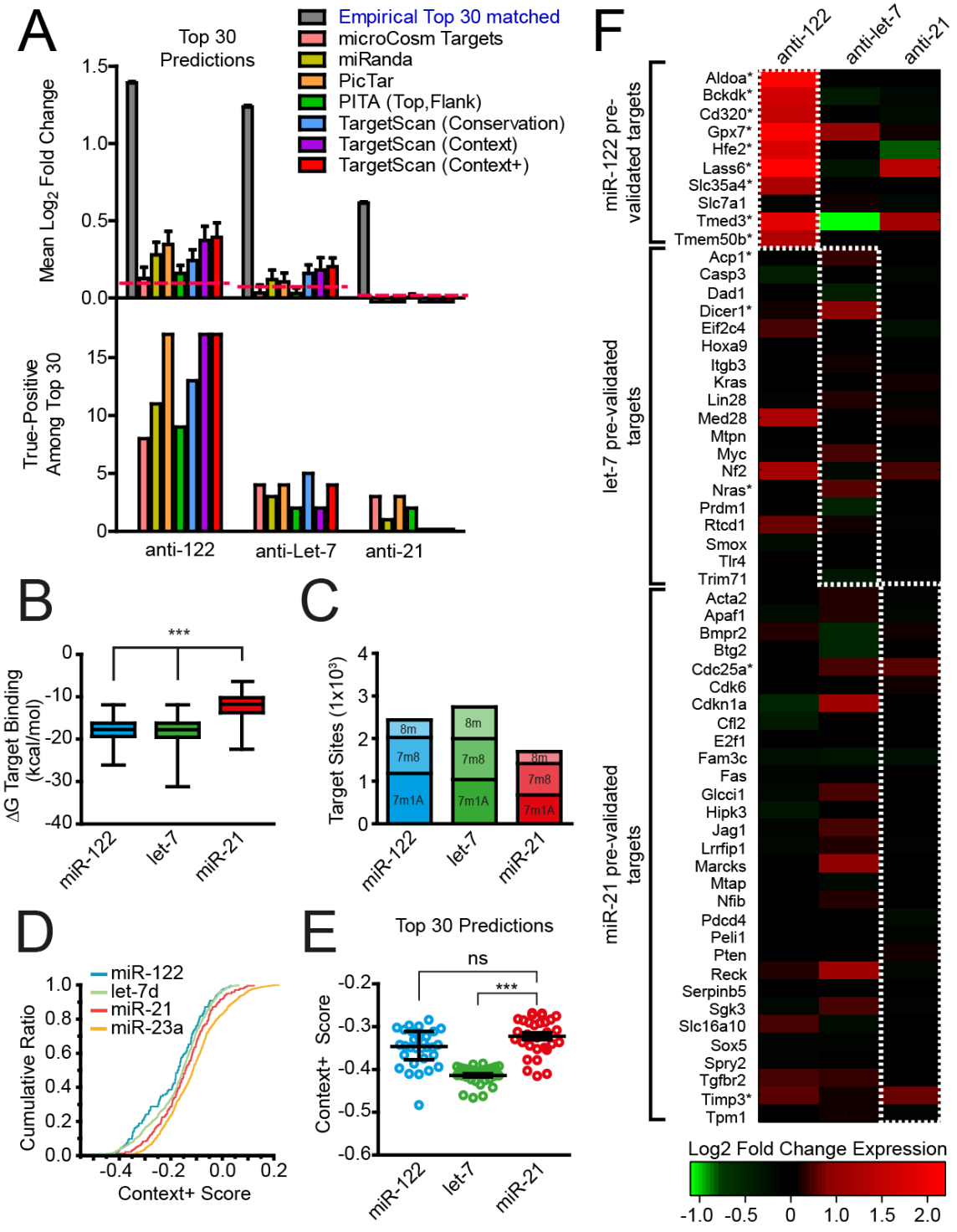


Figure 2.3 miR-21 inhibition does not derepress expected targets. (**A**) Comparison between bioinformatic target predictions and observed mRNA changes. Top, mean log2 fold-change of the top 30 ranked targets from each computational prediction algorithm shown for the corresponding miRNA of each treatment. Grey bars show the mean fold-change of the observed top 30 most up-regulated seed-matched mRNAs, and thus

represent the highest fold-change possible for any of the predictions. Dashed red lines show the mean fold-change for all seed-matched mRNAs for each miRNA. Bottom, the number of the top 30 predicted targets from each algorithm that were significantly upregulated (≥10% increase in mRNA with $p \leq 0.05$, student t-test). (**B**) Box-and-whisker plots of calculated free energies of target binding for each miR-122 (blue), let-7 (green) and miR-21 (red). Each bar represents five values: the median (center line), 25th percentile (bottom of box), 75th percentile (top of box), maximum free energy (bottom line), and minimum free energy (top line). (***) = $p < 0.001$, determined by Kruskal-Wallis test with Dunns post-test. (**C**) Predicted number of 3'UTR target sites for each miRNA binned by seed match type. (**D**) Cumulative distribution frequencies for TargetScan Context+ scores for the indicated miR-122 (blue), let-7 (green), miR-21 (red), and miR-23a (yellow). Negative scores are ranked higher. (**E**) Scatter plots of TargetScan Context+ scores for the top 30 predicted targets of the each miRNA. Each center line represents the median, whereas top and bottom lines denote the interquartile range. ns = not significant, (***) = $p < 0.001$, determined by Kruskal-Wallis test with Dunns post-test. (**F**) Heat map of observed changes in mRNA levels for previously validated targets of miR-122, miR-21, or let-7 curated in miRecords database. Expression changes for each gene are shown for all three anti-miR treatments, with the white-dashed line highlighting the treatment that corresponds to the miRNA of the pre-validated target. A star marks each gene that was significantly upregulated upon inhibition of the expected targeting miRNA.

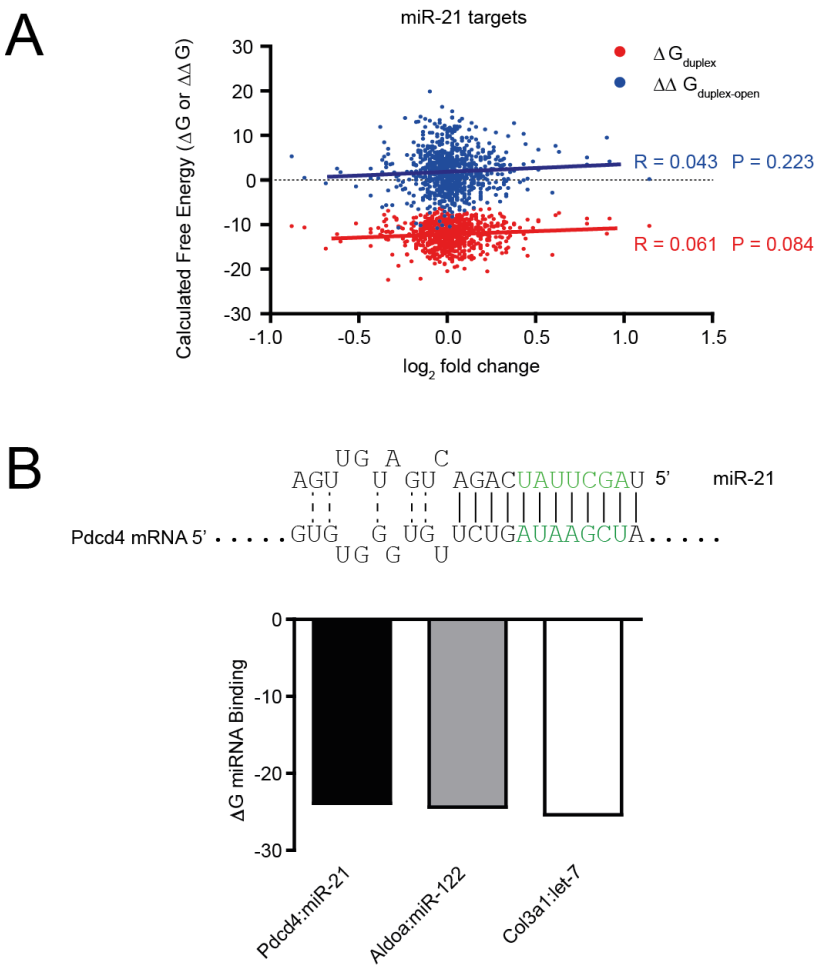


Figure 2.4 Derepression of seed-matched targets with anti-miR-21 treatment does not correlate with predicted thermodynamics of miR-21 binding. (**A**) Linear regression analysis of thermodynamic data from the PITA database[85] with measured expression changes for miR-21 seed-matched mRNAs. Two thermodynamic values are plotted on the y-axis for each seed-matched gene: free energy of duplex binding (ΔG_{duplex}, red) and free energy of duplex binding minus the cost of target unwinding ($\Delta\Delta G_{duplex\text{-}open}$, blue). Linear regression fits, correlation coefficients, and p values are shown for each. (**B**) miR-21 is expected to bind Pdcd4 mRNA, a previously validated target of miR-21, with particularly high stability. Top, predicted base pairing between miR-21 and Pdcd4 mRNA. The seed region, often the only region of the miRNA with perfect base pairing, is shown in green to illustrate the extended interaction. Bottom, extended base pairing with Pdcd4 is predicted to compensate for a high A/U rich seed region, increasing the stability of the interaction to levels comparable with those predicted for miR-122 and let-7. Shown are the RNAhybrid[179] calculated free energies for binding of miR-21 with Pdcd4 (black bar), of miR-122 with AldoA (grey bar), and of let-7d with Col3a1 (white bar).

Stress response genes are exceptionally sensitive to miR-21 inhibition

Since the limited transcriptome response to miR-21 inhibition corresponded well with neither miRNA target predictions nor previously reported targets, we next turned to gene ontology classifications to identify whether a biological relationship existed among the up-regulated genes. Closer inspection showed that the most up-regulated genes induced by anti-miR-21 treatment belonged to the DnaJ family of co-chaperones, with Dnaja1 as the single most up-regulated 8m-seed-matched target, and heat shock proteins, including several Hsp90 variants and Hsph1, as the most up-regulated genes in the entire profiling dataset (**Fig. 2.5.A**). Interestingly, stress response genes were significantly enriched ($p < 0.05$, binomial test[176,177] only in both the anti-miR-21 up-regulated seed-matched ($p = 0.028$) and non-matched ($p = 0.018$) datasets. The large majority of these genes were not significantly upregulated by the other anti-miRs, disfavoring the notion that the stress response pathway may have been triggered by a non-specific effect of the anti-miR treatment (**Fig. 2.5.A**). Many of the activated stress response genes do not contain a miR-21 seed-match, perhaps indicating that they are upregulated due to derepression of a transcription factor (**Fig. 2.5.A**, non-seed containing genes are marked in black). Notably, stress response proteins are coordinately transcribed by heat shock factors (HSFs), which trimerize in response to stress and activate transcription by releasing RNA polymerase II (pol II) from a TATA binding protein (TBP)-dependent TFIID transcription initiation complex, stalled at the promoter regions of DNA loci encoding chaperones and co-chaperones[194,195]. Although we did not find an upregulation of HSFs (**Fig. 2.5.B**), we did discover that several TBP-associated factors (TAFs) were strongly upregulated, particularly Taf7, which contains a miR-21 8m seed in its 3'UTR (**Fig. 2.5.B**). Recombinant Taf7 was previously found to associate with the HSF1 oligomerization domain in vitro and some models suggest an active role for TAF proteins in HSF transcriptional activation[195]. Thus, it emerges that miR-21 may regulate heat stress response cascades, possibly by directly repressing Taf7, although we cannot rule out alternative mechanisms.

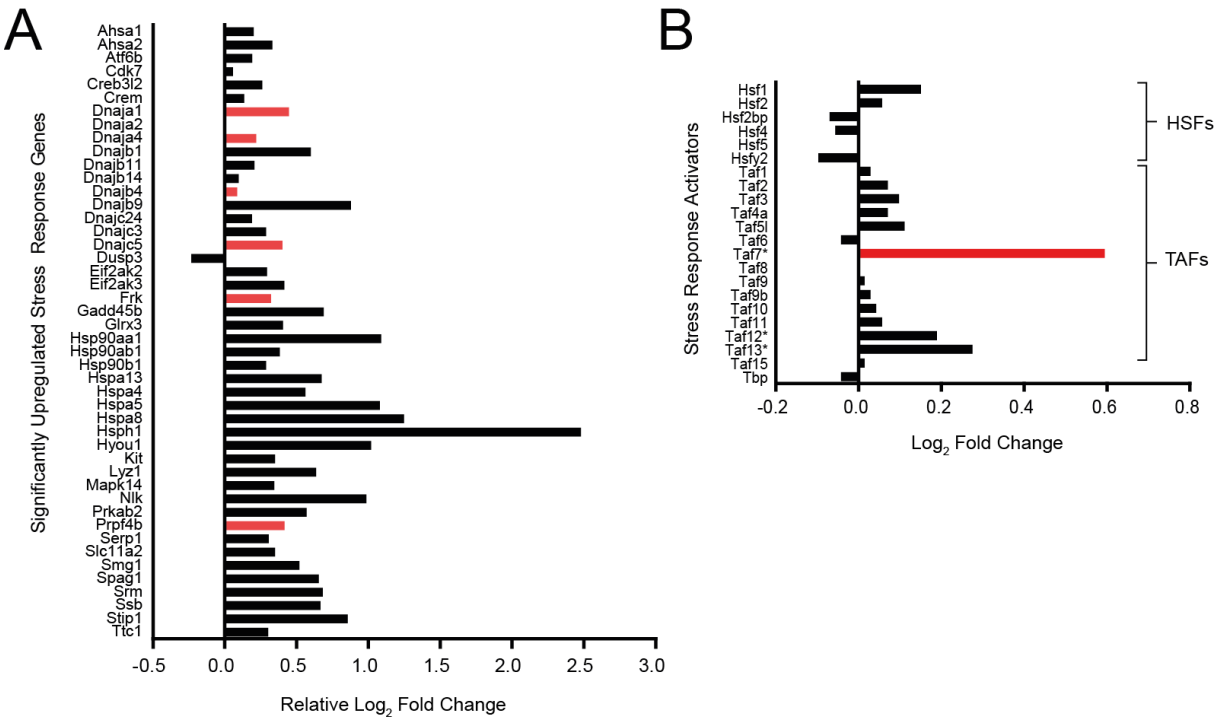


Figure 2.5 miR-21 inhibition induces expression of stress response genes. (**A**) mRNA changes for significantly up-regulated stress-response genes following anti-miR-21 treatment. Fold change levels are plotted relative to those observed for the other anti-miR treatments [fold-change$_{\text{anti-21}}$ - mean fold-change$_{\text{(anti-122, anti-let-7)}}$] to demonstrate the anti-miR-sequence-dependence of induction. Red bars indicate the gene transcript contains a seed match for miR-21. (**B**) mRNA changes for stress response activators (HSFs) or related transcriptional genes (TAFs and TBP). Fold change levels represent those observed for anti-miR-21 treatment alone. A star marks each gene that was significantly upregulated upon miR-21 inhibition. Red bars indicate that the gene contains a seed match for miR-21.

miR-21 is sparsely associated with translating polysomes

We next sought to investigate the functional basis for miR-21's diminished silencing activity. We hypothesized that miR-21 may be less associated with mRNA due to either weak target-binding or assembly of miR-21 into alternative RNA-protein complexes that prevent it from engaging target transcripts. To assess this possibility, we performed polysome analysis on linear sucrose gradients and compared the distribution of mature miR-21 to those of mature miR-122 and let-7. Such analyses have shown that miRNAs typically fractionate predominantly with translating ribosomes, likely by base pairing with target mRNAs[77-79,171]. Cytoplasmic lysates prepared from perfused liver tissue in a manner that preserves the majority of all tested miRNAs (**Fig. 2.6**) were fractionated by centrifugation on 20%-60% sucrose gradients. Analysis of the ribosomal RNA (rRNA) content of each fraction based on absorbance at $\lambda = 260$ nm (A_{260}) and electrophoresis of total RNA on denaturing agarose gels revealed that polysomes, representing mRNAs associated with multiple ribosomes, sedimented into the densest half of the gradient (fractions 11-20, **Fig. 2.8.A**), as expected. We note that cycloheximide, as an inhibitor of translation, was not necessary to prevent runoff under our conditions and its inclusion did not have an effect (**Fig. 2.7**); we therefore did not typically include it in our assays. western blots showed that the essential RISC proteins Argonaute 2 (Ago2) and TAR-RNA binding protein (Trbp) (MacRae et al., 2008) were present throughout the gradient, with the most significant proportion found in fractions 1-4 at the top of the gradient (**Fig. 2.8.A**). Using real-time quantitative PCR (RT-qPCR) we analyzed the abundance of miRNAs in each fraction and found that miR-21 displays a significantly different distribution compared to miR-122 and let-7d, a representative isoform of the let-7 family (**Fig. 2.8.B**). While ~50% of both miR-122 and let-7d were detected in the polysome-containing fractions, miR-21 was ~2.5-fold depleted in these fractions (**Fig. 2.8.C**), but enriched by a similar margin in the top fractions of the gradient (fractions 2-3) where it is likely assembled in a non-target bound RISC-like complex since these fractions are where RISC proteins are most abundant and the sedimentation velocity appears to be greater than expected for free miRNA alone (fraction 1-2). These results appear to be independent of overall miRNA abundance since the relative levels of miR-122 and let-7d are greater and less than that of miR-21, respectively (**Fig. 2.8.D**).

The association of miRNAs with polysomes was previously reported for cultured HeLa cells[78,79]. To confirm that these results are consistent *in vivo*, and that co-sedimentation was not coincidental, we tested whether the distribution of miRNAs in polysome-containing fractions could be disrupted with puromycin, a tRNA-mimic that blocks elongation of actively translating complexes. As expected, treatment of lysates with puromycin prior to loading on sucrose gradients caused a loss in polysomes and a resulting increase in 80S ribosomes (**Fig. 2.8.E**). miRNAs showed a similar shift away from the densest fractions towards the top of the gradient (Fig. **2.8.F**). In addition, we tested the effect of EDTA, which causes dissociation of polysomes and monosomes into 40S and 60S ribosomal subunits (**Fig. 2.9**). Again, polysome dissociation was matched by the loss of miRNA sedimentation into the densest fractions (**Fig. 2.9**, quantified in **Fig. 2.8.F**, inset). Interestingly, miR-122 under these conditions still sedimented with the 80S ribosome fractions, perhaps due to the presence of a RISC-loading complex of similar sedimentation properties[196].

Taken together, these results suggest that, in healthy mouse liver, miR-21 is associated with actively translating complexes to a significantly lesser extent than other miRNAs.

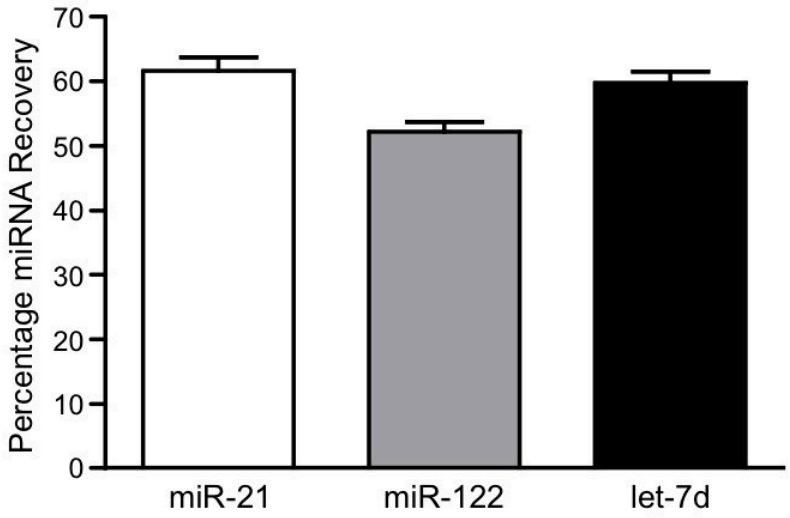


Figure 2.6 Percent recovery of miRNA in lysates used for sucrose gradient analyses. Homogenized liver tissue was pelleted briefly at 500 × g for 1 min at 4 °C to remove any remaining intact cells. The resulting S0.5 supernatant was then centrifuged at 16,000 × g for 20 min to yield S16 supernatant and P16 pellet. Total RNA was purified from S16 and P16 and equal volumes of each sample were assayed for each miRNA as described in the Materials and Methods section. Percent recovery was calculated as $N_{S16}/(N_{S16}+N_{P16})$, where N is the copy number detected. Error bars represent standard error of the mean (SEM).

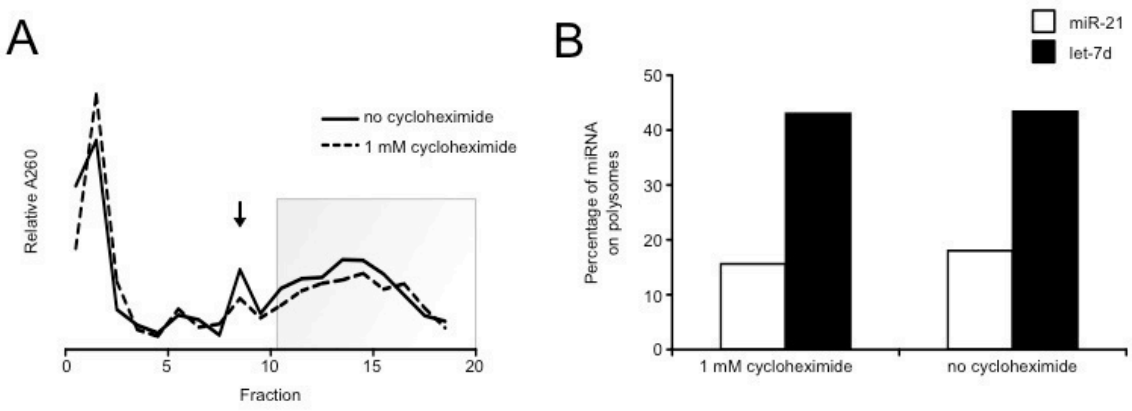


Figure 2.7 Comparison of lysates prepared in the presence or absence of cycloheximide. Liver sections of equal mass from the same animal were prepared as described in the Materials and Methods section. For sample prepared in the presence of cycloheximide, a final concentration of 1 mM cycloheximide was spiked into the sample during homogenization. (**A**) Relative A260 of purified RNA from each fraction. Arrow indicates the 80S fraction and the grey box indicates the fractions containing polysomes (PS). (**B)** Percentage of miR-21 and let-7d on polysomes from each preparation.

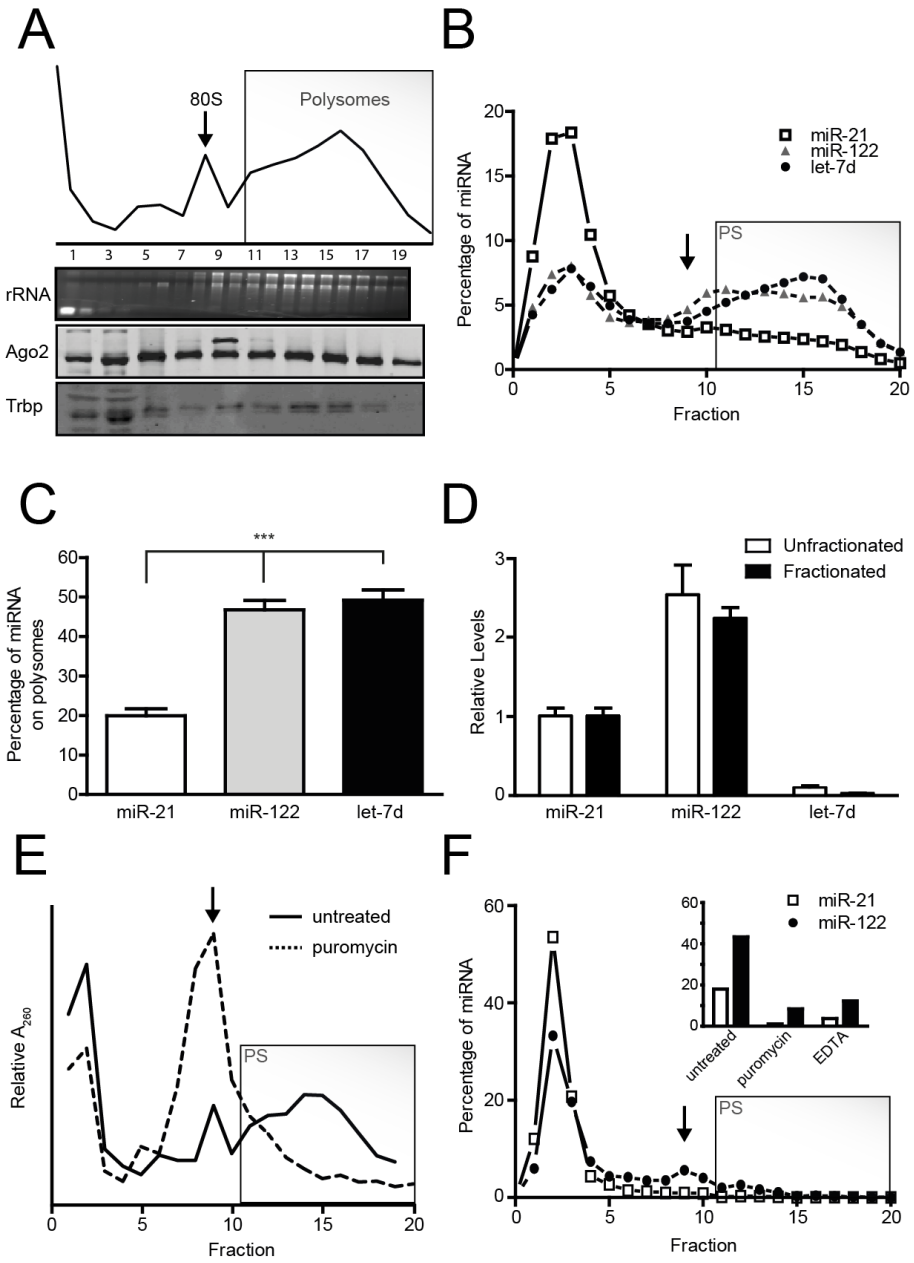


Figure 2.8 miR-21 is disproportionally lacking in polysomal complexes. (**A**) Top, A260 profile and denaturing agarose analysis of sucrose gradient fractions. Arrow indicates the 80S fraction and the grey box indicates the fractions containing polysomes (PS). Bottom, western blots of pooled adjacent fractions confirming the presence of RISC proteins throughout the gradient. Inputs for lanes 3-10 were concentrated 10-fold by precipitation prior to loading, while lanes 1-2 were not further concentrated. (**B**) Equal volumes of purified total RNA from each fraction were analyzed with RT-qPCR for the presence of miR-21 (white square), miR-122 (grey triangle), and let-7d (black circle). The fraction copy number for each miRNA is plotted as the percent total copy number detected from all fractions. Each data point shown is the mean from a total N = 7 from three independent experiments. (**C**) The mean summed percent total from *B* in fractions

11-20 (highlighted with grey box) for miR-21 (white bar), miR-122 (grey bar) and let-7d (black bar). (***) = $p < 0.001$, calculated by one-way analysis of variance (ANOVA) with Bonferroni post-test. Error bars represent standard error of the mean (SEM). (**D**) Relative miRNA levels before (unfractionated; white bar) or summed after (fractionated; black bar) sucrose gradient fractionation. Error bars represent standard error of the mean (SEM). (**E**) miRNA sedimentation is sensitive to translational drop-off. Lysates were treated with puromycin or EDTA to disrupt translation. A_{260} profiles for puromycin treated (dashed line) and untreated (solid line) samples. (**F**) miRNA distributions resulting from puromycin treatment. Inset, comparison of miR-21 (white bars) and miR-122 (black bars) summed percent totals in the densest fractions (normally taken to be polysome-containing fractions) for untreated, puromycin, and EDTA treated samples.

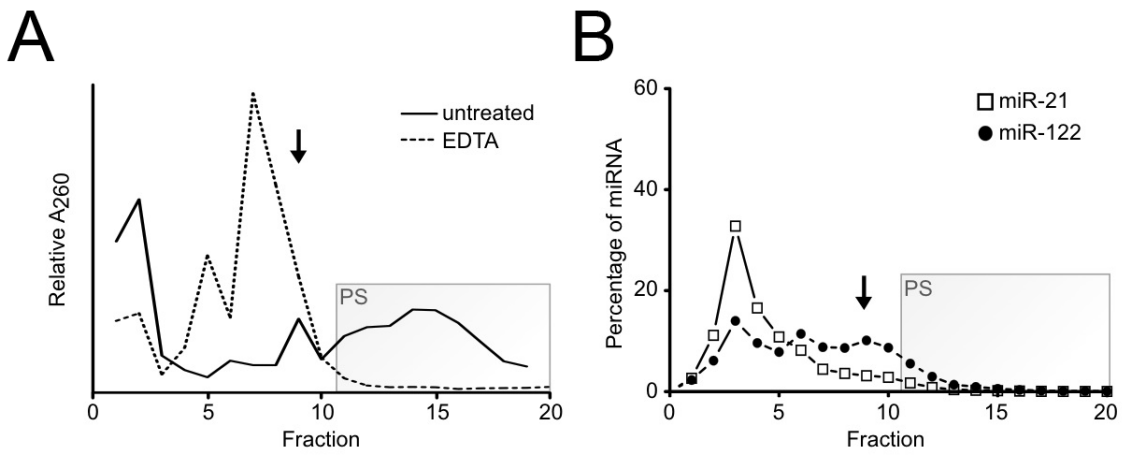


Figure 2.9 Comparison of lysates prepared in the presence or absence of EDTA. (**A**) A_{260} profiles for EDTA treated (dashed line) and untreated (solid line) samples. (**B**) Resulting miR-21 (white squares) and miR-122 (black circles) distributions for EDTA treatment. Arrows mark the 80S fraction. Grey box indicates the fractions containing polysomes in the untreated sample

miR-21 is able to bind complementary RNA and remains unassociated with polysomes upon administration of its anti-miR

To further probe for a correlation between miR-21's altered sub-cellular distribution and its mRNA silencing activity, we analyzed miRNA gradient distributions following anti-miR treatment. Mice were treated with a single dose of anti-miR 24 h prior to gradient analysis. This quick treatment, sufficient to elicit a measurable response using Sylamer analysis for anti-miR-122, manifested in the gradients as a nearly complete shift of miR-122 away from all but the lowest-density fractions (**Fig. 2.10.A**). In polysome-containing fractions, miR-122 enrichment was reduced ~20-fold with anti-miR-122 compared to saline and anti-miR-21 treated animals (**Fig. 2.10.A**, inset). This shift was sequence selective, thus is very likely due to complementary base pairing between anti-miR and miRNA. This observation suggests that miRNAs sediment with polysomes because they are directly bound to mRNAs rather than associated with ribosomes through interactions bridged by RISC components.

While miR-21 showed a similar shift to low-density fractions upon treatment with anti-miR-21, the effect was significantly minimized due to the fact that miR-21 was already preferentially distributed in the low-density fractions even in the absence of anti-miR (**Fig. 2.10.B**). As a result, anti-miR-21 induced only a ~4-fold change of miR-21 in the polysomal fraction, about five-fold less of an effect than that seen with miR-122 and anti-miR-122 (**Fig. 2.10.B**, inset).

Despite reports to the contrary, anti-miRs do not cause degradation of cognate miRNAs; they can, however, interfere with miRNA detection, making it appear as if the miRNA has been degraded[96]. To ensure that the observed anti-miR-induced shifts were genuine and not a detection artifact, the relative levels of miRNAs in treated and untreated lysates were compared prior to loading on gradients. Levels of miR-122 and miR-21 were not reduced following anti-miR treatment (**Fig. 2.10.C**), indicating that under our experimental conditions, anti-miR did not interfere with RT-qPCR measurements. Surprisingly, miR-21 instead showed about a two-fold increase in expression with anti-miR-21 treatment (**Fig. 2.10.C**). This observation may be attributed to either release of otherwise immeasurable miR-21 from sequestered populations or de

novo miR-21 expression triggered by a feedback loop in response to anti-miR mediated inhibition, perhaps suggesting that minimal silencing activity of miR-21 is still relevant to signaling pathways (see Discussion). Still, even with such increased levels, we infer that miR-21 is fully inhibited by anti-miR based on the extent of depletion from dense fractions of the sucrose gradient (**Fig. 2.10.B**).

Since the anti-miR inhibited miR-21 resembles the uninhibited miRNA on the sucrose gradient, we investigated whether miR-21 could at all bind complementary RNA. Liver lysates were prepared as usual and incubated with increasing concentrations of 5'-biotin-end-labeled 2'-O-methyl oligonucleotides fully complementary to either miR-122 or miR-21. After pull-down with streptavidin coated-beads, the depletion of each miRNA was quantified using northern blot. The entirety of measurable miR-21 was depleted with a half-titration point of $K_{1/2} = 0.82 \pm 0.09$ nM, while miR-122 was depleted with a $K_{1/2} = 1.4 \pm 0.1$ nM (**Fig. 2.10.D**). Considering that miR-122 levels in our lysates are ~2.5-fold those of miR-21 (**Fig. 2.8.D**), the binding results indicate that miR-21 is in principle capable of binding a complementary RNA, just as efficiently as miR-122.

The findings that systemically delivered anti-miRs cause an upregulation of corresponding mRNA targets and block the binding of their respective miRNAs to polysome-bound mRNAs indicates that the miRNA-polysome interaction is critical for miRNA-mediated silencing. Consequently, the severely reduced binding of miR-21 to polysomes is consistent with its decreased mRNA silencing activity and the reduced response of its mRNA targets to the administration of anti-miR-21.

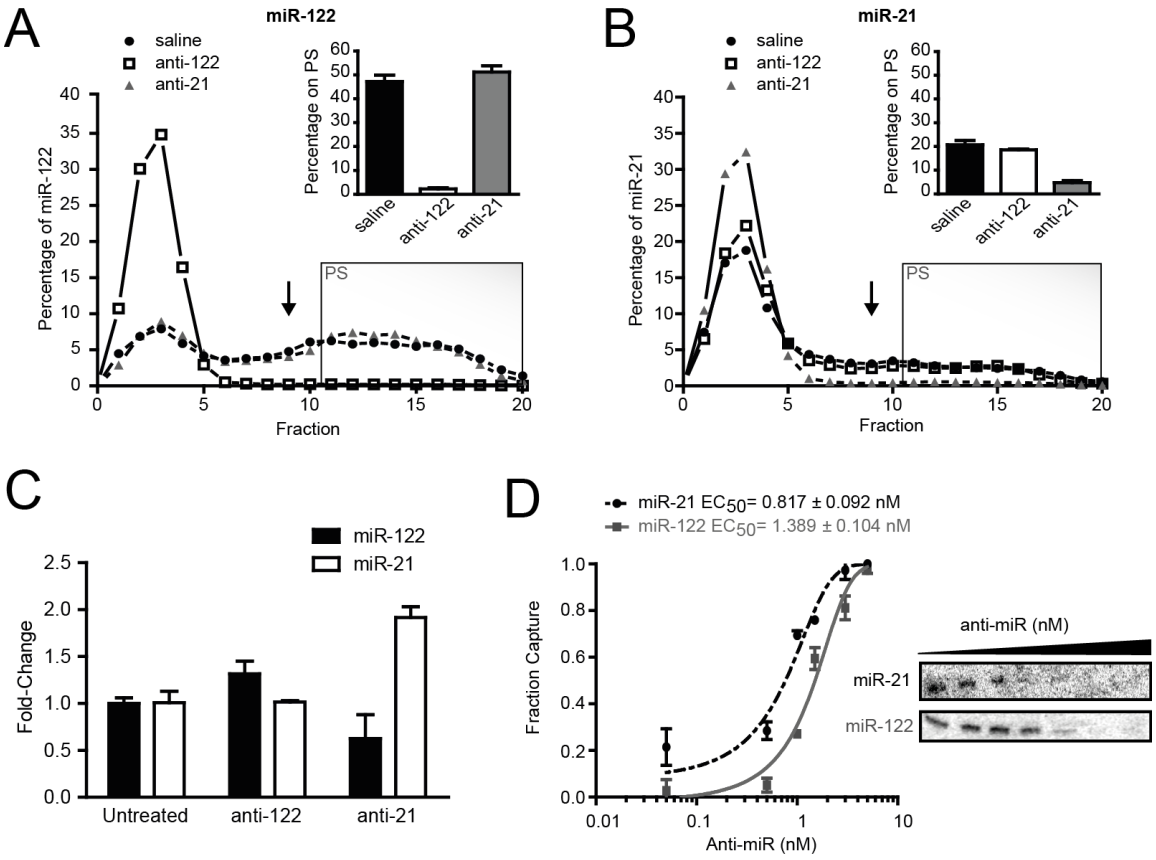


Figure 2.10 The sub-cellular distribution of miR-21 resembles that of inhibited miRNA. (**A**) Percent distribution profile of miR-122 in sucrose gradients following a single dose of saline (black circle or bar), anti-122 (white square or bar), or anti-21 (grey triangle or bar) 24 h prior to harvest. Arrow indicates the 80S fraction and the grey box indicates the fractions containing polysomes (PS). Inset, the corresponding mean summed percent total in polysomal fractions for each treatment. (**B**) As in *A* for miR-21. (**C**) Comparison of miRNA levels in anti-miR treated lysates prior to gradient fractionation. The relative miR-122 (black bars) and miR-21 (white bars) levels were calculated with the $2^{-\Delta\Delta Ct}$ method[197] using miR-22 as a reference. The normalized fold-change levels of miRNA were further set to 1.0 in saline treated animals. (**D**) miRNA binding assay with 2'O-methyl complementary capture RNA. Biotinylated 2'O-methyl oligos complementary to either miR-21 (black circle and dashed line) or miR-122 (grey square and solid line) were titrated into S16 liver extracts. Bound miRNA was depleted from the extract by precipitation of the capture RNA using streptavidin-coated beads, while unbound miRNA was detected by northern blot (shown here from a representative experiment). Data were fit with a sigmoidal dose-response curve with variable slope (for miR-21: $K_{1/2}$ = 0.817 +/- 0.092 nM, hill-slope = 1.080 +/- 0.2194, R^2 = 0.962; for miR-122: $K_{1/2}$ = 1.389 +/- 0.104 nM, hill-slope = 1.130 +/- 0.312, R^2 = 0.967). Data points were averaged from duplicate experiments and error bars represent SEM.

A known target mRNA is largely inaccessible to miR-21

Our interpretation that intracellular access and binding of miRNA to polysomes is important to miRNA-mediated silencing assumes that repressed transcripts are indeed associated with polysomes. While mRNA is necessary for polysome formation, previous cell culture studies have shown that some mRNA targets are not associated with polysomes during repression by miRNA[68,69]. Instead, these repressed mRNAs are translationally incompetent and sediment in low-density fractions of sucrose gradients.

To determine the location of repressed and non-repressed mRNAs under our experimental conditions, we measured the distribution of AldoA and Pdcd4 mRNA in the presence and absence of anti-miR. These mRNAs are previously validated targets of miR-122 and miR-21, respectively. Consistent with our profiling data, RT-qPCR showed a four-fold derepression (increase) of AldoA mRNA levels upon anti-miR-122 administration, confirming it as a miR-122 target. In contrast, we found no changes in Pdcd4 levels upon treatment with anti-miR-21, indicating the absence of a regulatory interaction with miR-21 (**Fig. 2.11.A**). Unlike the observations of earlier cell culture studies[68,69], the largest fraction of both mRNAs (~80%) was found in polysomes under saline treated conditions, demonstrating that repression by endogenous liver miRNAs does not prevent (at least partial) translation of target mRNAs (**Fig. 2.11.B-C**). In the presence of the corresponding anti-miR, both AldoA and Pdcd4 mRNAs shift slightly from fractions 16-20 to fractions 10-14; however, this shift does not significantly reduce the total percentage of mRNA contained in polysomal fractions (**Fig. 2.11.D**). This finding indicates that anti-miRs can cause slight changes to translation in a sequence-independent manner reminiscent of a mild interferon-like response[198,199], rather than the miRNA-mediated effect observed with miR-122 in Huh-7 cells[69] and with let-7 in HeLa cells[68].

Three Pdcd4 transcript variants are recorded in the NCBI Reference Sequence (RefSeq) database. All three are transcribed from the same locus and code for the same protein; however, the mRNAs differ in 5'- and 3'UTR length (**Fig. 2.11.E**, left). While variant 1 (NM_011050.4) and variant 2 (NM_001168491.1) have identical 3'UTRs and both contain a single miR-21 8m seed-match, variant 3 (NM_001168492.1) does not

contain a seed match due to the absence of an ~500 nt segment flanking the miR-21 site. Intriguingly, the remainder of the 3'UTR, both upstream and downstream of the miR-21 site, is identical to that of the other variants, suggesting that variant 3 may have evolved to avoid miR-21 regulation. If variant 3 were the predominant transcript in mouse liver, this would explain why we did not observe derepression of this highly referenced target. To test the relative abundance of the variants, 3' rapid amplification of cDNA ends (3'RACE) was employed to amplify the 3'UTRs of transcripts containing the Pdcd4 open reading frame from total RNA. Agarose gel electrophoresis of the amplified product displayed two bands with lengths of ~750 bp and ~250 bp, which matched the expected lengths of variants 1/2 and variant 3, respectively (**Fig. 2.11.E**, right). Densitometric analysis of the relative band intensities showed ≥ 90% of the Pdcd4 transcript population to be of the longer variety, indicating that the majority of Pdcd4 expressed in mouse liver contains a miR-21 binding site.

We conclude that mRNA targets of miRNAs retain translational capacity and reside in polysomes, regardless of the extent of their regulation. Taken together with miR-21's observed sub-cellular distribution in healthy liver, a known target mRNA such as Pdcd4 appears predominantly inaccessible to the abundant intracellular miR-21.

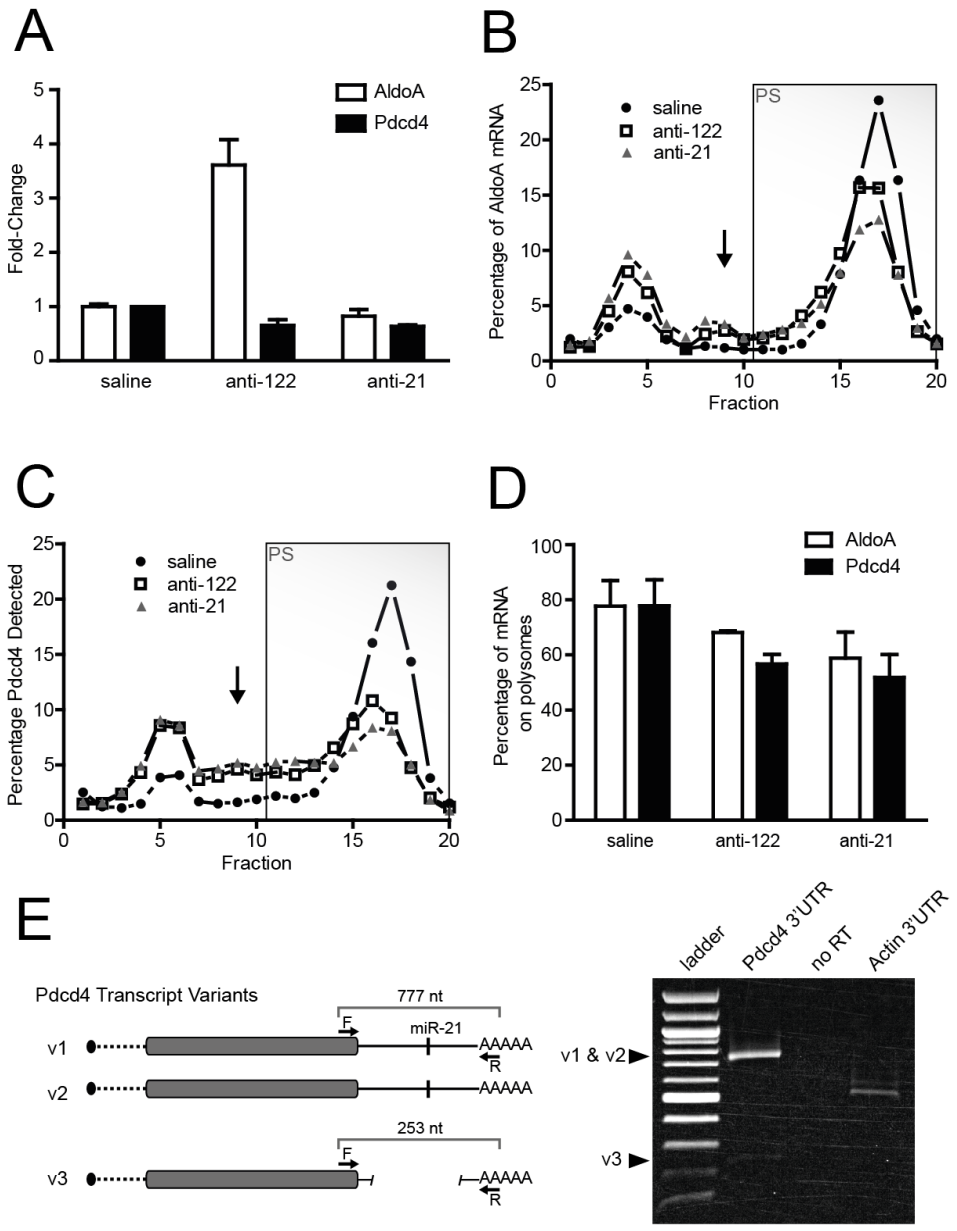


Figure 2.11 Predicted mRNA targets remain associated with polysomes in the presence and absence of miRNA-mediated repression. (**A**) Relative levels of AldoA (white bar) and Pdcd4 (black bar), following treatment with anti-miR. mRNA levels were calculated with the $2^{-\Delta\Delta Ct}$ method using GAPDH as a reference. (**B**) Percent distribution profile for AldoA mRNA, a predicted target for miR-122, in sucrose gradients following a single dose of saline (black circle), anti-122 (white square), or anti-21 (grey triangle) 24 h prior to harvest. Arrow indicates the 80S fraction and the grey box indicates the fractions containing polysomes (PS). (**C**) As in *B* for Pdcd4. (**D**) The corresponding mean summed percent totals of AldoA (white bars) and Pdcd4 (black bars) mRNA in polysomal fractions for each treatment. Data points were averaged from duplicate experiments and error bars represent SEM. (**E**) Three variants of the Pdcd4 transcript are encoded in the same locus in the mouse genome. Left, schematic of variant sequence

alignment. The transcripts, which code for the same protein, have variable 5'UTRs and identical open reading frames (grey box) and 3'UTRs, except for variant 3, where the miR-21 binding site is absent, but the upstream and downstream portions of the 3'UTR are retained. Arrows mark the sites for forward (F) and reverse (R) primers for sequence length determination using 3' rapid amplification of cDNA ends (RACE). Right, agarose gel of PCR products from 3'RACE. The expected lengths for the variants are marked by arrowheads. Based on intensity comparison between the slower migrating and faster migrating bands, the larger variants containing miR-21 seed matches make up ≥ 90% of the Pdcd4 transcript population. Reactions containing no reverse transcriptase (no RT) or primers for the actin 3'UTR served as negative and positive controls, respectively.

In cancer cells, miR-21 is highly associated with polysomes and has a strong repression signature

In notable contrast to our findings in liver, Nilsen and colleagues have previously reported miR-21 as highly associated with polysomes in cervical adenocarcinoma HeLa cells[78], where measurable miR-21 target repression is observed[127]. To verify these results, HeLa cell extracts were prepared as described by these authors, analyzed on sucrose gradients, and miR-21 was quantified by RT-qPCR. Consistent with the previous report, ~60% of miR-21 was distributed in polysome-containing fractions in similar abundance as that observed for liver let-7 and miR-122 (**Fig. 2.12.A**). Additionally, let-7a was also found to have enhanced polysome association in HeLa cells as compared to liver, with nearly all let-7a detected in polysomal fractions (**Fig. 2.12.A**). This observation suggests that miRNA target engagement may be globally enhanced in HeLa cells. As a control, only ~12% of liver miR-21 from gradients prepared using identical conditions were found in polysome fractions, even less than using our original preparation conditions (compare **Fig. 2.12.A** with **Fig. 2.8.B-C**).

Elevated levels of miR-21 are detected in both liver and cervical cancer cell lines relative to their corresponding healthy source tissues[200,201]. To determine whether a higher abundance of miRNA was responsible for the observed increase in polysome binding, we measured miR-21 copy numbers per amount of total input RNA isolated from liver or HeLa cells. Coincidentally, both samples contained comparable levels of miR-21 at 4,000-5,000 copies per 10 pg input RNA (**Fig. 2.12.B**). We conclude that miR-21's distinct polysome binding profile in liver and HeLa cells is not due to differences in expression levels.

Next, we investigated whether this increased association with polysome-bound mRNA translated to a stronger repression profile. HeLa cells transfected with either anti-miR-21 or saline were array profiled and analyzed with the Sylamer algorithm (**Fig. 2.12.C**). Contrasting with our findings for liver miR-21, inhibition of HeLa miR-21 resulted in a measurable derepression of mRNAs enriched with 7- or 8-nt seed matches. In addition, many of the previously validated targets such as Pdcd4 were strongly derepressed in HeLa (**Fig. 2.12.D**), in sharp contrast to our results in liver. Interestingly,

few targets were mutually derepressed in both contexts (**Fig. 2.13**), suggesting there may be different determinants for miR-21 target selection between HeLa and mouse liver.

These findings support the conclusion that polysome complexes are important for miRNA-mediated repression in cancer cells, as increased repression by miR-21 is associated with increased binding to polysome-bound mRNA. Our HeLa cell results are thus consistent with previous observations[77-79,171], and further underscore the unusually limited RNA silencing activity of the abundant miR-21 in healthy liver.

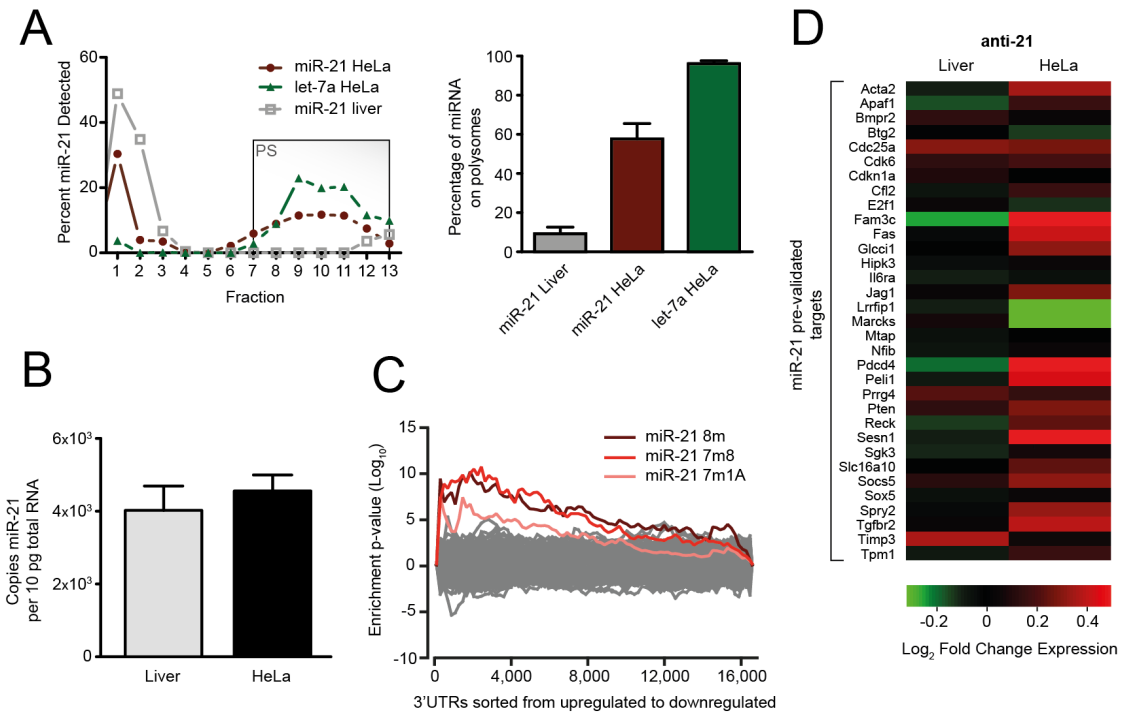


Figure 2.12 miR-21 is highly associated with polysomes and strongly represses a broad range of targets in HeLa cells. (**A**) Left, percent distribution profile of HeLa miR-21 (dark red circles), HeLa let-7a (dark green triangles), and liver miR-21 (grey open squares) in sucrose gradients loaded with mouse liver or HeLa extracts prepared under the same conditions. The grey box indicates the fractions containing polysomes (PS). Right, the mean summed percent total in polysome fractions 7-13 (highlighted with grey box) for each miRNA (colors are the same as in the left panel). Error bars represent SEM from triplicate experiments. (**B**) Absolute quantification of miR-21 copy numbers in liver and HeLa lysates normalized to total input RNA. Error bars represent SEM from N = 4 (HeLa) and N = 3 (liver) biological replicates assayed in two independent experiments. (**C**) Heptanucleotide Sylamer analysis of array profiling of HeLa cells transfected with anti-miR-21 (compared to saline mock transfection). The seed matches for each miRNA are highlighted. All other possible 7-nt sequences not related to the seed match are shown as grey lines and thus represent statistical background noise. The 8m seed match from the octanucleotide analysis is shown super-imposed. Heptanucleotide and octanucleotide analyses had similar backgrounds. The peaking of enrichment for miR-21 seed-matched sequences on the left hand side of the plot indicates that the genes most upregulated upon miRNA inhibition are enriched for the corresponding seed-matched sequence (compare with mmu liver in Fig. 1D, bottom left). (**D**) Heat map comparison of responses in mRNA levels for known targets from inhibition of miR-21 in mouse liver or HeLa.

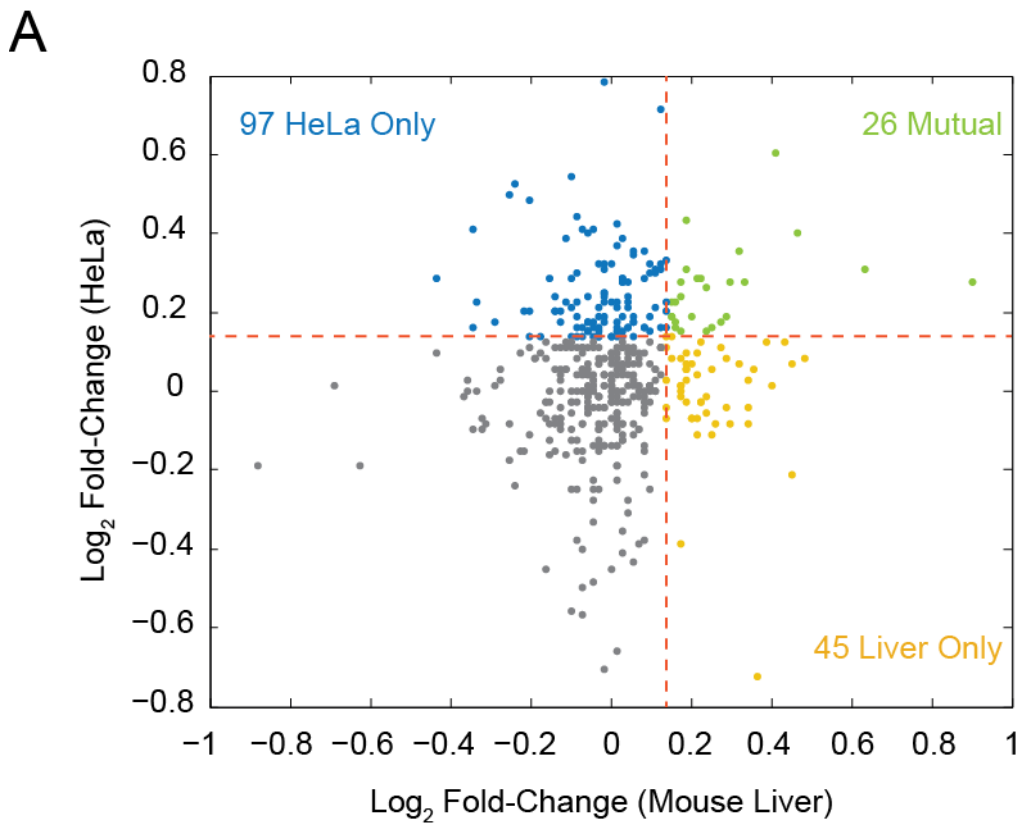


Figure 2.13 miR-21 represses different targets under different biological contexts. (**A**) Scatter plot of measured expression changes in miR-21 seed-matched transcripts following inhibition of miR-21 in HeLa (y-axis) versus mouse liver (x-axis). Transcripts with > 0.175 $\log_2$ fold-change expression (>10% increase on a linear scale) in inhibited versus saline treated samples were considered to be upregulated targets (i.e. derepressed due to miR-21 inhibition). The orange dotted lines represent these thresholds set for each dataset and they divide the plot into four quadrants: transcripts upregulated in HeLa only (top left, colored blue), transcripts mutually upregulated in both HeLa and mouse liver (top right, green), transcripts upregulated in liver only (bottom right, yellow), and transcripts not upregulated in either context (bottom left, grey). The number of transcripts in each quadrant is shown.

2.4 Discussion

In contrast to its oncogenic prowess in cancer cells, miR-21 is a surprisingly weak RNA silencer in healthy liver tissue

In the current study, we have examined the *in vivo* activity of miRNAs in healthy mouse liver tissue using array profiling and polysome analysis. Although testing of additional tissues and cell lines is desirable, our results strongly suggest that miR-21, relative to other miRNAs, is functionally limited under normal physiological conditions, yet not in cultured cancer cells. Evidence for this conclusion is provided by the following observations: (a) Pharmacological inhibition or knockout of miR-21 does not result in an upregulation of mRNA targets significantly enriched for seed sequence matches; (b) miR-21 exhibits reduced binding to translating target mRNAs, which represents the largest population of an mRNA in the cell; (c) the sucrose gradient distribution of miR-21 alone resembles that of the anti-miR inhibited miRNA, suggesting that miR-21 is impeded from interacting with polysome-bound mRNA equally in the absence as in the presence of its anti-miR. In contrast, by analyzing the global upregulation of mRNAs following miRNA inhibition we were able to observe trends for miR-122 and let-7 consistent with canonical miRNA-target prediction rules[82]. That miR-21 displays these trends so minimally in liver challenges the current assumption that the rules of engagement of highly expressed miRNAs are the same across different tissues and cell types. These results may extend to other healthy tissues as well, since previous studies have hinted that miR-21 is also surprisingly underactive in lung and heart in the absence of applied stress[112,138].

May idiosyncrasies in miR-21's mechanism of silencing explain our results? That miR-21 represses gene output by inhibiting translation directly without causing changes to mRNA levels cannot be unequivocally dismissed; however, we found little evidence to suggest an upregulation of translational capacity in the absence of miRNA-mediated repression. Moreover, a translation-based mechanism compared to a degradation-based mechanism arguably would necessitate an even more pronounced interaction of miRNA with translating complexes, which we found to be lacking.

miR-21's meager repression signature in healthy cells is surprising considering its reported oncogenic prowess, capable of inducing the formation of tumors that are dependent solely on its silencing activities[144]. Considering these apparently contradictory activities, our results suggest that miR-21 becomes "activated" during disease or other aberrant states, broadening both its repression footprint and its impact on the cellular state. In line with this conclusion, we found a large number of seed-matched targets to be derepressed upon inhibition of miR-21 in adenocarcinoma HeLa cells. Many of these affected transcripts have been found to be the downstream targets, underpinning miR-21's pathological contribution in tumor modulation[136], and similar derepression of targets has been reported upon inhibition of miR-21 in MCF-7 breast cancer cells[103]. Correspondingly, we observed starkly different sub-cellular locations of miR-21 under the two biological contexts: miR-21 in mouse liver predominantly sediments in light fractions apart from mRNAs, while in HeLa lysates miR-21 predominantly sediments in polysomal fractions with mRNA. Co-sedimentation of miRNA/mRNA was not coincidental, as it was dependent on miRNA base pairing and was sensitive to small-molecule translational inhibitors. Therefore, we posit that miR-21's increased repression proficiency in HeLa is at least partly due to enhanced binding to its targets. However, the mechanism(s) underlying this enhancement in HeLa cells — or alternatively, the diminished activity in liver tissue — are still unclear. Considering that let-7a also shows additional polysomal binding in HeLa cells suggests that this enhanced target loading capacity may be due to an increased abundance of a common protein factor like the receptor for activated C-kinase (Rack1), a 40S ribosomal protein and essential component for miRNA-mediated repression involved in loading miRISC onto translating mRNAs[49]. Proteomics analysis of miRNA-containing complexes first separated by sucrose gradient fractionation as done here could lend further insight into which protein complexes are less capable of assembling with miR-21.

What may cause miR-21's limited RNA silencing activity in healthy liver?

It is conceivable that miR-21's reduced binding and repression in liver tissue are attributable to thermodynamics. Indeed, computational analysis predicted miR-21 to bind targets less stably than other miRNAs largely due to high A/U richness in the seed region.

Binding assays showed that miR-21 binds a complementary 2'O-methylated RNA as effectively as does miR-122, however, these assays merely demonstrate miR-21's potential to bind target RNAs. The thermodynamics of miRNA:mRNA binding are not reflected due to the unusually high extent of base pairing and the 2'O-methyl modifications of the complementary RNA, which both raise the free energy of duplex binding[202]. Interestingly, miR-23 and lsy-6, miRNAs also with A/U rich seeds, have similarly been reported to have limited repressive activity[191]. However, unlike miR-21, these miRNAs additionally have high target abundance that further dilutes their effects on the transcriptome. Although experimental work demonstrated that these properties — weak seed pairing and high target abundance — can independently contribute to diminished proficiency, it is uncertain whether weak seed pairing alone could explain our observations for liver miR-21. The current "Context+" predictive model, which takes these properties into account, suggests not since the top predicted targets of miR-21 were scored equally high as those of miR-122, indicating that miR-21 is still predicted to repress targets similarly to miR-122, despite its weaker seed pairing. It should be noted, however, that the Context+ algorithm and its "Context only" predecessor were both trained with array profiling experiments from HeLa cell culture. Accordingly, and consistent with our own HeLa cell data, these algorithms score miR-21 targets highly. These considerations lead us to propose a model wherein miR-21 in liver tissue, despite its abundance, is expressed at a level below the threshold required to bind and repress a wide range of targets in this cellular environment. Interestingly, the same expression level is sufficient to elicit target repression in HeLa, which indicates that these cells are generally more conducive to miR-21 activity and perhaps that of other miRNAs as well. This favorable environment for miRNA-mediated repression may be due to increased availability of pathway components, alterations in the transcriptome landscape, or changes in the activity of other regulatory proteins (e.g., A/U-rich element RNA-binding proteins) that may compete with miRNAs for target binding. However, further investigation will be required to determine whether an increased abundance of miR-21, which frequently occurs under pathological conditions like liver and cervical cancer[201,203], is alone sufficient to activate target repression. This can be tested in the

future by experiments similar to those shown here, but using transgenic mice overexpressing liver miR-21.

Additionally or alternatively, miR-21's target abundance may be underestimated. Although the number of different transcripts containing miR-21 target sites is less than those for miR-122 or let-7, the abundances of these transcripts were not factored into our calculation. It is possible that miR-21 targets are expressed to higher levels, causing miR-21's effect on any one gene to be weakened. Yet, such a notion is inconsistent with the observed dominant localization of miR-21 in non-polysomal fractions, as increased target abundance would be expected to lead to greater relative miR-21 fractionation into polysomes due to increased mRNA binding.

Another possibility is that only specific abundant, yet translationally inactive transcripts play a role in miR-21's low silencing activity in liver. An increasing amount of recent research has concluded that transcripts of pseudogenes, evolutionarily miscopied gene duplicates that do not code for functional proteins, can strongly influence miRNA-mediated gene regulation by sequestering miRNAs away from real targets[204-207]. For example, the pseudogene PTENP1 can indirectly control expression of its ancestral gene PTEN by "sponging up" miRNAs that bind both transcripts; in this way, increases in the pseudogene's non-coding RNA levels cause increases in the ancestral gene's mRNA levels and vice versa. Since pseudogenes often contain non-sense mutations near the start of the open reading frame, these transcripts would be expected to largely reside in non-polysomal fractions, possibly explaining miR-21's localization if it were bound to a pseudogene(s)[207,208]. Intriguingly, miR-21 is presumed to bind PTENP1 in human cells[208]. Although a Ptenp1 mouse homolog has not been discovered to our knowledge, there remains the possibility that other yet-to-be identified miR-21 binding pseudotranscripts exist. The high sequence similarity between pseudogenes and their ancestral genes complicates a characterization of a large fraction of the pseudo-transcriptome by conventional means (i.e., by RT-qPCR and microarrays)[208,209]. The advent of deep-sequencing technology should rapidly enable the measurement of the presence and prevalence of pseudo-transcripts, and contribute to our understanding of how this endogenous network of miRNA-sponges affects miRNAs such as miR-21.

miR-21 regulation is specifically linked to stress response pathways

We found a select group of seed-matched targets to be upregulated upon miR-21 inhibition. An interesting question is why these transcripts respond while most do not. Importantly, we found stress response to be a common biological pathway among up-regulated genes both with and without miR-21 seed matches. Weakly binding miRNAs like miR-21 may be particularly suited to regulate response pathways, as they would be more sensitive to subtle transcriptomic fluctuations or other changes in their targets, allowing for a temporal control of miRNA-mediated repression, which in turn may be responsive to extracellular cues. We also found that miR-21 itself is up-regulated following anti-miR-21 treatment, suggesting the presence of a feedback loop that may act to contain a stress activation event by reestablishing the balance between miR-21 and its targets. Such a feedback loop may explain why miR-21 is overexpressed in chronically stressed, abnormal cell states such as cancer and cardiac disease[136]. We hypothesize that miR-21 exerts its influence on the activation of heat stress response cascades via Taf7, which has been shown to interact with HSF-activation complexes[195]. Additionally, Taf7 may act as a node for miR-21's integration into additional stress response pathways, as increasing amounts of Taf7 in HEK293 and COS cells correlate with enhanced transactivation of c-Jun, an AP-1 transcription factor activated by stress-activated protein kinases (SAPKs)[210-212]. In the future, it will be important to further test for links between miR-21 and stress to determine which stresses and stress-related pathways miR-21 may respond to in various tissues.

Finally, we note that we cannot entirely rule out the possibility that some or all of the seed-matched transcripts, despite containing miR-21 binding sites, are up-regulated as the result of secondary effects of miR-21 inhibition, not because they were directly repressed by miR-21. The current standard for validating miRNA:target interactions are reporter gene assays where luciferase constructs are engineered with the 3'UTRs of interest and expressed in cell lines co-expressing the targeting miRNA; selectivity of the interaction is demonstrated if mutation of the seed sequence results in increased luciferase output. However, our findings underscore that assays in cell culture may not be appropriate considering the drastically different miR-21 activities observed between liver

tissue and cultured HeLa cells. A better solution in general would therefore be to validate target specificity in the same biological context used for the original observation. For example, reporter gene assays mediated by in vitro translation have been previously used to study miRNA activity in Krebs-2 excites extract[46] and rabbit reticulocyte lysate[71]. Translation competent liver extracts have been reported[213]; thus, further validation using this type of system may be possible for miR-21 in the future.

In conclusion, we find unexpectedly narrow target specificity in healthy liver compared to cancer cells for one of the most abundant, disease-linked miRNAs, demonstrating the importance of biological context in RNA silencing. Future studies will need to test additional tissues and cell lines to confirm the generality of our findings as well as to address the mechanism(s) behind miR-21's regulatory activation in cancer cells in more depth. Focused approaches to compare miR-21 activity in normal and diseased tissue[138,144], where miR-21 is co-overexpressed with additional oncogenic or other disease promoting factors, may be particularly powerful to define the conditions required for miR-21 activation.

2.5 Acknowledgements

I would like to especially thank our collaborators from Regulus Therapeutics: B. Nelson Chau, who performed the mouse liver and HeLa array profiling experiments, and Peter Linsley, who originally suggested the sucrose gradient fractionation experiments. This work would not be possible without their contributions. We also thank S. Orkin and Jochen Hartner (Harvard Medical School) for the generous gift of miR-21 knockout mice, J. Lee and P. Tran for their technical assistance, S. Pitchiaya, E. Marcusson, A. Chang, and D. Hogan for their help and thoughtful insights, and members of Regulus R&D team and the Walter group for valuable discussions.

Chapter 3: Bioinformatic Analysis of the Contextual Features Influencing miRNA-21 Target Selection[1]

3.1 Introduction

In addition to protein coding, a messenger RNA encodes ancillary information on where it should be translated[214], to what levels[215], and under what conditions[69]. This information is encoded in recognition elements (REs)— short sequence motifs, often found in the 3' untranslated region (UTR), that are recognized by miRNA or RNA binding proteins (RBPs). A major challenge in the RNA field is determining how REs impart specificity with only little sequence information. The miRNA-RE — the seed site — is just 7 or 8 nucleotides (nts) long; perhaps not coincidentally, other RBP-REs are commonly of similar length. Given that the average 3'UTR contains ~700 nts, a single miRNA or RBP can theoretically bind and repress one out of approximately every 23 genes by chance alone. Estimating that ~20,000 protein coding genes are transcribed, each factor is then expected to regulate >800 genes. Experimental measurements, however, indicate that for miRNA only a minor portion of all possible targets are actually regulated. Array profiling of HeLa cells after miRNA transfection showed measurable changes in less than 25% and 50% of transcripts containing 7-mer and 8-mer sites, respectively[82]. Proteomic profiling using quantitative mass-spectrometry showed similar responses (<40%), indicating a comparable scope of regulation exists at both the protein and transcript levels[53]. From these data, it becomes apparent that there is more to REs than the short sequences themselves: additional determinants must make recognition elements actually recognizable.

Several miRNA target prediction methods are available to help experimental scientists narrow down large candidate pools of miRNA targets to sets that are more workable. The concept behind these tools is that the probability of a miRNA interaction

[1] The array profiling data analyzed in this chapter were obtained by Nelson Chau. All analyses were designed and performed by John Androsavich.

can be calculated based on mRNA features surrounding seed sites that have been experimentally determined to generally affect target repression. Contextual features can include thermodynamic parameters such as calculated energy of miRNA binding and extent of base pairing to the 3' end of the miRNA; accessibility of seed sites as estimated by local mRNA secondary structure or A/U richness; or other less intuitive features, such as relative location of the seed site. Together these features help to narrow down the list, but they often do not provide enough sensitivity or specificity. For this reason, most prediction measurements include a heavy weighting for site conservation across species in order to reduce noise. With the assumption that conserved sites are biologically relevant, this helps to account for the as-yet undiscovered features that could affect targeting. While unique combinations of these features and various scoring schemes have been implemented, independent experimental validations of these methods indicate that overall their performances are moderate at best. Depending on the prediction method and available experimental data, the rates of precision (correctly predicted/total predicted) can range from 24 − 51%, while the rates of sensitivity (correctly predicted/total correct) can range from 6 − 48%[86,216].

In **Chapter 2**, we observed that miR-21 in liver regulates an unexpectedly narrow portion of all possible targets: of 1,583 total targets, only 140 or ~8% showed significant derepression upon inhibition of miR-21. Surprisingly, only 8 of these (~6%) were considered highly probable targets across seven computational prediction algorithms, and only one was ranked twice, indicating that the algorithms did not converge on these targets. Therefore, it seems that even the minimal activity displayed by miR-21 does not follow the "rules" of target selection; or at least the same weighting of these rules as they apply to other miRNAs. As a result, in order to understand miR-21's proclivities, it may first be necessary to re-evaluate the known determinants of targeting to determine which features distinguish the select few responsive targets.

In the current **Chapter 3**, we systematically analyzed miR-21 array profiling data in healthy mouse liver for trends in both well-established and lesser-known sequence-based contextual features that may influence target selection. We found that miR-21 seemingly does not discriminate targets based on features that have been recognized as general determinants of miRNA-mRNA interactions. Instead, on average, miR-21

responsive targets have greater levels of expression and slightly shorter open reading frames (ORFs) than non-responsive targets, two features not taken into account by current prediction methods. Moreover, all miR-21 targets are highly enriched with RBP-REs, suggesting that crosstalk with RBP networks may explain miR-21's target selection. For comparison, we also analyzed miR-122 inhibition responses, which turned out to overall correspond better with predictions. For miR-122, two favorably scored features—multiple seed sites and seed sites proximal to ORFs — were detected to be preferential. Surprisingly, however, neither miR-21 nor miR-122 responses showed discernible trends relating to thermodynamics of target binding or site accessibility. We conclude that in addition to displaying a diminished repression capacity, liver miR-21 selects its few targets based on different features than other miRNAs, which have yet to be discovered. We also discuss potential divergences in studying miRNA targeting using anti-miR-induced inhibition compared to other methods of perturbation.

3.2 Materials and Methods

Many of the scripts and functions for running these analyses are provided in **Appendix A.**

Array Profiling Data. Array profiling data was the same as used in **Chapter 2**.

Comments on downregulated responses. Numerous seed-matched transcripts were significantly downregulated after miR-122 or miR-21 inhibition, suggesting that these miRNAs could have an activating effect on these genes. While Steitz and colleagues have found that miRNAs can enhance protein output of certain genes during cell cycle arrest[217], we found no evidence that this is a widespread occurrence for the studied miRNAs in liver tissue. If it were, we would expect enrichment of seed-sequences in the most downregulated genes with Sylamer analysis, which was not observed (**Fig 2.2.D**). We therefore attribute the majority of these changes to secondary effects. We binned these genes separately since they could otherwise be highly efficient targets if not for the stronger secondary effects on their expression levels.

Expression/Conservation Filtering. Expression level cutoffs were based on $\log_2$ transformed normalized labeled-cRNA array hybridization intensities from single-array profiling liver RNA samples of mouse C57BL/6. This expression filter eliminated genes with intensities less than the median and those not matched to anti-miR array profiling data. Target site conservation was determined by comparing mouse transcripts with human transcripts. A mouse target site was considered conserved if its human homolog, matched based on gene symbol, also contained a target site for the same miRNA.

Extension Scoring. Seed site sequences plus thirteen additional upstream nucleotides were extracted from 3'UTRs and aligned with the reverse complement of the miRNA sequence using Needleman-Wunsch global sequence alignment. Starting from the opposite end and up to the seed sequence, 1 point was scored for each base match and 0.5 points for each consecutive match with no intervening gap.

Statistical Analysis. Differences in response bins were tested for statistical significance using two-way ANOVA with Bonferroni post-tests using Prism Software 4.0 (GraphPad). Pearson correlation coefficients were calculated using Origin Software 8.6 (OriginLab). Kolmogorov-Smirnov tests were calculated using MatLab Software 2011a (Mathworks). Linear regression analysis was calculated using OriginPro 7.0 (OriginLab). Permutation tests were performed by searching for motifs in randomly scrambled sequences with the same base composition as the test sequence[218]. This process was repeated for n = 1000 and n = 500 rounds for 200 nt and full-length UTR sequences, respectively. p-values were calculated as $p = n(k) / n$, where n(k) is the number of rounds where the motif was found in the scrambled sequence more often than it was found in the test sequence, and n is the total number of rounds. The test sequence was considered enriched with a motif if $p < 0.05$.

3.3 Results

Sequence Database Development

In order to detect trends and compare sequence characteristics across different genes, we first needed to build a searchable local database of annotated information for each gene measured in our array profiling experiments. From the outset, we had in mind the following requirements for the database: flexibility to easily extract relevant information for hypothesis testing; adaptability to update or alter the database with additional experimental data or updated annotation releases; and compatibility both in terms of operating system independence and file I/O.

We chose to build our database using the Matlab (MathWorks, Natick, MA) bioinformatics toolkit as it met these requirements while still being easily accessible to bench scientists with only moderate coding experience. A schematic of the algorithm employed for database development and filtering is shown in **Figure 3.1**. In addition, many of the MatLab codes used for database development and sequence analysis are found in **Appendix A**.

Our local database (referred to as the Anti-miR Array Profiling (AMAP) database) was populated with curated mRNA (accession prefix 'NM_') entries from the National Center for Biotechnology (NCBI) Reference Sequence (RefSeq) Database. Although excluded from the analyses here due to their uncertainty, non-coding RNA ('NR_'), predicted mRNA ('XM_'), and predicted non-coding RNA ('XR_') were also added to the local database for future studies. In total, the AMAP database consisted of ~ $3.5x10^4$ accession files that were then clustered into $1.8x10^4$ groups – each group contained accession files for transcripts cross-detected by the same probe sets in the array profiling. For sequence analysis, we extracted 5'UTR, CDS, and 3'UTR sequences from each accession group. If an accession group contained >1 mRNA isoform then we gave priority to the isoform sequence that (1) contained the miRNA seed-match of that particular analysis and/or (2) had the longest 3'UTR.

Our aim was to determine whether unique sequence properties exist in transcripts found to have increased expression following inhibition of miR-21 (referred to as

responsive or upregulated transcripts). In addition to more general features like transcript length and estimated abundance, we searched for as many properties known to affect miRNA-mRNA interactions as possible. These characteristics included length, A/U-richness, non-canonical miRNA sites, number of seed sites, seed locations, 3' base pairing, RBP motifs, and expression level.

To reduce the influence of sequence background noise, our datasets were filtered based on measured expression levels and/or conservation of seed-sequence (**Fig. 3.1A**). Expression filtering was meant to reduce the number of false negatives by eliminating from the analysis genes not expressed in liver that might otherwise be effective targets; whereas conservation filtering, assuming that conserved targets are more likely to be regulated, was intended to reduce the number of false positives (i.e., genes containing a seed site that were upregulated through secondary effects). These filtered datasets were analyzed in parallel to the non-filtered datasets. The results of filtering on the number of seed-matched genes are shown in **Figure 3.1.B**.

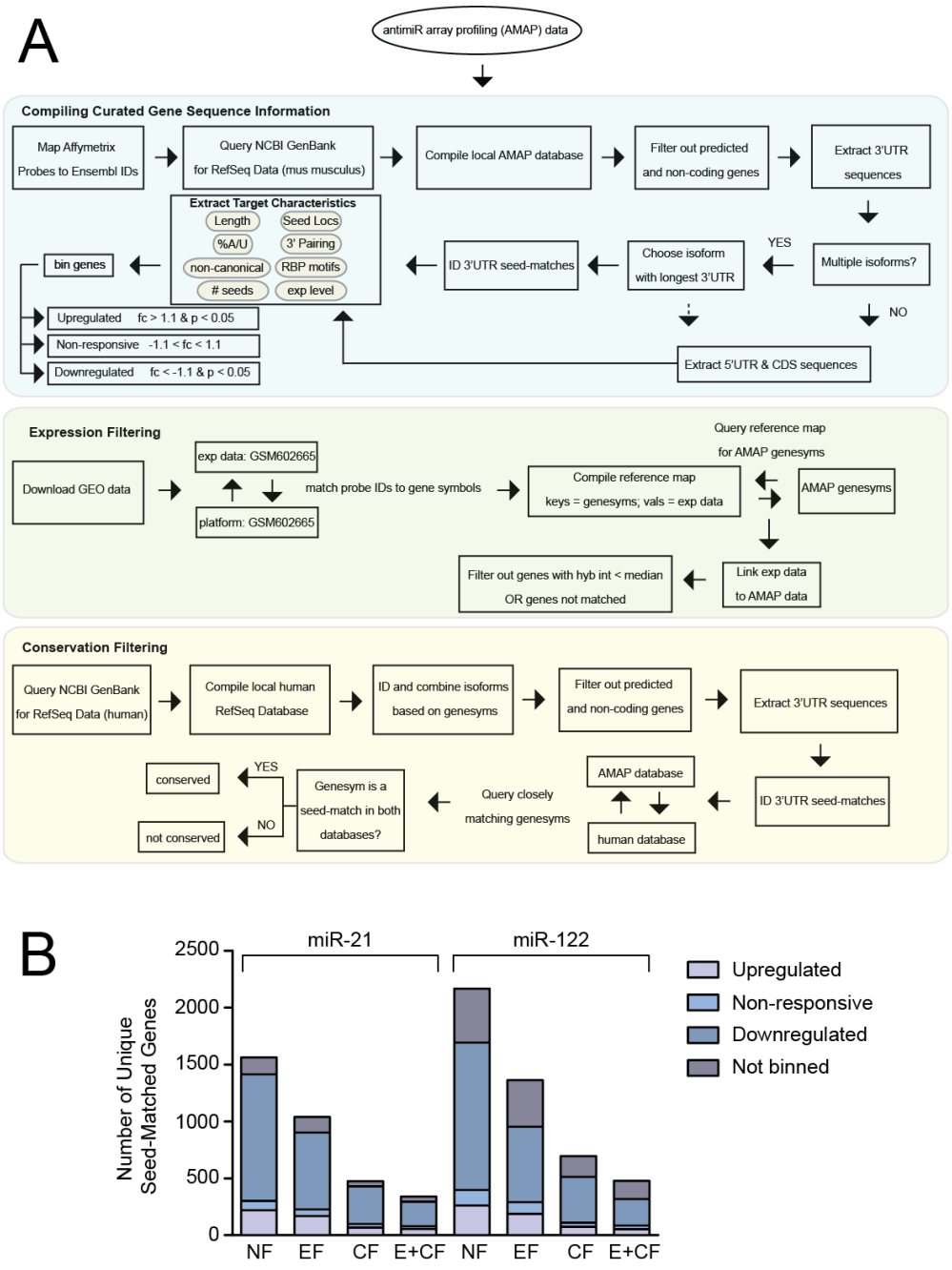


Figure 3.1 Bioinformatics flowchart and filtering results. (**A**) Schematic of pipeline built for local bioinformatics database. *Top,* Curated NCBI Ref-Seq gene/transcript information was collected for each gene measured in our anti-miR array profiling (AMAP) analysis. Seed-matched transcripts for each inhibited miRNA were identified based on the presence of 7m1A, 7m8, or 8m seed complementary sites in 3'UTRs. Each seed-matched transcript was characterized based on a number of properties including length, percent A/U base-composition, presence of non-canonical miRNA binding sites (non-canonical), number of miRNA seed sites (# seeds), miRNA seed site locations (seed locs), RNA-binding protein (RBP) motifs, and expression (exp) level. For the analyses described in this chapter, seed-matched transcripts were divided into 'upregulated', 'non-

responsive', and 'downregulated' bins and the characteristics of each bin were compared to assess miRNA-targeting preferences. *Middle, Bottom* To reduce the number of false-positives and false-negatives, datasets were filtered based on two criteria: expected expression levels in mouse liver (based on array profiling data, see Materials and Methods) and conservation in human homolog transcripts. (**B**) Results of filtering on the number of unique seed-matched genes for miR-21 and miR-122. The number of non-responsive genes, but not the number of upregulated genes, is reduced between the non-filtered (NF) and expression filtered (EF) gene sets for each miRNA. Application of the conserved filter (CF) alone, on the other hand, reduces the number of genes in each bin category. Use of both the expression filter and conservation filter (EF+CF) together gives the most specific gene set, and for miR-122, the largest ratio of #upregulated/#non-responsive genes.

Testing the influence of expression levels on target selection

We first examined whether responsive targets were more or less expressed than non-responsive targets. For seed-matched transcripts, we calculated the Pearson correlations for relative expression levels, as measured by the intensities from the microarray data used for expression filtering, and plotted them versus fold-change in expression. We found a significant, if small positive correlation for both miR-21 ($r = 0.091$, $p = 0.162 \times 10^{-3}$) and miR-122 ($r = 0.123$, $p = 6.25 \times 10^{-7}$) transcripts (**Fig. 3.2.A**). Similar results were obtained with the conserved datasets (miR-21-C: $r = 0.118$, $p = 0.0262$; miR-122-C: $r = 0.095$, $p = 0.032$). In order to gain a more intuitive perspective on these data, we binned genes based on their response to anti-miR-induced miRNA inhibition and calculated the average expression levels for each bin. Consistent with the observed correlations, mean expression levels of upregulated genes were significantly greater than non-responsive genes (**Fig. 3.2.B**). There were no significant differences between up and downregulated genes. The linear differences in mean hybridization intensities between upregulated and non-responsive genes were ~1.6-fold and ~1.8-fold for miR-21 and miR-122 targets, respectively. We consider these differences to be estimates rather than absolutes because microarray signal intensities can be biased towards probes with higher affinities[219]. However, use of processing methods that account for these biases should at least minimize these effects. Therefore, for both miRNAs analyzed we conclude that, on average, upregulated targets are slightly more highly expressed than their non-responsive counterparts.

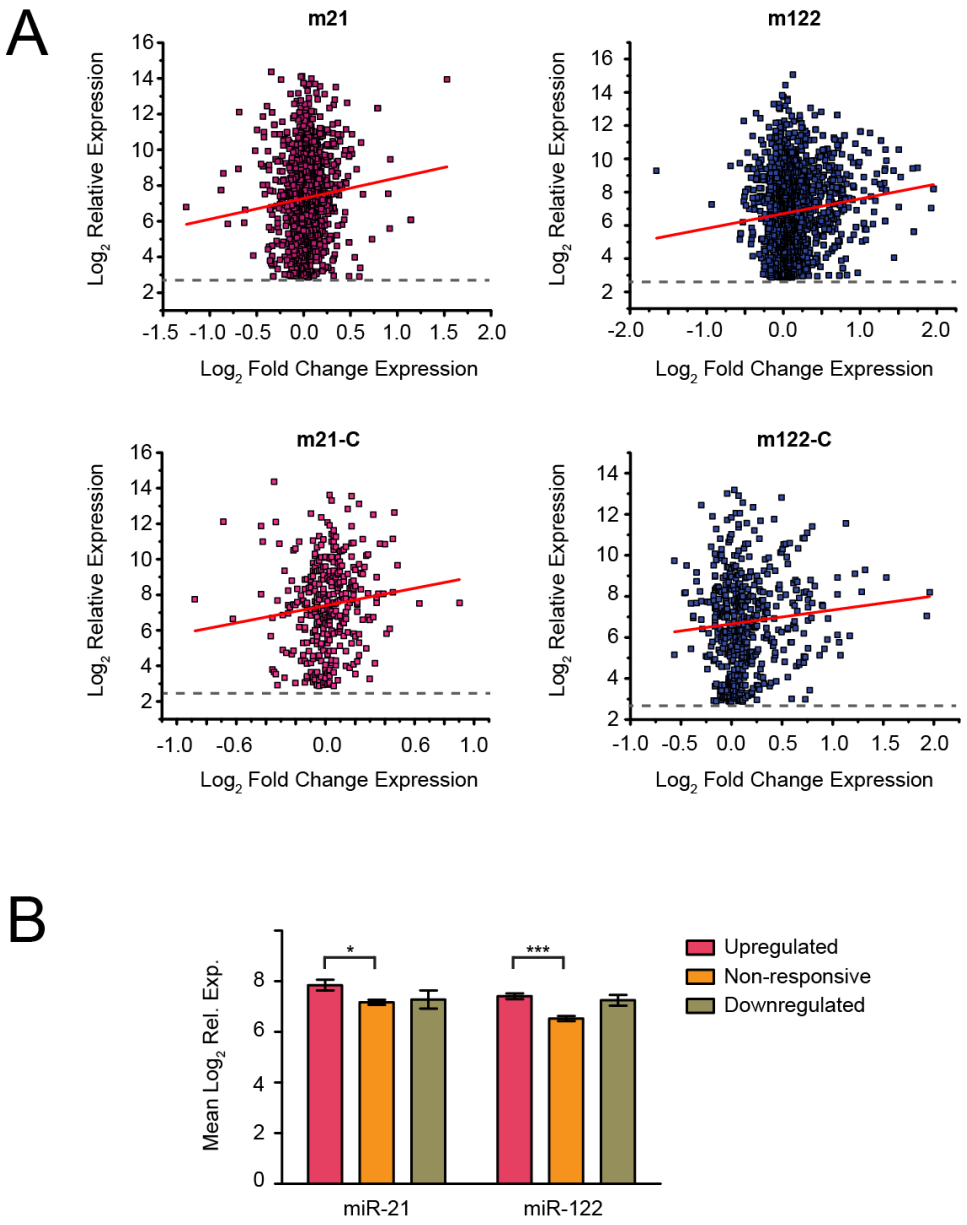


Figure 3.2 Influence of expression levels on target selection. (**A**) Scatter plots of seed-matched transcript expression levels for miR-21 (m21) and miR-122 (m122) expressed and expressed/conserved (m21-C; m122-C) gene sets. The grey dashed horizontal line represents the minimum cutoff for genes considered expressed (see **Materials and Methods**). (**B**) Mean relative expression levels of expressed miR-21 and miR-122 seed-matched genes binned based on their changes in expression following miR-21 or miR-122 inhibition, respectively. Error bars represent SEM. (*) = $p < 0.05$; (**) = $p < 0.001$, determined by two-way ANOVA with Bonferroni post-test.

Responsive messages tend to have shorter ORFs

Next, we tested the relationship between inhibition response and target length. We found 3'UTRs of targets upregulated upon miR-122 inhibition to be on average ~600 nt shorter than non-responsive targets, a significant deviation (**Fig. 3.3.B,** top). Although miR-21 showed a similar pattern, it was not statistically significant. We also found that this trend extended to the open reading frame (ORF) as miR-21, miR-21 conserved, and miR-122 upregulated target sets showed statistically shorter protein coding sequences than their non-responsive counterparts by a range of ~ 400 − 700 nt or ~ 130 − 230 codons (**Fig. 3.3.B,** middle). Less significant differences were observed for 5'UTRs (**Fig. 3.3.A,** bottom).

Seed site position in the 3'UTR has been shown to affect repression [82]. For miR-122, seed sites tended to be distributed closer to the end of the ORF in upregulated genes than non-responsive or down-regulated genes (**Fig. 3.3.B,** right). However, this trend was not observed in miR-21 targets (**Fig. 3.3.B,** left**)**. To determine whether 3'UTR length and seed site position were interrelated we normalized each absolute site position by the length of its 3'UTR. These relative positions were then binned, compiled into histograms, and plotted as trend lines for upregulated, non-responsive, and down-regulated targets (**Fig 3.3.C**). Seed sites were fairly uniformly distributed along the lengths of 3'UTRs with a maximum inter-bin range of ~10% (**Fig. 3.3.C**). However, upregulated sets for miR-21, miR-122, and miR-122 conserved showed slight bias towards the beginning of the 3'UTR, and also the end for miR-21, which showed a hyperbolic trend (**Fig. 3.3.C,** green line). This agrees with previous data indicating that seed sites at the beginning and near the end of the 3'UTR are most effective[82]. Non-responsive and down-regulated gene sets did not show dramatically different distributions, except in certain cases where a greater fraction appeared to be distributed near the middle and end of the 3'UTR as opposed to the beginning (**Fig. 3.3.C,** orange and purple lines). Surprisingly, upregulated miR-21 conserved targets appeared to resemble the non-responsive trends more so than they did the other upregulated datasets (**Fig. 3.3.C,** bottom left). This finding may be due to the limited number of targets classified in this dataset, resulting in poorer statistical sampling.

We next counted the number of targets in each dataset containing more than one seed site, a well-documented feature of robust targets[82,173,220]. Between miR-122 and miR-21, there were largely different responses measured in multi-seed targets. For miR-21, only 15 (~14%) multi-seed targets were responsive (**Fig. 3.3.D**), in-line with liver miR-21's narrow repression signature. Surprisingly though, these targets accounted for < 11% of the upregulated genes in the expressed dataset (**Fig. 3.3.E**), far less than expected if multiple seeds were a major influence on miR-21 target repression. In contrast, 93 (~49%) miR-122 multi-seed targets were responsive (**Fig. 3.3.D**), which accounted for ~20% of the upregulated genes (**Fig. 3.3.E**). Similar results were found with conserved datasets for each miRNA.

We conclude that miR-122 upregulated targets on average have shorter ORFs and 3'UTRs, seed sites closer to the ORF— due to both relative positioning and the shorter 3'UTR— and nearly 20% contain multiple seed sites. For miR-21, except for ORF length, these trends were not as clearly discernible.

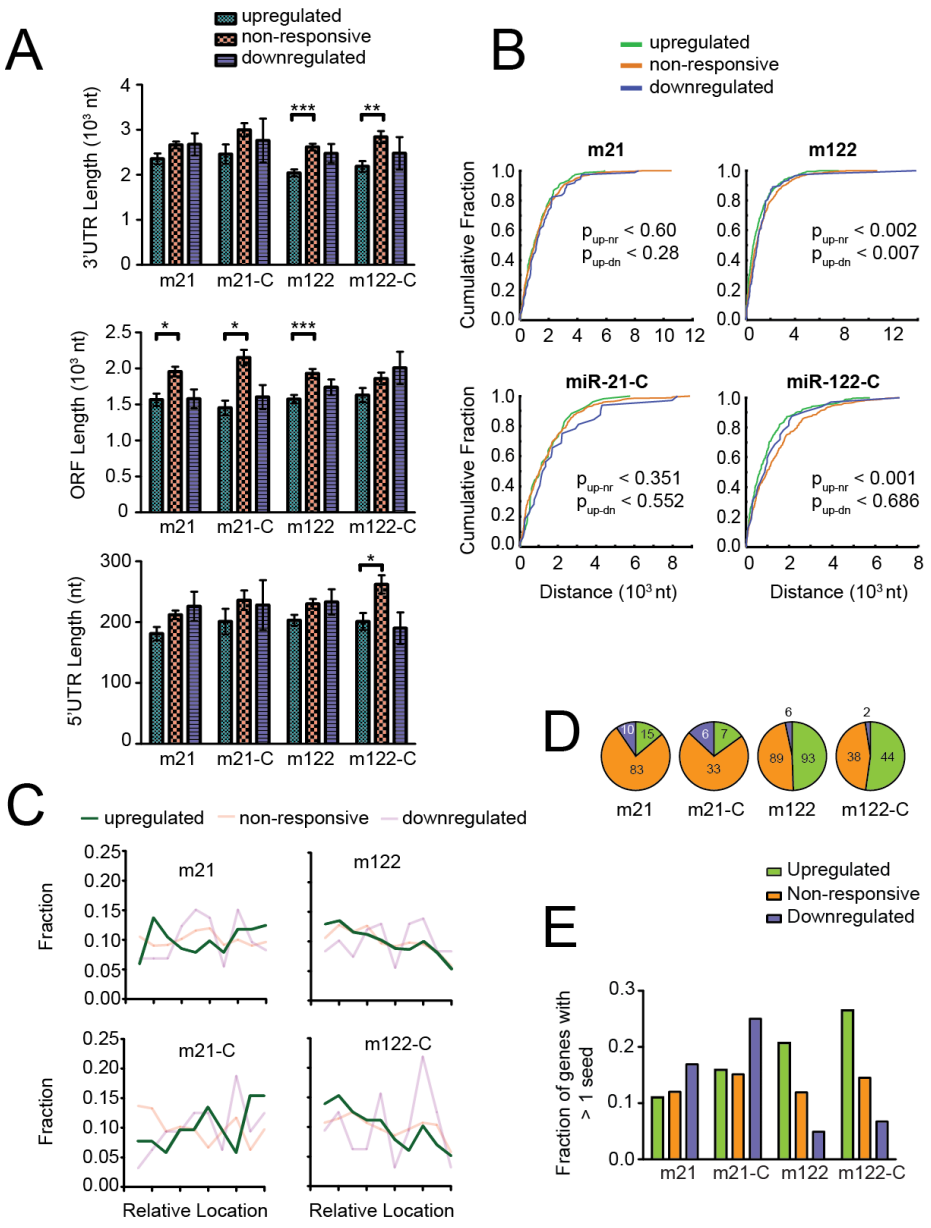


Figure 3.3 Influence of mRNA length, seed location, and number of seeds on target selection. (**A**) Mean 3'UTR (top), ORF (middle), and 5'UTR (bottom) lengths of binned expressed seed-matched transcripts for miR-21 (m21) and miR-122 (m122). Only expressed targets were included. Results from expressed, conserved transcripts for miR-21 (m21-C) and miR-122 (m122-C) are also shown. (*) = $p < 0.05$; (**) = $p < 0.01$; (***) = $p < 0.001$, determined by two-way ANOVA with Bonferroni post-test. (**B**) Cumulative distribution frequency plots of seed distances (in nucleotides) from ORFs. The significances of differences in distributions between upregulated-and-non-responsive and up-regulated-and-down-regulated gene bins are shown on each plot as Kolmogorov-Smirnov p-values (KS p). (**C**) Histograms of relative seed locations calculated as [seed position (nt) / UTR length (nt]. For clarity, the tops of each bar are plotted as a point in a connecting line rather than the traditional histogram bar plot. (**D**) Fraction of genes in each bin containing more than one corresponding miRNA seed site.

Responsive sites are not more A/U rich

It is now well established that target secondary structure in and around seed sites can dramatically impact miRNA-mediated regulation[85,221-223]. Because miRISC does not appear to be capable of actively unwinding target structure[224], there is a high enthalpic cost for a miRNA to access a highly structured seed site[85]. In lieu of secondary structure predictions, regions with greater A/U content should be thermodynamically less costly to disrupt compared to G/C rich regions due to the, on average, fewer hydrogen bonds formed between Watson-Crick base pairs of their secondary structure. Consequently, more effective miRNA sites have been reported to be found in regions with higher A/U base content[82]. In addition, A/U-rich mRNAs may be especially susceptive to miRNA-mediated degradation[223].

In our experiments, we did not find a significant difference in the base compositions of full length 3'UTRs between responsive and non-responsive or downregulated targets (**Fig. 3.4.A**). However, for certain upregulated targets we did notice that the seed sites occurred in local A/U-rich regions, irrespective of the overall full-length composition (**Fig. 3.4.B**). In order to see if this was a general characteristic of upregulated targets, we calculated local base compositions for regions surrounding the seed sites, varying the window size centered on the seed site from 20 nt to 200 nt (**Fig. 3.4.C**). Across these windows, miR-21 targets were slightly more A/U-rich compared to the full-length 3'UTRs. Surprisingly, miR-122 targets were more G/C-rich compared to full-length 3'UTRs. A/U content showed little change across the range of window sizes, indicating that the local base compositions around the seeds did not change at a uniform cutoff distance. Importantly, upregulated targets were not significantly more A/U rich: Upregulated targets showed a slightly higher A/U content in the miR-21 and miR-122 conserved datasets; however, these differences were statistically insignificant. We conclude that A/U base composition is not a distinguishing feature of responsive targets.

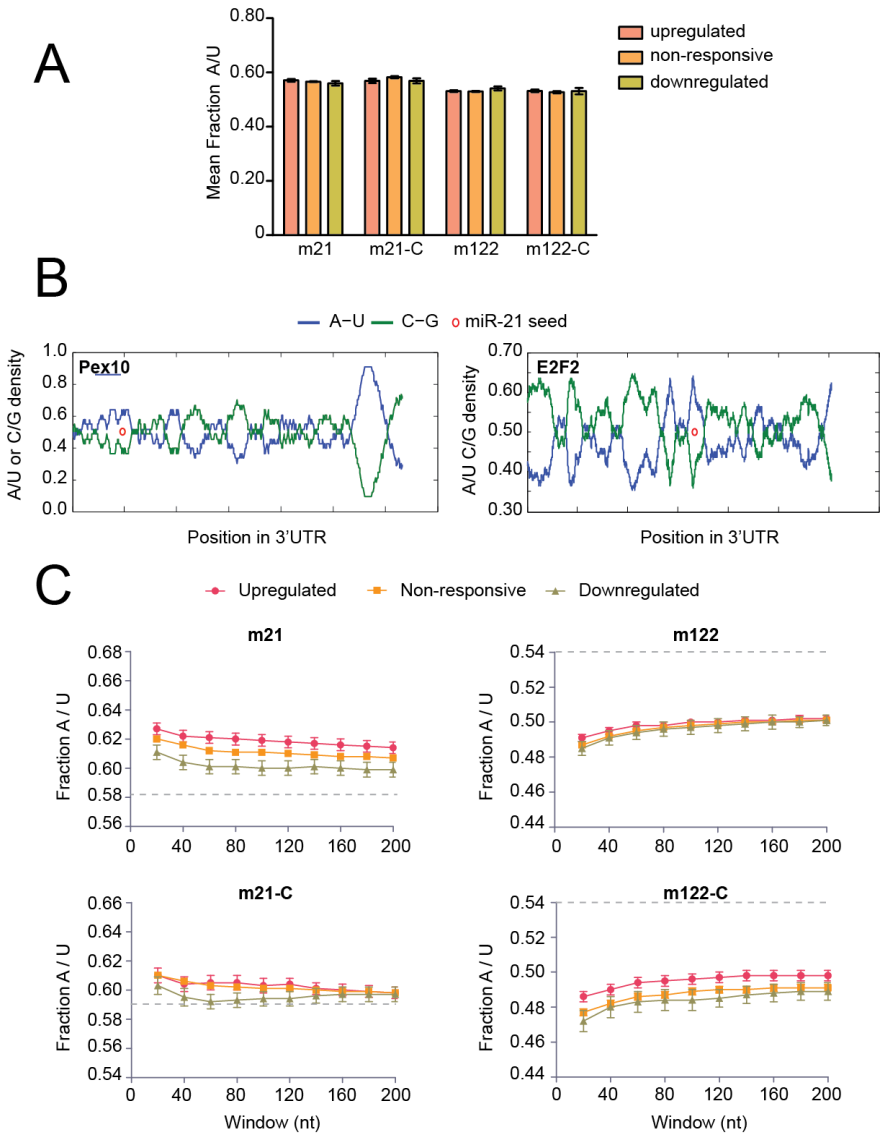


Figure 3.4 Influence of 3'UTR base composition on target selection. (**A**) Mean fraction A/U composition of 3'UTRs in each bin for each miRNA gene set. (**B**) Moving window analysis of 3'UTR base composition for two upregulated miR-21 seed-matched targets Pex10 (left) and E2F2 (right). The x-axis indicates the relative position in the 3'UTR with the left side of the axis representing the start of the 3'UTR and the right side of the axis representing its end. An open red circle in each plot marks the miR-21 8m seed position. For both these genes, the seed position falls inside an A/U rich window. (**C**) Mean fraction A/U for increasing window sizes centered on the miRNA seed-match sequence. For each window size, the average A/U compositions for upregulated (red circle and line), non-responsive (yellow square and line), and down-regulated (green triangle with line) binned genes are plotted. The grey dashed line represents the mean A/U composition of the full-length 3'UTRs of all seed-matched genes in that dataset (equal to the average of the three bins in each group shown in **A**).

Non-canonical sites are minimally effective

In addition to the 3'UTR, functional seed sites may also occur in the 5'UTR, and even more likely, in the mRNA coding sequence (CDS). CDS seed sites have been shown to be functional in both *Drosophila* and human cells[225,226]. Furthermore, cross-linking immunoprecipitation (CLIP) experiments have revealed that nearly 25% of identified Ago2 binding sites map back to mRNA CDSs[227], suggesting that miRISC can frequently access the CDS in spite of competition by ribosomes[228]. We thus wanted to investigate whether CDS targeting had any effect in our experiments. We were especially interested in seeing whether we could observe CDS sequence trends for miR-21 since, surprisingly, no trends were observed for the 3'UTR (**Fig. 2. 2.D**). Similar to the 3'UTR, however, Sylamer analysis of miR-21 inhibition data using CDSs showed no statistical enrichment for miR-21 7-mer seed sequences (**Fig. 3.5.A**). Accordingly, miR-21 6-mer seeds were found in < 20% of genes and most were not upregulated (**Fig. 3.5.B**). Similar results were found for miR-122 (**Fig. 3.5.B**), indicating that CDS seed sites are not effective sites for miR-21 and miR-122 in liver.

We also investigated alternative modes of 3'UTR binding. In brain, non-canonical sites have been validated with either a G-bulge between miRNA positions 5 and 6 [229] or centered pairing with miRNA positions 4-15 [230] (**Fig. 3.5.C**). These types of sites are often highly conserved, but the universalness of their function has not yet been shown. Surprisingly, in liver for miR-21, a greater number of genes containing G-bulge sites were upregulated than genes containing 8m canonical sites; and the majority of these did not contain a 7-mer or 8m canonical site (**Fig. 3.5.D,** left). G-bulge targets as a population also showed a greater median fold-change than both all miR-21 canonical seeds combined and a control population containing a random 8-mer sequence (**Fig. 3.5.D,** mid/right). However, G-bulge targets had a lower median fold-change than canonical 8m targets and, unlike the 8m population, the fold-change distribution for G-bulge targets was not significantly different from the distribution for all genes (**Fig. 3.5.D,** right). A comparable percentage of G-bulge genes was found in upregulated genes in the miR-122 inhibition data; however, owing to the strong activity of miR-122, this was ~2.5-fold less than the number of up-regulated miR-122 canonical 8m genes (**Fig. 3.5.E,** left). Furthermore, G-bulge targets as a population were not found to be

upregulated overall with a median fold-change less than canonical miR-122 seeds (**Fig. 3.5.E,** mid/right). As for centered pairing sites: Only two of these types of sites were found for miR-122 and only one for miR-21; neither were up-regulated following miRNA inhibition.

We conclude that non-canonical miRNA sites account for few of the expression changes in our data. For miR-21, G-bulge sites are less effective than 8m sites but perhaps just as effective as 7-mer sites; but these effects are still quite small, on par with miR-21's weak canonical activity.

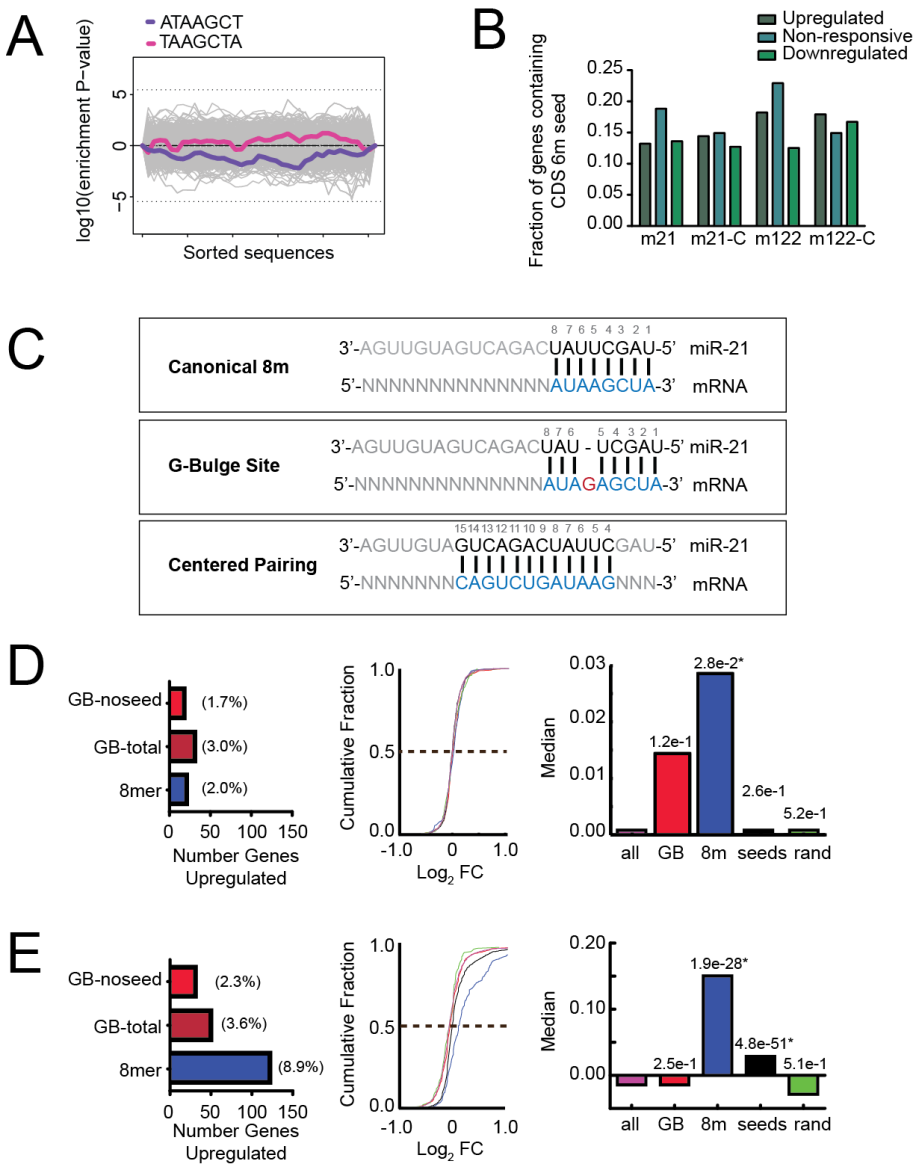


Figure 3.5 Influence of non-canonical miRNA binding sites on target selection. (**A**) Heptanucleotide Sylamer analysis of coding sequences from all genes (non-filtered) in anti-miR-21 profiling experiments. 7m8 and 7m1A miR-21 seed sequences are highlighted in purple and pink, respectively. All other seven-nucleotide sequences are shown in grey and represent statistical background. Significance thresholds are marked by horizontal dotted lines. No sequence surpasses these thresholds indicating a lack of enrichment for seven-nucleotide sequences on either side of the ranked gene list. (**B**) Fraction of genes in each bin containing a 6m seed in its coding sequence. Results from expressed seed-matched transcripts for miR-21 (m21) and miR-122 (m122) as well as expressed, conserved transcripts (m21-C & m122-C) are shown. (**C**) Schematic illustrating the differences in canonical 8m seed-target binding and non-canonical G-bulge site binding and centered pairing. The sequence of miR-21 is used as an example. miRNA and mRNA bases involved in target sequence recognition are highlighted in

black and blue, respectively. Watson-crick base pairing is illustrated as dark vertical lines. Nucleotides shown in grey are considered less important for target sequence recognition and have variable complementarity between the miRNA and target. The G-bulge site is characterized as having an additional non-pairing guanine between base 5 and 6 of the miRNA (highlighted in red). Centered pairing involves an extended region of complementarity between the target and the middle of the miRNA sequence (bases 4 – 15). (**D**) *Left,* Number of expressed upregulated genes in anti-miR-21 profiling data containing a G-bulge site (GB-total; red), G-bulge site and no canonical miR-21 seed (GB-noseed; dark red), or canonical 8m miR-21 seed (8mer; blue). The percent that each population makes up of the total number of upregulated genes is indicated above each bar in parentheses. *Middle*, cdf plot of expression changes following miR-21 inhibition for all expressed genes (pink) and those containing a G-bulge site (red), miR-21 8m seed (blue), any miR-21 seed (black), and a random 8m seed with the same base composition as the miR-21 8m seed (green). The dotted brown line marks the median. *Right* Bar plot of median values from the middle cdf plot. Value above each bar is the KS p-value calculated between that distribution and the distribution of all genes. Significant differences ($p < 0.05$) between two distributions are marked with a star (*). (**E**) Same as *D* for anti-miR-122 array profiling.

Target responses do not correlate with thermodynamics of miRNA binding

A poor seed interaction can be compensated for by extensive base pairing to the 3' end of the miRNA[82,229]. Since miR-21 has a weak seed-sequence due to a high A/U base composition, we hypothesized that responsive targets may be able to supplement binding energetics by forming more base pairs with the 3' end of miR-21. To test this hypothesis, we used an alignment based approach to score contiguous base pairs starting from the 3'end of the miRNA, similar to what has been done previously[82]. Neither miR-21 nor miR-122 upregulated targets were scored higher than non-responsive targets, suggesting that these targets were not more likely to form additional base pairs outside of the seed region (**Fig 3.6.A**). More robust thermodynamic calculations also showed no correlation between response and thermodynamics of binding (**Fig 3.6.B-C**).

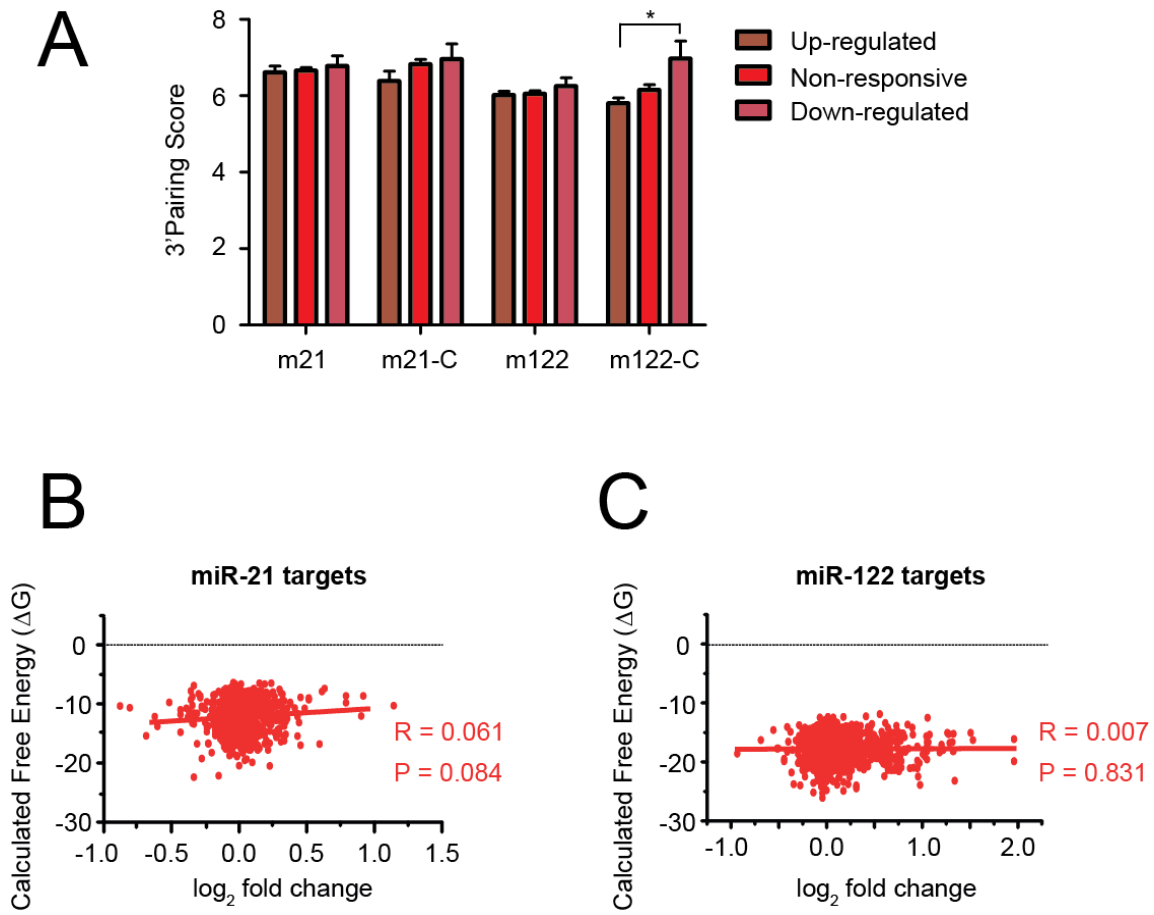


Figure 3.6 miRNA 3'base pairing and binding thermodynamics. (**A**) 3' pairing scores for miR-21 (m21) and miR-122 (m122) expressed genes and miR-21 (m21-C) and miR-122 (m122-C) expressed, conserved genes binned based on response to miRNA inhibition. Scores were computed based on the number of contiguous miRNA-mRNA base pairs that could form starting from the 3'end of the miRNA (see **Materials and Methods**). (**B**) Linear regression analysis of thermodynamic data for duplex binding from the PITA database[85] with measured expression changes for miR-21 seed-matched mRNAs. Linear regression fits, correlation coefficients, and p-values are shown for each. (**C**) Same as **B** for miR-122.

Targets are commonly enriched in RBP motifs

Hundreds of RNA binding proteins (RBPs) are encoded in mammalian cells and several have been identified as capable of augmenting or inhibiting miRNA-mediated gene regulation by binding to mRNAs proximal to or overlapping with miRNA binding sites[218,231-233]. We therefore sought to gain perspective on the extent to which known RBPs could affect miR-21/miR-122 targeting networks. Transcripts were searched for four RBP consensus motifs linked to miRNA pathways:

— U-rich motif (URM), UUUUAAA, which has been correlated with repressed transcripts in several small RNA transfection experiments[218];

— Hu (ELAV) binding site, UUUNUUU, which can antagonize miRNA activity by stabilizing mRNA[69,234,235]. In other cases, it can also enhance miRNA activity[236].

— PUM motif, UGUANAUA, often found adjacent to conserved miRNA seed sites[233]. PUM proteins have also been shown to interact with Ago containing complexes[237]; and

— A/U rich element (ARE), UAUUUAU, which affects mRNA stability and is recognized by nearly 20 different RBPs[218].

We first searched for the presence of these motifs in regions proximal (within 200 nt) to miR-21 and miR-122 seed sites, where the RBPs could directly compete or even cooperatively bind with miRISC (**Fig 3.7.A**). The Hu motif was the most commonly found among both miR-21 and miR-122 targets. It occurred more frequently near miR-21 seeds (~30 − 40% per bin) than miR-122 seeds (~15 − 25%), suggesting its antagonistic effects on miR-21 may partially explain miR-21's limited repression signature; however, it appeared just as frequently among upregulated genes as it did among non-responsive genes, arguing against this hypothesis. The other motifs were not as common, occurring at <10% for most bins. In general, the RBP motifs occurred in proximal regions with similar frequencies across the response bins.

This analysis was then extended to the full-length 3'UTRs (**Fig. 3.7.B**). Again, the Hu motif was the most common, found in ~90% of miR-21-targeted genes and a slightly lower fraction of miR-122 genes. The UR motif was next most common at >50%, followed by PUM and ARE, which occurred at similar frequencies ~20 − 40%. Because

3'UTR sequences are often long and A/U rich, these motifs would be expected to occur at high rates due to chance alone. We therefore used the permutation method (see Materials and Methods) to test for statistical enrichment of the motifs[218]. For Hu and URM, approximately half of the genes containing these motifs were significantly enriched for them, indicating that these motifs often occur repetitively throughout each of the 3'UTRs. ARE and PUM sites were often less enriched. Similar to the results with proximal sequences (**Fig 3.7.A**), each motif occurred at similar frequencies between the response bins.

Much like miRNA seed sites, not all RBP motif sites are expected to be functional. Since the rules for RBP binding are even less clear than those for miRNA, it is impossible to gauge the true biological relevance of these results. However, the prevalence of these motifs, especially Hu and URM, in sequences both proximal and distal to miR-21 and miR-122 seed sites strongly suggests that RBPs could significantly influence target repression by these miRNAs.

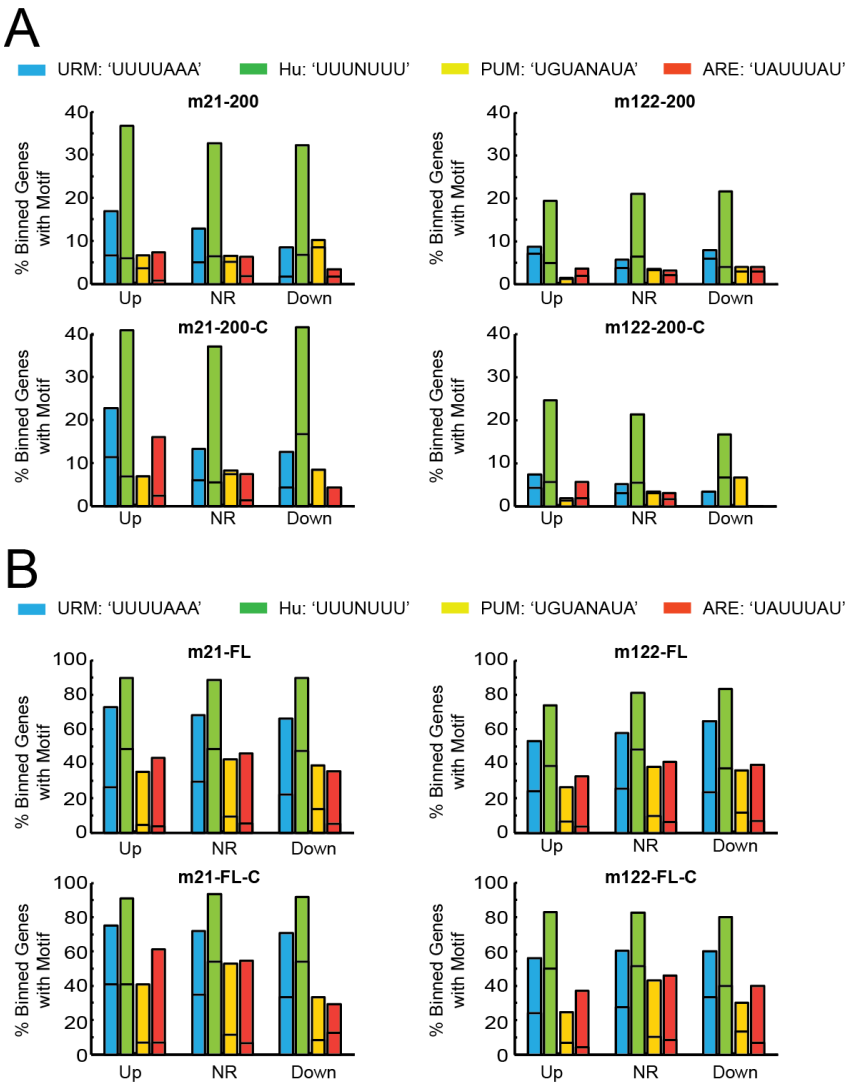


Figure 3.7 Influence of RNA binding protein sites on target selection. (**A**) Percent of genes in each upregulated (Up), non-responsive (NR), and downregulated (Down), bin containing known RBP motifs within 200 nt of the seed sequence. Sequences were analyzed for four different RBP motifs that have previously been shown to affect miRNA activity: URM (blue), Hu (green), PUM (yellow), and ARE (red). Transcripts containing an RBP motif were assessed for statistical enrichment of the motif with consideration of sequence base composition. Enrichment scores (permutation test, see Materials and Methods) were assigned to each gene transcript and the percent of genes with significant enrichment for each motif is indicated on each bar by a horizontal line. (**B**) Same as **A**, but for full-length (FL) 3'UTRs.

3.4 Discussion

We have recently reported (**Chapter 2** and [238]) that liver miR-21 represses only a narrow set of targets containing canonical 7- or 8-mer seed sites. In the current study, we have examined in depth the sequences of these targets to better understand what makes them exceptional. We found little evidence to support that the minimal activity of miR-21 follows the current canonical or even known non-canonical modes of target selection. Instead, we uncovered a new possible determinant, ORF length, and found significant enrichment in motifs for RBPs suggesting these regulators may influence target selection.

Length and Expression

miR-21 responsive targets as a whole were distinguishable from non-responsive targets by two features: length and expression. Neither of these are known features affecting miRNA-mRNA interactions. We found that responsive miR-21 targets, on average, had significantly shorter ORFs. This was also observed for miR-122 targets, in addition to having shorter 3'UTRs. The effects of transcript length on miRNA-mediated repression are not well studied. One possible explanation is that shorter transcripts, with fewer bases to cleave, may be more quickly degraded, enabling the miRNA to reduce steady-state levels to a greater extent. In addition, a shorter ORF may speed up the rate of decay if ribosome runoff must be completed before transcript degradation can occur. This is supported by models where miRNAs first inhibit translation at the initiation step, before inducing transcript destabilization[239]. We also found a weak, but detectable correlation with expression levels, with responsive targets slightly more highly expressed. This result was unexpected since miRNAs generally are less effective as the abundance of target sites increases [240]. While we do not discount a biological relevance, alternatively this effect may be caused by a bias in the array profiling data. Measurements on more abundant (or better hybridizing) transcripts may be less prone to noise, resulting in lower standard deviations between biological replicates and greater significance between treated and untreated groups. Because we based our cutoffs on p-values, more abundant genes may then be more likely to be binned as upregulated than genes with lower

expression. Still, we expect any bias to be minimal and do not believe that these effects had a major contribution on our other results.

Inhibition versus Transfection of miRNA

Array profiling of HeLa cells after different transfections of eleven miRNAs indicated that target repression generally correlated with four independent features of 3'UTR sites: A/U base composition near the seed site, the number of seed sites, extent of base-pairing to 3' non-seed regions of the miRNA, and position of the site towards the ends of the 3'UTR[82]. These results were subsequently built into a quantitative model for target prediction called TargetScan(Context)[82]. Here, these features were among those tested in an attempt to better understand miR-21 target selection. As a positive control, we chose to analyze miR-122 in parallel with miR-21. Unlike miR-21, miR-122 responses showed moderate correspondence with TargetsScan predictions (TargetScan performed better than four other prediction methods; **Fig 2.3.A**). miR-122 was in fact one of the eleven miRNAs used for training TargetScan[82]. In agreement, miR-122 showed clear trends for two TargetScan features — responsive targets showed enrichment for having multiple sites and sites near the ORF. These features were not detected for miR-21. Unexpectedly, trends for A/U-richness and 3'-pairing were not detected for either miRNA. One simple explanation for this discrepancy is technical— our bioinformatic analyses are similar, but not identical to those of Grimson et al.[82]. Alternatively, miRNAs may target different transcripts in HeLa cells, which demonstrate enhanced miRNA activity, than liver cells (**Fig. 2.12** and **Fig 2.13**). A third possibility is that targets, and targeting preferences, identified by miRNA transfection may not be the same as those identified by inhibition.

Fundamentally, transfection and inhibition are different methods for perturbing miRNA-mRNA interactions. In the former, an abundant amount of miRNA is introduced into the repression network, while in the latter a miRNA is effectively removed. We posit that these differences may cause distinctive changes in the transcriptome, even with the same miRNA. On the one hand, transfected miRNA may repress any and all preferred targets due to shear abundance, irrespective of competing factors. On the other hand, inhibition would be expected to only cause derepression of actively repressed targets (aside from secondary effects), which may be far fewer than the total number of preferred

or possible targets. In support of this notion, compared to values reported for transfection (repression/downregulation)[82], a lower percentage of seed-matched genes displayed detectable responses (derepression/upregulation) in our inhibition experiments. In addition, the magnitudes of inhibition responses appear to be more modest. It is therefore possible that, upon inhibition, the same targeting trends exist, but are not as easily discernible since it is more difficult to, proverbially, separate the wheat from the chaff. Alternatively, endogenous competitive effects may reshape target selection: a more abundant, stronger binding miRNA could block all of the most favorable sites of a lower-abundance miRNA that shares many of the same targets, and as a result shunt the lesser miRNA to otherwise less-favorable targets. Likewise, RBPs, which can induce conformational changes in RNA secondary structure[241], could convert an otherwise 'bad' target into a more favorable one. Importantly, because transfection could dramatically shift equilibrium in favor of a single miRNA, miRNA copy number may drown out the influence of endogenous competition on target selection. Future studies directly comparing transfection and inhibition of various miRNAs will be necessary to shed more light on this question and help assess the relative biological validity of these two experimental approaches.

Influence of RNA binding proteins on target selection

An expanding number of reports have demonstrated that RBPs can directly modulate the activities of certain miRNAs. The ubiquitously expressed Hu family protein HuR may be a particularly important liaison between RBP and miRNA networks. Filipowicz and colleagues have found that HuR reversibly, in response to stress, prevents miR-122 from repressing the cationic amino acid transporter 1 (Cat-1) mRNA in liver cells[69]. This antagonism, the result of HuR oligomerization on mRNA to prevent miRISC binding, can protect other mRNAs harboring the HuR motif, regardless of proximity to seed sites and even in the absence of stress[242]. HuR can also enhance miRNA-repression in other cases, as it and let-7 interdependently work together to repress c-Myc expression in HeLa cells[236].

We posit that RBPs may have a broad impact on miR-21 targeting networks. Both miR-21 and miR-122 targets were highly enriched in RBP-REs, especially the Hu

motif, which also occurred frequently proximal to miR-21 sites. Although direct correlations between RBP motifs and response were not uncovered, their frequency of occurrence, in particular of the Hu motif, and emerging insights into their abilities to modulate miRNA activity make RBPs an alluring candidate for the focus of future experiments to understand the targeting dispositions of miR-21 and other miRNAs upon anti-miR inhibition.

The miRNA network is seemingly built on surprisingly little sequence information. Single molecule fluorescence RNA annealing studies suggest that 7-nt of complementary sequence, the length of a miRNA seed, is the minimum number of contiguous base-pairs required for rapid Watson-Crick paired annealing[243]. In addition, Nature may have selected for miRNAs the minimum amount of specificity as to cast a wide net on gene regulation. RBPs may play the part of the fishermen, choosing exactly where and when to cast the net of repression. Further, HuR and Pum1 RBPs are phosphorylated in response to stress or growth factor stimulation, which alters their sub-cellular localization and interactions with miRNAs[244,245]. RBPs may therefore serve to connect the miRNA network to cell signaling cascades, allowing for temporal, transcript specific alterations in gene expression in response to extracellular stimuli. Since miR-21 has been directly linked to stress response pathways[238], it activities may be especially coupled to that of RBPs.

Chapter 4: iSHiRLoC: A single-molecule method for visualizing and measuring heterogeneities in microRNA processes[1]

4.1 Introduction

The mechanisms by which miRNAs reduce expression output are debated[62,246]. Some have concluded that translational inhibition is the primary pathway[65,68,80], while others claim it is transcript destabilization[53,81,170]. Adding to the confusion, several modes of translational inhibition have been proposed: blocking of the initiation step[68,170], stalled elongation[80], ribosome drop-off[76] and even nascent chain proteolysis[79]. Supporting evidence has been derived from experimental conditions and techniques as diverse and varied as the mechanisms themselves, making it difficult to compare and contrast these seemingly contradictory findings. This leads to the question: Is there a unitary mechanism underlying miRNA activity or can different miRNAs, or even individual molecules of the same miRNA, divergently enact repression?

Recently, we have developed and published a method for probing the heterogeneities of miRNA-mediated processes by visualizing single miRNAs as they proceed through the repression pathway— from diffuse mature duplex, to miRISC assembly, to target engagement, to P body docking, to release of the miRNA (the full publication is provided in **Appendix B**)[247]. The method is based on a two-pronged approach: high resolution tracking of fluorophore labeled miRNAs in live cells and step-wise single-molecule photobleaching in fixed cells. Together, these tools enabled us to measure the dynamic localization and stoichiometry of miRNA complexes, respectively. Complex assembly and disassembly could be kinetically determined by utilizing the imaging techniques at varying time increments after introducing the miRNA into the cell.

[1] This chapter includes John Androsavich's contributions to Pitchiaya, S., Androsavich, J.R., and Walter, N.G (2012), *EMBO Reports*, the full text of which is provided in Appendix B. Results referred to in Appendix B were largely the work of Sethu Pitchiaya. In addition, Katelyn Doxtader, an undergraduate, performed the Ago2-WT/Y529E mutagenesis, imaging, and particle counting analysis under John Androsavich's supervision.

We have termed this method iSHiRLoC, or intracellular single-molecule high-resolution localization and counting, and propose its broad applicability, but find it particularly well suited for studying the miRNA pathway in hopes of establishing a unifying model for repression.

In its first implementation, we used iSHiRLoC to study the complex assembly of let-7a miRNA in HeLa cells. The miRNA was 3' fluorophore labeled and microinjected into the cytoplasm. Microinjection provided a convenient means to introduce a relatively consistent number of miRNAs—preventing both overloading of the pathway and too high of miRNA densities to track single molecules—and unlike transfection, a better defined '0 h' time point for kinetic analysis. Control experiments demonstrated that neither fluorophore labeling nor microinjection hampered the miRNA's ability to repress expression of reporter genes. Following microinjection, we were able to detect miRNAs as well-resolved foci as early as at the 2-h time point. Step-wise photobleaching revealed that the majority of these foci contained single fluorophore labeled miRNAs. We attribute this extraordinary sensitivity of detection largely to the use of highly inclined laminar optical sheet (HILO) illumination[248], wherein the angle of incidence of the illuminating laser was adjusted to just above the critical angle, causing the excitation light to pass diagonally, rather than perpendicularly, through the sample; this reduced the volume of excitation above and below the focal plane, and as a result, increased signal-to-noise.

The microinjected miRNA progressed through a series of time-dependent changes in diffusion and multimerization. Immediately after microinjection, the miRNA could only be detected as a hazy background in the cell due to rapid diffusion too fast for our time resolution of 100 ms. At 2 h post-injection, detectable complexes were formed and their diffusion coefficients found distributed into two Gaussian populations — a slow-moving population with diffusion coefficient equal to 0.26 $\mu m^2 s^{-1}$, and a fast population with a coefficient of ~ 0.034 $\mu m^2 s^{-1}$. Importantly, these diffusion coefficients are similar to those of mRNPs and P bodies, respectively, key intermediates of the miRNA pathway that we found to be colocalized with the miRNA. These populations, on average, diffused progressively faster through the 4 h and 8 h time points, suggesting that the complexes decreased in mass over time. This trend appeared to extend out to 32 h, the latest time point measured, where foci were no longer trackable due to large inter-frame

displacements. Step-wise photobleaching of fixed cells at these time points showed that multimeric assembly followed the same course as diffusion, with the number of multimers rapidly increasing after injection, peaking at ~4 h, and then slowly diminishing over time. The rates of these processes lead us to propose a model for let-7a miRNA assembly that is characterized by two distinct kinetic processes: an initial fast assembly step (with a rate constant of 1.2±0.2 h^{-1}), where miRNAs undergo rapid assembly into first mRNPs and then P bodies, followed by a slower disassembly step (with a rate constant of 0.14±0.08 h^{-1}). An artificial CXCR4 miRNA, which has very few endogenous targets, only exhibited similar rates of assembly/disassembly when an exogenous artificial mRNA target was co-microinjected, thus providing strong evidence for target-dependent miRNA assembly.

In this chapter, I detail a few of the key developments for establishing the iSHiRLoC system for studying miRNA. In addition, we discuss tracking of the central RISC component Ago2; recent inhibition experiments that further support the functionality of trackable complexes; and lastly, the synthesis and testing of a biarsenical fluorophore that may be useful for iSHiRLoC.

4.2 Materials and Methods

Cell culture and media. HeLa cells (CCL-2) were acquired from ATCC and maintained under growth conditions in Dulbecco's modified eagle medium (DMEM) with high glucose, 10% (v/v) fetal bovine serum (FBS), and formulated non-essential amino acids (MEM-NEAA). Several (1 − 4) hours before imaging, to reduce autofluorescence background, growth media was replaced with DMEM containing reduced (2% v/v) FBS and no phenol red pH indicator. This reduced media was again replaced immediately before imaging with a buffered saline solution (BSS) containing 20 mM HEPES pH 7.4, 135 mM NaCl, 5 mM KCl, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, and 5.6 mM glucose.

Fluorescence Repression Assays. For microinjection repression assays, $1.5x10^5$ cells/dish were plated on deltaT (Bioptech) dishes the day prior to microinjection. Cells were microinjected in the nucleus with 1X PBS solutions containing 0.1 μg/μl pEF6-

mCherry-mH3Um, 0.1 μg/μl pEGFP-C1, and 1.5 μm wt- or mut-let-7a1. Following injection, cells were maintained in growth media for 24 h to allow for expression and maturation of the fluorescent reporter proteins. Widefield microscopy images were acquired for mCherry and GFP and intensities were measured using ImageJ software (NIH). For individual cells, the background subtracted mCherry signal was normalized with that of GFP, and the reported repression levels were taken from the mean mCherry/GFP ratio over all cells from two independent experiments. When necessary, signals were corrected for channel cross-talk (bleed-through) by subtracting the fraction of signal from the opposite channel that is empirically determined to bleed-through:

$$Ia' = Ia - Ib \times C$$

where Ia' is the cross-talk corrected intensity for channel A, Ia is the non-corrected intensity, Ib is the intensity for channel B, and C is the fraction of Ib that enters channel A. This fraction is pre-determined by imaging each color separately using both channels. In our experiments, we found ~14% of mCherry signal bleed-through to the GFP channel. For transfection repression assays (**Fig. 4.1.B**), $4x10^5$ cells were plated on 6-well glass bottom plates the day before transfection. Cells were co-transfected with 4.0 μg DNA consisting of an equimolar ratio of pEF6-mCherry-mH3Um and pEGFP-C1 plasmids and 10 pmol miRNA duplex. Cells were incubated with transfection complex for 6h before replacing the media. 48 h post-transfection, cells were either assayed using fluorescence microscopy or by western blot. For fluorescence microscopy, to facilitate counting of cells, nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI) for 1 h at 37 ºC. Cells were imaged using a Nikon TE-2000U widefield microscope with a CoolSNAP HQ^2 CCD camera (Photometrics) and fluorescence intensities were quantified in the same manner as for microinjection repression assays. For western blots, wells transfected in triplicate were pooled and the cells were pelleted at 1,000 x g for 10 min at 4 ºC. After washing, pellets were resuspended in 75 μl lysis buffer (20 mM Tris pH 7.2, 100 mM NaCl, and 2.5 mM $MgCl_2$) and lysed by four cycles of freeze-thaw in liquid nitrogen with rigorous vortexing between each cycle. Lysates were cleared by benchtop centrifugation at 9,000 rpm for 10 min at 4 ºC, supernatants were transferred to fresh tubes and 5 μl protease inhibitor cocktail (Sigma) was added to each. Equal volumes of each lysate were resolved with SDS-PAGE and immunoblotted following a standard protocol. Rabbit monoclonal antibody to dsRed/mCherry was purchased from Clontech. Mouse monoclonal antibody to β-actin was purchased from Sigma. Primary antibodies

were detected using Cy3/Cy5 conjugated secondary antibodies and membranes were scanned using a Typhoon imaging system (GE).

Plasmids and miRNA oligonucleotides

Fluorescent reporter plasmids. Fluorescence reporter plasmid pEF6-mCherry-mH3Um was constructed by first subcloning the mCherry ORF from the pRSET-mCherry vector, a gift from Roger Tsien (UCSD), into the multiple cloning site of the pEF6-myc-His-B (Invitrogen) vector using BamHI and EcoRI restriction sites. Then, using PCR cloning, the mutant HMGA2 3' UTR, a gift from David Bartel (Whitehead Institute, MIT) was inserted downstream of the mCherry ORF between XhoI and NotI sites. pEGFP-C1 was purchased from Clontech.

Ago2 and Dcp1a plasmids. pEGFP-hAgo2 was contributed by Phillip Sharp (MIT). pEGFP-hAgo2-Y529E was created by site-directed mutagenesis using the Quikchange XL kit (Stratagene) as per manufacturers recommendations. pRFP-Dcp1a was a gift from Nancy Kedersha (Brigham and Women's Hospital).

MS2 plasmids. Basic plasmids encoding the EGFP fused MS2 coat protein (pMCP-EGFP) and 24 copies of the MS2 stem-loop (pSL-MS2_24x) were received as gifts from Robert Singer (Albert Einstein College of Medicine). The pMCP-GFP plasmid was used as is for initial experiments (**Fig. 4.2**); however, it suffered from poor expression in our hands likely due to having a weak promoter (derived from the large subunit of RNA polymerase II[162]). To improve expression, the MCP-GFP sequence was PCR amplified and inserted into the pEF6-myc-His-B (Invitrogen) parent plasmid under control of the EF1α promoter. This plasmid, pEF6-MCP-GFP, showed a slight improvement in expression and was used for the remainder of the MS2 experiments (**Fig. 4.3**). The hairpin loops of pSL-MS2_24x were incorporated into the tail end of a luciferase reporter construct, engineered in the pmirGLO (Promega) backbone, consisting of an optimized firefly luciferase coding DNA sequence (CDS) and the mutant Hmga2 3'UTR. This plasmid, pmG-mH3UM-MS2, was created in two steps as detailed elsewhere.

miRNA. RNA oligonucleotides were ordered from Keck Oligo Synthesis facility (Yale University School of Medicine). All oligos contained a 5'phosphate and guide strands contained 3' amino-C7 linkers (*).

wt-guide-let-7a1: P-UGAGGUAGUAGGUUGUAUAGUU*
wt-passenger-let-7a1: P-CUAUACAAUCUACUGUCUUUCC*
mut-guide-let-7a1: P-UGCGUUAGUAGGUUGUAUAGUU*
mut-passenger-let-7a1: P-CUAUACAAUCUACUGUCGUUCC*

RNA was purified as described [249] and the 3' amine groups were labeled with Cy5 N-succinimidyl-ester and repurified with HPLC to remove all unlabeled RNA. Guide and passenger strands were annealed in a 1:1 ratio in RNase-free H2O by placing tubes in a 65 ºC bath that was then allowed to cool to room temperature.

Anti-miRs. LNA/DNA antisense oligonucleotides were purchased from Exiqon and used as provided. For injection with miRNA, 3.0 μm anti-miR was mixed with 1.5 μm duplex miRNA in 1x PBS. For labeling, Cy5 was conjugated to the 5' ends using the same procedure for labeling miRNAs. Labeled anti-miR was injected at 3.0 μm concentration in 1x PBS. The anti-miR sequences were as follows, without showing proprietary (Exiqon) LNA positions, and marking the amino-linker (*) positions:

anti-let-7a: *ACTATACAACCTACTACCTC
anti-control:*GTGTAACACGTCTATACGCCCA

Fluorescence In Situ Hybridization (FISH). Probes against the MS2 hairpin-loop containing four amino-C6 modified deoxythymidine bases[250] were ordered from the Keck Oligo Synthesis facility (Yale University School of Medicine) with the sequence: A*TGTCGACCTGCAGACA*TGGGTGATCCTCA*TGTTTTCTAGGCAAT*TA, where *T indicates the modified bases. The probes were labeled with Cy3 fluorophores by reacting the amino-handels with Cy3-N-succinimidyl-ester (GE) at room temperature for 72 h, rotating and protected from light (final labeling efficiency ~ 93%). Dye-labeled

probes were subsequently HPLC purified. FISH was performed on cells co-transfected with pMCP-GFP and pMS2-mHM3Um-24xMS2 using previously published methods [250].

Microinjection. Microinjection was performed using an Eppendorf Femtojet injection system and InjectMan micromanipulator (Eppendorf). All injection solutions were prepared in 1x PBS and backfilled into prepulled Femtotip microinjection needles (Eppendorf). Injection parameters were typically set at 100 hPa injection pressure, 20 hPa constant pressure, and 0.5 s injection time.

Microscopy. Imaging was performed using a customized Olympus IX81 microscope. The illumination module consisted of a cell-TIRF system coupled with solid-state lasers with wavelengths of 488 nm (25 mW), 532 nm (100 mW) and 640 nm (100 mW). Lasers were directed through an acousto-optical tunable filter before being coupled into the cell-TIRF module. All laser lines had 10 nm bandwidth clean-up filters to ensure monochromatic illumination. Net powers of ~1.2 mW, 7 mW, and 8 mW were achieved at the objective for the 488 nm, 532 nm and 640 nm laser lines, respectively. Laser beams were focused on the back-focal plane of the objective at distances from the optical axis such that the incident angle at the dish-media interface created a highly inclined laminar optical sheet (HILO) of illumination. This provided significantly greater signal-to-noise, while also enabling greater penetration depths. The detection module consisted of a 60X 1.45 NA oil-immersion objective (Olympus), 1x - 4x magnification changer (Olympus), and an Evolve EM-CCD camera (Photometrics). A dual-band filter cube consisting of a z491/639rpc dichroic filter (Chroma) and z491/639m emission filter (Chroma) was used for detection of GFP and Cy5, while a broadband filter cube with Q570LP dichroic and HQ610/75m emission filters (Chroma) was used for detecting Cy3 and mCherry. Unless otherwise specified, all tracking videos and images were recorded at 100 ms camera exposure time using 120x magnification (60x objective plus 2x magnification changer). For live-cell imaging, cells were maintained on-stage at 37 ºC using the DeltaT open dish system and heated lid (Bioptechs).

Tracking image analysis. Centroid tracking was performed using Imaris as described in **Appendix B**. Gaussian-fit tracking was performed using the Jaqaman et al. Matlab routine[251]. In both cases, tracking data were similarly analyzed with in-house Matlab routines to calculate the mean squared displacement (MSD) and diffusion coefficients.

Colocalization image analysis. For mRNA-miRNA colocalization experiments, HeLa cells were co-microinjected with the MS2/MCP plasmids and the appropriate Cy5 labeled miRNA, incubated for 24 h under normal growth conditions and formaldehyde fixed just prior to imaging. Images for each channel were acquired sequentially and corrected for chromatic positional offset with an in-house Matlab routine. An object based approach[252] was used to quantitatively assess co-localization based on sub-pixel localization of the centers of mass (COMs) of isolated objects from the GFP (green circles) and Cy5 (red circles) channels (**Fig. 4.3.B**). Two objects, one from each channel, were considered 'perfectly' co-localized if their COMs were localized in the same pixel (yellow circles; pixel size ≈ 133 nm) or simply co-localized if the radial distance between the COMs was less than the resolution limit (~281 nm for Cy5 dye). In the latter case, the COMs of the two co-localized objects occur on adjacent, but separate pixels; these were mapped as cyan and magenta circles for Cy5-COMs and GFP-COMs, respectively, for Fig. 4.3.B. Statistical significance was estimated using simulated images created in Matlab with the same particle densities as experimental images. Images were simulated 5000 times and p-values were computed from the number of iterations where colocalization was equal to or exceeded experimental results.

Purification of recombinant IF2 proteins. pIF2-His plasmid encoding the bacterial initiation factor 2 (IF2) was received as a gift from the Rachel Green lab (Johns Hopkins Medical School). A modified version, pIF2-TC-His containing a C-terminal tetracysteine (TC) tag situated between the end of the IF2 ORF and the His tag was engineered by ligating a short adapter sequence between the HindIII and NotI restriction sites. Both constructs were expressed in *Escherichia coli* strain BL21 grown in 2L of lysogeny broth (LB) media at 37 ºC for 3 h after induction with 1 mM IPTG. IF2 expressing cells were pelleted, resuspended in cold lysis buffer (20 mM Tris pH 7.5, 300 mM NaCl, 5 mM

imidazole, and 1 mM 2-mercaptoethanol) containing protease inhibitors, and lysed in a microfluidizer at 80-90 psi. Clarified lysate was stirred with pre-equilibrated nickel sepharose resin at 4 ºC for 30 min, after which, the slurry was transferred to a glass column and the filtrate passed through by gravity flow. The resin was generously (5 – 6 column volumes each) washed with lysis buffer followed by binding buffer (lysis buffer w/ 500 mM NaCl), and then eluted with elution buffer (lysis buffer with 250mM imidazole and raised pH 8.5). Collected fractions were analyzed with SDS-PAGE, and IF2 containing fractions were pooled and dialyzed overnight against 100 mM Tris pH 7.5, 140 mM NaCl, 60 mM KCl, 14 mM $MgCl_2$, and 2 mM DTT. Protein concentrations were determined by Bradford assay (Promega) before being diluted 1:1 with 100% glycerol and stored at –20 ºC.

Cy3AsH binding assays. Binding assays were performed similarly as described by Cao et al.[253]. Proteins (IF2, IF2-TC, or BSA) were pre-incubated at room temperature in a buffer containing 50 mM HEPES pH 7.5, 10% DMSO (v/v), and 140 mM KCl in the presence of 100 μm ethanedithiol (EDT), 1 mM *tris*(2-carboxyethyl)phosphine (TCEP), and 1 mM 2-mercaptoethanol (BME). After 1 h, Cy3AsH solubilized in DMSO was added to a 1 μM final concentration and the labeling proceeded for an additional 2 h. Fluorescence was recorded using an Aminco-Bowman S2 fluorimeter with 545 nm excitation and a 550 − 600 nm detection range scanned with 1 nm stepsizes at a rate of 1 step/sec. Apparent dissociation constants (K_{app}) were calculated based on changes in fluorescence observed with increasing protein levels using nonlinear regression to the equation,

$$\Delta F = \frac{\Delta F_{\max}}{2[F]_{total}}\left([F]_{total}+[P]_{total}+K_{app}) - \sqrt{([F]_{total}+[P]_{total}+K_{app})^2 - 4[F]_{total}[P]_{total}}\right)$$

where ΔF is the observed fluorescence enhancement measured for each total concentration of protein ([P]), and [F] is the total concentration of Cy3AsH. For SDS-PAGE analysis, labeling reactions were mixed with SDS-PAGE loading buffer with 200 mM DTT containing no loading dye and run on 4-12% Bis-Tris NuPage polyacrylamide gels (Invitrogen) for 25 min at 200 V. Gels were imaged on a Typhoon scanner using Cy3 settings.

Synthesis of Cy3AsH:

General Comments. The synthesis was based on Cao et al.[253] with modifications in the solvents used for the second reaction and the chromatography purification for the final product. All reagents were obtained from Sigma-Aldrich without further purification. 1H-NMR spectra were acquired using a Varian Inova 400 MHz spectrometer. Time of flight electrospray ionization mass spectrometry (TOF/ESI/MS) was performed on a Micromass LCT instrument.

1--Sulfobutyl-2,3,3-trimethyl-indolenine (*Compound 1*). To a 25 ml round bottom flask, 0.8 ml 2,3,3-trimethylindolene, 1.6 ml butane sulton, and 7.0 ml 1,2-dichlorobenzene were added, in order, through a rubber septum. The mixture was refluxed for 15 h at 120 ºC and the product was precipitated by adding the reaction to 100 ml ice-cold acetone with gentle swirling, yielding a burnt red-orange solid. The solid was dried to completion in a desiccator containing calcium chloride. Yield 60%. 1H NMR (DMSO): 1.54 (*s, 6H*), 1.74 (*m, 2H*), 1.96 (*m, 2H*), 2.5 (*m, 10H*), 2.84 (*s, 3H*), 4.48 (*t, 2H*), 7.62 (*m, 2H*), 7.82 (*m, 2H*), 8.03 (*m, 1H*). TOF/ESI/MS: m/z calcd. for [M+H]+ 296.1, found 296.1.

3-(3,3-dimethyl-1-(4-sulfobutyl)indolin-2-ylidene)prop-1-enyl)-3,3-dimethyl-1-(4-sulfobutyl)-3H-indolium (Cy3, *Compound 2*). 300 mg *compound 1* (1.0 mmol) was dissolved in 1.5 ml pyridine solvent in a 25 ml round bottom flask equipped with a reflux condenser. The solution was heated to 120 ºC and a total of 300 µl triethylorthoformate (2 mmol) was added in 40 µl increments every 15 min over the course of 2 h. Following the last addition, the reaction was continued for another 2 h (4 h total). The reaction was cooled to room temperature, yielding a viscous dark purple substance with a metallic sheen. This substance was transferred to a 50 ml flask and several volumes of ether were added, producing a cloudy orange suspension with visible precipitate collected at the bottom of the flask. The ether was removed by rotary evaporation and the remaining solid was dissolved in a small volume of an ethylacetate/methanol (6:5) mixture. The solution was transferred to a 250 ml round-bottom containing silica gel (5-6 scoops using a standard curved scoopula) and the solvent was removed by rotatory evaporation. The

adsorbed silica was dry loaded on top of a 7 inch packed silica column and an ethylacetate/methanol (6:5) mixture was passed through as mobile phase. The products resolved as two bands: a faster migrating pink band and a slower migrating dark purple band. Eluted fractions containing the pink product were collected and dried, first with rotatory evaporation to remove the solvent and second with phosphorous pentoxide desiccation to remove remaining water. Yield 68%. 1H NMR (DMSO): 1.54 (*t, 2H*), 1.73 (*m, 16H*), 1.82 (*m, 4H*), 4.1 (*t, 4H*), 6.5 (*d, 2H*), 7.29 (*t, 2H*), 7.44 (*t, 2H*), 7.50 (*d, 2H*), 7.62 (*d, 2H*), 8.35 (*t, 1H*). TOF/ESI/MS: m/z calcd. for [M]- 599.23, found 599.0.

Cy3-bis-mercuric trifluoroacetate (***Compound 3***)**.** 131 mg mercury oxide was dissolved in 4 ml trifluoroacetic acid (TFA) with mild heat. After cooling, 180 mg of the Cy3 dye (*compound 2)* was added. The reaction was stirred overnight at room temperature. The following day, the TFA was evaporated by gently streaming nitrogen gas over the surface of the uncapped reaction at 40 ºC for 1 h, resulting in the formation of dark purple/red precipitate around the sides of the flask. The flask was placed in a desiccator containing phosphorous pentoxide and a vacuum was pulled for 2 days to ensure complete drying. The yield was 83%. No analytical data available due to safety concerns regarding the mercury.

3-(5-(1,3,2-dithiarsolan-2-yl)-3,3-dimethyl-1-(4-sulfobutyl)indolin-2-ylidene)prop-1-enyl)-5-(1,3,2-dithiarsolan-2-yl)-3,3-dimethyl-1-(4-sulfobutyl)-3H-indolium (Cy3AsH, ***Compound 4)*****.** In an oven-dried 10 ml round-bottom flask, 300 mg *compound 3* and 18 mg palladium II acetate were mixed with 3.5 ml anhydrous N-methylpyrrolidinone (NMP), 0.52 ml arsenic trichloride (6.21 mmol, 27 eq), and 0.40 ml diisopropylethylamine (DIEA; 2.3 mmol, 10 eq). The reaction proceeded under nitrogen at 60 ºC. After 3h, the solution was transferred to a 150 ml flask containing 0.25M phosphate buffered saline (PBS; pH 6.9) and 840 µl ethanedithiol (EDT) was added. The solution was extracted with 3 x 30 ml chloroform, and the combined organic layers were dried by addition of copious amounts of sodium sulfate. After 30 min, the sodium sulfate was removed by filtration and washed with several portions of chloroform until the filtrate ran clear. The collected filtrate was distilled at 40 ºC under vacuum for several hours until only reddish oil remained. This was then purified *twice* by chromatography on silica gel using a methylenechloride/methanol/water (20:6:1 v/v) mixture. The yield was 3%. 1H NMR (DMSO): 1.69-1.82 (*m, 17H*), 2.33 (*m, 1H*), 3.21 (*m, 4H*), 4.13 (*t, 4H*), 6.61 (*d, 2H*), 7.53 (*d, 2H*), 7.72 (*d, 2H*), 7.87 (*s, 2H*), 8.33 *(t, 2H*). TOF/ESI/MS: m/z calcd. for [M]- 931.00, found 930.5.

4.3 Results

Orthogonal fluorescence repression assay for measuring the activity of microinjected miRNA

iSHiRLoC is based on the premise that miRNAs injected into cells can be incorporated into the miRNA pathway and function to repress targets. In order to validate this premise, we developed a fluorescence-based repression assay that could measure miRNA activity in single cells. A fluorescence reporter construct encoding mCherry, a red fluorescence protein, was engineered to contain the 3'UTR of the mouse high mobility group A2 gene (mHmga2). The mHmga2 3'UTR is approximately 3 kilobases (kb) in length and contains seven seed sequences to the let-7 miRNA. Pairwise mutagenesis has demonstrated that each let-7 site pair can independently contribute to repression, having an additive effect in combination, and although the 3'UTR contains predicted seed sites for other miRNAs, let-7 appears to be the predominating repressor[173]. HeLa cells, the cell line used for the current iSHiRLoC work (**Appendix B**), express high amounts of endogenous let-7 (~20,000 copies/cell, **Appendix B.3. Materials and Methods**). To distinguish the activity of injected let-7 from that of endogenous let-7, we utilized orthogonal mutations in the mHmga2 3'UTR and let-7 miRNA[173] (**Fig. 4.1.A**).

We first tested the assay with transfection. The mCherry-mHmga2 3'UTR mutant reporter construct (mCherry-mHM3Um) was transfected with either no miRNA, wt-let-7a, or mut-let-7a. In order to control for transfection efficiency, an EGFP plasmid was co-transfected and EGFP and mCherry fluorescence were quantified ratiometrically. While cells transfected with wt-let-7a or no miRNA showed a similar ratio of mCherry/EGFP, cells transfected with mut-let-7a showed an ~70% decrease in mCherry expression, thus demonstrating the specificity of the system (**Fig. 4.1.B**). To validate that the fluorescence levels were reflective of protein levels, transfected cells were analyzed with western blots (**Fig. 4.1.C**). Again, mut-let-7a caused an ~70% decrease in mCherry protein compared to no miRNA. However, unlike the fluorescence measurements, wt-let-7a also showed ~50% decreased mCherry levels compared to no miRNA. This discrepancy is likely due to using endogenous actin as a loading control in the western blot, which does not account for changes in transfection efficiency. Under the same

transfection conditions, co-transfection of plasmid DNA with additional nucleic acids is likely less efficient than transfection of plasmid DNA alone due to changes in the charge:lipid ratio. This is supported by the fact that a random single-stranded DNA (ssDNA) oligo also caused decreased expression of mCherry (data not shown); thus, emphasizing the necessity for ratiometric measurements with dual exogenous reporters.

After validating the specificity of the orthogonal system and ratiometric quantification, we tested whether microinjected miRNA could cause repression. Cells were microinjected into the nucleus with the dual reporter plasmids in combination with either wt- or mut-let-7a miRNA. Since only faint mCherry expression could be detected ~2.5 h post-transfection in the absence of miRNA (data not shown), cells were maintained for 24 h post-injection to allow the proteins to fully mature and reach steady state. Similar to transfection, injected mut-let-7a caused a decrease in mCherry/GFP ratios compared to wt-let-7a (**Fig. 4.1.D**). This strongly indicates that at least a fraction of injected miRNA is functional, and also suggests that mature miRNA is transported out of the nucleus to effect repression.

These data, taken together with luciferase assays showing that 3'end fluorophore labeling does not inhibit miRNA activity (**Fig. B.1.A**), validate the biological relevance of our iSHiRLoC approach.

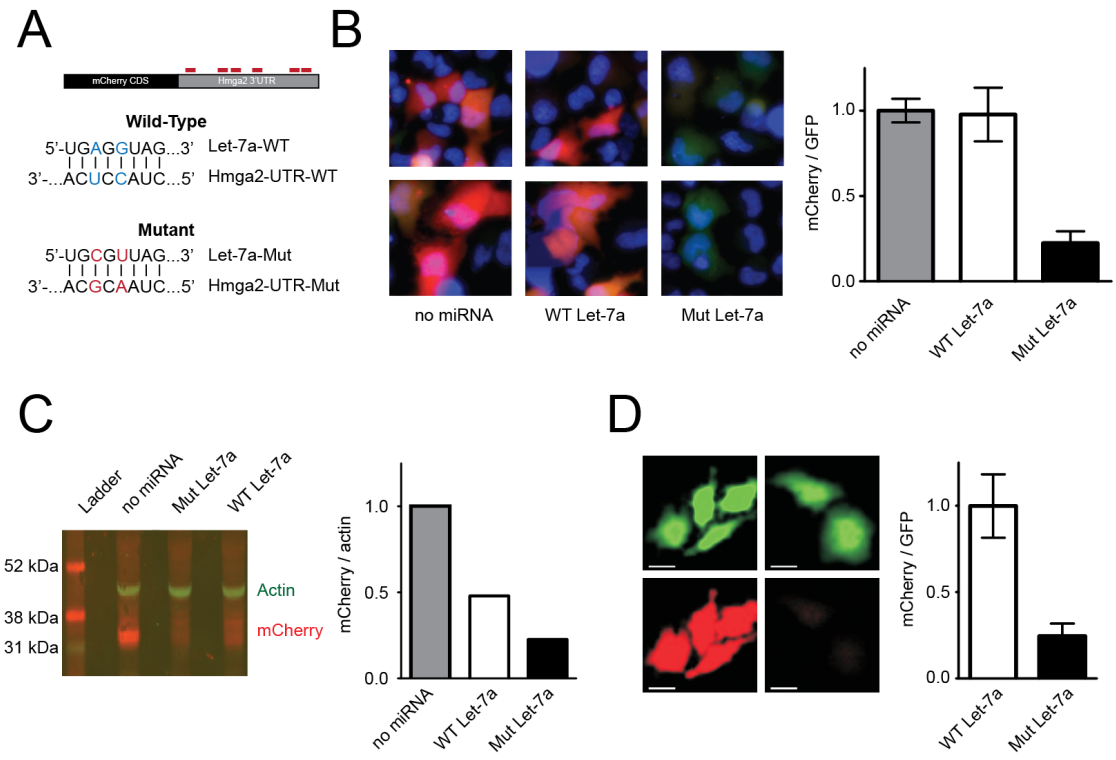


Figure 4.1 Fluorescence-based miRNA repression assay. (**A**) Schematic of mCherry reporter with Hmga2 3'UTR, which has x7 seed sequences for let-7 miRNA (positions indicated with red blocks). Two point mutations in let-7 seed sites were introduced to make the reporter sensitive to an orthogonal mutant miRNA. (**B**) *Left,* Representative composite images of mCherry (red) and GFP (green) fluorescence in cells co-transfected with a mCherry-Hmga2 mutant reporter, EGFP control plasmid, and the indicated miRNA. Nuclei are stained with DAPI (blue) for visualization. Two examples are shown for each group. *Right*, quantification of mCherry fluorescence relative to GFP, normalized with respect to 'no miRNA' sample. (N = 2 independent trials, 75 cells per group) Error bars, SEM. (**C**) Quantification of mCherry-Hmga2 mutant reporter expression by western blot. Endogenous actin was used as a loading control. (**D**) Representative images of GFP fluorescence (top) and mCherry fluorescence (bottom) in cells co-injected with an mCherry reporter plasmid, a GFP control plasmid and, either the wild-type (WT) or mutant let-7-a1 (MUT) miRNA are shown. Scale bar, 20 µm. (*C*) Quantification of mCherry fluorescence relative to GFP fluorescence, normalized with respect to the WT sample (N = 3 independent trials, 50 cells per group). Error bars, standard error of the mean.

Intracellular colocalization of miRNA and messenger RNA

Diffusion analysis of injected wt-let-7a miRNA revealed two normally distributed populations 4 h after microinjection: a more slowly diffusing population, with a mean diffusion constant of ~0.0034 $\mu m^2 s^{-1}$, and a faster moving population, with a mean diffusion constant of ~0.26 $\mu m^2 s^{-1}$ (**Fig. B.3.G**). These diffusion rates are similar to those of P-bodies (PBs), cytoplasmic sites of RNA degradation[165], and mRNA ribonucleoprotein complexes (mRNPs)[162]; thus, suggesting that the tracked miRNAs were assembled in these two respective complexes. Indeed, by co-expressing GFP-Dcp1a, a decapping protein that localizes to P-bodies, we could observe colocalization of miRNA and P-bodies (**Fig. B.3.A**).

To investigate the interactions with mRNPs, we utilized the MS2 RNA labeling method developed by the Singer group[162]. A luciferase reporter construct (mG-mH3UM-MS2) was engineered with the mutant Hmga2 3'UTR followed by 24 tandemly repeated MS2 hairpin-loops (**Fig. 4.2.A**). These loops acted as binding sites for the MS2 coat protein (MCP), which was co-expressed as a GFP fusion such that binding of multiple MCP-GFPs to the reporter construct enabled its detection above background (**Fig. 4.2.A-B**). To decrease background, the MCP-GFP was appended with a nuclear localization signal (NLS), thus removing most non-bound protein from the cytoplasm.

We generally experienced poor expression of these constructs with very few (<1%) cells having a green nucleus as evidence that the MCP-GFP-NLS is being expressed (**Fig. 4.2.B**). In these few cells, well resolved cytoplasmic and nuclear particles could be detected, and these were verified to contain the MS2-mRNA by fluorescence *in situ* hybridization (FISH) (**Fig. 4.2.B**).

Next, we measured the extent of colocalization with miRNA. Due to poor expression, we made two modifications that were slight improvements: first, the MCP-GFP promoter was swapped for a slightly stronger variant, and second, rather than using transfection that gave variable results, the MS2/MCP plasmids were microinjected into the nucleus in order to deliver them at a more consistent quantity and ratio. The latter step had the added convenience that the fluorophore-labeled miRNA could be introduced simultaneously since nuclear-injected miRNA was confirmed to be functional (**Fig. 4.1.D)**. At 24 h post-injection, several cells could be imaged that contained both the

miRNA and expressed, labeled mRNA (**Fig. 4.2.C**). Following deconvolution, an image processing method that quantitatively removes out of focus light, colocalization of the mut-Hmga2 mRNA with mut-let-7a could be qualitatively discerned by overlaying the two channels (**Fig. 4.2.C**). For quantitative analysis, we used an object-based method[252]. Images were segmented into objects by thresholding, and the sub-pixel localizations of these objects' centers of mass (COMs) were determined. Objects were then assigned as colocalized taking into account that, due to the diffraction limit of light, two objects cannot be resolved, and are thus considered colocalized, if their COMs are separated by a distance less than ~1/2 the emission wavelength (~250 nm for GFP) (**Fig. 4.2.C**). Using this approach, we observed miRNA-specific colocalization: ~18% of mut-let-7a-Cy5 was determined to be colocalized with the mut-Hmga2-GFP mRNA, compared to ~7% for wt-let-7a-Cy5 and <3% for control cells not expressing mRNA (where colocalization simply arises from coincidental colocalization background) (**Fig. 4.2.D**). The wt- and mut-let-7a samples had comparable particle densities, thus ruling out the possibility that increased colocalization was due to coincidence alone (**Fig. 4.2.E**). Furthermore, statistical analysis using simulated images showed that colocalization was non-random for the matched mutant mRNA-miRNA samples ($p = 0.0138$). The wt-let-7a-Cy5 likely shows a higher colocalization than background due to the fact that it has many endogenous targets; this may lead to its localization to common cellular RNA foci, such as PBs, where the mutant mRNA may also reside due to either its regulation by endogenous factors or natural decay.

Lastly, we combined the colocalization data with stepwise photobleaching analysis to reveal the types of complexes that were being formed between miRNA and mRNA. For mut-let-7a-Cy5, colocalized spots contained more multimers (~62%) than monomers (38%) while independent, non-colocalized spots had nearly an even distribution (**Fig. 4.2.F**). Taken together, these data indicate that colocalized miRNA:mRNA are interacting in complexes that can contain more than one miRNA bound to a single mRNA. Alternatively, they may be assembled into larger degradation complexes, such as PBs, that contain multiple mRNAs and multiple miRNAs.

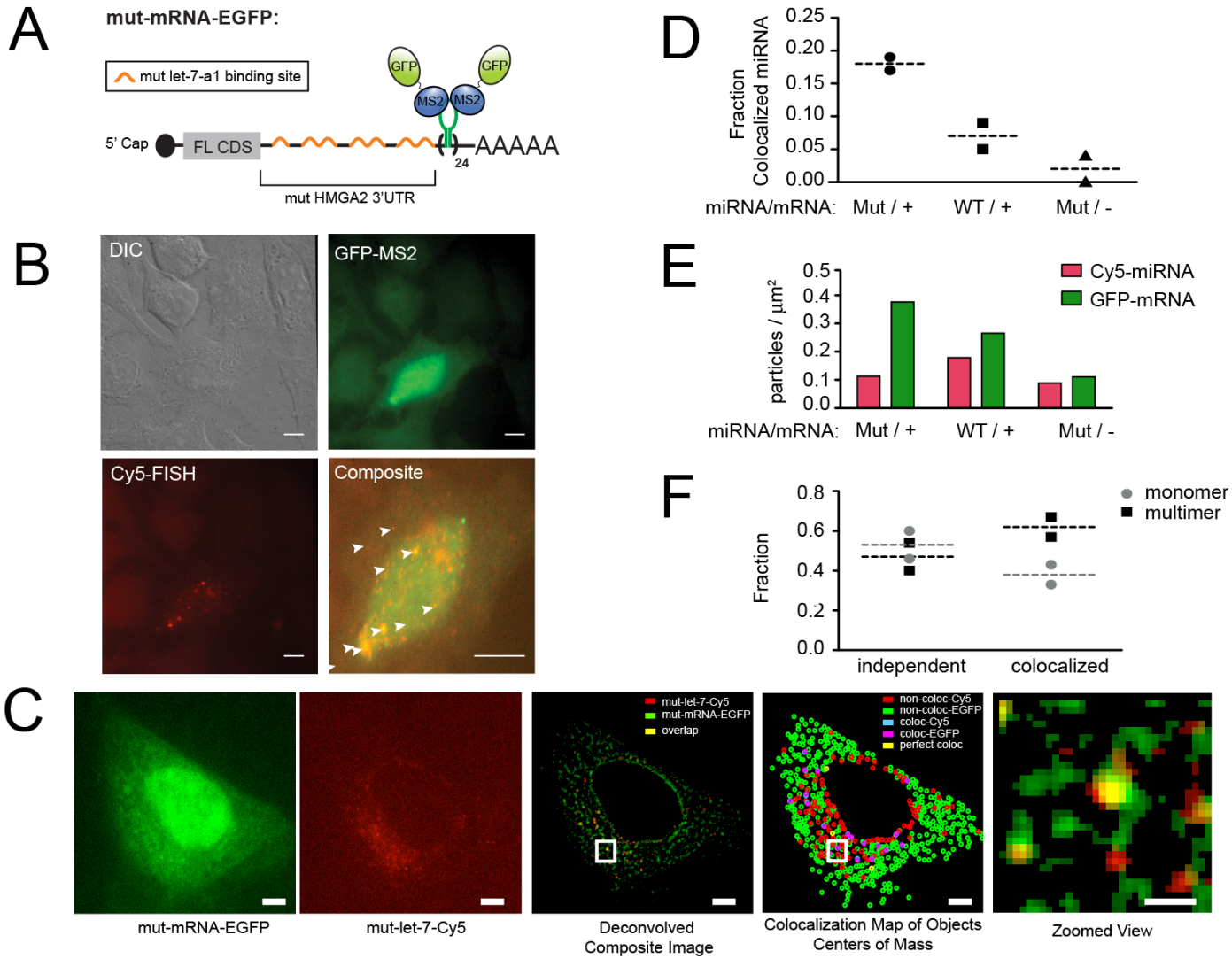

A
mut-mRNA-EGFP:
mut let-7-a1 binding site
GFP
GFP
MS2
MS2
5' Cap
FL CDS
AAAAA
24
mut HMGA2 3'UTR
B
DIC
GFP-MS2
Cy5-FISH
Composite
C
mut-mRNA-EGFP
mut-let-7-Cy5
mut-let-7-Cy5
mut-mRNA-EGFP
overlap
Deconvolved
Composite Image
non-coloc-Cy5
non-coloc-EGFP
coloc-Cy5
coloc-EGFP
perfect coloc
Colocalization Map of Objects
Centers of Mass
Zoomed View
D
Fraction
Colocalized miRNA
0.25
0.20
0.15
0.10
0.05
0.00
miRNA/mRNA:
Mut / +
WT / +
Mut / -
E
particles / µm²
0.5
0.4
0.3
0.2
0.1
0.0
Cy5-miRNA
GFP-mRNA
miRNA/mRNA:
Mut / +
WT / +
Mut / -
F
Fraction
0.8
0.6
0.4
0.2
0.0
monomer
multimer
independent
colocalized

Figure 4.2. Injected let-7a colocalizes with MS2-labeled reporter transcripts. (**A**) Schematic of the MS2 reporter mRNA containing a firefly luciferase coding sequence (FL CDS), mutant Hmgs2 3'UTR, and 24 tandem repeats of the MS2 hairpin-loop. When co-expressed, up to 48 MS2(coat protein)-GFP fusions can bind to the reporter hairpin loops, enabling the transcript to be detected by fluorescence microscopy. (**B**) Confirmation of MS2 reporter expression with fluorescence in situ hybridization (FISH). Cells transfected with the MS2 plasmids were hybridized with Cy5-labeled probes antisense to the MS2 hairpin loops. Several colocalizing particles (yellow; marked with white arrows) of GFP (green) and Cy5 (red) could be observed in a cell expressing the MS2-GFP fusion and reporter, identifiable by the GFP signal in the nucleus. Scale bars = 10 μm. (**C**) Injected Cy5-labeled mutant let-7 colocalizes with the reporter mRNA. From left: (1) Unprocessed image of GFP (green) fluorescence from a cell injected with the MS2 plasmids and mut-let-7-Cy5; (2) Unprocessed image of Cy5 (red) fluorescence from the same cell; (3) GFP (green) and Cy5 (red) images were deconvolved to reduce background and overlaid into a composite image. Yellow particles indicate co-localization. (4) Co-localization map representing independent miRNA (red) and mRNA (GFP) particles, and either perfectly co-localized (yellow) or adjacent miRNA-mRNA localizations with an inter-particle center-center distance less than the lateral resolution limit of the image (cyan and magenta). (5) Zoomed view of the region in the preceding images marked by a white box. Yellow circles indicate co-localized spots. Scale bars = 5 μm (1-4) or 1 μm (5). (**D**) Fraction of mutant (mut / +; circle) or wild-type (WT / +; square) miRNA colocalizing with the mutant reporter. To test for the contribution of background fluorescence, control cells were injected with mutant let-7-a and MS2-GFP plasmid without the MS2-mRNA expression plasmid (mut / - ; triangle). Data points within each group are explicitly shown and their associated means (dashed lines) are shown (N = 2 cells for each group). (**E**) Mean densities of detected Cy5 (red bars) and GFP (green bars) particles for images used for quantifying colocalization in **D**. (**F**) Distribution of monomers and multimers of mutant let-7-a miRNA non-colocalized (independent) or colocalized with MS2-GFP tagged target mRNAs. The coordinates of stepwise-photobleached particles in the mut-let-7-Cy5 channel were mapped to the mut-mRNA-GFP channel to assess the extent of colocalization. N = 2 cells containing a total of 60 independent and 16 colocalized let-7-a particles. Data points within each group are explicitly shown (grey circles, monomers; black squares, multimers) in addition to their mean values (grey dotted line, monomers; black dotted line, multimers). The statistical variation within each group was minimal between replicates.

Dcp1a and Ago2 tracking suggest miRISC assembles into intermediates of P body formation

Argonaut (Ago) proteins are the central protein components of the miRNA pathway. They directly bind and unwind mature duplex miRNAs[254], stabilize target binding[43], and recruit additional factors to effect target repression[255]. Ago also serves to protect miRNA from nuclease degradation[256]. Therefore, it is likely that all functional, mature miRNAs are escorted by Ago from the time they are handed off by Dicer until they are released for decay.

We thus sought to measure the dynamics and localizations of Ago to compare it with miRNA. Transiently expressed Ago2-EGFP localized into bright granules the diffusion of which fell into three Gaussian populations with mean diffusion coefficients of ~0.0264 $\mu m^2 s^{-1}$, ~0.00313 $\mu m^2 s^{-1}$, and ~1.81e-5 $\mu m^2 s^{-1}$ (**Fig. 4.3.A**, *top*). These coefficients are highly similar to those of Dcp1a (**Fig. B.S.2.C**), and in agreement with several previous reports of Ago2 colocalizing with P body complexes[257-261]. Alignment of the histograms for Ago2 and miRNA tracking at 4 h post-injection, when multimer formation peaks (**Fig. B.2.E**,*top*), revealed that the slow diffusing population of let-7a, attributed to association with P bodies, matched the fastest diffusing population of Ago2 (**Fig. 4.3.A**, *top and middle*). However, the fastest let-7a population, attributed to incorporation into mRNPs, was not detected for Ago2. This is likely due to the fact that these complexes have few Ago protein molecules, which when combined with the fast diffusion and the high background at the wavelength of EGFP emission would make them difficult to detect with the sensitivity of our current microscope. Surprisingly, the most prevalent Ago2 population (~0.00313 $\mu m^2 s^{-1}$) was not explicitly observed for let-7a (**Fig. B.2.G**). Reexamination of the original let-7a data set provided evidence, however, for a third Gaussian at 1.38×10^{-3} $\mu m^2 s^{-1}$, and that this actually fit slightly better to the data (R^2 = 0.9862 versus 0.9816); this Gaussian population is, however, still narrower and slightly offset compared to that of Ago2 (**Fig. 4.3.A**, *top and middle*).

Recently, we have begun testing a new particle tracking routine based on Jaqaman et al.[251], which detects particles with Gaussian fitting and determines a particle's localization based on the peak of its fit. This method may be more robust than our current routine that uses a center of mass (centroid) based method to detect particles. Reanalysis

of the same let-7a images at 4 h post-injection with this new method resolved three well fit (R^2 = 0.9850) Gaussian populations: a slow population (~ 0.1740 $\mu m^2 s^{-1}$), an intermediate population (~0.0230 $\mu m^2 s^{-1}$), and a fast population (~0.0053 $\mu m^2 s^{-1}$) (**Fig. 4.3.A**, *bottom*). Importantly, the new tracking routine maintained the fast diffusing population from the original results, but split the previous 'slow' population into two distributions that better agree with the Ago2 results. Therefore, we may observe three populations of miRNA, not two: a fast population that diffuses similarly to mRNPs, and an intermediate and slow population which both diffuse similarly to Dcp1a/Ago2.

The exact nature of the Dcp1a and Ago2 populations is unclear. Based on similar distributions between Dcp1a and Ago2 and their known colocalization, we hypothesize that all the observed populations are P body-like. The differences in diffusion coefficients are then likely due to differences in size, as we observe a broad size distribution in detectable Dcp1a foci (**Fig. 4.3.B**). According to the Stokes-Einstein equation, objects with smaller radii undergo faster Brownian diffusion than objects with larger radii, and we generally observe this trend (although we have not yet directly quantified the correlation of size and diffusion for P bodies). We propose that smaller foci may actually be immature P bodies or intermediate complexes. In line with this hypothesis, we often observe smaller Dcp1a foci merging with larger foci, suggesting that these complexes may be delivering cargo to the larger assembly (**Fig. 4.3.C**).

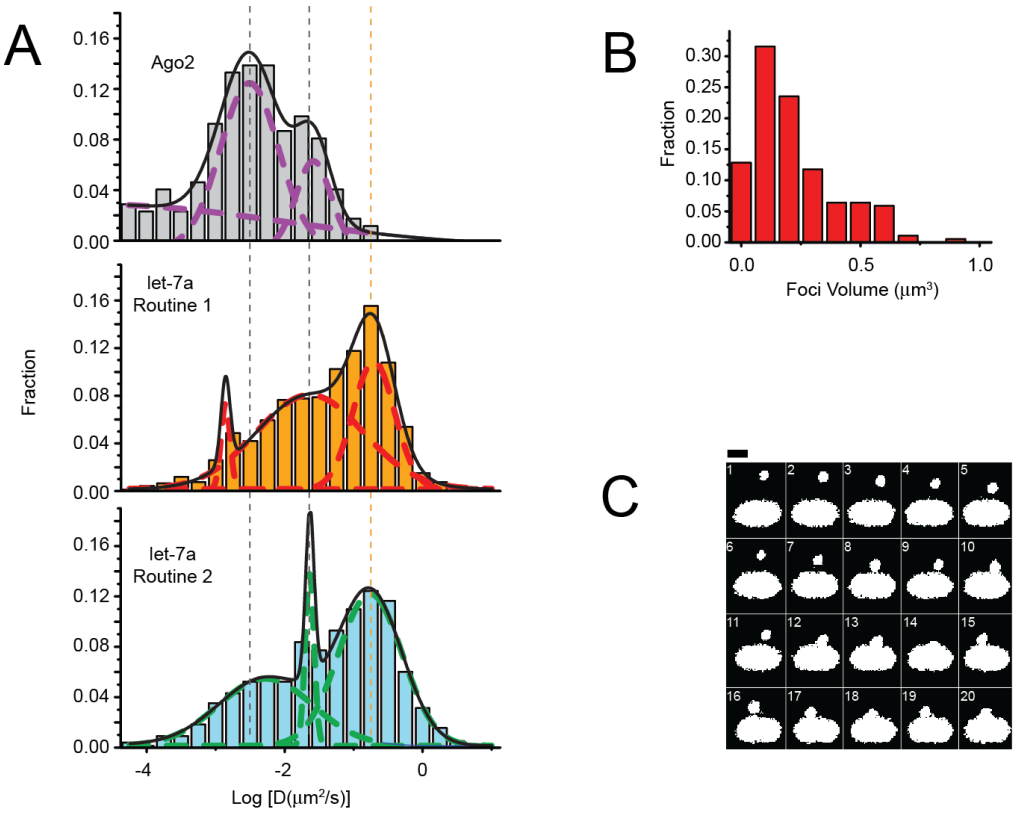


Figure 4.3 Ago2 tracking suggests the presence of a smaller intermediary miRISC/P body complex. (**A**) Comparison of Ago2 and miRNA diffusion coefficient distributions. *Top*, Histogram of the distribution of diffusion coefficients for Ago2-EGFP particles, calculated assuming Brownian diffusion. The distributions can be fit (purple dashed lines) to three normal populations (R^2 = 0.974). The median bins for the two most prevalent populations are demarcated with a grey dotted line to facilitate comparison with the other histograms. *Middle*, let-7a diffusion histogram from the 4h timepoint (**Fig. B.2.G**), fitted with three Gaussian functions (red dotted lines, R^2 = 0.9862) instead of two (see **Fig. B.2.G)**. These data are the result of centroid-based tracking (Routine 1). The median bin for the fastest population is demarcated with a dashed yellow line. *Bottom*, the same 4h images were reanalyzed with a Gaussian-based tracking method (Routine 2) and fitted into three populations (green dashed line, R^2 = 0.9850). (**B**) Volumetric measurements of Dcp1a-containing foci. Z-stack images of fixed cells expressing RFP-Dcp1a were acquired and 3D volumes of detected particles were calculated using Huygens Software. Ago2 foci, which colocalizes with Dcp1a, shows a similar distribution. (**C**) Montage of a smaller Dcp1a particle sampling and docking with a larger particle. Each frame is a 0.1 s snapshot taken every 2.0 s (20 frames/40 s total). Images were thresholded and made binary for visualization. Scale bar (black bar above top left hand corner) = 1 μm.

Inhibited miRNA and miRISC seldom form trackable foci

That microinjected miRNA can repress reporter gene expression and colocalizes with on-pathway complexes demonstrates the biological relevancy of our approach. For additional confirmation, we performed several additional control experiments that include inhibition of injected miRNA and mutation of Ago2.

Microinjection of let-7a-Cy5 with locked-nucleic acid (LNA)-modified DNA antisense (LNA/DNA) anti-miR resulted in complete loss of punctate formations 4 h post-injection (**Fig. 4.4.A**). Importantly, mismatched anti-miR control did not decrease the number of foci formed (**Fig. 4.4.A**, *left*). Additionally, fluorophore labeled anti-let-7 and anti-miR control did not themselves form foci, as expected (**Fig. 4.4.A**). This indicates that inhibited miRNA or modified RNA is not recognized by miRISC and suggests that only functional miRNA can assemble into complexes detectable as foci.

Rüdel et al.[262] have reported that a tyrosine in the 5' miRNA binding pocket in the mid-domain of Ago2 can be phosphorylated by an as-yet unknown kinase, resulting in a loss of miRNA-binding capacity due to charge-charge repulsion. This phosphorylation can be constitutively mimicked by an Y529E mutation, causing both loss of miRNA binding and localization to P bodies. We have now reproduced these results using our system (**Fig. 4.4.B**). As mentioned above (4.3.3), transient expression of Ago2-EGFP in HeLa cells induces formation of bright P body-like cytoplasmic punctae (**Fig. 4.4.B**, *left*). The number of foci per cell varies widely, with most showing 10-14 particles. In order to test whether these foci contain miRNA, as suggested by the overlap between our miRNA and Ago2 tracking results (**Fig. 4.4.A**), we expressed the Ago2-Y529E mutant for comparison. Unlike wild-type, expression of the Ago2-Y529E mutant forms few foci; instead, fluorescence was more evenly distributed throughout most cells (**Fig. 4.4.B,** *right*). Foci counting reveals, however, that the mutation has incomplete penetrance for foci formation, which was not previously reported[262]. The great majority of cells had 0-4 particles, but a fraction (~32%) had several more (**Fig. 4.4.B,** *right*). While the reason(s) behind this cell-to-cell variation are unclear, we conclude that the majority of tracked Ago2 is probably bound to miRNA or in miRNA-related complexes since a mutant Ago2 incapable of loading miRNA only irregularly assembles into trackable particles.

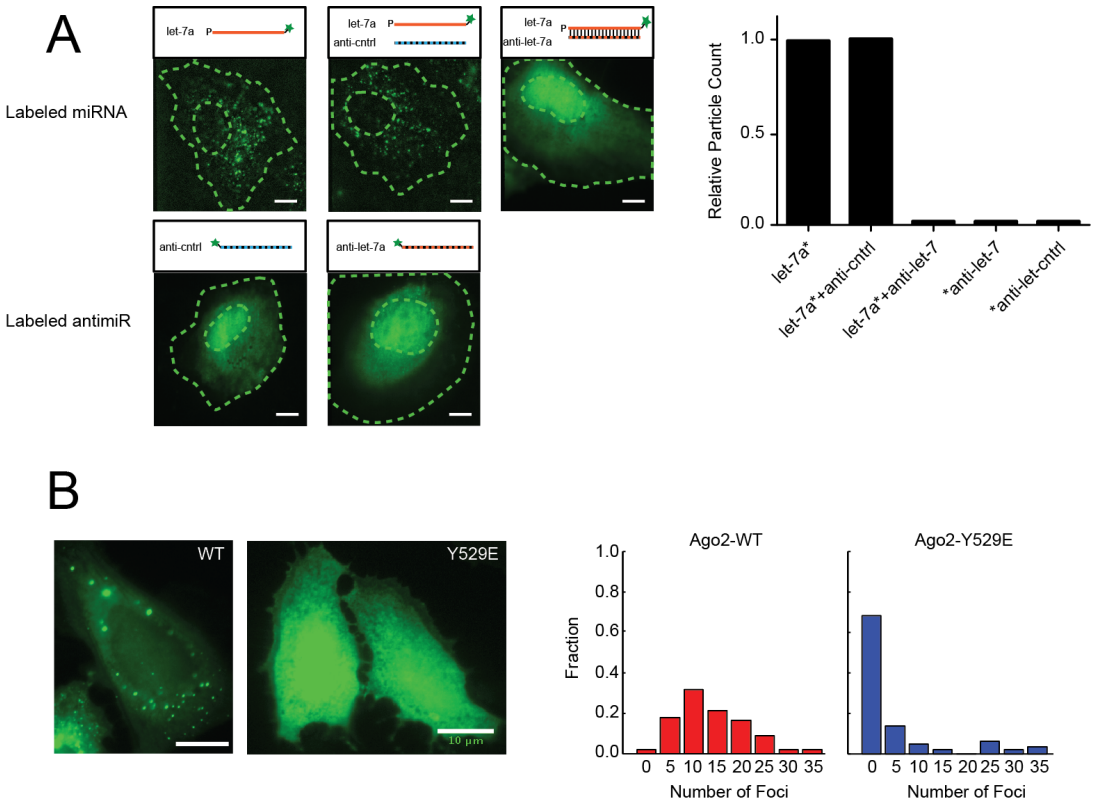


Figure 4.4 Inhibition of miRNA or miRISC alters their subcellular distributions. (**A**) Anti-miR inhibited miRNA does not form well-resolved particles at 4h post-injection. *Left top row,* HeLa cells were microinjected with let-7a-Cy5 alone or in the presence of anti-let-7 oligonucleotide or mismatched anti-control. *Left bottom row,* cells injected with labeled anti-let-7 or anti-control. *Right*, The relative number of detectable particles for each image to the left normalized to let-7a-Cy5 alone. Images were analyzed with Imaris software, using the same quality threshold. (**B**) miRNA loading impaired Ago2-Y529E mutant forms fewer particles than Ago2-WT when expressed as a GFP fusion. *Left*, representative images of Ago2-WT and Ago2-Y529E. *Right*, Histograms of the number of Ago2-WT or Ago2-Y529E foci detected per cell (N = 67 cells for wt; N = 66 cells for Y529E). Numbers on the x-axis below each bin represent the bottom edge for that bin.

Synthesis and testing of Cy3AsH: a potential new tool for iSHiRLoC

To apply iSHiRLoC to answer questions on how miRNAs effect translation requires new fluorescence tools. GFP and its derivatives are the workhorses of modern-day fluorescence microscopy. The ability to clone and express, with relative ease, a fluorescent protein in the same open reading frame (ORF) as virtually any other protein of interest has enabled the determination of the dynamic intracellular localization of perhaps thousands of proteins. Yet, fluorescent proteins have their drawbacks. At a mass of nearly 30 kilodaltons (kDa) (238 amino acids), they can sterically interfere with protein-protein interactions and obfuscate endogenous protein localization [263]. Moreover, GFP-based chromophores, which are autocatalytically formed by oxidation in the center of a beta-barrel tertiary structure, can take at least several minutes to mature even after the full-length protein is translated[264]. As a consequence, neither the localization of protein synthesis nor its rates can easily be studied intracellularly using fluorescent proteins[265].

A new class of biarsenical fluorophores has been developed with the intent of eliminating these problems[266]. These fluorophores are small molecular weight organic dyes modified to recognize genetically encoded tags that are orders of magnitude shorter than fluorescent proteins. The approach is based on the ability of organoarsenicals to form reversible covalent bonds bridging two adjacent thiols. This property, ironically, is also responsible for the toxic effects of arsenic: pyruvate catabolism and ultimately ATP production are inhibited by arsenic binding to the dithiolane ring of lipoic acid, a cofactor of pyruvate dehydrogenase[267]. To prevent this toxic effect and impart specificity to the labeling method, organic dyes are appended with dual *cis*-facing arsenic handles, and the protein to be labeled is expressed with an engineered tetracysteine tag containing two pairs of cysteines spaced at the same distance from one another as are the arsenic handles on the fluorophore. This arrangement allows for cooperative binding of the fluorophore to the tag with an affinity great enough that the antidote for arsenic poisoning, ethanedithiol, can be coadministered during labeling to prevent cytotoxicity and non-specific binding.

The first biarsenical fluorophore developed was based on the common green fluorescent dye fluorescein, and was termed FlAsH or fluorescein arsenical helix binder

since the tag, consisting of six amino acids in the configuration 'CCXXCC', was believed to form an α-helix structure[266]. Further improvements to the FlAsH tag, however, found the optimal motif to be 'CCPGCC', which is more consistent with a β-turn[268] and NMR studies have confirmed this to be the case[269]. Despite the misnomer, a second biarsenical compound, ReAsH, was subsequently developed that recognizes the same FlAsH tag but is red in fluorescence[268].

FlAsH and ReAsH, facilitated by their commercial availability, have been used in a number of different applications. In addition to their small size, another advantage of these fluorophores is that they are cell permeable and exhibit a several-fold enhancement in fluorescence only after tag binding; therefore, by pre-incubating cells in the dye, transient expression events or changes in protein conformation can be detected nearly instantaneously with minimal background[270,271].

Most relevant to our studies, ReAsH has been used for intracellular live-cell localization of active translation sites[265]. In this method, a TC tag was engineered just after the start codon in a reporter construct encoding a GFP-β-actin fusion protein. Because only the TC tag had to be translated, and not the full-length protein, sites of translation could be detected as cytoplasmic punctae before the nascent chain was released. Importantly, the formation of these punctae were sensitive to translational inhibitors: puromycin, a translation initiation inhibitor, caused rapid drop-off of punctae, while cycloheximide, a translation elongation inhibitor, increased the persistence of punctae as if translation were frozen in place.

Translation site labeling combined with colocalization and tracking of miRNA and mRNA could provide for direct kinetic measurements of translational repression and simultaneous determination of the inhibited steps (initiation or elongation). However, these experiments would require single-molecule localizations, which in-turn requires fluorophores with optimal photophysical properties. FlAsH and ReAsH have relatively low extinction coefficients and fast photobleaching and/or –blinking rates, and are largely untested in this application. ReAsH localization has been measured with high precision (5 nm) *in vitro*[272], but it is unclear whether this is translatable to inside cells where far more photons need to be detected in order to overcome higher backgrounds from cellular autuofluorescence. Translation site labeling with ReAsH reported by Rodriguez et al.[265]

appeared to produce well resolved, bright particles, but their localization analysis was only semi-quantitative, thus providing little benchmark to judge its potential.

Cao et al.[253] have reported on the synthesis of a promising new biarsenical, Cy3AsH, which features arsenic handles appended to a scaffold Cy3, a commonly used fluorophore for single molecule methods including iSHiRLoC. This dye has improved photophysical properties over FlAsH/ReAsH including a higher extinction coefficient and brightness (quantum efficiency×extinction coefficient) as well as increased photostability. Although its fluorescence enhancement upon binding is several fold lower, this may be less a characteristic of the fluorophore than its tag —CCKAEAACC— which, compared to the FlAsH tag, is extended between the cysteines by three amino acids in order to account for the longer inter-arsenic distance on Cy3AsH. The Cy3AsH tag, unlike the FlAsH tag, has not gone through extensive rounds of modification[273], thus there may be significant room for improvement.

Given its potential, we wanted to test Cy3AsH's fluorescence properties and its ability to specifically label tagged proteins. Since it is not commercially available, we synthesized the molecule using a modified four-step procedure (**Fig. 4.5**) based on that of Cao et al.[253]. We were unable to produce the Cy3 intermediate (*compound 2*) in the second reaction using the previously described acidic buffer conditions (acetic acid/acetic anhydride mix), but instead were successful using pyridine as the solvent [274]. We also had to alter the conditions for chromatographic purification following the transmetallation of the final product, substituting the methanol/toluene (1:2 v/v) mobile phase with methylenechloride/methanol/water (20:6:1 v/v). Yields for the first three steps were moderate at 60 − 83%, however, the final product was recovered at much lower rates (3%), likely due to two rounds of purification necessary to eliminate an additional spot on the thin layer chromatography (TLC).

Previous testing of Cy3AsH labeling was done using a short helical peptide with a terminal TC tag[253]. For preliminary testing, we wanted to see whether Cy3AsH could be used to label a full-length protein with a more complex structure. We chose to use the 100 kDa bacterial translation initiation factor IF2 since it is easy to express recombinantly and purify, and engineered the Cy3AsH-TC tag 'KLCCKAEACCKA' at the C-terminus (**Fig. 4.6.A**). Our Cy3AsH had a maximum absorbance at λ = 565 nm and

a maximum emission at λ = 571 nm. Addition of IF2-TC protein resulted in a bathochromic shift in fluorescence of 11 nm (**Fig. 4.6.B-C**). Titration of Cy3AsH with increasing concentrations of IF2-TC in the presence of 100 μm EDT resulted in a maximum enhancement of ~5-fold and an apparent K_d = 6.31±0.99 μM (**Fig. 4.6.B-C**). Dye binding was highly stable and not completely reversed with reducing SDS-PAGE (**Fig. 4.6.D**).

Next, we tested the specificity of labeling. Cy3AsH was incubated with either 5 or 20 μM IF2-TC, or unlabeled IF-2 or bovine serum albumin (BSA) as controls, and run on SDS-PAGE (**Fig. 4.6.E**). The fraction of bound Cy3AsH was quantified by measuring the intensities of bands containing bound Cy3AsH (slow migrating) versus non-bound free Cy3AsH (fast migrating) (**Fig. 4.6F**). IF2-TC at 20 μM concentration showed the greatest extent of labeling (~26%), approximately 5-fold greater than non-tagged IF2 (~5%). Unexpectedly, BSA control at the same concentration showed a similar binding efficiency as IF2-TC (~20%). Fluorescence scans, however, show a maximal emission at λ = 596 nm in the presence of 20 μM BSA, which is 14 nm and 25 nm red-shifted to the maximal emissions for IF2-TC and no protein, respectively. This suggests that Cy3AsH has a different mode of binding to BSA compared to IF2-TC.

We conclude that Cy3AsH can be used to fluorescently label TC tagged proteins and that binding causes a change in its spectral properties, causing a red-shift and modest enhancement of emission. However, labeling is not tag-specific as considerable amounts of Cy3AsH could also bind non-tagged proteins.

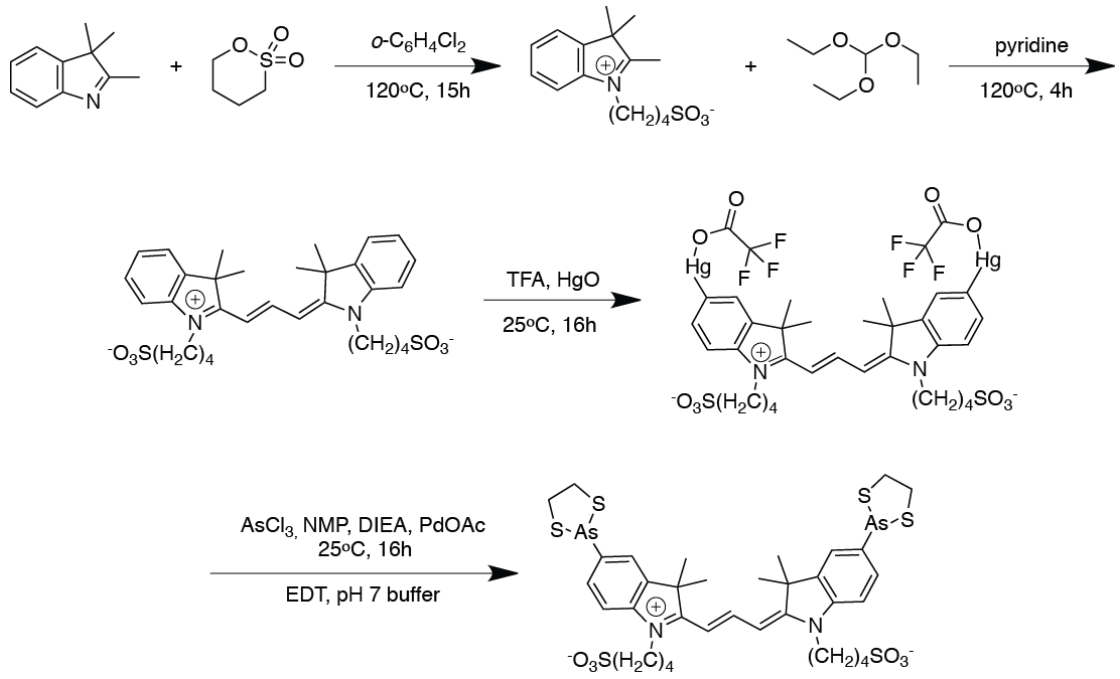


Figure 4.5 Scheme of Cy3AsH Synthesis.

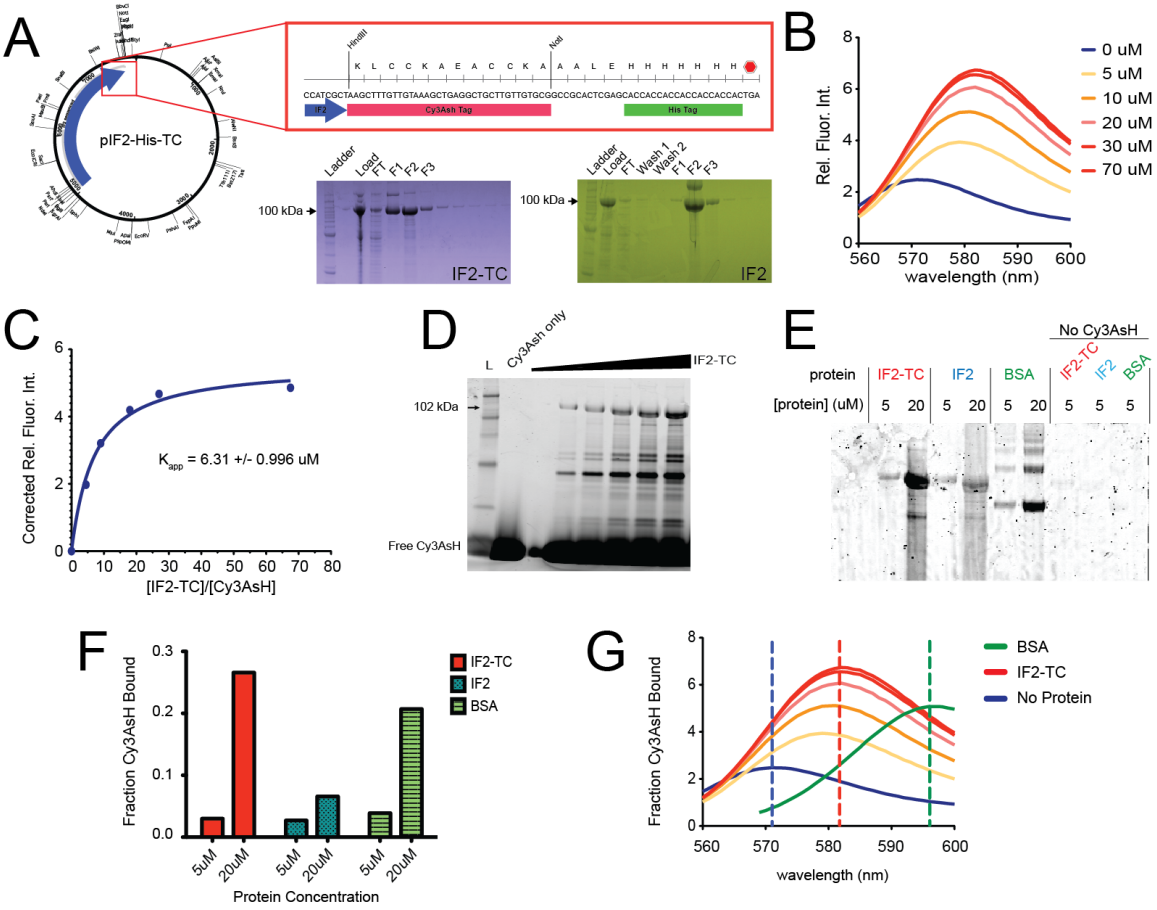


Figure 4.6 Testing Cy3AsH labeling of tetracysteine-tagged protein. (**A**) IF2 and IF2-TC plasmid cloning and protein purification. The Cy3AsH-TC tag was cloned downstream of the IF2 CDS before the histidine tag. Both IF2 and IF2-TC were purified by the histidine tag with nickel-sepharose. Fractions collected during purification of each protein are shown resolved by SDS-PAGE. (**B**) Fluorescence wavelength scan of 1 µm Cy3AsH in the presence of varying concentrations of IF2-TC. (**C**) Maximum intensities from ***B*** plotted as a function of IF2-TC concentration. Intensities are baseline corrected for non-zero fluorescence in the absence of protein. Data points are fitted with a non-linear regression (see Materials and Methods). (**D**) SDS-PAGE of Cy3AsH labeled IF2-TC. Labeled full-length protein is observed at 102 kDa. Additional bands are likely protein decay products resulting from harsh labeling conditions. (**E**) SDS-PAGE of Cy3AsH labeling reactions containing 1 µm Cy3AsH and 5 µm or 20 µm IF2-TC, IF2, or BSA. (**F**) Quantification of Cy3AsH bound to each protein in the gel shown in ***E***. The fraction bound was calculated as the summed density of slow migrating bands divided by the sum of all bands (including free dye). (**G**) Fluorescence wavelength scan of Cy3AsH in the presence of no protein (blue), IF2-TC (yellow-to-red), or BSA (green). The maximum emission wavelengths are marked with dashed lines.

4.4 Discussion

Summary

Single molecule analysis can access depths of biological information previously inaccessible by ensemble methods[275]. We have now developed and implemented, in iSHiRLoC, a single molecule analysis method for studying miRNA in real-time, in live-cells. In its current form, iSHiRLoC is able to distinguish kinetically distinct steps in complex assembly and disassembly of a miRNA[247]. We have also characterized the diffusion of Dcp1a (P bodies) and Ago2, two important pathway components. We envision future experiments where these various components are tracked simultaneously, and their interactions are timed and mapped to build a systems biology framework[276] for understanding miRNA processes.

Protein repression assay

Specific repression of an mCherry-Hmga2 mutant reporter by seed-matched injected miRNA demonstrated that microinjection does not hamper cellular RNA silencing and that duplexed mature miRNA can be effectively incorporated into repressive complexes. At 24 h post-microinjection, we measured ~70% repression of mCherry levels. In order to place our complex assembly kinetics into context, it will be important to repeat these measurements at earlier time points to determine when the onset of repression occurs. Based on our kinetics, we predict this may be as soon as the 2 h time point, when miRNA was first reliably detected in high molecular weight RNPs. However, the reporter assay in its current form may be limited at these early time points as the rate of protein expression and maturation, to the levels required for detection, may be slower than the kinetics of repression. Indeed, we were only able to observe faint mCherry fluorescence as early as ~2.5 h post-injection in the absence of miRNA. GFP may even take longer to detect due to its even slower maturation[264,277]. A better approach would be to inject miRNA into cells already expressing the reporters at steady state. However, because fluorescent proteins have long half-lives (~24 h), moderate attenuation to protein output would have little if any effect on steady-state levels, making the assay insensitive. One possible

solution is to destabilize the fluorescent proteins; while another is to adopt a luminescence based reporter gene system suitable for microscopy.

Colocalization with MS2-labeled mRNA

Using a quantitative object-based analysis, we measured sequence-specific colocalization of fluorophore labeled mutant let-7a with MS2-labeled mutant Hmga2 reporter transcripts. These data support the conclusion that microinjected miRNAs bind targets and assemble into mRNPs. In addition, stepwise photobleaching showed that colocalized miRNA was more often multimeric in nature than non-colocalized miRNA. Due to the high number of GFPs (~30-40 [162]) loaded onto mRNAs containing 24 MS2 hairpin-loops, stepwise photobleaching cannot be used to determine the ratio of miRNA:mRNA (the upper limit for photobleaching is approximately 7-8 before the steps are no longer discernible). Since Hmga2 contains seven let-7 seed sites, it may be expected that several miRNAs could be bound to a single transcript; however, while the presence of multiple seed sites can enhance repression[220], this enhancement is often simply additive[82,173]. This lack of cooperativity suggests that multiple sites enhance repression by increasing the chance for a single binding event, rather than enabling cooperative binding. An alternative possibility is that the colocalized multimers are in sites with several target transcripts, like P bodies. Immunofluorescence against P body proteins could be used to further character these complexes. In addition, MS2-labeled mRNPs can be co-tracked in live cells with miRNA or Dcp1a.

Identification of an intermediary complex

In agreement with their colocalization in P bodies, Dcp1a and Ago2 exhibited similar diffusive properties when expressed as fluorescent fusions. Tracking revealed that for both proteins there are at least two populations. We attribute these differences in diffusion to size, since both small and larger foci are detected, but we expect there may be additional distinctions in the molecular architecture of these complexes — they may consist of different factors or stoichiometry of factors. While a large list of P body components has been identified[145], it is still unclear what exactly makes a P body a ‘P

body'; are they consistent assemblies with defined factors or is there heterogeneity between the granules? Consistent with the latter possibility, Poly(-rC) Binding Protein 2 (PCBP2) and TNGW1, an isoform of GW182, have been shown to only localize in a subset of P bodies with unknown determinants[278,279]. We have observed that smaller Dpc1a foci interact and even join with larger foci, and we speculate that the smaller versions may be immature aggregates or satellite assemblies that form around and transport destabilized mRNPs to be degraded in the larger sites. We also note the possibility that some of these foci, both small and large, may be non-functional aggregates resulting from transient protein overexpression[280]. It will be important to see if both the distributions in diffusion as well as size persist in stably transfected cell lines selected for low-to-moderate expression more akin to endogenous levels.

Comparing Ago2 diffusion with that of miRNA, we found only one corresponding population — the slow miRNA population — matched that of the fastest Ago2. This was surprising considering that Ago is required for all known miRNA functions, and thus is expected to exhibit similar dynamics. While sensitivity limits may explain why Ago2 cannot be detected diffusing as fast as the most mobile miRNA complexes (which we assign to mRNPs), it does not explain why few miRNAs are observed to have very slow diffusion coefficients coincident with the most prevalent Ago2 population. Importantly, this very slow population could be detected for miRNA when the same movies were analyzed with a Gaussian-based, in place of a centroid-based, method for particle detection[251]. Why might this be the case? The centroid-based method may be more sensitive to low signal-to-noise, causing localization error to be compounded with diffusion[281]. This error would arguably have the heaviest weight on dim particles with low diffusion coefficients, making stagnant particles appear more mobile than they actually are. It will thus be critical to directly compare the two routines using simulated images with varying signal-to-noise levels to determine their relative accuracies and limitations.

Inhibition of miRNA and miRISC-loading

Co-microinjection of LNA/DNA anti-miR with fluorophore-labeled duplexed miRNA inhibited the formation of trackable foci. Due to the very high base-pairing affinity of

LNA [282], we suspect that, even in the absence of heat-annealing, the anti-miR displaced the passenger strand and was bound to the guide strand before microinjection. Loss of foci then suggests that miRNA:anti-miR duplexes cannot be incorporated into the miRNA pathway. This is supported by previous evidence that modifications, including 2'deoxy substitutions that distort the A-form structure of small RNA duplexes decrease the effectiveness of gene repression[283]. That functional and non-functional miRNA duplexes display such dissimilar subcellular distributions serves as an important control for our experiments — we are unlikely tracking non-functional complexes — but it may also be useful as a readout to screen oligonucleotide chemistries for the development of anti-miRs or chemically modified siRNA/miRNA mimics[282].

We also observed a similar loss of foci when Ago2 was mutationally inhibited to prevent miRNA loading. Surprisingly, this effect varied from cell-to-cell, with most cells expressing Ago2-Y529E having no or very few foci, but a few having as many as Ago2-WT. This incomplete penetrance may trivially be due to residual Ago2-WT plasmid in the Ago2-Y529E DNA (although we did not detect this in our sequencing results). Alternatively, the mutant glutamate residue, which conceivably inhibits miRNA binding by repulsing the 5' phosphate of the RNA[262], could become protonated under certain situations due to perturbations in the pKa induced by local conformational changes[284]. These changes may be induced by cell cycle dependent protein-protein interactions, seeing as Ago2 can directly interact with genes involved in cell cycle transition[285]. It is also possible that under certain conditions Ago2 could localize to PBs without being loaded with miRNA.

Cy3AsH

Biarsenical fluorophores are a relatively new tool in the microscopist's toolbox. As such, they have not undergone the same extensive testing and rounds of improvement as the steadfast fluorescent protein. While FlAsH and ReAsH provide unobtrusive fluorophore labeling with added versatility, their largest drawback currently is non-specific binding[273]. This can be especially problematic with endogenous proteins that contain several naturally occurring cysteines[286]. Cy3AsH has only been previously tested using short TC-tagged peptide fragments that were designed to be helical in nature and

contained no other cysteines[253]. We found that Cy3AsH could be used to label IF2-TC, a much larger protein the structure of which near the TC-tag is unknown. Yet, Cy3AsH also bound BSA with nearly the same affinity as IF2-TC. A closer inspection of the amino acid sequence for BSA shows that it contains 35 cysteines including 8 cysteine pairs. It is possible that at least two of these pairs could fold close to one another creating a TC-like site. Interestingly though, the wavelength of emission from BSA-bound Cy3AsH was red-shifted relative to that of IF2-TC, indicating that there are differences in binding. It is unclear what exactly these differences are; they could be the distance between cysteine pairs or the result of binding to single cysteines (with the other arsenic unbound). Regardless, the utility of Cy3AsH for intricate intracellular experiments like translation site labeling is put into question with the observed potential for non-specific binding. To get an idea of how many BSA-like proteins exist in the cell, we searched ORFs in the mouse genome and found that more than thirty different proteins contain >10 cysteine pairs. Moreover, 22 genes contain a 'CCXXXXXCC' sequence. Therefore, we may expect significant non-specific binding of Cy3AsH inside cells.

There are several discrepancies between our results and those of Cao et al[253]: our molecule, relative to theirs, had slightly red-shifted maximum absorption and emission wavelengths and a higher (i.e., weaker) K_{app}. These differences may point to underlying structural differences in the small molecules. We characterized our compound by both 1H-NMR and mass spectrometry, but unfortunately had too low a yield for 13C-NMR. Without such analytical data, we cannot rule out the possibility that mercuration of the cyanine was not regioselective for the *para* position relative to the nitrogen group of the indoline ring. Alternative positioning of the arsenics would consequently reduce tag-binding affinity and could even change photospectral characteristics. Additional research is needed to confirm the structure of our molecule.

In its current form, Cy3AsH may be useful for certain single molecule experiments. It is best suited for labeling *in vitro* small recombinant proteins with few natural cysteines or synthetic peptides. One potential application may be for labeling peptide ligands for use in single molecule cell membrane or lipid bilayer protein binding assays[287,288].

4.5 Acknowledgements

We thank D. Bartel, C. Mayr, R. Tsien, N. Kedersha, and R. Singer for generous gifts of plasmids containing the 3' UTR of HMGA2, mCherry ORF, Dcp1a ORF, and the MS2 system of plasmids, A. Manzo for MSD software development, K. Jaqaman for the Gaussian-fit tracking routine, H. Ding, A. Gafni and D. Steel for sharing their intensity analysis routine, and the Single Molecule Analysis in Real Time (S.M.A.R.T) center at the University of Michigan for access to Imaris. Special thanks to E. Vedejs, J. Montgomery, R. Baxter, H. Malik, and J. Nelson for generously sharing lab space and equipment and for their assistance in the Cy3AsH synthesis.

Chapter 5: Summary, Conclusions, and Future Experiments

5.1 Summary and Conclusions

The complexity of higher organisms demands stringent control over gene expression. miRNAs have evolved in metazoans to provide an additional layer of control at the post-transcriptional level. By binding to mRNAs, miRNAs can regulate the levels of hundreds of genes and modulate expression programs[17]. During development, miRNAs can shape cellular identity and control cell growth and differentiation; while in post-development, it is thought that miRNAs function to maintain homeostasis and act as buffers against stress[17]. In many diseases, miRNA regulatory networks become dysfunctional, and the system in place to protect the cell turns harmful.

Over 1,500 miRNAs have been identified in humans and hundreds of these are highly conserved across mammals[289]. Each miRNA has its own unique set of targets based on seed pairing rules and, therefore, each has its own influence on various pathways. It is uncertain whether other distinctions exist among miRNA class members. Relatively few miRNAs have been extensively studied to date, and only in limited contexts. Computational target prediction methods, commonly used for inferring miRNA function, assume a certain universality: One tool fits all -- an abundant miRNA is likely to repress highly favorable targets x, y, and z irrespective of cell or tissue type. However, few of these predictions or assumptions have been validated across different cell types, particularly *in vivo* where little is known about miRNA activity. In contrast, seemingly contradictory mechanistic data suggests there may be diversity in miRNA activity, with certain miRNAs enacting different modes of repression on different targets and/or in different contexts.

Here, I have investigated in healthy mouse liver the functions of three highly abundant disease-linked miRNAs: miR-21, an oncogenic miRNA upregulated in most cancers; let-7, a tumor suppressor; and miR-122, a mostly liver specific miRNA linked to cancer and hepatitis C viral (HCV) infection. I find that in contrast to other miRNAs,

genetic deletion or inhibition of miR-21 with a specific antimiR oligonucleotide has surprisingly little effect on mRNA levels of predicted and known targets. This limited activity is further contrasted by miR-21's activity in the cervical cancer HeLa cell line, where inhibition has a strong effect on targets. These findings indicate both that miR-21 is functionally restricted in non-stressed tissue relative to cancer cells, and current miRNA target engagement rules do not apply equally to all miRNAs and cell states.

Using both biochemical and computational approaches, we sought to better understand the reasons behind liver miR-21's reduced activity. I discovered using sucrose gradients with fresh liver extracts that, comparatively, miR-21 is significantly underrepresented within polysome-complexes, where the majority of target mRNAs are found. Binding assays indicated, however, that miR-21 is not sequestered or actively prevented from binding targets. Noticing that the sequence of miR-21 is relatively A/U rich, I next studied its theoretical target-binding energies. Thermodynamic calculations with bioinformatically predicted targets showed miR-21 to bind targets with significantly lower stability than other miRNAs. Thus, we conclude that unfavorable thermodynamics and possibly other factors elevate the threshold required for binding and repressing most targets, and miR-21, despite its abundance, is maintained at a level below this threshold in healthy mouse liver.

Interestingly, the same expression level is sufficient to elicit target repression in HeLa, which indicates that these cells are generally more conducive to miR-21 activity and perhaps that of other miRNAs as well. In agreement, we found that HeLa miR-21 is highly enriched in polysomal fractions, indicating that increased association with polysome-bound mRNA enhances target repression. Considering that increased miR-21 activity is strongly implicated in the development and progression of cancer, and miR-21 is required for maintenance of HeLa cell replication, we propose that enhanced association with polysome-bound mRNA is a critical step towards the gain of miR-21 function observed in many stressed or diseased cells. Additionally, we conclude that for the measured miRNA/target combinations, binding to targets while the mRNA resides in polysomes is critical for miRNA-mediated regulation; a conclusion supported previously using cultured cells. For the first time, however, we provide *in vivo* evidence from a non-stressed tissue for the significance of miRNA-polysome association. Unlike previous

reports of mRNA targets being removed from polysomes during miRNA repression[68,69], we did not observe large shifts in mRNA relative to polysomal fractions in the presence or absence (inhibition) of miRNA. We note, however, that subtle changes in translation rates would likely not be detected with the resolution of our sucrose gradient preparations, especially in combination with the observed sequence-independent effects of antimiRs on translation.

miR-21 inhibition in liver induces upregulation of genes significantly enriched in stress response including heat shock factors and co-chaperones. The majority of these genes do not contain a miR-21 seed site, indicating that their induction is a secondary effect of inhibition. I propose induction of the stress cascade may be mediated through Taf7, a heat shock transcription factor associated protein that was among the few strongly upregulated miR-21 seed-matched targets. I also found that miR-21 itself is upregulated following antimiR-21 treatment, suggesting the presence of a feedback loop that may act to contain a stress activation event. Taf7 may also mediate this loop since it has previously been found to activate transcription factor AP-1[210], a known promoter of miR-21 expression[137,142]. This loop has not yet been characterized, and it should be further explored along with its connection to the heat shock cascade.

Taken together, I propose a model wherein miR-21 plays a critical role as a regulator of cellular stress response (**Fig. 5.1**). In this model, miR-21 first serves under basal conditions as a pressure valve for stress response by repressing a narrow set of 'threshold 1' targets that are more easily repressible than most other seed-matched transcripts (referred to as 'threshold 2' targets); the reasons behind this favorability are unknown, but are likely due to contextual features and/or association with additional regulatory factors. In the event of stress, one or more of these threshold 1 targets may transition to threshold 2 to temporally evade miR-21 regulation, and in turn, trigger the stress response cascade. This step of the model is reflected in our experimental data: when miR-21 targets were released from repression by antimiR inhibition, molecular chaperones, co-chaperones, and other stress-response factors were upregulated as a secondary effect. Following induction of stress response, miR-21 may next act as a brake to contain the response and restore the cell to homeostasis. To do so, miR-21 may undergo an activation event to enable it to regulate threshold 2 targets in order to restore

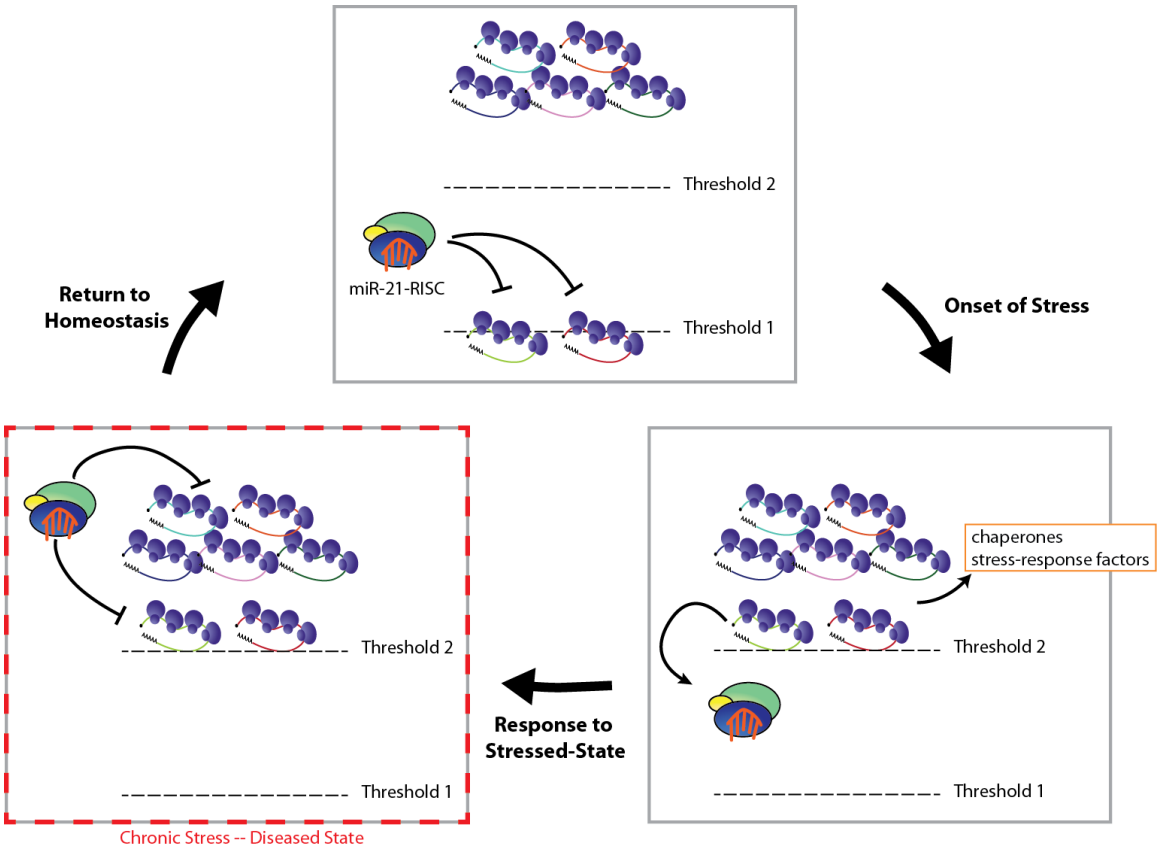


Figure 5.1 Proposed model for miR-21's role in stress response. Under normal conditions (top center panel) miR-21 operates at a threshold below that of most targets (threshold 2), instead repressing a limited set of more favorable targets (threshold 1). At the onset of stress, certain threshold 1 targets may evade miR-21 regulation in order to activate a stress response cascade that functions to respond to and resolve the stress (bottom right panel). In order to return to homeostasis, miR-21 repressive activity is enhanced above threshold 2 (bottom left panel). However, in the event of chronic stress, as is typical in cancer, miR-21 may remain in its activated form (bottom left panel, marked with red). See text for more details.

regulatory balance. Increased expression of miR-21 via the putative Taf7/AP-1 feedback loop discussed above may contribute to this activation. Augmented miR-21 activity may also enable miR-21 to temporally regulate additional targets in order to, for instance, prevent premature cell death in response to otherwise resolvable stress events. Consistently, although absent in non-stressed liver, several apoptotic factors including Pdcd4 are regulated by miR-21 in cancer cells including HeLa where miR-21 repression is enhanced and where cellular environments are assailed by chronic stress from various sources including metabolic, misfolded proteotoxic, mitotic, oxidative, and DNA damage stress[18]. In healthy cells, miR-21 would be expected to return to its more restricted repression capacity following stress mediation; however, in the chronically stressed cancer state, miR-21 would remain activated, and as a result, contribute to tumorigenesis by persistently inhibiting intrinsic fail-safe mechanisms like apoptosis. Although largely speculative, this model takes into account the experimental data reported here while helping to explain the seemingly contradictory activity of miR-21 between healthy and diseased cells. It can further serve as an experimental framework for future investigations.

Lastly, bioinformatics was used to assess the determinants underlying miR-21's target selection in liver. It was hypothesized that Taf7 and other responsive targets have optimal features that facilitate miR-21 binding (so called threshold 1 targets). To test this hypothesis, a local database of cDNA sequences was built and sequence-based features were extracted for each gene measured by array profiling. Features included A/U richness, number of seed sites, seed locations, extent of miRNA base pairing, and estimated binding thermodynamics. These features are known determinants of target selection, and have previously been found to correlate with target repression induced by transfected miRNA in HeLa cells[82]. Additionally, we searched for lesser known determinants including transcript length, expression level, non-canonical miRNA binding sites, and the presence of recognition sites for RNA binding proteins (RPBs) that are recently emerging as modulators of miRNA networks[218,231-233]. Unexpectedly, clear trends for known determinants were not detected in miR-21 responsive targets, indicating that miR-21 may select targets differently than other miRNAs. Instead, upregulated targets were found on average to have greater levels of expression and shorter open

reading frames (ORFs). Neither of these features are known determinants. Whereas expression may be a bias in detection, transcripts with shorter ORFs could potentially be more rapidly destabilized by miRNA. This can be tested in cell culture using reporters with varying ORF lengths and detecting destabilization with RT-qPCR. Moreover, all miR-21 targets are highly enriched with RBP-REs, suggesting that crosstalk with RBP networks may explain miR-21's target selection. For comparison, miR-122 targets were also tested for the same sequence based trends. In addition to length and expression, additional known determinants were detected, consistent with better overall agreement of miR-122 inhibition responses with prediction algorithms. Surprisingly, however, neither miR-21 nor miR-122 responses showed discernible trends relating to thermodynamics of target binding or site accessibility. We conclude that *in vivo* targeting may be more complicated than depicted by current models of target prediction, due to crosstalk with other regulatory networks.

In summary, this work demonstrates the complexity of miRNA regulatory networks and suggests that a great deal of diversity exists in miRNA activity, especially in terms of biological context. In addition, it provides important functional and mechanistic insights into the underlying changes in miRNA-mediated repression in the transition between health and disease. Despite the scientific enthusiasm for miRNA over the last decade (**Fig. 1.1**), little research has focused on the *in vivo* physiological roles of mammalian miRNA most likely due to the fact that using cell culture as an experimental model is both more convenient and experimentally more flexible. Because research using cell culture has become commonplace throughout all fields of biology, it is easy to forget that these cells, even those that were once human in origin, are not representative of 'normal' cells; instead they often have altered morphologies and karyotypes and have cellular environments defined by chronic stress-- hallmarks of their diseased state. This work demonstrates that considerable differences exist between the physiological *in vivo* activities and the pathological *in cellulo* activities of miRNA. While we focused on miR-21, we also found that *in vivo* let-7 regulated few of the targets previously validated in cell culture (**Fig. 2.3**) suggesting that these differences extend past the idiosyncrasies of miR-21. Further work is needed to truly understand what/how context-dependent factors influence targeting. In addition, miR-21 may not be an isolated anomaly, but rather just

one example of the diversity that exists among members of this class of regulatory non-coding RNA. Other miRNAs, like miR-21, may be less active than their counterparts or display other divergences in target selection, binding, or repression mechanisms. Importantly, some of these differences may be normalized in cell culture due to globally exaggerated miRNA pathways; therefore, extensive diversity between miRNA class members may only be uncovered with comparative studies in healthy animal tissue as done here.

5.2 Future Experiments

In this thesis I propose that miR-21 undergoes an overactivation event to transition from its normal activity as a weak repressing, tightly controlled miRNA in healthy cell states to a strong repressing dysregulated oncogenic miRNA in diseased states. In the following, I outline future experiments to further test this hypothesis and to probe the underlying mechanisms/causes. Additionally, experiments to better understand miR-21 target selection in healthy liver are proposed. These experiments supplement the many other proposed experiments mentioned throughout this work and are listed by priority.

Test miR-21 activity in additional contexts

In our analysis we found context specific activity of miR-21 in HeLa cells and healthy liver tissue. Future studies will need to extend these analyses (antimiR array profiling / polysome analysis) to additional biological contexts to confirm the generality that miR-21 is moderately active in healthy cells relative to other miRNAs and overactive in diseased cells (and that repression corresponds with polysome binding). Literature suggests that miR-21 is moderately active in other tissues besides liver: Hatley et al.[138] have noted, in brief, the absence of robust miR-21 target repression in miR-21 knockout lung tissue. Additionally, Patrick et al.[112] have shown that Pdcd4 is only upregulated following anti-miR-21 treatment in heart tissue following transaortic constriction (TAC), but not in the healthy sham control group. We would therefore expect to find similar results in lung and heart as we found in liver. However, interestingly, miR-21 may have increased activity in healthy keratinocytes as Ma et al.[290] have reported that several

predicted and known miR-21 targets including the tumor suppressors Spry1, Pten, and Pdcd4 are upregulated in keratinocytes from miR-21 knockout mice relative to wild-type. Therefore, miR-21 activity could be tissue and/or growth dependent, and given its role in controlling apoptosis, its activity may vary based on normal cell turnover rates, and vice versa.

We suspect based on the well-documented activity of miR-21 in cancer cells that most transformed cell lines show irregularly enhanced miR-21 activity similar to HeLa. Because our *in vivo* studies have focused on liver, hepatocarcinoma cell lines such as Huh7 and HepG2 would make for a better comparison than cervical HeLa cells. Huh7/HepG2 cells replication has previously been shown to be miR-21 dependent[126], therefore, in these cells we would expect to find miR-21 enriched in polysomal fractions. Additionally, primary hepatocytes and non-transformed immortalized cells such as NIH-3T3 could be tested. If these cells can replicate miR-21 activity observed in liver tissue, it would greatly facilitate research on non-activated miR-21 since it would allow for greater experimental manipulation than what is realistically feasible in mice. Comparing cultured cells grown to confluence with actively dividing cells could also assess the influence of growth on miR-21 activity.

Compare levels of miRNA related factors between HeLa and liver

We found that comparable levels of miR-21 are expressed between liver and HeLa, yet miR-21 activity is less well controlled in the latter. This suggests that other factors besides miR-21 abundance affect the strength of miR-21 activity. We also found enhanced polysome binding for let-7 in HeLa, suggesting that the miRNA machinery in general is overactive, rather than just miR-21 specifically. One possibility is that a certain miRNA-related factor(s) is more abundantly expressed in HeLa. To test this idea, quantitative western blots could be used to compare levels of core miRNA components between liver and HeLa. One factor of particular interest that may be upregulated in HeLa is the receptor for activated C-kinase (Rack1), which is necessary for loading miRNA onto translating mRNAs[49]. Another factor is the RBP HuR, whose binding motif frequently occurs near miR-21 sites. Since HuR can have both positive and negative effects on miRNA-mediated repression, enhanced miR-21 activity could result from

increased or decreased HuR expression, respectively. In cancer, HuR is often found to be overexpressed[291], suggesting that if HuR is responsible for miR-21's activity in HeLa, it is due to positive cooperativity. If one of these factors is found to be differentially expressed in HeLa, functional studies could be carried out to assess its role. For example, miR-21 polysome binding, relative to that of a stronger binding miRNA, could be assessed after Rack1 partial knockdown (Rack1 is essential for miRNA activity). Alternatively, although technically more challenging, Rack1 could be overexpressed in mouse liver to see if it enhances miR-21 polysome binding.

Directly induce miR-21 activation in liver

A more direct approach to studying miR-21 and its contributions to disease processes would be to try to stimulate miR-21 activation in liver. miR-21 is upregulated in most cancers, therefore it could be tested whether increased miR-21 levels, expressed from an inducible promoter in a transgenic animal, are sufficient to increase the repression footprint and enhance target binding / polysome sedimentation. We would suspect this not to be the case since miR-21 expression alone is not sufficient to induce tumorigenesis in solid tissues including brain[144] and lung[138]. The oncogene Ras has been shown to induce miR-21 expression, and in lung, Ras is necessary for miR-21 to contribute and reinforce cancerous phenotypes[138]. We would expect Ras to increase miR-21's repression signature and enhance polysome binding. Since miR-21 appears to be stress-responsive, various stressors such as chemical stimulants (e.g. carbon tetrachloride, a highly hepatotoxic carcinogen[292]) could also be screened for their ability to trigger miR-21 overactivation.

Use iSHiRLoC to compare miR-21 complexes in transformed and primary culture

As discussed in **Chapter 4**, iSHiRLoC is a single molecule technique for studying miRNA complex assembly in live cells. iSHiRLoC could be especially useful for studying miR-21 or other miRNAs in a variety of contexts including primary hepatocytes. Of particular interest will be to co-track miR-21 with putative target mRNAs in both healthy and diseased cells. My prediction is that specific differences in

diffusion will be observed as long as the miR-21 concentration remains below its repression threshold. Once such a signature is discovered, the influence of varying some of the cofactors discussed above by either RNAi knockdown or controlled expression will then reveal their influence on miR-21-mediated target repression.

Utilize deep sequencing to further probe miR-21 targeting

A major limitation of our current study is that we cannot distinguish between primary and secondary effects of miRNA inhibition. This is especially problematic in the case of seed-matched targets, and may be why few clear feature trends were detected in **Chapter 3**. Unfortunately, validation of these targets using routine luciferase assays requires cell culture, where miR-21 activity is altered.

High throughput sequencing crosslinking immunoprecipitation (HITS-CLIP) could be useful for validating targets and further exploring miR-21 target binding[227]. In this method, cDNA sequenced target fragments that are cross-linked and immunoprecipitated with Ago precisely identify miRNA binding sites. HITS-CLIP could be used in conjunction with array profiling to obtain a set of high confidence targets that exhibit both derepression and loss of a mapped miRNA binding site in the presence of antimiR. These data could then be reanalyzed for sequence-based trends or ontology classification to identify new pathways regulated by miR-21 and other miRNAs.

Additionally, deep sequencing is capable of providing additional sequence information on the transcriptome that cannot be ascertained by array profiling or accounted for *a priori* using annotated sequence data. There is a growing appreciation that a great deal of context-dependent sequence variation exists among transcriptomes. Variation arising from alternative polyadenylation, which can lengthen or shorten 3'UTRs, and non-templated RNA editing can significantly affect miRNA regulatory networks by creating, altering or deleting seed sites in certain transcripts. Alternative polyadenylation is pervasive in cancer, as well as are other genetic aberrations such as chromosomal translocations[173]. Therefore, these changes in the targeting landscape could partially explain why miR-21 represses different genes in healthy and cancer cells, a hypothesis which could be explored using deep sequencing to compare -- with nucleotide resolution -- transcriptomes derived from healthy liver and HeLa cells.

Appendix A

Matlab Functions and Scripts for Bioinformatic Analysis of Target Contextual Features

The following scripts were used to extract various mRNA features. In general, the scripts follow the same pattern:
— array profiling data is called for a particular miRNA;
— seed-matched genes are expression or expression/conservation filtered;
— Genbank gene annotation data is called
— feature data is extracted; and
— data is binned by response and outputted.

All of these scripts require certain database variables to be pre-loaded in the workspace:

```
load gen_refmap2 –> loads gen_refmap (mouse database)
load mmu_liver_DataMatrices –> loads an_21DM, an122DM, anlt7DM (AMAP data)
load mmuliver_expressedgenes –> loads expressedgenes
load human_gen_refmap2 –> loads hsaGBmap2 (human database)
```

Local A/U Composition

```
%%AUwindowscript2: Computes A/U compositions of sequences flanking
%%seed-sites with growing window size.
%% For expressed m21 anyseed
m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);

%% For expressed m21 conserved anyseed
m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);

%% For expressed m122 anyseed
m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);

%% For expressed m122 conserved anyseed
m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);

%% For expressed let7 anyseed
let7data = mtxseed2(an122DM,'let7','anyseed',gen_refmap);
explet7 = let7data(expressedgenes,:);
utrs = getLongUtrSeqs2(explet7,'let7',gen_refmap);
accs = getLongUtrData2(explet7,'let7',gen_refmap);
binidx = binDM(explet7);
seedlocs = cellfun(@(x) x.UTR3.let7.index, accs,'UniformOutput', false);

%% For expressed let7 conserved anyseed
let7data = conservedsites(an122DM,'let7','anyseed',gen_refmap,hsaGBmap2);
explet7 = let7data(expressedgenes,:);
utrs = getLongUtrSeqs2(explet7,'let7',gen_refmap);
```

```
accs = getLongUtrData2(explet7,'let7',gen_refmap);
binidx = binDM(explet7);
seedlocs = cellfun(@(x) x.UTR3.let7.index, accs,'UniformOutput', false);
%% Calcs

utrlengths = cell2mat(cellfun(@(x) length(x.UTR3.Sequence), accs, 'UniformOutput',
false));
pad = 'n';
for i = 1:10
    pad = strcat(pad, 'nnnnnnnnnn');
end
padutrs = cellfun(@(x) strcat(pad, x,pad),utrs, 'UniformOutput', false);


seeds = cellfun(@(x) x(x>0),seedlocs,'UniformOutput',false);
seeds = cellfun(@(x) x+numel(pad), seeds, 'UniformOutput', false);


windows = 10:10:100;
FAUbasecount = zeros(length(seeds),length(windows));

for i = 1:length(seeds);
    for j = 1:length(windows)
        win = windows(j);
        Useq = padutrs{i};
        for k = 1:numel(seeds{i})
            seq = Useq(seeds{i}(k)-win:seeds{i}(k)+win);
            bases = basecount(seq);
            FAU(k) = (bases.A+bases.T)./(bases.A+bases.T+bases.G+bases.C);
        end
        FAUbasecount(i,j) = mean(FAU);
    end
end

% Output
for i = 1:length(windows);
test = FAUbasecount(:,i);

AUupmean = mean(test(binidx(:,1),:),1);
AUncmean = mean(test(binidx(:,2),:),1);
AUdnmean = mean(test(binidx(:,3),:),1);

Nup = numel(find(binidx(:,1) == 1));
Nnc = numel(find(binidx(:,2) == 1));
Ndn = numel(find(binidx(:,3) == 1));

AUupstd = std(test(binidx(:,1),:),0,1)/sqrt(Nup);
AUncstd = std(test(binidx(:,2),:),0,1)/sqrt(Nnc);
AUdnstd = std(test(binidx(:,3),:),0,1)/sqrt(Ndn);

dataout(i,:) = [AUupmean' AUupstd' Nup AUncmean' AUncstd' Nnc AUdnmean' AUdnstd' Ndn];

end
```

Full Length A/U Composition

```
%% AUrichscript2: Computes A/U composition of full length 3'UTRs

dataout = zeros(6,9);
for i = 1:6
    disp(i)
    if i == 1
        % For expressed m21 anyseed
        m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
        expm21 = m21data(expressedgenes,:);
        utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
        binidx = binDM(expm21);
    elseif i == 2
        % For expressed m21 conserved anyseed
        m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
        expm21 = m21data(expressedgenes,:);
        utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
```

```
        binidx = binDM(expm21);

    elseif i == 3
        % For expressed m122 anyseed
        m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
        expm122 = m122data(expressedgenes,:);
        utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
        binidx = binDM(expm122);

    elseif i == 4
        % For expressed m122 conserved anyseed
        m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
        expm122 = m122data(expressedgenes,:);
        utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
        binidx = binDM(expm122);
    elseif i == 5
        % For expressed let7 anyseed
        let7data = mtxseed2(an122DM,'let7','anyseed',gen_refmap);
        explet7 = let7data(expressedgenes,:);
        utrs = getLongUtrSeqs2(explet7,'let7',gen_refmap);
        binidx = binDM(explet7);

    elseif i == 6
        % For expressed let7 conserved anyseed
        let7data = conservedsites(an122DM,'let7','anyseed',gen_refmap,hsaGBmap2);
        explet7 = let7data(expressedgenes,:);
        utrs = getLongUtrSeqs2(explet7,'let7',gen_refmap);
        binidx = binDM(explet7);
    end
    %
    Atailidx = cellfun(@(x) regexp(x,'a{5,}\>'),utrs,'UniformOutput',false);
    utrsnotail = cell(size(utrs));
    for j = 1:length(utrs)
        if isempty(Atailidx{j})
            utrsnotail(j) = cellstr(utrs{j}(1:end));
        else
            utrsnotail(j) = cellstr(utrs{j}(1:Atailidx{j}));
        end
    end
    utrBaseComps = cellfun(@(x) basecount(x),utrsnotail,'UniformOutput', false);
    utrFAU = cellfun(@(x) (x.A+x.T)./(x.A+x.T+x.G+x.C),utrBaseComps)';
    test = utrFAU';

    AUupmean = mean(test(binidx(:,1),:),1);
    AUncmean = mean(test(binidx(:,2),:),1);
    AUdnmean = mean(test(binidx(:,3),:),1);

    Nup = numel(find(binidx(:,1) == 1));
    Nnc = numel(find(binidx(:,2) == 1));
    Ndn = numel(find(binidx(:,3) == 1));

    AUupstd = std(test(binidx(:,1),:),0,1)/sqrt(Nup);
    AUncstd = std(test(binidx(:,2),:),0,1)/sqrt(Nnc);
    AUdnstd = std(test(binidx(:,3),:),0,1)/sqrt(Ndn);

    dataout(i,:) = [AUupmean' AUupstd' Nup AUncmean' AUncstd' Nnc AUdnmean' AUdnstd'
Ndn];
end
```

5'UTR Length

```
%% UTR5length_script computes lengths of 5'UTRs
dataout = zeros(6,9);
for i = 1:6
    if i == 1
        % For expressed m21 anyseed
        m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
        expm21 = m21data(expressedgenes,:);
        accs = getLongUtrData2(expm21,'m21',gen_refmap);
        binidx = binDM(expm21);
    elseif i == 2
        % For expressed m21 conserved anyseed
        m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
        expm21 = m21data(expressedgenes,:);
```

```
        accs = getLongUtrData2(expm21,'m21',gen_refmap);
        binidx = binDM(expm21);
    elseif i == 3
        % For expressed m122 anyseed
        m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
        expm122 = m122data(expressedgenes,:);
        accs = getLongUtrData2(expm122,'m122',gen_refmap);
        binidx = binDM(expm122);
    elseif i == 4
        % For expressed m122 conserved anyseed
        m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
        expm122 = m122data(expressedgenes,:);
        accs = getLongUtrData2(expm122,'m122',gen_refmap);
        binidx = binDM(expm122);
    elseif i == 5
        % For expressed let7 anyseed
        let7data = mtxseed2(an122DM,'let7','anyseed',gen_refmap);
        explet7 = let7data(expressedgenes,:);
        accs = getLongUtrData2(explet7,'let7',gen_refmap);
        binidx = binDM(explet7);
    elseif i == 6
        % For expressed let7 conserved anyseed
        let7data = conservedsites(an122DM,'let7','anyseed',gen_refmap,hsaGBmap2);
        explet7 = let7data(expressedgenes,:);
        accs = getLongUtrData2(explet7,'let7',gen_refmap);
        binidx = binDM(explet7);
    end
    % Calcs and Output
    cdsidx = cell2mat(cellfun(@(x) x.CDS.indices, accs,'UniformOutput', false));
    lengths = cdsidx(:,1);

    upmean = mean(lengths(binidx(:,1),:),1);
    ncmean = mean(lengths(binidx(:,2),:),1);
    dnmean = mean(lengths(binidx(:,3),:),1);

    Nup = numel(find(binidx(:,1) == 1));
    Nnc = numel(find(binidx(:,2) == 1));
    Ndn = numel(find(binidx(:,3) == 1));

    upstd = std(lengths(binidx(:,1),:),0,1)/sqrt(Nup);
    ncstd = std(lengths(binidx(:,2),:),0,1)/sqrt(Nnc);
    dnstd = std(lengths(binidx(:,3),:),0,1)/sqrt(Ndn);


    dataout(i,:) = [upmean' upstd' Nup ncmean' ncstd' Nnc dnmean' dnstd' Ndn];
end
```

CDS/ORF Length

```
%% CDSlengths_script2 computes CDS/ORF lengths
%% For expressed m21 anyseed
m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
expm21 = m21data(expressedgenes,:);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);

%% For expressed m21 conserved anyseed
m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
expm21 = m21data(expressedgenes,:);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);

%% For expressed m122 anyseed
m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
expm122 = m122data(expressedgenes,:);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);

%% For expressed m122 conserved anyseed
m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
expm122 = m122data(expressedgenes,:);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
```

```
%% For expressed let7 anyseed
let7data = mtxseed2(an_21DM,'let7','anyseed',gen_refmap);
explet7 = let7data(expressedgenes,:);
accs = getLongUtrData2(explet7,'let7',gen_refmap);
binidx = binDM(explet7);

%% For expressed let7 conserved anyseed
let7data = conservedsites(an_21DM,'let7','anyseed',gen_refmap,hsaGBmap2);
explet7 = let7data(expressedgenes,:);
accs = getLongUtrData2(explet7,'let7',gen_refmap);
binidx = binDM(explet7);
%% Calcs and Output
cdsidx = cell2mat(cellfun(@(x) x.CDS.indices, accs,'UniformOutput', false));
lengths = cdsidx(:,2)-cdsidx(:,1);

upmean = mean(lengths(binidx(:,1),:),1);
ncmean = mean(lengths(binidx(:,2),:),1);
dnmean = mean(lengths(binidx(:,3),:),1);

Nup = numel(find(binidx(:,1) == 1));
Nnc = numel(find(binidx(:,2) == 1));
Ndn = numel(find(binidx(:,3) == 1));

upstd = std(lengths(binidx(:,1),:),0,1)/sqrt(Nup);
ncstd = std(lengths(binidx(:,2),:),0,1)/sqrt(Nnc);
dnstd = std(lengths(binidx(:,3),:),0,1)/sqrt(Ndn);

dataout = [upmean' upstd' Nup ncmean' ncstd' Nnc dnmean' dnstd' Ndn];
```

3'UTR Length

```
%% UTRlengthscript Computes 3'UTR lengths
%% For expressed m21 anyseed
m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
%% For expressed m21 conserved anyseed
m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);


%% For expressed m122 anyseed
m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);

%% For expressed m122 conserved anyseed
m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);

%% For expressed let7 anyseed
m7data = mtxseed2(anlt7DM,'let7','anyseed',gen_refmap);
expm7 = m7data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm7,'let7',gen_refmap);
binidx = binDM(expm7);

%% For expressed let7 conserved anyseed
m7data = conservedsites(anlt7DM,'let7','anyseed',gen_refmap,hsaGBmap2);
expm7 = m7data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm7,'let7',gen_refmap);
binidx = binDM(expm7);
%%
utrlengths = cellfun(@(x) length(x),utrs);
test = utrlengths;


AUupmean = mean(test(binidx(:,1),:),1);
AUncmean = mean(test(binidx(:,2),:),1);
```

```
AUdnmean = mean(test(binidx(:,3),:),1);

Nup = numel(find(binidx(:,1) == 1));
Nnc = numel(find(binidx(:,2) == 1));
Ndn = numel(find(binidx(:,3) == 1));

AUupstd = std(test(binidx(:,1),:),0,1)/sqrt(Nup);
AUncstd = std(test(binidx(:,2),:),0,1)/sqrt(Nnc);
AUdnstd = std(test(binidx(:,3),:),0,1)/sqrt(Ndn);

dataout = [AUupmean' AUupstd' Nup AUncmean' AUncstd' Nnc AUdnmean' AUdnstd' Ndn];
```

Extended 3'base-pairing

```
%% extendedbasepairingscript computes scores for 3'miRNA base-pairing.
% miRNA is aligned to target sequence and starting from the 3'end, each
% 'base-pair' is given a score of 1 point and each consecutive 'base-pair'
% is given a bonus 0.5 points
%%
%% For expressed m21 anyseed
m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
expm21 = m21data(expressedgenes,:);
DM = expm21;
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);
%% For expressed m21 conserved anyseed
m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);
DM = expm21;
%% For expressed m122 anyseed
m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);
DM = expm122;
%% For expressed m122 conserved anyseed
m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);
DM = expm122;
%%
firstseeds = cell2mat(cellfun(@(x) min(x(x>0)), seedlocs, 'UniformOutput',false));
utrlengths = cell2mat(cellfun(@(x) length(x), utrs, 'UniformOutput', false));
tocloseidx = firstseeds < 14 | (utrlengths-firstseeds)<8;
firstseeds = firstseeds(~tocloseidx);
bindsites = cell(length(firstseeds),1);
utrs2 = utrs(~tocloseidx);
for i = 1:length(utrs2)
    bindsites{i} = utrs2{i}(firstseeds(i)-13:firstseeds(i)+8);
end

%m21 = lower(rna2dna('UAGCUUAUCAGACUGAUGUUGA')); %uncomment for m21
m122 = lower(rna2dna('UGGAGUGUGACAAUGGUGUUUG')); %uncomment for m122
miR = m122;
count = cell(length(utrs2),1);
for i = 1:length(utrs2)
    target = bindsites{i};
    [~,b] = nwalign(target(1:end-1), seqrcomplement(miR));
    idx = zeros(1, numel(b(2,:)));
    idx(strfind(b(2,:),'|')) = 1;
    score = 0;
    for j = 1:length(idx(1:13))
        if idx(j) == 1 && idx(j+1) == 1
            score = score+0.5;
```

```
        end
    end
    score = score+sum(idx(1:13));
    count{i} = score;
end

count = cell2mat(count);
newDM = DM(~tocloseidx,:);
binidx = binDM(newDM);

upnumel = sum(binidx(:,1));
upmean = mean(count(binidx(:,1)));
upsem = std(count(binidx(:,1)))/sqrt(upnumel);

ncnumel = sum(binidx(:,2));
ncmean = mean(count(binidx(:,2)));
ncsem = std(count(binidx(:,2)))/sqrt(ncnumel);

dnnumel = sum(binidx(:,3));
dnmean = mean(count(binidx(:,3)));
dnsem = std(count(binidx(:,3)))/sqrt(dnnumel);

dataout = [upmean upsem upnumel ncmean ncsem ncnumel dnmean dnsem dnnumel];
```

G-bulge search. In its current form presented here, it is written only for miR-122 as an example.

```
%% gbulgescript (for miR-122)
expDM = an122DM(expressedgenes,:);
expDM = sortrows(expDM);
accs = getLongUtrData3nomir(expDM, gen_refmap);
datanotfound1 = ~cellfun(@(x) isstruct(x), accs);
datanotfound2 = cellfun(@(x) length(x), accs)<1;
datanotfound = datanotfound1 | datanotfound2;
accs = accs(~datanotfound);
cdsidx = cellfun(@(x) x.CDS.indices, accs,'UniformOutput', false);
%%
for i = 1:length(accs);
    fullseq = accs{i}.Sequence;
    utrseqs(i,1) = cellstr(fullseq(cdsidx{i}(2):end));
end

%%
expDMfltr = expDM(~datanotfound,:);
genes = rownames(expDMfltr);
%%
gbulgeidx = ~cellfun(@isempty, strfind(utrseqs,'acagctcc'));
m8idx = ~cellfun(@isempty, strfind(utrseqs,'acactcca'));
randseq = lower(randseq(8,'FromStructure',basecount('acactcca')));
randidx = ~cellfun(@isempty, strfind(utrseqs,randseq));
%m7idx = ~cellfun(@isempty, strfind(utrseqs,'ataagct'));
binidx = binDM(expDMfltr);

%%
bulge = getlog2(expDMfltr(gbulgeidx,:));
m8 = getlog2(expDMfltr(m8idx,:));
rand = getlog2(expDMfltr(randidx,:));
seed = getlog2(mtxseed2(an122DM,'m122','anyseed', gen_refmap));
all = getlog2(expDMfltr);

%%
meanbulge = median(getlog2(expDMfltr(gbulgeidx,:)));
mean8m = median(getlog2(expDMfltr(m8idx,:)));
meanrand = median(getlog2(expDMfltr(randidx,:)));
meanseed = median(getlog2(mtxseed2(an122DM,'m122','anyseed', gen_refmap)));
meanall = median(getlog2(expDMfltr));

%%
[f1 x1] = ecdf(bulge);
[f2 x2] = ecdf(m8);
[f3 x3] = ecdf(rand);
[f4 x4] = ecdf(seed);
[f5 x5] = ecdf(all);
figure; hold on
```

```
plot(x1,f1,'r');
plot(x2,f2,'b');
plot(x3,f3,'g');
plot(x4,f4,'k');
plot(x5,f5,'m');
%%
[~,p(1)] = kstest2(all, m8);
[~,p(2)] = kstest2(all, all);
[~,p(3)] = kstest2(all, rand);
[~,p(4)] = kstest2(all, bulge);
[~,p(5)] = kstest2(all, seed);
```

RBP motif search for full-length 3'UTRs.

```
%% motifsearch2script_FL searches the full length 3'UTR for RBP motifs
%% For expressed m21 anyseed
m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);
seed = 'taagct';
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);

%% For expressed m122 anyseed
m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);
%% For expressed m21 conserved anyseed
m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);
%% For expressed m122 conserved anyseed
m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);
%%
firstseeds = cellfun(@(x) min(x(x>200)), seedlocs,'UniformOutput', false);
toclose2startidx = ~cellfun(@(x) length(x), firstseeds);
firstseeds(toclose2startidx) = {0};
firstseeds = cell2mat(firstseeds);
utrlengths = cellfun(@(x) length(x), utrs);
% seedlocs = strfind(utrs, seed);
% firstseeds = cellfun(@(x) x(1), seedlocs);
% toclose2startidx = firstseeds < 200;
toclose2endidx = (utrlengths - firstseeds) < 200;
tocloseidx = toclose2startidx | toclose2endidx;
%%
flankidx = zeros(length(utrs),2);
flankidx(~tocloseidx,1) = firstseeds(~tocloseidx)-100;
flankidx(~tocloseidx,2) = firstseeds(~tocloseidx)+100;
flankidx(toclose2startidx,:) = repmat([1 201], numel(find(toclose2startidx == 1)), 1);
flankidx(toclose2endidx,1) = utrlengths(toclose2endidx)-200;
flankidx(toclose2endidx,2) = utrlengths(toclose2endidx);
%find utrs smaller than 200 nt
tosmallidx = flankidx(:,1)<1;
flankidx(tosmallidx,1) = 1;
flankidx(tosmallidx,2) = utrlengths(tosmallidx);
%
% %%
utrflanks = cell(length(utrs),1);
for i = 1:length(utrs)
    seq = utrs{i};
```

```
    utrflanks(i,1) = cellstr(seq(flankidx(i,1):flankidx(i,2)));
end
%%
motifs = {'ttttaaa'; 'ttt\wttt'; 'tgta\wata'; 'tatttat'};
motifhit = cell(length(motifs),2);
for i = 1:length(motifs)
    numhits = cellfun(@numel, regexp(utrs, motifs{i}));
    hits = find(~cellfun(@isempty, regexp(utrs, motifs{i})));
    P = zeros(length(hits),1);
    for j = 1:length(hits)
        seqcall = utrs{hits(j)};
        rseq = cell(500,1);
        for k = 1:500
            rseq(k) = cellstr(lower(randseq(length(seqcall), 'fromstructure',
basecount(seqcall))));
        end
        rhits = find(~cellfun(@isempty, regexp(rseq, motifs{i})));
        numrhits = cellfun(@numel, regexp(rseq, motifs{i}))>=numhits(hits(j));
        P(j) = sum(numrhits)/500;
        fprintf('round %d/%d: %d iterations to go\n', i, length(motifs), length(hits)-j);
    end
    motifhit{i,1} = [hits P];
end

%Convert significant hits from linidx to index
%%
motifsigidx = zeros(length(accs), length(motifs));
for i = 1:length(motifhit)
    test = motifhit{i};
    sigidx = test(:,2)<=0.05;
    linidx = test(sigidx,1);
    motifsigidx(linidx,i) = 1;
end

motiftotidx = zeros(length(accs), length(motifs));
for i = 1:length(motifhit)
    test = motifhit{i};
    sigidx = true(length(test),1);
    linidx = test(sigidx,1);
    motiftotidx(linidx,i) = 1;
end
%%
dataout = zeros(6,length(motifs));
for i = 1:length(motifs);
    dataout(1,i) = sum(motiftotidx(binidx(:,1),i)/sum(binidx(:,1))); %UP-Total
    dataout(2,i) = sum(motiftotidx(binidx(:,2),i)/sum(binidx(:,2))); %NR-Total
    dataout(3,i)  = sum(motiftotidx(binidx(:,3),i)/sum(binidx(:,3))); %DN-Total
    dataout(4,i) = sum(motifsigidx(binidx(:,1),i)/sum(binidx(:,1))); %UP-SIG
    dataout(5,i) = sum(motifsigidx(binidx(:,2),i)/sum(binidx(:,2))); %NR-SIG
    dataout(6,i)  = sum(motifsigidx(binidx(:,3),i)/sum(binidx(:,3))); %DN-SIG
end
dataout = dataout*100;

%%
bar(dataout(1:3,:))
hold on; bar(dataout(4:6,:))
axis([0.5 3.5 0 100])
```

RBP motif search for 200 nt local sequences centered on seed-sites.

```
%% motifsearch2script searches for RBP motifs within 200 nt of seed-sites
%% For expressed m21 anyseed
m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);
seed = 'taagct';
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);

%% For expressed m122 anyseed
```

```
m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);

%% For expressed m21 conserved anyseed
m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);
%% For expressed m122 conserved anyseed
m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);
%%
firstseeds = cellfun(@(x) min(x(x>200)), seedlocs,'UniformOutput', false);
toclose2startidx = ~cellfun(@(x) length(x), firstseeds);
firstseeds(toclose2startidx) = {0};
firstseeds = cell2mat(firstseeds);
utrlengths = cellfun(@(x) length(x), utrs);
% seedlocs = strfind(utrs, seed);
% firstseeds = cellfun(@(x) x(1), seedlocs);
% toclose2startidx = firstseeds < 200;
toclose2endidx = (utrlengths - firstseeds) < 200;
tocloseidx = toclose2startidx | toclose2endidx;
%%
flankidx = zeros(length(utrs),2);
flankidx(~tocloseidx,1) = firstseeds(~tocloseidx)-100;
flankidx(~tocloseidx,2) = firstseeds(~tocloseidx)+100;
flankidx(toclose2startidx,:) = repmat([1 201], numel(find(toclose2startidx == 1)), 1);
flankidx(toclose2endidx,1) = utrlengths(toclose2endidx)-200;
flankidx(toclose2endidx,2) = utrlengths(toclose2endidx);
%find utrs smaller than 200 nt
tosmallidx = flankidx(:,1)<1;
flankidx(tosmallidx,1) = 1;
flankidx(tosmallidx,2) = utrlengths(tosmallidx);
%
% %%
utrflanks = cell(length(utrs),1);
for i = 1:length(utrs)
    seq = utrs{i};
    utrflanks(i,1) = cellstr(seq(flankidx(i,1):flankidx(i,2)));
end
%%
motifs = {'ttttaaa'; 'ttt\wttt'; 'tgta\wata'; 'tatttat'};
motifhit = cell(length(motifs),2);
for i = 1:length(motifs)
    numhits = cellfun(@numel, regexp(utrflanks, motifs{i}));
    hits = find(~cellfun(@isempty, regexp(utrflanks, motifs{i})));
    P = zeros(length(hits),1);
    for j = 1:length(hits)
        seqcall = utrflanks{hits(j)};
        rseq = cell(1000,1);
        for k = 1:1000
            rseq(k) = cellstr(lower(randseq(length(seqcall), 'fromstructure',
basecount(seqcall))));
        end
        rhits = find(~cellfun(@isempty, regexp(rseq, motifs{i})));
        numrhits = cellfun(@numel, regexp(rseq, motifs{i}))>=numhits(hits(j));
        P(j) = sum(numrhits)/1000;
        fprintf('round %d/%d: %d iterations to go\n', i, length(motifs), length(hits)-j);
    end
    motifhit{i,1} = [hits P];
end

%Convert significant hits from linidx to index
```

```
%%
motifsigidx = zeros(length(accs), length(motifs));
for i = 1:length(motifhit)
    test = motifhit{i};
    sigidx = test(:,2)<=0.05;
    linidx = test(sigidx,1);
    motifsigidx(linidx,i) = 1;
end

motiftotidx = zeros(length(accs), length(motifs));
for i = 1:length(motifhit)
    test = motifhit{i};
    sigidx = true(length(test),1);
    linidx = test(sigidx,1);
    motiftotidx(linidx,i) = 1;
end
%%
dataout = zeros(6,length(motifs));
for i = 1:length(motifs);
    dataout(1,i) = sum(motiftotidx(binidx(:,1),i)/sum(binidx(:,1))); %UP-Total
    dataout(2,i) = sum(motiftotidx(binidx(:,2),i)/sum(binidx(:,2))); %NR-Total
    dataout(3,i)  = sum(motiftotidx(binidx(:,3),i)/sum(binidx(:,3))); %DN-Total
    dataout(4,i) = sum(motifsigidx(binidx(:,1),i)/sum(binidx(:,1))); %UP-SIG
    dataout(5,i) = sum(motifsigidx(binidx(:,2),i)/sum(binidx(:,2))); %NR-SIG
    dataout(6,i)  = sum(motifsigidx(binidx(:,3),i)/sum(binidx(:,3))); %DN-SIG
end
dataout = dataout*100;

%%
bar(dataout(1:3,:))
hold on; bar(dataout(4:6,:))
axis([0.5 3.5 0 60])
```

Relative seed location script

```
%% relativeseedlocationscript2 finds seed locations and normalizes to 3'UTR length

%% For expressed m21 anyseed
m21data = mtxseed2(an_21DM,'m21','anyseed',gen_refmap);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);
%% For expressed m122 anyseed
m122data = mtxseed2(an122DM,'m122','anyseed',gen_refmap);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);

%% For expressed m21 conserved anyseed
m21data = conservedsites(an_21DM,'m21','anyseed',gen_refmap,hsaGBmap2);
expm21 = m21data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm21,'m21',gen_refmap);
accs = getLongUtrData2(expm21,'m21',gen_refmap);
binidx = binDM(expm21);
seedlocs = cellfun(@(x) x.UTR3.m21.index, accs,'UniformOutput', false);

%% For expressed m122 conserved anyseed
m122data = conservedsites(an122DM,'m122','anyseed',gen_refmap,hsaGBmap2);
expm122 = m122data(expressedgenes,:);
utrs = getLongUtrSeqs2(expm122,'m122',gen_refmap);
accs = getLongUtrData2(expm122,'m122',gen_refmap);
binidx = binDM(expm122);
seedlocs = cellfun(@(x) x.UTR3.m122.index, accs,'UniformOutput', false);
%%
utrlengths = cell2mat(cellfun(@(x) length(x.UTR3.Sequence), accs, 'UniformOutput',
false));
seeds = cellfun(@(x) x(x>0),seedlocs,'UniformOutput',false);
relseeds = cell(length(utrlengths),1);
```

```
for i = 1:length(utrlengths)
    relseeds(i,1) = cellfun(@(x) x/utrlengths(i), seeds(i),'UniformOutput', false);
end

test = relseeds;
up = cell2mat(test(binidx(:,1)));
nc = cell2mat(test(binidx(:,2)));
dn = cell2mat(test(binidx(:,3)));

uphist = histc(up,0:0.1:1.0);
nchist = histc(nc,0:0.1:1.0);
dnhist = histc(dn,0:0.1:1.0);

dataout = [uphist nchist dnhist];
```

This script was used to build the local human genbank database. A similar method was used for the mouse database.

```
%%Build human genref map
cd('/Users/Shared/Genbank_refseq_HSA')
accString = ls;
idxN = strfind(accString, 'N');
fileList = cell(1,1);
for i = 1:length(idxN)
    if i == length(idxN)
        fileList(i) = cellstr(deblank(accString(idxN(i):end)));
    else
        fileList(i,1) = cellstr(deblank(accString(idxN(i):idxN(i+1)-1)));
    end

end

for i = 1:length(fileList)
    hsaGB(i) = genbankread(fileList{i});
    if rem(i,100) == 0
        disp(i)
    end
end

%% Find genbank data with same gene symbol
LocusNames = arrayfun(@(x) x.LocusName, hsaGB,'UniformOutput',false);
geneSym = values(ref_gene_master_map, LocusNames);
uniqueGeneSym = unique(geneSym);
multEntryIdx = cell(length(uniqueGeneSym),1);
for i = 1:length(uniqueGeneSym)
    multEntryIdx(i) = {find(~cellfun(@isempty, strfind(geneSym, uniqueGeneSym{i})))};
    if rem(i,100) == 0
        disp(i)
    end
end

%% Create new GB map structure with same genesym accessions grouped together
hsaGBmap = containers.Map(uniqueGeneSym{1}, hsaGB(multEntryIdx{1}));
for i = 1:length(uniqueGeneSym)
    hsaGBmap(uniqueGeneSym{i}) = hsaGB(multEntryIdx{i});
end

%%
genbankdata = values(hsaGBmap);
%Create 3UTR field in genbankdata structures
for i = 1:length(genbankdata)
    if strcmp(genbankdata{i},'No Accession Number') == 1
    else
        count = numel(genbankdata{i});
        for j = 1:count
            if isempty(genbankdata{i}(j).CDS) ~= 1
                CDSstring = genbankdata{i}(j).CDS.location;
                CDSend_idx = strfind(CDSstring,'..');
                CDSend = str2num(CDSstring(CDSend_idx+2:end));
                UTR3 = genbankdata{i}(j).Sequence(CDSend+1:end);
                genbankdata{i}(j).UTR3.Sequence = UTR3;

                %Search for m21 seeds
```

```
                m21hit = seedmatch(UTR3, 'hsa-miR-21');
                if isempty(m21hit) == 1
                    genbankdata{i}(j).UTR3.m21.match = 'no';
                    genbankdata{i}(j).UTR3.m21.index = m21hit;

                elseif sum(m21hit(1,:)) >= 1
                    genbankdata{i}(j).UTR3.m21.match = '8m';
                    genbankdata{i}(j).UTR3.m21.index = m21hit;

                elseif sum(m21hit(2,:)) >= 1
                    genbankdata{i}(j).UTR3.m21.match = '7m8';
                    genbankdata{i}(j).UTR3.m21.index = m21hit;

                elseif sum(m21hit(3,:)) >= 1
                    genbankdata{i}(j).UTR3.m21.match = '7m1A';
                    genbankdata{i}(j).UTR3.m21.index = m21hit;

                end

                %Search for m122 seeds
                m122hit = seedmatch(UTR3, 'hsa-miR-122');
                   if isempty(m122hit) == 1
                    genbankdata{i}(j).UTR3.m122.match = 'no';
                    genbankdata{i}(j).UTR3.m122.index = m122hit;

                elseif sum(m122hit(1,:)) >= 1
                    genbankdata{i}(j).UTR3.m122.match = '8m';
                    genbankdata{i}(j).UTR3.m122.index = m122hit;

                elseif sum(m122hit(2,:)) >= 1
                    genbankdata{i}(j).UTR3.m122.match = '7m8';
                    genbankdata{i}(j).UTR3.m122.index = m122hit;

                elseif sum(m122hit(3,:)) >= 1
                    genbankdata{i}(j).UTR3.m122.match = '7m1A';
                    genbankdata{i}(j).UTR3.m122.index = m122hit;

                end

                %Search for let7 seeds
                let7hit = seedmatch(UTR3, 'hsa-let-7a');
                if isempty(let7hit) == 1
                    genbankdata{i}(j).UTR3.let7.match = 'no';
                    genbankdata{i}(j).UTR3.let7.index = let7hit;

                elseif sum(let7hit(1,:)) >= 1
                    genbankdata{i}(j).UTR3.let7.match = '8m';
                    genbankdata{i}(j).UTR3.let7.index = let7hit;

                elseif sum(let7hit(2,:)) >= 1
                    genbankdata{i}(j).UTR3.let7.match = '7m8';
                    genbankdata{i}(j).UTR3.let7.index = let7hit;

                elseif sum(let7hit(3,:)) >= 1
                    genbankdata{i}(j).UTR3.let7.match = '7m1A';
                    genbankdata{i}(j).UTR3.let7.index = let7hit;

                end

            else
                genbankdata{i}(j).UTR3.Sequence = {};
            end
        end
    end
    if rem(i,100) == 0
        fprintf('Iterations Remaining:%0.0f\n',length(genbankdata) - i)
    end
end

%% Create another new GB map structure with modified genbank data
hsaGBmap2 = containers.Map(keys(hsaGBmap),genbankdata);
```

These functions are commonly called in the preceding scripts.

```
function idxout = binDM(dataMtx)
%binDM Returns indices for upregulated, no change, and downregulated genes from input
datamatrix

%dataMtx = sortrows(dataMtx);
binUPidx = dataMtx(:,1) >= 1.1 & dataMtx(:,2) <= 0.05;
binNCidx = dataMtx(:,1) < 1.1 & dataMtx(:,1) > -1.1;
binDNidx = dataMtx(:,1) <= -1.1 & dataMtx(:,2) <= 0.05;

idxout = [binUPidx binNCidx binDNidx];
end


function newdataMtx = conservedsites(dataMtx, miRNA, seed, gen_refmap, hsaGBmap2)
%conservedsites Accepts a mouse datamatrix,  filters for a particular miRNA and
%returns the genes with verified site conservation in human

mmudata = mtxseed2(dataMtx, miRNA, seed, gen_refmap);
mmugenes = rownames(mmudata);
mmugenesclean = cellfun(@(x) genecleanup(x, ' /// '), mmugenes,'UniformOutput',false);
hsagenes = cnvgenes_human(mmugenesclean, hsaGBmap2);
emptyidx = cellfun(@isempty, hsagenes);
hsaGB = values(hsaGBmap2, hsagenes(~emptyidx));

conservedidx1 = zeros(1,1);
for i = 1:length(hsaGB)
    hasutridx = cellfun(@length, arrayfun(@(x) fieldnames(x.UTR3),hsaGB{i},
'UniformOutput', false)) == 4;
    tempGB = hsaGB{i}(hasutridx);
    if strcmp(seed, 'anyseed')
        yesidx =  ~arrayfun(@(x) isempty(eval(sprintf('x.UTR3.%s.index',
miRNA))),tempGB);
    else
        matchanswers = arrayfun(@(x) eval(sprintf('x.UTR3.%s.match', miRNA)),
tempGB,'UniformOutput',false);
        yesidx = strcmp(matchanswers, seed(1:end-2));
    end
    if sum(yesidx) >= 1
        conservedidx1(i,1) = 1;
    else
        conservedidx1(i,1) = 0;
    end
end

conservedidx2 = zeros(size(emptyidx));
conservedidx2(~emptyidx) = conservedidx1;
conservedidx2 = logical(conservedidx2);

newdataMtx = sortrows(mmudata(unique(mmugenes(conservedidx2)),:));
end


function [newDataMatrix] = mtxseed2(datamat, miRNA, seed, gen_refmap, varargin)
%mtxseed Will search a datamatrix for genes matching a specific miRNA
%(currently limited to miR-21 (m21), miR-122 (m122), or let-7 (let7)). The
%seed type should be specified as '8mer' '7m8' '7m1A' or 'anyseed'. Input
%must also include the gen_refmap containing genes keys and genbank
%structure values for each gene, prematched to the miRNAs. Varargin
%includes 'sig' for only genes having significant (P < 0.05) change and/or
%a threshold value (example: '> 1.6'). Threshold cannot be applied without
%specifying significant filter (must be varargin{2}). Specify 'nosig' if
%significant filter should not be on when threshold is applied.

n = nargin;
%error check
if strcmp(seed, '8mer') || strcmp(seed,'7m8') || strcmp(seed,'7m1A') ||
strcmp(seed,'anyseed') || strcmp(seed,'noseed')

    genes = rownames(datamat);
```

```
    switch seed
        case '8mer'
            seed_idx = '1';
        case '7m8'
            seed_idx = '2';
        case '7m1A'
            seed_idx = '3';
        case 'anyseed'
            seed_idx = ':';
        case 'noseed'
            seed_idx = ':';
    end

    k = 0;
    for i = 1:length(genes)
        key = genes{i};
        entry = gen_refmap(key);
        entry_count = numel(entry);
        j = 0;
        while j < entry_count
            j = j+1;
            if ischar(entry(j))
            elseif isempty(entry(j).UTR3.Sequence) == 1
            elseif eval(sprintf('sum(sum(entry(j).UTR3.%s.index(%s,:)))',miRNA,
seed_idx)) > 0
                k = k+1;
                newgenes(k,1) = genes(i,1);
                j = entry_count+1;
            end
        end
    end
    newDataMatrix = datamat(unique(newgenes),:);
    newDataMatrix = sortrows(newDataMatrix,'Fold Change','descend');
    if strcmp(seed,'noseed')
        newDataMatrix = datamat;
        newDataMatrix(unique(newgenes),:) = [];
    end
    if n >= 5 && strcmp(varargin{1},'sig') == 1
        newDataMatrix = newDataMatrix(newDataMatrix(:,2)<=0.05,:);
    end
    if n == 6
        threshold = varargin{2};
        rows = eval(sprintf('newDataMatrix(:,1) %s',threshold));
        newDataMatrix = newDataMatrix(rows,:);
    end

else
    fprintf('Invalid Input. Use 8mer, 7m8, 71mA, anyseed or noseed');

end
end



function UtrCell = getLongUtrSeqs2(dataMatrix, miRNA, genrefmap)
%getLongUtrSeqs Returns the longest seed-matched UTR sequences
%   can accept datamatrix or gene list as input

%check if input is a cell list of genes or dataMatrix
inputcheck = iscell(dataMatrix);
if inputcheck == 0
    genes = rownames(dataMatrix);
else
    genes = dataMatrix;
end

%find empty values
emptyidx = cellfun(@isempty, genes);
UtrCell = cell(size(emptyidx));
genes = genes(~emptyidx);
data = values(genrefmap, genes);
UTRs = cell(1,1);
for i = 1:length(data)
```

```
    if length(data{i}) == 1
        if length(fieldnames(data{i}.UTR3)) == 1
            UTRs(i,1) = cellstr('NUM');
        else
            UTRs(i,1) = cellstr(data{i}.UTR3.Sequence);
        end
    else
        %Find the longest UTR for that genesymbol
        LocusName = arrayfun(@(x) x.LocusName, data{i},'UniformOutput',false);
        options = data{i};
        options = options(~cellfun(@isempty, regexp(LocusName,'\wM')));
        idx = ~arrayfun(@(x) isempty(eval(sprintf('x.UTR3.%s.index', miRNA))),options);
        if numel(idx) == 0
            UTRs(i,1) = cellstr('NUM');
        elseif numel(idx) == 1
            UTRs(i,1) = cellstr(options(idx).UTR3.Sequence);
        elseif numel(idx) > 1
            options = options(idx);
            [~, idx] = max(arrayfun(@(x) length(x.UTR3.Sequence),options));
            UTRs(i,1) = cellstr(options(idx).UTR3.Sequence);
        elseif numel(idx) < 1
            disp(i)
        end
    end
end
UtrCell(~emptyidx) = UTRs;
UtrCell(emptyidx) = cellstr('NGM');
end



function DataCell = getLongUtrData2(dataMatrix, miRNA, genrefmap)
%getLongUtrSeqs Returns DATA for the longest seed-matched UTR sequences
%   can accept datamatrix or gene list as input

%check if input is a cell list of genes or dataMatrix
inputcheck = iscell(dataMatrix);
if inputcheck == 0
    genes = rownames(dataMatrix);
else
    genes = dataMatrix;
end

%find empty values
emptyidx = cellfun(@isempty, genes);
DataCell = cell(size(emptyidx));
genes = genes(~emptyidx);
data = values(genrefmap, genes);
UTRs = cell(1,1);
for i = 1:length(data)
    if length(data{i}) == 1
        if length(fieldnames(data{i}.UTR3)) == 1
            UTRs(i,1) = cellstr('NUM');
        else
            UTRs{i,1} = data{i};
        end
    else
        %Find the longest UTR for that genesymbol
        LocusName = arrayfun(@(x) x.LocusName, data{i},'UniformOutput',false);
        options = data{i};
        options = options(~cellfun(@isempty, regexp(LocusName,'\wM')));
        idx = ~arrayfun(@(x) isempty(eval(sprintf('x.UTR3.%s.index', miRNA))),options);
        if numel(idx) == 0
            UTRs(i,1) = cellstr('NUM');
        elseif numel(idx) == 1
            UTRs{i,1} = options(idx);
        elseif numel(idx) > 1
            options = options(idx);
            [~, idx] = max(arrayfun(@(x) length(x.UTR3.Sequence),options));
            UTRs{i,1} = options(idx);
        elseif numel(idx) < 1
            disp(i)
```

```
        end
    end
end
DataCell(~emptyidx) = UTRs;
DataCell(emptyidx) = cellstr('NGM');
end


function [seed_index_matrix] = seedmatch(sequence, miRNA)
%seedmatch Searches input sequence for a match to a given seed
% User must specify sequence to search and miRNA using miRbase nomenclature
% PrintValue == 1 will return a prompt with the number of seed_matches
% found

seed_index_cell = cell(3,1);
seed_8mer = miRmatchlookup(miRNA, '8mer');
seed_7m8 = miRmatchlookup(miRNA, '7m8');
seed_7m1A = miRmatchlookup(miRNA, '7m1A');

s8mer_idx = strfind(sequence, seed_8mer);

s7m8_idx = strfind(sequence, seed_7m8);

s7m1A_idx = strfind(sequence, seed_7m1A);


s8hits = numel(s8mer_idx);
seed_index_matrix(1,1:s8hits) = s8mer_idx;

s7m8hits = numel(setdiff(s7m8_idx,s8mer_idx));
seed_index_matrix(2,1:s7m8hits) = setdiff(s7m8_idx,s8mer_idx);

s7m1Ahits = numel(setdiff(s7m1A_idx-1, unique([s8mer_idx s7m8_idx]))+1);
seed_index_matrix(3,1:s7m1Ahits) = setdiff(s7m1A_idx-1, unique([s8mer_idx s7m8_idx]))+1;


end


function [seed_match] = miRmatchlookup(miRbase_miRNA_name, seed_type)
%miRmatchlookup Finds seed_matches (on mRNA) to a particular miRNA
%   Input miRNA name should use miRbase nomenclature standards, for
%   example: mmu-miR-21 for mouse miRNA-21. Seed_type options include 6mer,
%   7m8, 7m1A, and 8mer.

load miRfamilymatch_map

seed7m8 = miRfamilymatch_map(miRbase_miRNA_name);

if strcmp(seed_type, '6mer')
    seed_match = seed7m8(2:end);

elseif strcmp(seed_type, '7m8')
    seed_match = seed7m8;

elseif strcmp(seed_type, '7m1A')
    seed_match = sprintf('%sa',seed7m8(2:end));

elseif strcmp(seed_type, '8mer')
    seed_match = sprintf('%sa',seed7m8);

end


function [array] = getlog2(datamat)
%getlog2 Computes the log2 data from linear fold change in a datamatrix

values = double(datamat(:,1));
values(values>0) = log2(values(values>0));
values(values<0) = log2(-1./(values(values<0)));
array = values;

end
```

```
function DataCell = getLongUtrData3nomir(dataMatrix, genrefmap)
%getLongUtrSeqs Returns DATA for the longest transcripts
%   can accept datamatrix or gene list as input

%check if input is a cell list of genes or dataMatrix
inputcheck = iscell(dataMatrix);
if inputcheck == 0
    genes = rownames(dataMatrix);
else
    genes = dataMatrix;
end

%find empty values
emptyidx = cellfun(@isempty, genes);
DataCell = cell(size(emptyidx));
genes = genes(~emptyidx);
data = values(genrefmap, genes);
UTRs = cell(1,1);
for i = 1:length(data)

    if ~isstruct(data{i})
        UTRs(i,1) = cellstr('NUM');
    elseif length(data{i}) == 1
        if length(fieldnames(data{i}.UTR3)) == 1
            UTRs(i,1) = cellstr('NUM');
        else
            UTRs{i,1} = data{i};
        end
    else
        %Find the longest UTR for that genesymbol
        LocusName = arrayfun(@(x) x.LocusName, data{i},'UniformOutput',false);
        options = data{i};
        options = options(~cellfun(@isempty, regexp(LocusName,'\wM')));
            [~, idx] = max(arrayfun(@(x) length(x.Sequence),options));
            UTRs{i,1} = options(idx);
    end
end
DataCell(~emptyidx) = UTRs;
DataCell(emptyidx) = cellstr('NGM');
end

function [newstr] = genecleanup(str, delimiter)
%genecleanup Cleans up compound gene names containing a delimiter
delim = delimiter;
delim_count = numel(strfind(str, delim));
if delim_count >= 1
holder = cell(delim_count+1,1);
for i = 1:delim_count
    [first str] = strtok(str, delim);
    holder{i} = first;
end
holder{end} = str(6:end);
[a b] = min(cellfun(@numel,holder));
newstr = holder{b};
else
    newstr = str;
end

function [output_genes] = cnvgenes_human(input_genes, hsaGBmap2)
%cnvgenes Converts a cell array of strings of gene names to compatible
%versions for indexing into the anti-miR mouse liver datamatrix arrays
%   IMPORTANT: Input must be a cell array of strings!

output_genes = cell(length(input_genes),1);
mygenes = keys(hsaGBmap2);

for i = 1:length(input_genes)
    match_test = strcmpi(input_genes{i},mygenes);
    if sum(match_test) > 0
        output_genes(i) = mygenes(match_test);
    else
```

```
        search = strfind(lower(mygenes), lower(input_genes{i}));
        idx = cellfun(@isempty, search);
        if sum(idx) == length(mygenes)
        else
            output_genes{i} = mygenes{~idx};
        end
    end

end
end
```

Appendix B

Intracellular single molecule microscopy reveals time and mRNA dependent microRNA assembly[4]

B.1 Abstract

MicroRNAs (miRNAs) associate with components of the RNA induced silencing complex (RISC) to assemble on messenger RNA (mRNA) targets and regulate protein expression in higher eukaryotes. Here, we describe a method for the intracellular single molecule, high-resolution localization and counting (iSHiRLoC) of miRNAs. Microinjected, singly-fluorophore labeled, functional miRNAs were tracked within diffusing particles, a majority of which contained single such miRNA molecules. Observed mobility and mRNA dependent assembly changes suggest the existence of two kinetically distinct pathways involving miRNAs, revealing the dynamic nature of an important gene regulatory pathway and paving the way towards its single molecule systems biology.

B.2 Introduction

Gene regulation by microRNAs (miRNAs) is an evolutionarily conserved RNA silencing pathway wherein ~22-nucleotide-short non-coding RNAs assemble with components of the RNA induced silencing complex (RISC) so that their guide strand can bind partially complementary sequences in mRNA 3' untranslated regions (UTRs) to repress protein expression[17,293]. To date, over 1,500 mammalian miRNAs have been identified that, collectively, are predicted to regulate over 60% of all protein coding genes[291]. The ubiquitous nature of miRNAs implicates their involvement in all aspects of multi-cellular life[17,293], from general cellular processes such as cell differentiation and survival to pathologies such as cancer, thus creating an urgent need to understand the mechanism(s)

[4] Reproduced in part with permission from Pitchiaya, S., Androsavich, J. R., & Walter, N. G. (2012). Intracellular single molecule microscopy reveals two kinetically distinct pathways for microrna assembly. *EMBO Reports.*

by which miRNAs function. While standard ensemble assays, including intracellular fluorescence microscopy[68,294], have revealed a wealth of information[17,293], the mechanism of gene repression by miRNAs is still debated[81,246]. We have developed an incisive tool to quantify the intracellular diffusion and assembly of single functional miRNAs, which are parameters fundamental to their biological function, in pursuit of their still elusive mode of action.

Single molecule fluorescence microscopy (SMFM) has emerged as a powerful tool to quantify properties of biomolecules not accessible to conventional ensemble-averaging techniques[275,295]. In particular, SMFM has been applied to living cells to assess diffusive properties by single particle tracking (SPT)[295] and the stoichiometry of molecular complexes by stepwise photobleaching[198]. The strategies available so far, however, either decorate the target molecule with multiple fluorescent probes, thereby creating high-molecular weight appendages that may impede function, and/or have only been applied to proteins or large RNA-protein complexes[295-297]. Our intracellular single molecule, high-resolution localization and counting (iSHiRLoC) method overcomes these caveats by combining microinjection (which defines an experimental start point) and low-background illumination with SPT and stepwise photobleaching to visualize the diffusive motions and distribution of singly fluorophore labeled functional small RNAs inside cells. We observe two kinetically distinct pathways involving mRNA dependent miRNA assembly processes, providing a direct look at the complexity of the intracellular processes involved in RNA silencing.

B.3 Materials and Methods

Plasmids, DNA and RNA olignucleotides. Luciferase reporter plasmids pmG-mH3U (wt) and pmG-mH3UM (mut) were engineered by inserting the wild-type and mutant mouse HMGA2 3' UTR sequences, respectively, downstream of the firefly luciferase ORF in the pmirGLO dual-luciferase expression vector (Promega). The wild-type and mutant 3' UTRs were PCR amplified from plasmid templates received as a gift from David Bartel (Whitehead Institute, MIT)[173] using primers containing at their 5' ends XhoI and NotI or XhoI and XbaI restriction sequences, respectively. Following

restriction digestion, the amplicon insert was ligated to complementary ends in the pmirGLO vector. Fluorescence reporter plasmid pEF6-mCherry-mH3Um was constructed by first subcloning the mCherry ORF from the pRSET-mCherry vector, a gift from Roger Tsien (UCSD), into the multiple cloning site of the pEF6-myc-His-B (Invitrogen) vector using BamHI and EcoRI restriction sites. Then, using PCR cloning, the mutant HMGA2 3' UTR was inserted downstream of the mCherry ORF between XhoI and NotI sites. pEGFP-C1 was purchased from Clontech. pMCP-EGFP, a plasmid bearing the ORF of the MS2 coat protein fused to EGFP and an SV-40 nuclear localization signal (NLS), and pSL-MS2_24x, a plasmid bearing 24 copies of the MS2 stem loops[162], were received as gifts from Robert Singer (Albert Einstein College of Medicine). A Luciferase reporter plasmid bearing the MS2 stem loops was created in two steps. First, the ORF of IF2, PCR amplified using a forward primer bearing SbfI and EcoRI restriction enzyme sites and a reverse primer bearing a NotI restriction enzyme site, was cloned into the corresponding SbfI and NotI sites in plasmid pmG-mH3UM. 24 copies of the MS2 stem loop (from pSL-MS2_24x) were then cloned into the EcoRI-NotI restriction enzyme sites of the resultant plasmid, pmG-mH3UM-IF2, to generate pmG-mH3UM-MS2. We had to resort to this two-step procedure because linearized plasmids containing the MS2 stem loops often recombined with the genome of the bacterial competent cells we were using, thus resulting in clones bearing smaller plasmids. Clones containing the MS2 stem loops were created in SURE2 bacterial cells (Stratagene) to minimize recombination of the MS2 repeats with the bacterial genome. Plasmid pEGFP-hDcp1a was created by removing the mRFP1 ORF from pmRFP1-hDcp1a (gift from Nancy Kedersha, Brigham Women's hospital) using AgeI and XhoI restriction enzyme digestion and replacing it with the similarly digested ORF of EGFP, which was PCR amplified from pEGFP-C1 (Clontech) using DNA primers bearing AgeI and XhoI restriction enzyme sites. Plasmid pRL-TK-cxcr4-6x [71] containing the renilla luciferase coding sequence and an artificial 3' UTR with six binding sites for an artificial cxcr4 miRNA downstream of the T7 promoter, was purchased from Addgene.

Negative control siRNA used in transfection experiments was purchased from Ambion. SiLuc2 siRNA, a positive control siRNA, was designed towards the coding sequence of the firefly luciferase gene (*luc2*) and purchased from Dharmacon as a duplex.

All other RNA oligonucleotides were purchased from the Keck Biotechnology Resource Laboratory at the Yale University School of Medicine with a 5' phosphate (P) and, in the case of amine modified RNA, a 3' amino group on a C7 carbon linker. RNA was purified as described[249] and the 3' amine groups were labeled with Cy3 or Cy5 succinimidyl ester as described[249,298]. RNA sequences were as follows:

let-7-a1 guide: P-UGAGGUAGUAGGUUGUAUAGUU
let-7-a1-passenger: P-CUAUACAAUCUACUGUCUUUCC
let-7-a1 mut guide: P-UGCGUUAGUAGGUUGUAUAGUU
let-7-a1 mut passenger: P-CUAUACAAUCUACUGUCGUUCC
cxcr4 guide: P-UGUUAGCUGGAGUGAAAACUU
cxcr4 passenger: P-GUUUUCACAAAGCUAACACA
siLuc2 guide: GAAGUGCUCGUCCUCGUCCUU
siLuc2 passenger: GGACGAGGACGAGCACUUCUU

Guide and passenger strands were heat-annealed in a 1:1 ratio, resulting in duplex miRNAs, and frozen for further use. DNA oligonucleotide MS2 with four internal amine modifications was purchased from the Keck Biotechnology Resource Laboratory at the Yale University School of Medicine, purified and labeled as described above. A DNA oligonucleotide completely complementary to MS2 (C-MS2) with a 5' phosphate (P) was purchased from Invitrogen. The labeled and complementary strands were annealed in a 1:1 ratio and frozen for further use. DNA sequences were as follows:

MS2: AXGTCGACCTGCAGACAXGGGTGATCCTCAXGTTTTCTAGGCAATXA
(X represents an amine modified deoxy-uridine)
C-MS2: P-
TAATTGCCTAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACAT

mRNA synthesis. pRL-TK-cxcr4-6x was linearized either with XhoI or NotI restriction enzyme and purified by phenol-chlorofom extraction followed by ethanol precipitation. *In vitro* transcriptions were performed with 1 µg linearized plasmids using the MegaScript T7 kit (Ambion) according to manufacturer‘s protocol. Transcription

reactions were then DNase treated (turbo DNase, Ambion) and the respective RNAs, RL (XhoI digest) and RL-cxcr4-6x (NotI digest), purified by LiCl precipitation. The RNAs were size selected from upon 5% (w/v) urea poly-acrylamide gel electrophoresis, 5'capped with m7G using the ScriptCap capping kit and polyadenylated using the polyA tailing kit (Epicentre). Capped and tailed RNAs were purified by gel filtration (Roche© RNA columns). The length of the polyA tails was estimated based on electrophoretic mobility on a 1.2% formaldehyde agarose gel.

Cell culture and media. HeLa cells (CCL-2, ATCC) were cultured in DMEM (GIBCO, Invitrogen) supplemented with 10% (v/v) fetal bovine serum (FBS) and non-essential amino acids (NEAA) at 37°C. 1 - 1.25e5 cells were seeded onto delta-T dishes (Bioptechs) one day before microinjection, such that they were ~80% confluent at the time of microinjection. Regular medium was replaced with phenol red-free medium 4 h prior to microinjection and subsequently with a minimal medium (HBS), without serum and vitamins, but containing 20 mM HEPES-KOH pH 7.4, 135 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$, 1.8 mM $CaCl_2$ and 5.6 mM glucose, immediately before microinjection. After microinjection, cells were incubated in phenol red-free DMEM containing 2% (v/v) FBS and 1x NEAA and in the presence of a 5% CO2 atmosphere for the indicated amounts of time prior to imaging. For long incubations (4 h and greater), the medium was replaced every 4 h with fresh growth medium containing 2% (v/v) FBS. We found that cellular autofluorescence was a function of serum concentration; higher serum concentrations resulted in higher background and medium with no serum exhibited the least background. We used 2% (v/v) FBS to prevent conditions of serum starvation and simultaneously minimize cellular autofluorescence.

Microinjection. Microinjections were performed using a Femtojet pump (Eppendorf) and a microscope mounted Injectman NI2 micromanipulator (Eppendorf). Solutions to be microinjected were centrifuged at 16,000 x g for 15 min at 4°C just prior to microinjection. For a majority of the tracking and fixed cell experiments, the micropipette

(Femtotips, Eppendorf) was loaded with 1.5 μM fluorophore labeled miRNA and 0.05% 10 kDa Fluorescein dextran (Invitrogen) in PBS. Fluorescein, which is spectrally distinct from Cy5, served as a marker to locate microinjected cells, mainly because Cy5 photobleached relatively quickly (within ~1.5 s). The focal plane to be imaged was adjusted during 488 nm (fluorescein) excitation and then quickly switched to 640 nm (Cy5) excitation for image acquisition. For (negative) control experiments, the miRNA was omitted and either 0.05% Fluorescein dextran (Invitrogen) in PBS or plain PBS were used. Microinjection of fluorescein dextran only resulted in very low background and false-positive particles. In some experiments, the fluorescein dextran was omitted and 2 μM Cy3 labeled miRNA in PBS was used. Although Cy3 labeled particles were visible for a longer time than those of Cy5, the majority of experiments were performed with Cy5 labeled miRNAs in conjunction with fluorescein dextran for two reasons: (i) cellular autofluorescence was higher upon 532 nm (Cy3) excitation as compared to 640 nm excitation; and (ii) it was easier to choose microinjected cells and check for microinjector clogging without photobleaching the dye conjugated to the miRNA. Co-microinjecting let-7-a1-Cy3 along with either fluorescein dextran or Alexa-647 dextran resulted in high background fluorescence and strong Cy3 fluorescence quenching, respectively, when illuminated with at 532 nm. The former background is largely contributed by the spectral overlap between fluorescein and Cy3. For MS2 DNA microinjections, 0.5 μM duplex DNA mixed with 0.05% (w/v) of 10 kDa fluorescein dextran in PBS was microinjected. For co-microinjections of miRNA with mRNA, 1.5 μM cxcr4 miRNA and 0.25 μM mRNAs (RL or RL-cxcr4-6x) were mixed with 0.05% (w/v) of 10 kDa fluorescein dextran in PBS. All of the above microinjections were performed at 100 hPa microinjection pressure for 0.5 s with 20 hPa compensation pressure. Using these conditions, microinjection volume was estimated to be ~0.02 pL or ~0.5% - 5% of the total cell volume, which translates to ~18,000 molecules of miRNA (at 1.5 μM working concentration). For plasmid microinjections, compensation pressure was increased to 40 hPa and plasmids were co-microinjected along with the appropriate amounts of miRNA at a working concentration of 0.1 μg/μl in PBS (see **Repression Assays** for further details).

Fluorescence microscopy. Imaging was performed using a cell-TIRF system based on an Olympus IX81 microscope equipped with a 60x 1.45 NA oil-immersion objective (Olympus), nanometer-precision motorized stage (ASI Imaging), focal drift control module (zero drift control, Olympus), 1x – 4x magnification changer (Olympus) and an EM-CCD camera (Evolve, Photometrics). Solid state lasers with wavelengths of 488 nm (25 mW), 532 nm (100 mW) and 640 nm (100 mW), were directed through an acousto-optical tunable filter, split into different fiber-optic cables and then coupled to the cell-TIRF module. All laser lines had 10 nm bandwidth clean-up filters to ensure monochromatic illumination. Net powers of ~1.2 mW, 7 mW and 8 mW were achieved at the objective for the 488 nm, 532 nm and 640 nm laser lines, respectively. Laser beams were focused on the back-focal plane of the objective and made to travel parallel to the optic axis such that changing the distance from the optic axis controlled incident angle at the dish-media interface. Highly inclined laminar optical sheet (HILO) microscopy [248] was used to achieve deeper penetration into cells without compromising on signal-to-noise ratio. A dual band filter cube consisting of a z491/639rpc dichroic filter (Chroma) and z491/639m emission filter (Chroma) was used to detect GFP and Cy5 emission. Cy3 emission was detected using a Q570LP dichroic filter (Chroma) and HQ610/75m emission filter (Chroma). All videos were recorded at 100 ms camera exposure time using 120x magnification (60x objective and 2x additional magnification using the magnification changer) unless otherwise mentioned. Cells were maintained at 37°C on the microscope stage while imaging using the DeltaT open dish system (Bioptechs) and a heated lid (Bioptechs). 5% CO2 was supplemented to cells using the side port of the heated lid.

Live cell imaging data analysis. The dish was washed several times with HBS immediately after microinjection. Cells were imaged in HBS consisting of 0.3 U/mL Oxyfluor (Oxyrase), 20 mM sodium succinate, 2 mg/mL ascorbic acid and 200 μM trolox (HBS-OSS). Oxyfluor and its substrate succinate act as an oxygen scavenging system, whereas ascorbic acid and trolox served as strong triplet state quenchers, effectively increasing the lifetime of the fluorophores inside the cells.

Single particles were tracked using the tracking module of Imaris (Bitplane). Only particles that lasted for more than 9 frames, in which case the error in the calculated diffusion coefficient will be ~50%[299], were used for further analysis. All tracks were visually inspected to ensure that they arose from well-isolated particles. In-house MATLAB routines were then used to calculate the mean squared displacement (MSD) and diffusion coefficients. Brownian diffusion coefficients were calculated by fitting the mean squared displacement for the first three time intervals to the equation[300]:

$$\left\langle r^2 \right\rangle = 4D\Delta t + offset$$

An offset was used to account for the error in localization at t = 0. Localization precision was measured by tracking the variation in position of fixed miRNA particles (**Supplementary Fig. S1A**). Relative deviation analysis[301] was performed on individual trajectories, based on the equation:

$$RD = \frac{MSD(n)}{4Dn\Delta t}$$

where D is the instantaneous diffusion coefficient and n is a data point at a later time lag, not included in the original calculation of D. An RD value of 1 corresponds to pure Brownian motion. If the MSD time course were to deviate from linearity, such as those in anomalous, directed and corralled motions, RD values will deviate from 1. Thus, by comparing the distribution of RD values between multiple sample sets, we can gauge if a change in mobility of particles is due to the preponderance of one type of diffusive pattern over another.

The analytical forms of MSD time courses for different forms of motion:

$\left\langle r^2 \right\rangle = 4D\Delta t + offset$ (Brownian motion)

$$\left\langle r^2 \right\rangle = \left\langle r_c^2 \right\rangle \left[1 - A_1 e^{\left(-4A_2 D\Delta t / \left\langle r_c^2 \right\rangle\right)} \right] + offset$$ (Corralled motion)

$$\left\langle r^2 \right\rangle = 4D\Delta t + \left(vt\right)^2 + offset$$ (Biased / directed motion with diffusion)

were used to derive diffusion coefficients from data points in Fig. 1e respectively.

Calculation of background intensity was done using ImageJ software (NIH). Briefly, a binary mask, using the background subtracted-thresholded image created by our intensity analysis software (see **Fixed cell imaging and data analysis**), was used to subtract all miRNA particles from an image and the total intensity from the resultant image was used to calculate the average background intensity. Intensity of regions outside the cell was used as the baseline to calculate the relative average background intensity.

The 1 h and 32 h datasets were background subtracted, thresholded and deconvolved to accurately localize single particles. These images were subsequently also used in **Supplementary Fig. S4** to calculate the average number of particles visible for at least nine consecutive frames.

Fixed cell imaging and data analysis. Cells were first washed five times with PBS, fixed using 4% (w/v) paraformaldehyde in PBS for 20 min, washed five times with PBS after fixing and imaged in PBS consisting of the same OSS-antioxidant mix as mentioned above (PBS-OSS).

Intensity analysis to determine number of photobleaching steps was performed using a custom written LabView (National Instruments) code[302]. A Chung-Kennedy non-

linear filter[303] was used to effectively average out the noise within intensity traces, yet preserving fast and sudden transitions, which aided in better visualization of photobleaching steps. The number of fluorophores per spot was estimated based on the number of photobleaching steps. The density of spots within cells was low enough (~0.1 - 0.25 spots/μm^2) to be confident that single particles were discerned. We also note that the number of miRNAs we observed in the focal plane of a microinjected cell (~500 on average per fixed cell, considering both the number of distinct particles and their individual photobleaching steps; **Supplementary Table B.S.1** and **Fig B.4.D**) correlates well with the volume fraction of the cell in focus multiplied with the total number of microinjected miRNAs (~18,000), suggesting that we can account for a significant fraction of the microinjected miRNAs.

For mRNA-miRNA colocalization experiments, HeLa cells were co-microinjected with the MS2/MCP plasmids and the appropriate Cy5 labeled miRNA, incubated for 24 h under normal growth conditions and formaldehyde fixed just prior to imaging. An object based approach[252] was used to quantitatively assess co-localization based on sub-pixel localization of the centers of mass (COMs) of isolated objects from the GFP (green circles) and Cy5 (red circles) channels (**Fig B.3.C**). Two objects, one from each channel, were considered 'perfectly' co-localized if their COMs were localized in the same pixel (yellow circles) or simply co-localized if the radial distance between the COMs was less than the resolution limit. In the latter case, the COMs of the two co-localized objects occur on adjacent, but separate pixels; these were mapped as cyan and magenta circles for Cy5-COMs and GFP-COMs, respectively, for **Fig. B.3.C**. We typically observed a density of 0.08 Cy5 particles/μm^2, 0.28 GFP particles/μm^2 and 0.13 Cy5 particles/μm^2, 0.2 GFP particles/μm^2 in the specific (in the presence of target mRNA) and non-specific (with a negative control miRNA) samples, respectively. The density of green particles arising from background fluorescence was 0.08 GFP particles/μm^2 in samples that were injected with mutant let-7-a1-Cy5 (0.07 particles/μm^2) and MCP plasmid without the MS2-mRNA expression plasmid. We chose wild-type let-7-a1 as a non-specific control miRNA for two reasons: (i) This miRNA is incapable of repressing a reporter target with mutant let-7-a1 seed sequences (**Fig B.1.A**) and (ii) we expect maximal non-specific co-localization of this miRNA with

our reporter mRNA in common cellular RNA foci, such as PBs, since let-7-a1 has many endogenous targets that it would localize to PBs, where the mutant mRNA may also reside due to its regulation by endogenous factors. Fluorophore labeled wild-type let-7-a1 thus serves as a robust negative control miRNA, setting an upper limit for sequence independent co-localization.

Repression assays. For transfection repression assays 24 h prior to transfection, 100 μL of 15,000 HeLa cells were seeded per well of a 96 well plate. Cells were transfected with 60 ng of the indicated plasmid, 10 nM of the indicated dsRNA, 0.4 μL of Lipofectamine 2000 (Invitrogen) and 50 μL of OptiMEM (GIBCO). 6 h after transfection the growth medium was replaced with fresh medium. 24 h after transfection, medium was replaced with phenol red-free DMEM. Dual luciferase assays were performed using the Dual-Glo luciferase assay reagents (Promega) as per the manufacturer's protocol and luminescence was detected using a Genios Pro (Tecan) plate reader. For microinjection repression assays, 0.1 μg/ul of both pEF6-mCherry-mH3UM and pEGFP-C1 plasmids were microinjected with 1.5 μM WT or MUT let-7 miRNA and 0.025% Alexa647 dextran (Invitrogen) in PBS. Following microinjection, cells were maintained under normal growth conditions for 24 h prior to imaging. Fluorescence signal were quantified using ImageJ software (NIH). The background subtracted mCherry signal was normalized with that of GFP for individual cells, which was then averaged over all cells. Using our filter sets (refer to **Microscopy**) we observed a small amount of mCherry bleed-through signal, ~14%, while imaging GFP. This bleed through was corrected for in our calculations. Repression was measured by normalizing the average relative intensity of mutant let-7-a1 injections with respect to that of wild-type let-7-a1 injections.

Estimating the cellular abundance of mRNA targets for let-7 and cxcr4. We first used targetscan 5.2© to identify potential mRNA targets of the let-7 and cxcr4 miRNAs. These targets were then compared to data from a HeLa cell mRNA expression profiling array[304] (dataset accessible at NCBI GEO database[304,305], accession GSM650992) using a custom written MATLAB code to identify endogenously expressed targets of both miRNAs. We found that ~10% of all predicted targets in either datasets were not expressed in HeLa cells. Comparing mRNA abundances of all potential targets for both

miRNAs, we deduced that let-7 is predicted to have ~10-fold more expressed target molecules than cxcr4.

Calculation of miRNA copy number in HeLa cells. We calculated the number of let-7 miRNA molecules per cell based on miRNA expression profiling data[306] and previously reported abundance of let-7a[307] and miR-21[308] in HeLa cells. Fluorescence[307] and qRT-PCR [308] assays have estimated the copy number of let-7a and miR-21 to be ~4,000 copies and ~20,000 copies per cell, respectively. Deep sequencing data[306] (http://www.mirz.unibas.ch/cloningprofiles/) suggest that let-7a, which comprises ~16.5% of the total let-7 population, and miR-21 represent ~1.6% and ~13.2%, respectively, of the total miRNA pool in HeLa cells (~150,000 copies). Taking both of these data sets into consideration, we estimate the total let-7 population to be between 15,000 and 24,000 molecules per HeLa cell.

B.4 Results and Discussion

iSHiRLoC – a biologically relevant, intracellular, single molecule method

To study intracellular miRNA transport and localization, we microinjected mature double-stranded let-7-a1 or artificial cxcr4 miRNA, labeled on the 3'-end of the guide strand with one Cy3 or Cy5 fluorophore, into human HeLa cervical cancer cells. Let-7-a1 miRNA is a member of the highly conserved, tumor suppressing let-7 family whose function is critical to cell proliferation[173]. Let-7-a1 and cxcr4 have ~700 and ~70 predicted mRNA targets in HeLa cells, respectively and both miRNAs have been reported to localize to processing bodies (PBs)[68,153], sub-cellular ribonucleoprotein (RNP) complexes enriched in RNA processing enzymes. Microinjection allowed us to control the number of molecules introduced into the cytoplasm (~18,000) to be similar to the endogenous levels of let-7 miRNAs, minimizing cell perturbation and facilitating single miRNA detection as distinct point-spread functions upon imaging by highly inclined laminar optical sheet (HILO) microscopy[248]. We supplemented cell medium with an oxygen scavenger system to delay photobleaching. Microinjected cells divided normally

after 24 h and remained viable for at least 32 h, attesting to the low invasiveness of iSHiRLoC.

We further sought to ensure that microinjected, fluorophore-labeled miRNAs retain functionality. Transfection experiments showed that a luciferase reporter gene with seven let-7 target sequences was strongly repressed even without adding exogenous let-7-a1, most likely due to the fact that let-7 is one of most abundant miRNA families in HeLa cells. Accordingly, repression was moderately enhanced (only ~2.8 fold) upon co-transfection with let-7-a1 (**Fig B.1.A)**. To enhance the sensitivity of our assay, we included a mutant let-7-a1 that was previously shown to specifically repress a compensatory mutant of the target sequence[173]. As expected, the repression of mutant target by mutant let-7-a1 was strong (~5.6 fold) and specific, consistent with previous observations[173]. In this more sensitive assay, we then showed that attaching either a single Cy3 or Cy5 to the 3'-end of the let-7-a1 guide strand does not significantly compromise repression (**Fig B.1.A**). Finally, we microinjected either mutant or wild-type let-7-a1 along with both an mCherry reporter plasmid containing seven mutant let-7 target sequences and a control GFP reporter plasmid, and found that the intended target protein was specifically and strongly repressed, supporting the biological relevance of microinjecting miRNAs as part of iSHiRLoC (**Fig B.1.B-C**).

Upon RNP assembly, microRNAs diffuse within well-resolved single particles

We next analyzed SPT trajectories of miRNA particles 4 h after microinjecting cells with let-7-a1-Cy3 or let-7-a1-Cy5 (**Fig B.2.A-B**) and found individual particles to undergo (at least) four distinct diffusive patterns (**Fig B.2.C-E**). Control experiments showed that microinjection of PBS buffer alone or fluorescein-conjugated dextran, used as a marker for microinjection, contributed neither to the fluorescence signal nor impacted the function and mobility of miRNA particles (**Fig B.1.B** and **Fig B.2.F**). Our SPT analysis reached a super-accuracy of ~30 nm at 100 ms temporal resolution with an observation window of, on average, ~3 s for let-7-a1-Cy3 (and cxcr4-Cy3) and ~1.5 s for let-7-a1-Cy5, before an individual particle went out of focus or photobleached (**Supplementary**

Fig B.S.1.B). Only very few particles had large changes in intensity over their trajectory (**Supplementary Fig B.S.1.B**) that may report diffusion along the z-axis. However, we cannot rule out that miRNA particles diffuse along the z-axis between successive imaging frames, taken every 100 ms.

Time-averaged mean squared displacement (MSD) analysis of individual SPT trajectories allowed us to characterize the distinct diffusive patterns as biased, corralled, fast or very slow Brownian diffusion, leading to a wide range of diffusion coefficients (**Fig B.2.E-F**, **Supplementary Fig B.S.1.C-E**). Analysis of a large number of trajectories (**Supplementary Table B.S.1**) 4 h after microinjection revealed two distinct Gaussian distributions of microscopic diffusion constants, with average diffusion coefficients of ~0.26 μm^2/s and ~0.034 μm^2/s for both let-7-a1-Cy3 (**Fig B.2.F**) and let-7-a1-Cy5 miRNA (**Fig B.2.G**), smaller than those reported for RISC by fluorescence correlation spectroscopy, ~5.4 μm^2/s[294]. We did not observe more rapidly diffusing particles possibly due to limits in our time resolution (100 ms camera integration time) and/or since they may not be visible for all of 9 frames, our threshold for diffusion coefficient calculation. The main particles we did observe had diffusion coefficients that resemble those of messenger RNPs[162] and PBs (**Supplementary Fig B.S.2**)[165], key mechanistic intermediates of mRNA repression by miRNAs[17,293] that have much higher molecular masses than RISC. These assignments were further supported by the observed co-localization of microinjected miRNAs both with messenger RNPs (mRNPs) containing matching miRNA target sites and with PBs (**Fig B.3**). Taken together, our microinjected miRNAs are functionally fully active (**Fig B.1**), and are associated with mRNAs and PBs (**Fig B.3**), strongly suggesting that the observed particles must contain miRNAs bound to and actively repressing target mRNAs.

The diffusion constants of miRNA containing RNPs increase over an 8-hour period

Next, we asked whether the diffusive properties of miRNA containing particles changed over time. Such changes may be expected for RISC-mediated loading of miRNAs onto mRNA targets and subsequent processing of the resulting mRNPs. To this end, we incubated let-7-a1-Cy5 microinjected cells for 1, 2, 4, 8 and 32 h prior to imaging. After 1 h and 32 h, high background and large inter-frame displacements of individual

particles, respectively, precluded accurate tracking. This observation, together with further analysis (**Supplementary Fig B.S.3**), is consistent with the presence of fast diffusing free miRNAs and smaller miRNA-protein complexes, respectively, that become blurred at our 100-ms time resolution. Notably, after 32 h the 3' fluorophore labeled miRNAs, whose 3' ends are blocked against poly-uridylation - a key step in the miRNA degradation pathway[309], remained detectable as intracellular particles of relatively slow diffusion, suggesting that they were neither (fully) degraded nor exported. For the intermediate time points (2 h, 4 h and 8 h) we again clearly discerned diffusing particles within two distinct Gaussian distributions, whose average Brownian diffusion constants systematically increased over time (**Fig B.2.G**). This increase in mobility over time was not caused by an increase in the fraction of particles undergoing directed motion (**Supplementary Fig B.S.4**), suggesting that, on average, the molecular weights of the RNPs decrease in the 2 to 8 h timeframe. We note that the shifted distributions measured at later time points became increasingly curtailed above a diffusion coefficient of ~1 μm^2/s (**Fig B.2.F-G** and **Supplementary Fig B.S.1.E**) due to our limited time resolution, somewhat diminishing the observed effect by narrowing the fast diffusing population.

Intracellular fluorophores can be counted accurately by stepwise photobleaching

The goal of iSHiRLoC is to measure both diffusion properties and assembly states of individual small RNA containing particles. To assess assembly, we utilized the relatively rapid stepwise photobleaching[198] of single Cy5 fluorophores to quantify the number of miRNA molecules per particle. HeLa cells microinjected with either let-7-a1-Cy5 or cxcr4-Cy5 were formaldehyde-fixed and continuously illuminated until all fluorescent particles had photobleached (**Fig B.4.A-B**). By counting the number of photobleaching steps for many particles (**Fig B.4.C, Supplementary Fig B.S.5.A** and **Supplementary Table B.S.1**), we were able to deduce the distribution of their assembly states (**Fig 4D**). While close to 50% of all particles contained single fluorophore labeled miRNAs 2 h after microinjection, a significant fraction held up to seven labeled miRNAs (**Fig B.4.C-D** and **Supplementary Fig B.S.5.A**). To test for bias from 'dark' (prebleached) Cy5, we

microinjected a control DNA labeled with an average of 3.45 Cy5 fluorophores per molecule and counted a majority of 3 and 4 photobleaching steps in each particle (**Supplementary Fig B.S.5.B-D**), as expected if prebleaching is negligible. In addition, we observed a strong overlap in the distribution of fluorescence intensities between particles in fixed and living cells 2 h after microinjection (**Supplementary Fig B.S.5.E**), when most miRNAs are assembled into RNPs, suggesting that our counting results in fixed cells closely reflect the miRNA assembly states in living cells.

Time and mRNA dependent changes in miRNA assembly support a multiple target turnover model

We monitored miRNA assembly over time at the same time points as in our diffusion measurements. Similar to our observations in living cells, high background made it impossible to accurately localize single miRNA particles in cells that were fixed only 20 min after microinjection (**Supplementary Fig B.S.6.A**). At later time points (1 h to 32 h), we found two distinct assembly phases for let-7-a1 miRNA that fit well with a double-exponential function, suggesting the existence of two kinetically distinct processes (**Fig B.4.E** and **Supplementary Fig B.S.6**). The faster phase is characterized by a rate constant of 1.2 ± 0.2 h^-1, during which the number of RNPs containing more than one miRNA increases; we note that this time evolution resembles that of the initial assembly of miRNAs into RNPs observed over the earliest (1-2 h) live cell time points. The slower phase with a rate constant of 0.14 ± 0.08h^-1 leads to a modest increase in RNPs containing just a single miRNA and matches the timing of the increase in mobility of miRNA containing particles in live cells (**Fig B.2.G** and **Supplementary Fig B.S.3.B**) as well as that of miRNA-induced mRNA target degradation[81]. Based on these observations, we propose the following time-dependent model for miRNA assembly: At initial time points, such as 20 min and up to 1 h (**Supplementary Figs B.S.3.A** and **B.S.6.A**) after microinjection, a significant fraction of our miRNAs remains free, thus contributing to a large background in our live and fixed cell experiments. Between 1 and 2 h, most miRNAs have assembled into miRISC-mRNA complexes, thereby depleting the free miRNA population. This assembly leads to relatively slowly diffusing complexes in

living cells and a corresponding decrease in the population of single miRNAs containing particles in fixed cells. Still later, miRNAs mediate mRNA degradation via association with PBs, eventually releasing miRISC complexes to possibly elicit further rounds of repression, which is represented by a relative increase in the mobility of miRNA particles in living cells and an increase in the population of monomeric miRNAs in fixed cells (**Fig B.5**). Consistent with this model, cells microinjected with the artificial cxcr4 miRNA, which is predicted to find 10-fold fewer mRNA target molecules in a HeLa cell, do not show these time dependent changes in the fractions of single and multiple miRNA containing particles (**Fig B.4.E**; since the decrease in monomeric miRNAs is expected to be ~10-fold smaller for cxcr4 than let-7-a1, it becomes indiscernible in our experiments with an estimated standard deviation about the mean of ~5-10%). By contrast, when cxcr4 was co-microinjected with ~3,000 specific mRNA target molecules, miRNA assembly adopted again two kinetically distinct phases (with rate constants of 1.14 ± 0.12 h^-1 and 0.30 ± 0.07 h^-1; **Fig B.4.E** and **Supplementary Fig B.S.7.A-B**). Additionally, and as expected from our model, miRNAs shifted towards multimeric assemblies when we found them co-localized with target mRNA, but were more monomeric in the absence of a co-localized target (**Supplementary Fig B.S.7.C**). Taken together, we find strong evidence for mRNA dependent miRNA assembly.

In summary, in iSHiRLoC we have developed a broadly applicable method that visualizes both diffusive properties and assembly states of functional small non-coding RNAs in cultured cells. We find that miRNAs exhibit different types of Brownian diffusion in live HeLa cells with (at least) two widely ranging Gaussian distributions of diffusion coefficients, representing two complexes of distinct molecular mass, consistent with messenger RNPs and PBs. Stepwise photobleaching revealed that the largest fraction of these particles contains single fluorophore labeled miRNA molecules. A still significant fraction of particles, however, contains multiple miRNA molecules, strongly invoking the formation of higher-order miRNA complexes either assembled on mRNA targets or associated with PBs. Time dependent changes in diffusion and assembly of miRNAs were observed, supporting the model in **Fig B.5,** featuring two kinetically distinct processes wherein miRNAs first assemble into large RNPs (such as a miRISC bound to an mRNA or PB), then are released from these complexes. In the future,

iSHiRLoC can be extended by multicolor imaging[295] to co-track a miRNA with its corresponding target mRNA, RISC or PB proteins and map the dynamic interaction network of RNA silencing, with the ultimate goal of understanding its mechanism through single molecule systems biology.

B.5 Acknowledgements

We thank D. Bartel, C. Mayr, R. Tsien, N. Kedersha, and R. Singer for generous gifts of plasmids containing the 3' UTR of HMGA2, mCherry ORF, Dcp1a ORF, and the MS2 system of plasmids, A. Manzo for MSD software development, H. Ding, A. Gafni and D. Steel for sharing their intensity analysis routine, the Microscopy and Image Analysis Laboratory at the University of Michigan for access to Imaris, and A. Mapp for access to her plate reader. This work was supported by NIH grant GM081025 to N.G.W. and an NIH Cellular Biotechnology training grant fellowship to J.R.A.

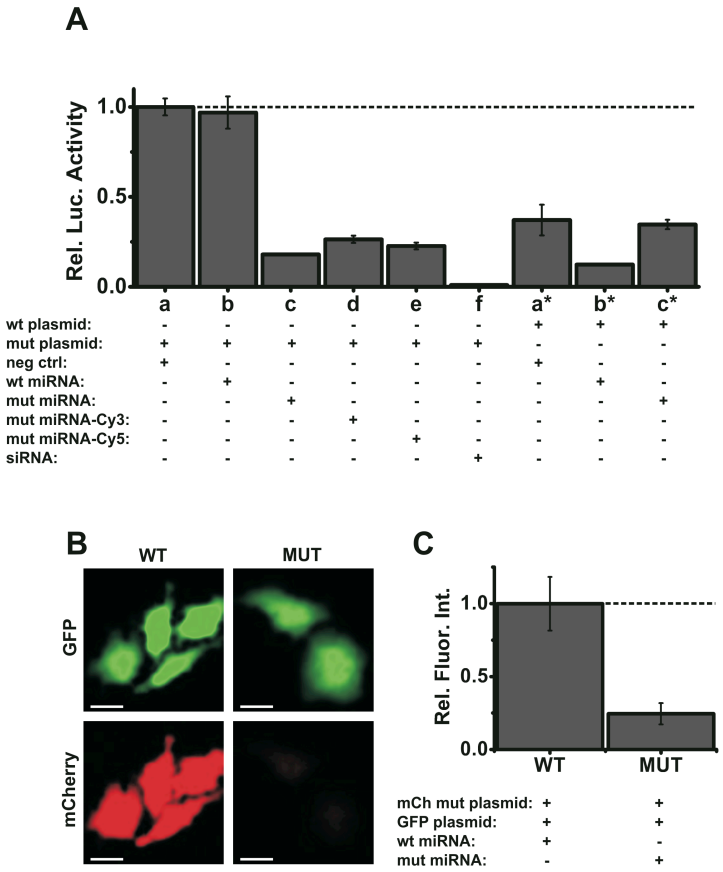


Figure B1. Effect of fluorophore modification and microinjection on miRNA function. (*A*) Luciferase reporter assays of HeLa cells co-transfected with luciferase reporter plasmids bearing the wild-type (wt) or a mutant (mut) 3`UTR of mouse HMGA2, and either a negative control siRNA (neg ctrl), wild-type let-7-a1 (wt miRNA) or mutant let-7-a1 (mut miRNA) miRNA. An siRNA, Siluc2, was used as a positive control for repression. Renilla luciferase activity was used for internal normalization of firefly luciferase activity within each sample. All samples were normalized with respect to negative control (a). Results presented are from four replicates. Error bars, standard deviations. (*B*) Representative images of GFP fluorescence (top) and mCherry fluorescence (bottom) in cells co-injected with an mCherry reporter plasmid, a GFP control plasmid and, either the wild-type (WT) or mutant let-7-a1 (MUT) miRNA are shown. Scale bar, 20 μm. (*C*) Quantification of mCherry fluorescence relative to GFP fluorescence from panel *B*, normalized with respect to the WT sample (N = 3 independent trials, 50 cells per group). Error bars, standard error of the mean.

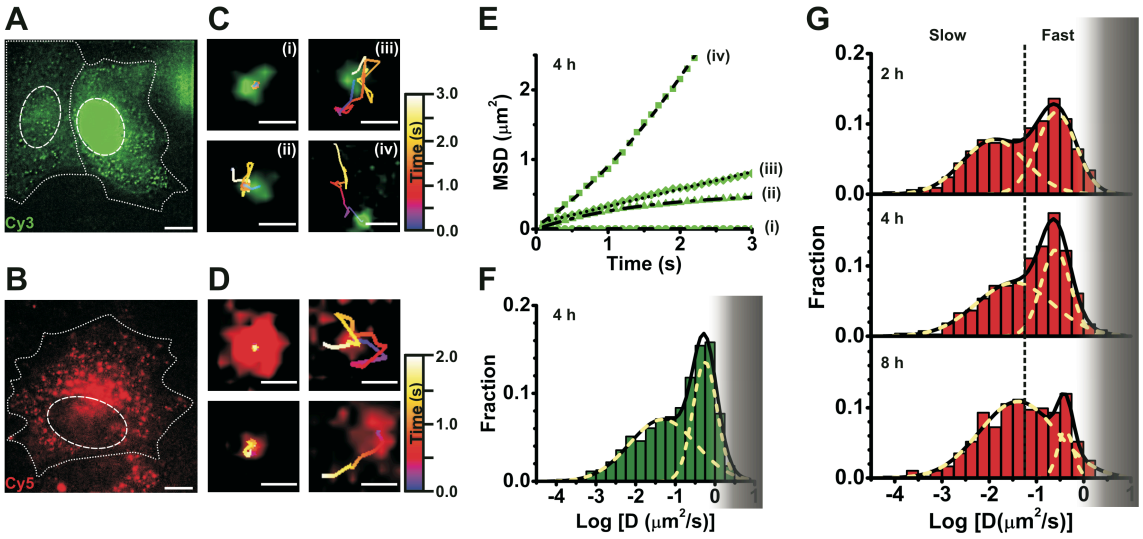


Figure B.2. Single molecule high-resolution localization and tracking of miRNAs diffusing in living HeLa cells. Pseudo-colored images of cells microinjected with (**A**) let-7-a1-Cy3, (left, cytoplasmic injection; right, nuclear injection) imaged 4 h after microinjection and (**B**) let-7-a1-Cy5 imaged 2 h after microinjection, showing distinct particles containing miRNAs. Dashed and dotted lines indicate nuclear and cellular boundaries, respectively. Scale bar, 10 μm. Different types of diffusive motions exhibited by (**C**) let-7-a1-Cy3 and (**D**) let-7-a1-Cy5 miRNAs. Scale bars, 0.5 μm. (**E**) MSD plots of the let-7-a1-Cy3 particles shown in panel **C**. Data were fit with equations representing biased (iv), corralled (ii), fast (iii) and very slow (i) Brownian diffusion. Diffusion coefficients as derived from the fits are: D(i) = 0.0001 μm2/s; D(ii) = 0.06 μm2/s (corral radius = 0.52 μm); D(iii) = 0.062 μm2/s; and D(iv) = 0.16 μm2/s (average velocity = 0.46 μm/s). (**F**) Distribution of diffusion coefficients calculated from individual MSD plots of let-7-a1-Cy3 particles assuming Brownian diffusion. Cells were imaged 4 h after microinjection (N = 4 cells, **Supplementary Table B.S.1**). The gray shaded region represents diffusion coefficients of particles that are increasingly lost due to our limited time resolution of tracking. (**G**) Distribution of diffusion coefficients of let-7-a1-Cy5 at different time points after microinjection. Dotted lines represent demarcations of fast and slow particles to guide the eye, estimated based on segregation of the two Gaussian distributions 2 h after microinjection. Histograms represent data from multiple cells (N = 4, 4, 6 cells for data points corresponding to 2 h, 4 h and 8h, **Supplementary Table B.S.1**). Gray shaded region, as in (**F**).

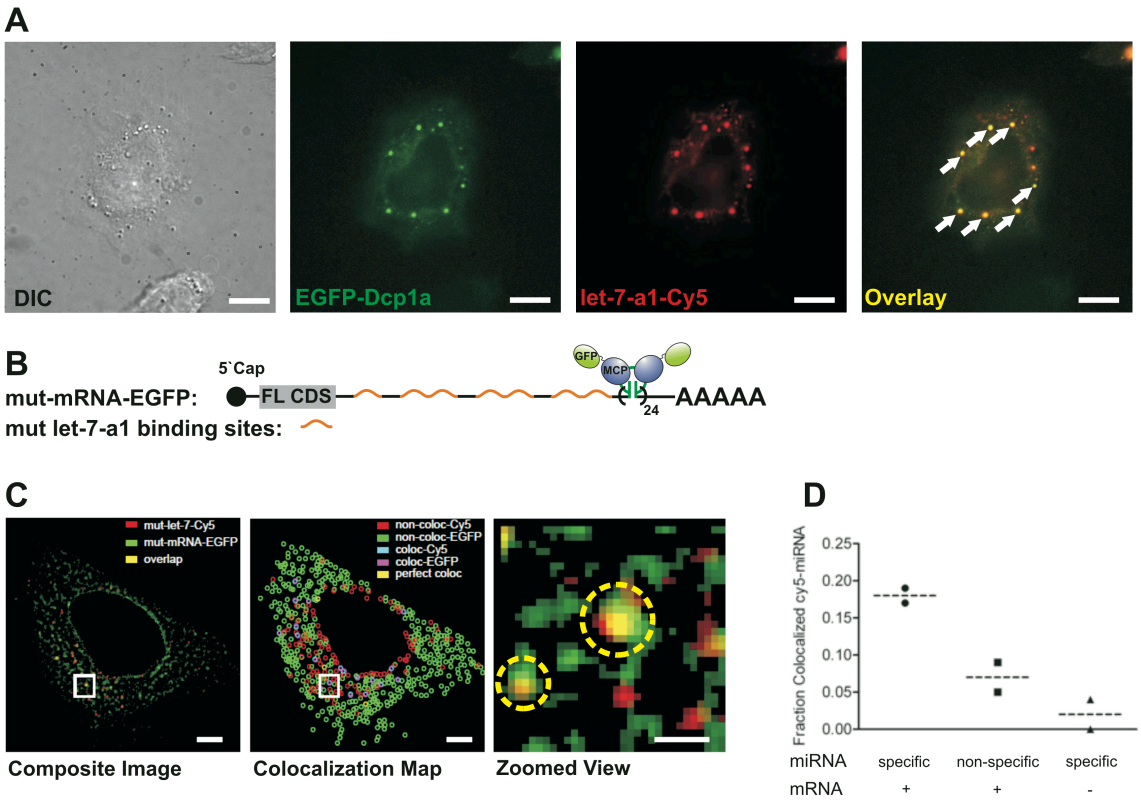


Figure B.3. Co-localization of microinjected miRNAs with PBs and cognate mRNA targets. (**A**) miRNA-PB co-localization. DIC (left) and pseudo-colored images depicting the co-localization (overlay, right) of ectopically expressed EGFP tagged Dcp1a, a component of PBs (green), and microinjected let-7-a1-Cy5 miRNA (red). Wide-field epi-illumination was used to visualize these cells. Arrows show representative foci containing both Dcp1a and miRNA. Scale bar, 10 µm. (**B**) Schematic of a GFP labeled mRNA (top) based on the MS2-MS2 coat protein (MCP) mRNA labeling system [16]. (**C**) miRNA-mRNA co-localization. Left, composite image of GFP (green) and Cy5 (red) channels from a representative cell expressing the mutant-mRNA-EGFP and mutant let-7-a1-Cy5 miRNA. Yellow particles indicate co-localization. Scale bar = 5 µm. Center, co-localization map representing independent miRNA (red) and mRNA (GFP) particles, and either perfectly co-localized (yellow) or adjacent miRNA-mRNA localizations with an inter-particle center-center distance less than the lateral resolution limit of the image (cyan and magenta). Right, zoomed view of the region in the preceding images marked by a white box. Yellow circles indicate co-localized spots. Scale bar = 1 µm. (**D**) Fraction of specific or non-specific miRNA co-localizing with mRNA-EGFP. Cells were injected with the MS2/MCP plasmids and either mutant (specific; circle) or wild-type (non-specific; square) let-7-a1. To test for the contribution of background fluorescence, control cells were injected with mutant let-7-a1 and MCP plasmid without the MS2-mRNA expression plasmid (specific; triangle). Data points within each group and their associated means (dashed lines) are shown (N = 2 cells for each group).

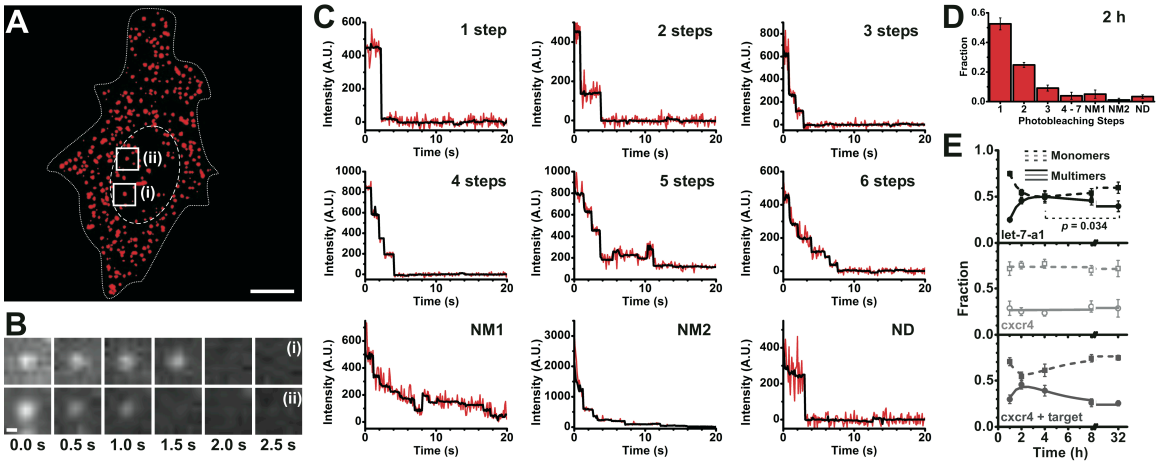


Figure B.4. Counting of single miRNA molecules in fixed HeLa cells. (**A**) A pseudo-colored (background corrected) image of a formaldehyde-fixed HeLa cell showing the intracellular distribution of miRNAs 2 h after microinjection. Dashed and dotted lines indicate nuclear and cellular boundaries, respectively. Scale bar, 10 µm. (**B**) Sets of frames showing photobleaching of two different miRNA particles over the indicated time. The locations of the particles are labeled in panel **A**. Scale bar, 300 nm. (**C**) Representative stepwise photobleaching traces of individual miRNA particles within the cell shown in panel **A**. For some traces, the number of steps could not accurately be determined. These traces were further classified as either non-determinable multimer with indistinct steps (NM1); non-determinable multimer with initial peak intensity exceeding that expected from up to 7 fluorophores based on an average intensity of ~250 arbitrary units (A.U.) per fluorophore bleaching step (NM2); or non-determinable as either monomer or multimer due to large fluctuations in intensity and rapid (< 3 frames) photobleaching (ND). (**D**) Distribution of miRNA photobleaching steps observed 2 h after microinjection (N = 4, **Supplementary Table B.S.1**). Error bars, standard deviation. (**E**) Change in the fraction of monomers (particles undergoing a single step of photobleaching, dotted line) and multimers (particles that bleach in two or more steps, solid line) over time for let-7-a1 (black), cxcr4 (gray) and cxcr4 miRNAs co-microinjected with specific mRNA targets (dark gray, **Supplementary Fig B.S.7A,B**), respectively. Each data point is derived from multiple cells (nlet-7-a1 = 4, 4, 3, 4, 8 cells; ncxcr4 = 5, 5, 5, 7, 9 cells; and ncxcr4 + mRNA = 4, 4, 4, 6 and 8 for 1 h, 2 h, 4 h, 8 h and 32 h, respectively; **Supplementary Table B.S.1**). Error bars, standard deviation. The relative changes marked by asterisk are statistically significant with greater than 95% confidence based on a paired t-test (p = 0.034).

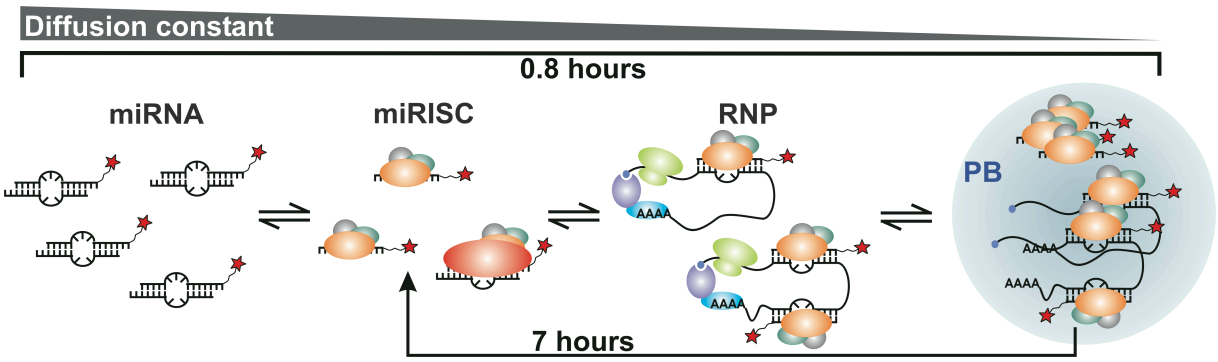
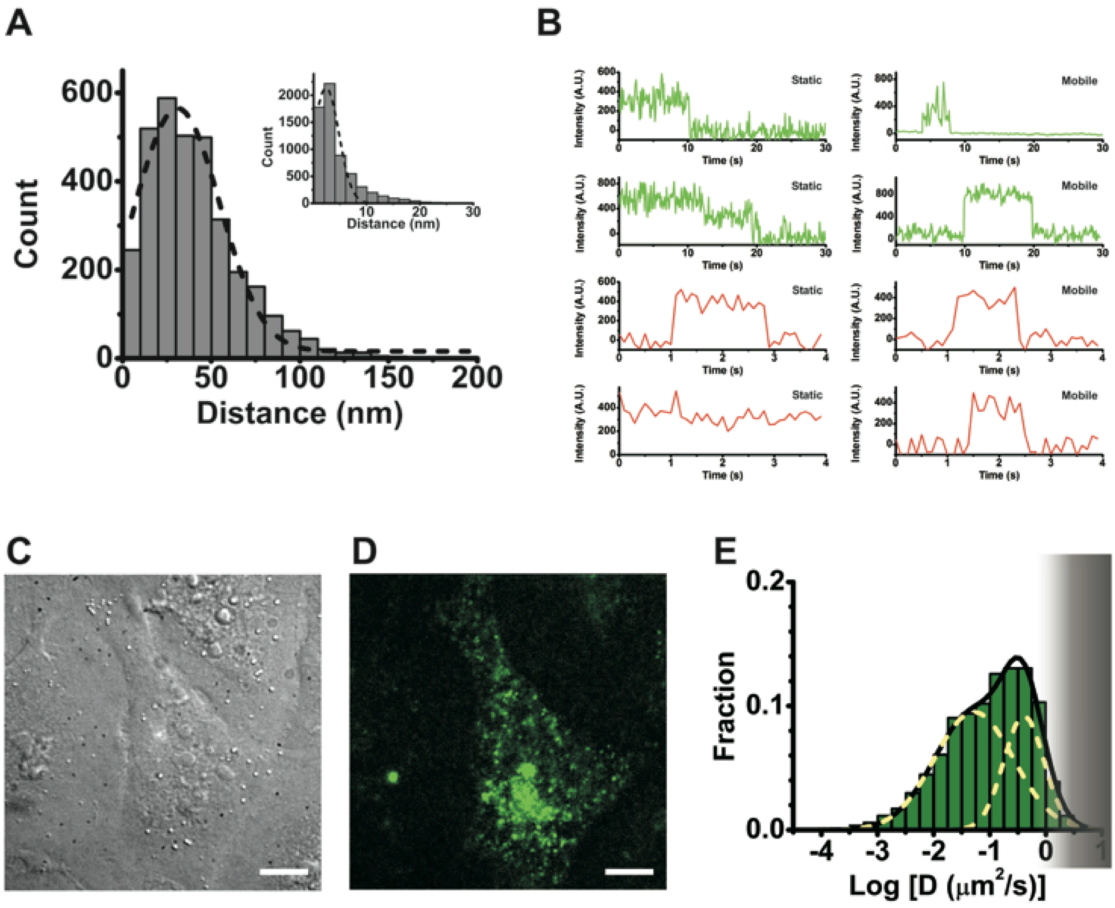
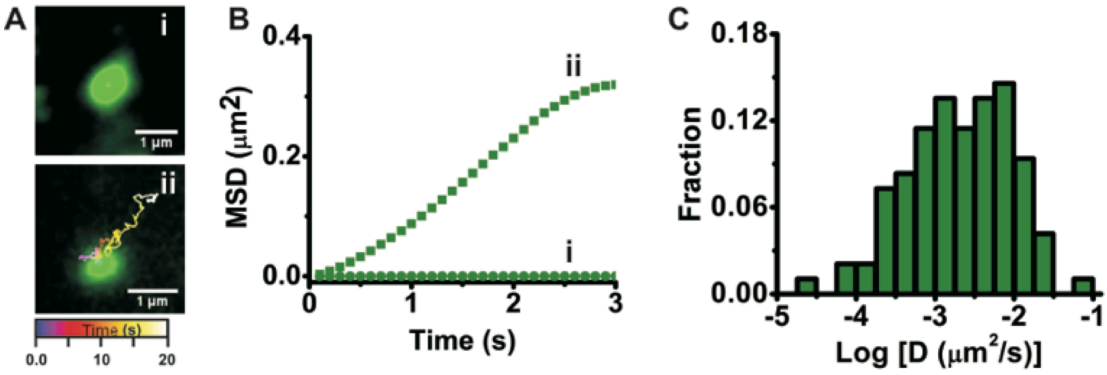


Figure B.5. Model illustrating intracellular miRNA assembly as derived from our data. miRNAs assemble into miRISC complexes that bind translationally active mRNA targets to form large RNPs, associate with PBs, and are finally released upon PB-mediated mRNA degradation to potentially engage in multiple turnover of targets. The indicated timeframes were derived from the kinetics measured in the current study for let-7-a1.



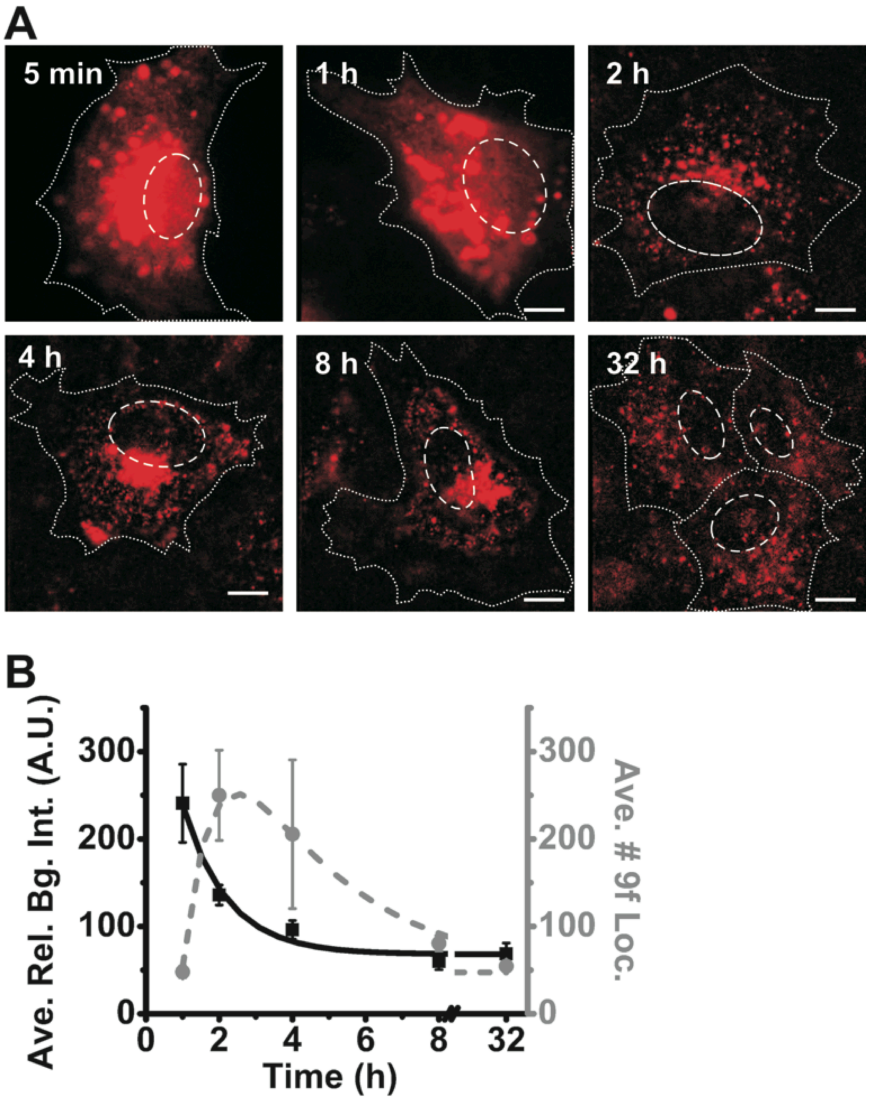
Supplementary Figure B.S.1 Precision and signal intensity of live cell imaging, and cxcr4 control experiments. (**A**) Precision of single particle localization. Histogram depicting the displacement of formaldehyde-fixed let-7-a1-Cy5 particles about their origin (N = 100 randomly selected particles). The location of each particle was tracked over time until it photobleached. The histogram was fit with a Gaussian function, which resulted in a mean of ~30 nm and a standard deviation of ~50 nm. Large jumps (>100 nm) were observed only when there were large fluctuations of intensity. Localization precision of immobilized beads (inset) was ~4 nm, largely owing to their higher signal-to-noise ratio (N = 100 randomly selected particles). (**B**) Signal intensities of particles undergoing diffusion in living cells. Intensity time trajectories of let-7-a1-Cy3 (green) and let-7-a1-Cy5 (red) particles are shown. Trajectories of both mobile and static particles are shown, as indicated in each trajectory. A significant proportion of particles had a relatively constant intensity during diffusion before disappearing. We infrequently observed large fluctuations in intensity as a particle was diffusing, possibly due to diffusion in the axial dimension (z-axis). We also occasionally observed stepwise changes in intensity of static particles, possibly due to photobleaching (see, for example, second trajectory from the top on the left side). (**C-E**) Intracellular diffusion of cxcr4 miRNA, with (**C**) DIC image and (**D**) the corresponding pseudocolored image of a cell microinjected with cxcr4-Cy3 and imaged 4 h after injection. Scale bar, 10 μm. (**E**) Distribution of diffusion coefficients calculated from individual MSD plots of cxcr4-Cy3.

We observed two major Gaussian distributions, with different fractional abundance but average diffusion coefficients similar to those obtained with let-7-a1-Cy3.



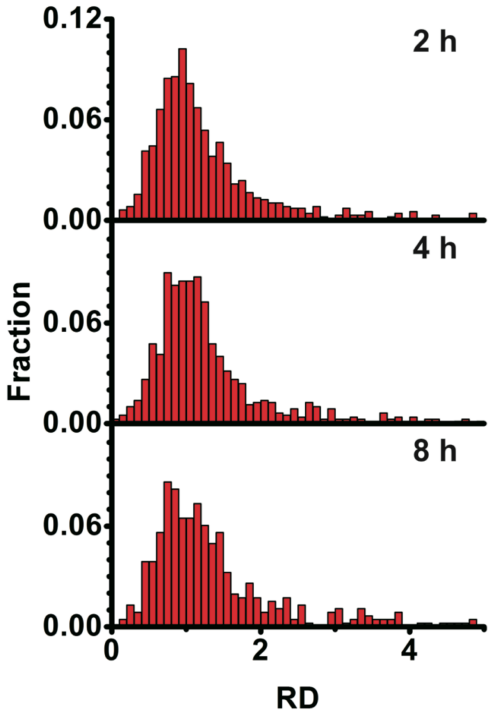
Supplementary Figure B.S.2. Tracking the diffusion of individual PBs in live HeLa cells. (**A**) Representative PBs demonstrating confined (i) and Brownian motion (ii). (**B**) MSD over time of the particles in A. Diffusion coefficients calculated from the plot were D(i) = 0.0007 µm2/s and D(ii) = 0.026 µm2/s. (**C**) Distribution of diffusion coefficients of Dcp1a foci in live HeLa cells (N = 15 cells, 192 particles). The distribution bears strong resemblance to the slowly diffusing miRNA population (**Fig 2F, G**).

z

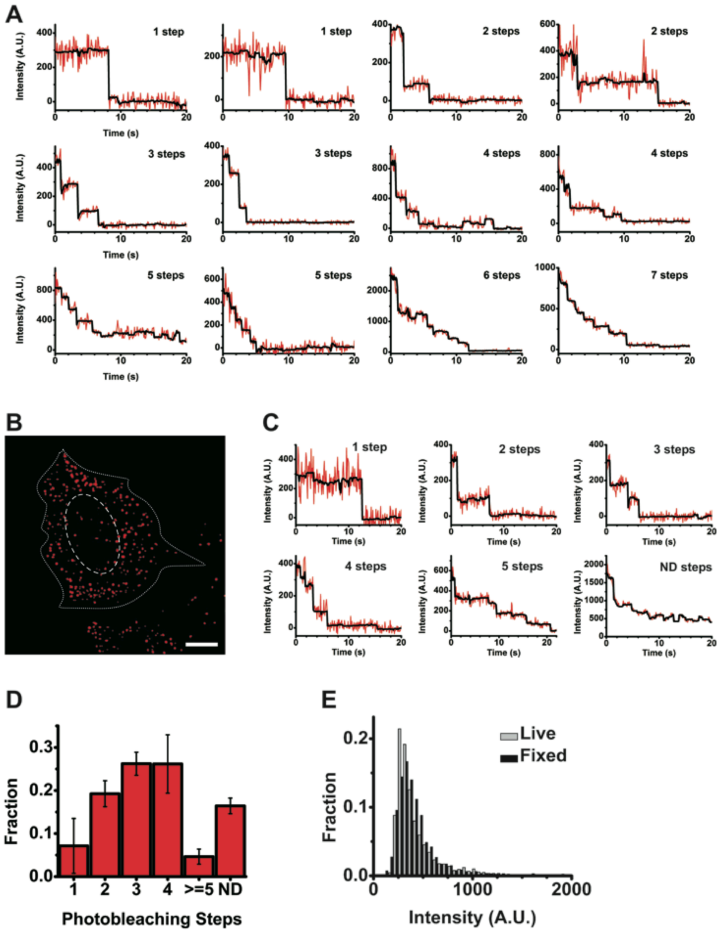


Supplementary Figure B.S.3. Live cell images after incubation for varying amounts of time after microinjection. (**A**) Representative images of cells incubated for 5 min to 32 h after microinjection, as indicated. Dashed and dotted lines indicate nuclear and cellular boundaries, respectively. Scale bar, 10 μm. (**B**) Analysis of background intensity and average number of nine-consecutive-frame localizations. The relative average background intensity (black, solid), which reflects blurred, rapidly diffusing particles, was calculated for each time point and the resultant curve fitted with a single-exponential decay function, yielding a rate constant of 0.82 ± 0.18 h-1. The highest background intensity was found at the earliest time point and continually decreased, consistent with the notion that miRNAs increasingly assemble into large, slowly diffusing, and thus less blurred (at a time resolution of 100 ms) RNPs over this time frame. To quantify assembly further, we determined the average number of particles at each time point that were visible for at least nine consecutive video frames (grey, dashed), a lower bound we used for our diffusion coefficient calculation. The data fit well with a double-exponential function with a rapid-rise phase characterized by a rate constant of 1.18 ± 0.62 h-1,

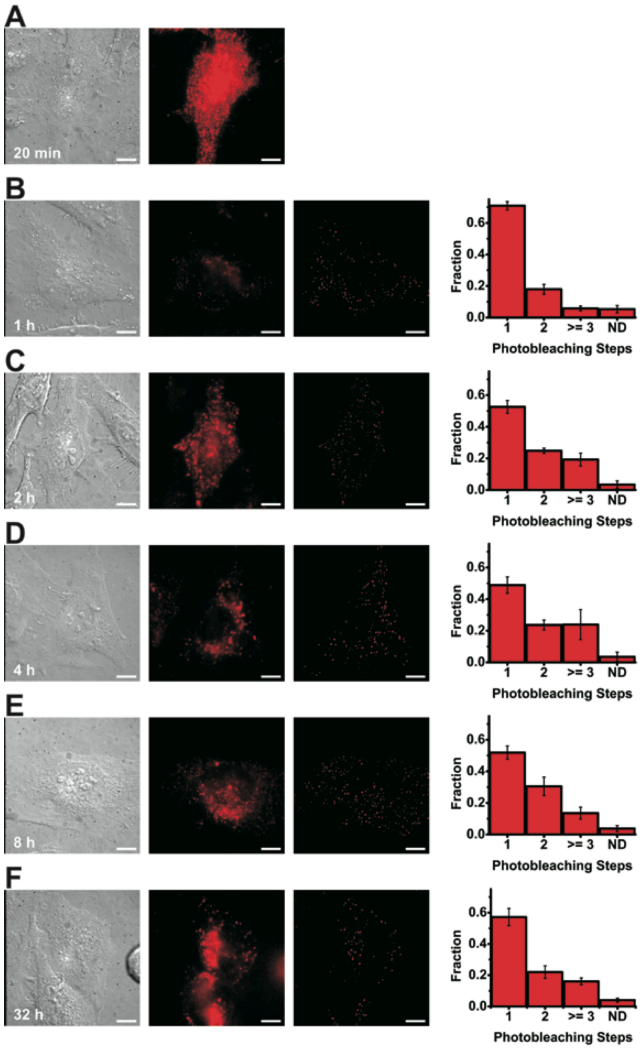
indicating assembly of miRNAs into larger RNPs (especially over the first 2 h of observation), and a slower decrease phase with a rate constant of 0.34 ± 0.13 h-1, indicating an increase in RNP mobility over a longer time frame (2 h – 32 h). These kinetically distinct processes and their rate constants are similar to those obtained in our fixed-cell photobleaching analysis (1.18 ± 0.18 h-1, 0.14 ± 0.08 h-1 and **Fig. 4E**, black lines), further supporting our model that a relatively rapid initial assembly of miRNAs into large RNPs is followed by slow disassembly at later times. Error bars, standard error of the mean (N = 4, 4, 4, 6, 4 cells for 1, 2, 4, 8, 32 h samples respectively).



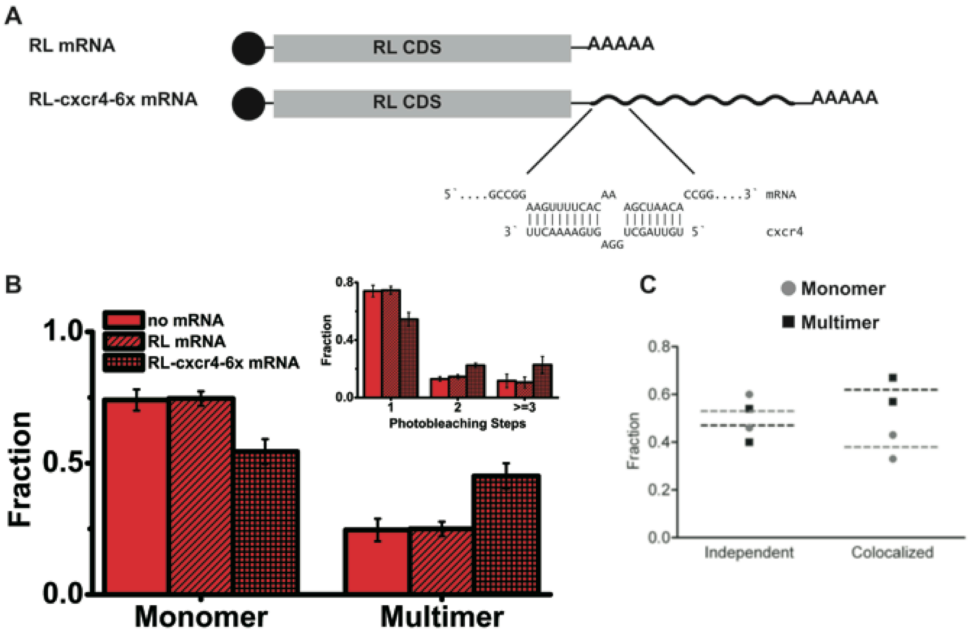
Supplementary Figure B.S.4. Relative deviation (RD) analysis of diffusion coefficients over time. Histograms depict the distribution of RD values at different times after let-7-a1-Cy5 microinjection (2 h, 4 h and 8 h, as indicated). RD values were calculated using MSD values corresponding to the sixth localization point and diffusion coefficients extracted from the first three points of the MSD time course. Although the distributions differ slightly, there is no significant increase in the fraction of higher RD values which, if present, would predict an increase in the fraction of particles undergoing directed motion (N = 4 cells for each, **Supplementary Table 1**).



Supplementary Figure B.S.5. Stepwise photobleaching in fixed cells. (**A**) Additional examples of stepwise photobleaching traces. (**B-D**) Control experiments. (**B**) Pseudo-colored, background subtracted image of cell microinjected with double-stranded DNA probe MS2 labeled with ~3.45 Cy5s per molecule. Outlines of the cell (dotted) and its nucleus (dashed) are also shown. Scale bar, 10 μm. (**C**) Types of photobleaching observed from experiment in **B**. (**D**) Distribution of photobleaching steps for MS2-Cy5 injections (N = 3 cells, 702 particles). Error bars, standard deviation. (**E**) Comparison of fluorescence intensities of Cy5-labeled miRNA-containing particles between living (grey) and fixed cells (black), imaged 2 h after microinjection. Only diffusing particles visible in live cells for 9 frames or longer were included in the analysis. The intensity observed in the first frame of each video (100 ms exposure time) was used to create this histogram (N = 4 cells each).



Supplementary Figure B.S.6. Representative images and resulting stepwise photobleaching distributions from fixed cells incubated for various amounts of time after microinjection. Shown are (from left to right) DIC, pseudo-colored, pseudo-colored/background-subtracted images and photobleaching step distributions of cells injected with let-7-a1-Cy5 and fluorescein dextran. Images and distributions were taken from cells fixed (**A**) 20 min, (**B**) 1 h, (**C**) 2 h, (**D**) 4 h, (**E**) 8 h and (**F**) 32 h after microinjection (nB-F = 5, 4, 3, 4, 8 cells and 1,125, 1,509, 940, 1,168, 688 particles respectively). Error bars, standard deviation. Scale bar, 10 μm. The background was very high in cells incubated for 20 min after microinjection, thus it was practically impossible to locate single molecules.



Supplementary Figure B.S.7. miRNAs assembly is mRNA dependent. (**A**) Schematic of 5' capped (black dot) and 3' poly-A tailed mRNAs. RL mRNA (top) is a negative control that contains only the coding sequence of renilla luciferase. RL-cxcr4-6x mRNA (bottom) is the target that contains the same renilla luciferase coding sequence plus six cxcr4 binding sites as indicated in the magnified view. (**B**) Distribution of monomers and multimers in cells that were microinjected with cxcr4 miRNA and no mRNA, with cxcr4 miRNA and RL mRNA, or with cxcr4 miRNA and RL-cxcr4-6x mRNA. The inset shows the distribution further divided into one, two and three or more photobleaching steps in these samples. Even though a significant increase in the number of multimers was observed upon co-microinjection of cxcr4 miRNA with its target RL-cxcr4-6x mRNA, a specific enrichment of particles bearing six miRNAs was not seen, suggesting sub-stoichiometric binding of miRNAs to target mRNA binding sites. Error bars, standard deviation (N = 4 cells for each of the samples). (**C**) Distribution of monomers and multimers of mutant let-7-a1 miRNA non-colocalized (independent) or colocalized with MS2-GFP tagged target mRNAs (see **Fig 3B, C**). The coordinates of stepwise-photobleached particles in the mut-let-7-Cy5 channel were mapped to the mut-mRNA-EGFP channel to assess the extent of colocalization. N = 2 cells containing a total of 60 independent and 16 colocalized let-7-a1 particles. Data points within each group are explicitly shown (grey circles, monomers; black squares, multimers) in addition to their mean values (grey dotted line, monomers; black dotted line, multimers). The statistical variation within each group was minimal between replicates

References

1. Hooke, R. *Micrographia: or some physiological descriptions of minute bodies made by magnifying glasses : with observations and inquiries thereupon* .
2. Schwann, T. *Microscopic Researches into the Accordance in the Structure and Growth of Animals and Plants* (trans Smith, H.) 1847).
3. Jennings, H. S. The Cell in Relation to its Environment. *Journal of the Maryland Academy of Sciences* **2,** (1931).
4. Ephrussi, B. *Nucleo-cytoplasmic Relations in Micro-Organisms: Their Bearing on Cell Heredity and Differentiation* 1953).
5. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849-857 (1983).
6. Crick, F. Central dogma of molecular biology. *Nature* **227,** 561-563 (1970).
7. Cech, T. R. The RNA Worlds in Context. *Cold Spring Harbor Perspectives in Biology* **4,** (2012).
8. Collins, F. & Galas, D. A new five-year plan for the U.S. Human Genome Project. *Science* **262,** 43-46 (1993).
9. Hahn, M. W. & Wray, G. A. The g-value paradox. *Evol Dev* **4,** 73-75 (2002).
10. International Human Genome Sequencing Consortium Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931-945 (2004).
11. Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799-816 (2007).
12. Riddihough, G. In the forests of RNA dark matter. *Science's STKE* **309,** 1507 (2005).
13. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75,** 843-854 (1993).
14. Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403,** 901-906 (2000).
15. Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408,** 86-89 (2000).
16. Friedman, R. C., Farh, K. K. -H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19,** 92-105 (2009).
17. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215-233 (2009).
18. Mendell, J. T. & Olson, E. N. MicroRNAs in Stress Signaling and Human Disease. *Cell* **148,** 1172-1187 (2012).
19. Broderick, J. A. & Zamore, P. D. MicroRNA therapeutics. *Gene Ther* **18,** 1104-1110 (2011).
20. Ding, X., Weiler, J. & Großhans, H. Regulating the regulators: mechanisms controlling the maturation of microRNAs. *Trends in Biotechnology* **27,** 27-36 (2009).

21. Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* **11,** 228-234 (2009).
22. Saini, H. K., Griffiths-Jones, S. & Enright, A. J. Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 17719-17724 (2007).
23. Cai, X., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna-a Publication of the Rna Society* **10,** 1957-1966 (2004).
24. Yoontae Lee, M. K. J. H. K. -H. Y. S. L. S. H. B. V. N. K. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal* **23,** 4051 (2004).
25. Han, J., Lee, Y., Yeom, K. H., Nam, J. W., *et al.* Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125,** 887-901 (2006).
26. Young-Kook Kim, V. N. K. Processing of intronic microRNAs. *The EMBO Journal* **26,** 775 (2007).
27. Morlando, M., Ballarino, M., Gromak, N., Pagano, F., *et al.* Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* (2008).
28. Ruby, J. G., Jan, C. H. & Bartel, D. P. Intronic microRNA precursors that bypass Drosha processing. *Nature* **448,** 83-86 (2007).
29. Katahira, J. & Yoneda, Y. Nucleocytoplasmic transport of microRNAs and related small RNAs. *Traffic* **12,** 1468-1474 (2011).
30. Macrae, I. J., Zhou, K., Li, F., Repic, A., *et al.* Structural basis for double-stranded RNA processing by Dicer. *Science* **311,** 195-198 (2006).
31. MacRae, I. J., Ma, E., Zhou, M., Robinson, C. V. & Doudna, J. A. In vitro reconstitution of the human RISC-loading complex. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 512-517 (2008).
32. Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., *et al.* TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436,** 740-744 (2005).
33. Haase, A. D., Jaskiewicz, L., Zhang, H., Laine, S., *et al.* TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* **6,** 961-967 (2005).
34. Lau, P. W., Potter, C. S., Carragher, B. & MacRae, I. J. Structure of the human Dicer-TRBP complex by electron microscopy. *Structure* **17,** 1326-1332 (2009).
35. Wang, H. W., Noland, C., Siridechadilok, B., Taylor, D. W., *et al.* Structural insights into RNA processing by the human RISC-loading complex. *Nature Structural & Molecular Biology* (2009).
36. Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115,** 199-208 (2003).
37. Noland, C., Ma, E. & Doudna, J. siRNA Repositioning for Guide Strand Selection by Human Dicer Complexes. *Molecular Cell* **43,** 110 - 121 (2011).
38. Farazi, T. A., Juranek, S. A. & Tuschl, T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* **135,** 1201-1214 (2008).
39. Schirle, N. T. & Macrae, I. J. The Crystal Structure of Human Argonaute2. *Science* (2012).

40. Kawamata, T. & Tomari, Y. Making RISC. *Trends Biochem Sci* **35,** 368-376 (2010).
41. Kwak, P. B. & Tomari, Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat Struct Mol Biol* **19,** 145-151 (2012).
42. Lambert, N. J., Gu, S. G. & Zahler, A. M. The conformation of microRNA seed regions in native microRNPs is prearranged for presentation to mRNA targets. *Nucleic Acids Res* **39,** 4827-4835 (2011).
43. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Molecular cell* **33,** 204-214 (2009).
44. Zipprich, J. T., Bhattacharyya, S. N., Mathys, H. & Filipowicz, W. Importance of the C-terminal domain of the human GW182 protein TNRC6C for translational repression. *Rna-a Publication of the Rna Society* (2009).
45. Chekulaeva, M., Mathys, H., Zipprich, J. T., Attig, J., *et al.* miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. *Nat Struct Mol Biol* **18,** 1218-1226 (2011).
46. Fabian, M. R., Svitkin, Y. V. & Sonenberg, N. An Efficient System for Let-7 MicroRNA and GW182 Protein-Mediated Deadenylation In Vitro. *Methods Mol Biol* **725,** 207-217 (2011).
47. Chu, C. Y. & Rana, T. M. Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biology* **4,** e210 (2006).
48. Meister, G., Landthaler, M., Peters, L., Chen, P. Y., *et al.* Identification of novel argonaute-associated proteins. *Curr Biol* **15,** 2149-2155 (2005).
49. Jannot, G., Bajan, S., Giguère, N. J., Bouasker, S., *et al.* The ribosomal protein RACK1 is required for microRNA function in both C. elegans and humans. *EMBO Rep* **12,** 581-586 (2011).
50. Chendrimada, T. P., Finn, K. J., Ji, X., Baillat, D., *et al.* MicroRNA silencing through RISC recruitment of eIF6. *Nature* **447,** 823-828 (2007).
51. Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., *et al.* mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & Development* **20,** 1885-1898 (2006).
52. Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149,** 515-524 (2012).
53. Baek, D., Villén, J., Shin, C., Camargo, F. D., *et al.* The impact of microRNAs on protein output. *Nature* **455,** 64-71 (2008).
54. Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433,** 769-773 (2005).
55. Anokye-Danso, F., Trivedi, C. M., Juhr, D., Gupta, M., *et al.* Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* **8,** 376-388 (2011).
56. Baccarini, A., Chauhan, H., Gardner, T. J., Jayaprakash, A. D., *et al.* Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. *Curr Biol* **21,** 369-376 (2011).
57. Ameres, S. L., Horwich, M. D., Hung, J. H., Xu, J., *et al.* Target RNA-directed trimming and tailing of small silencing RNAs. *Science* **328,** 1534-1539 (2010).

58. Heo, I., Joo, C., Cho, J., Ha, M., *et al.* Lin28 Mediates the Terminal Uridylation of let-7 Precursor MicroRNA. *Molecular Cell* **32,** 276-284 (2008).
59. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* **11,** 597-610 (2010).
60. Cannell, I. G., Kong, Y. W. & Bushell, M. How do microRNAs regulate gene expression? *Biochemical Society Transactions* **36,** 1224-1231 (2008).
61. Jackson, R. J. & Standart, N. How do microRNAs regulate gene expression? *Sci STKE* **2007,** re1 (2007).
62. Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Review Genetics* **9,** 102-114 (2008).
63. Pillai, R. S., Bhattacharyya, S. N. & Filipowicz, W. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends in Cell Biology* **17,** 118-126 (2007).
64. Béthune, J., Artus-Revel, C. G. & Filipowicz, W. Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* (2012).
65. Djuranovic, S., Nahvi, A. & Green, R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336,** 237-240 (2012).
66. Fabian, M. R. & Sonenberg, N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat Struct Mol Biol* **19,** 586-593 (2012).
67. Mishima, Y., Fukao, A., Kishimoto, T., Sakamoto, H., *et al.* Translational inhibition by deadenylation-independent mechanisms is central to microRNA-mediated silencing in zebrafish. *Proc Natl Acad Sci U S A* **109,** 1104-1109 (2012).
68. Pillai, R. S., Bhattacharyya, S. N., Artus, C. G., Zoller, T., *et al.* Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309,** 1573-1576 (2005).
69. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125,** 1111-1124 (2006).
70. Humphreys, D. T., Westman, B. J., Martin, D. I. & Preiss, T. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 16961-16966 (2005).
71. Wang, B., Love, T. M., Call, M. E., Doench, J. G. & Novina, C. D. Recapitulation of short RNA-directed translational gene silencing in vitro. *Mol Cell* **22,** 553-560 (2006).
72. Kiriakidou, M., Tan, G. S., Lamprinaki, S., De Planell-Saguer, M., *et al.* An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell* **129,** 1141-1151 (2007).
73. Frank, F., Fabian, M. R., Stepinski, J., Jemielity, J., *et al.* Structural analysis of 5'-mRNA-cap interactions with the human AGO2 MID domain. *EMBO Rep* **12,** 415-420 (2011).
74. Wang, B., Yanez, A. & Novina, C. D. MicroRNA-repressed mRNAs contain 40S but not 60S components. *Proceedings of the National Academy of Sciences* **105,** 5343-5348 (2008).

75. Eulalio, A., Huntzinger, E. & Izaurralde, E. GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay. *Nat Struct Mol Biol* **15,** 346-353 (2008).
76. Petersen, C. P., Bordeleau, M. E., Pelletier, J. & Sharp, P. A. Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* **21,** 533-542 (2006).
77. Nelson, P. T., Hatzigeorgiou, A. G. & Mourelatos, Z. miRNP:mRNA association in polyribosomes in a human neuronal cell line. *RNA* **10,** 387 (2004).
78. Maroney, P. A., Yu, Y., Fisher, J., Nilsen, T. W., *et al.* Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat Struct Mol Biol* **13,** 1102 (2006).
79. Nottrott, S., Simard, M. J. & Richter, J. D. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol* **13,** 1108-1114 (2006).
80. Olsen, P. H. & Ambros, V. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental biology* **216,** 671-680 (1999).
81. Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466,** 835-840 (2010).
82. Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27,** 91-105 (2007).
83. John, B., Enright, A. J., Aravin, A., Tuschl, T., *et al.* Human MicroRNA targets. *PLoS Biol* **2,** e363 (2004).
84. Krek, A., Grün, D., Poy, M. N., Wolf, R., *et al.* Combinatorial microRNA target predictions. *Nature Genetics* **37,** 495-500 (2005).
85. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat Genet* **39,** 1278-1284 (2007).
86. E, K. K. W. K. Practical aspects of microRNA target prediction. *Current molecular medicine* **11,** 93 (2011).
87. Vergoulis, T., Vlachos, I. S., Alexiou, P., Georgakilas, G., *et al.* TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res* **40,** D222-D229 (2012).
88. Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., *et al.* Widespread changes in protein synthesis induced by microRNAs. *Science* **454,** (2008).
89. Fabian, M. R., Mathonnet, G., Sundermeier, T., Mathys, H., *et al.* Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell* **35,** 868-880 (2009).
90. Kloosterman, W., Lagendijk, A., Ketting, R., Moulton, J. & Plasterk, R. H. Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development. *PLoS Biology* **5,** e203 (2007).
91. Choi, W. Y., Giraldez, A. J. & Schier, A. F. Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. *Science* **318,** 271-274 (2007).
92. Ebert, M. S., Neilson, J. R. & Sharp, P. A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* **4,** 721-726 (2007).

93. Iba, H. Vectors expressing efficient RNA decoys achieve the long-term suppression of specific microRNA activity in mammalian cells. *Nucleic Acids Research* **37,** e43-e43 (2009).
94. Haraguchi, T., Nakano, H., Tagawa, T., Ohki, T., *et al.* A potent 2'-O-methylated RNA-based microRNA inhibitor with unique secondary structures. *Nucleic Acids Res* (2012).
95. Stenvang, J., Petri, A., Lindow, M., Obad, S. & Kauppinen, S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* **3,** 1 (2012).
96. Davis, S., Propp, S., Freier, S. M., Jones, L. E., *et al.* Potent inhibition of microRNA in vivo without degradation. *Nucleic Acids Research* **37,** 70-77 (2009).
97. Davis, S., Lollo, B., Freier, S. & Esau, C. Improved targeting of miRNA with antisense oligonucleotides. *Nucleic Acids Res* **34,** 2294-2304 (2006).
98. Lennox, K. A. & Behlke, M. A. A direct comparison of anti-microRNA oligonucleotide potency. *Pharm Res* **27,** 1788-1799 (2010).
99. Elmén, J., Lindow, M., Schütz, S., Lawrence, M., *et al.* LNA-mediated microRNA silencing in non-human primates. *Nature* **452,** 896-899 (2008).
100. Muñoz-Alarcón, A., Guterstam, P., Romero, C., Behlke, M. A., *et al.* Modulating Anti-MicroRNA-21 Activity and Specificity Using Oligonucleotide Derivatives and Length Optimization. *ISRN Pharm* **2012,** 407154 (2012).
101. Krützfeldt, J., Kuwajima, S., Braich, R., Rajeev, K. G., *et al.* Specificity, duplex degradation and subcellular localization of antagomirs. *Nucleic Acids Research* **35,** 2885-2892 (2007).
102. Yang, Y., Chaerkady, R., Beer, M. A., Mendell, J. T. & Pandey, A. Identification of miR-21 targets in breast cancer cells using a quantitative proteomic approach. *Proteomics* **9,** 1374-1384 (2009).
103. Frankel, L. B., Christoffersen, N. R., Jacobsen, A., Lindow, M., *et al.* Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J Biol Chem* **283,** 1026-1033 (2008).
104. Obad, S., dos Santos, C. O., Petri, A., Heidenblad, M., *et al.* Silencing of microRNA families by seed-targeting tiny LNAs. *Nat Genet* **43,** 371-378 (2011).
105. Frankel, L. B., Wen, J., Lees, M., Høyer-Hansen, M., *et al.* microRNA-101 is a potent inhibitor of autophagy. *The EMBO Journal* **30,** 4628-4641 (2011).
106. Krützfeldt, J., Rajewsky, N., Braich, R., Rajeev, K., *et al.* Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438,** 685-689 (2005).
107. Esau, C., Davis, S., Murray, S. F., Yu, X. X., *et al.* miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab* **3,** 87-98 (2006).
108. Elmén, J., Lindow, M., Silahtaroglu, A., Bak, M., *et al.* Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. *Nucleic Acids Res* **36,** 1153-1162 (2008).
109. Huynh, C., Segura, M., Gaziel-Sovran, A., Menendez, S., *et al.* Efficient in vivo microRNA targeting of liver metastasis. *Oncogene* (2010).
110. Rodriguez, A., Vigorito, E., Clare, S., Warren, M. V., *et al.* Requirement of bic/microRNA-155 for normal immune function. *Science* **316,** 608-611 (2007).
111. Park, C. Y., Choi, Y. S. & McManus, M. T. Analysis of microRNA knockouts in mice. *Hum Mol Genet* **19,** R169-R175 (2010).

112. Patrick, D. M., Montgomery, R. L., Qi, X., Obad, S., *et al.* Stress-dependent cardiac remodeling occurs in the absence of microRNA-21 in mice. *J Clin Invest* **120,** 3912-3916 (2010).
113. Ruepp, A., Kowarsch, A., Schmidl, D., Buggenthin, F., *et al.* PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol* **11,** R6 (2010).
114. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nature Reviews Cancer* **6,** 857-866 (2006).
115. Henke, J. I., Goergen, D., Zheng, J., Song, Y., *et al.* microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO J* **27,** 3300-3310 (2008).
116. Jopling, C. L., Yi, M., Lancaster, A. M., Lemon, S. M. & Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309,** 1577-1581 (2005).
117. Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., *et al.* Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet* **41,** 609-613 (2009).
118. Janssen, H. L., Reesink, H. W. & Zeuzem, S. A randomized, double-blind, placebo controlled safety and anti-viral proof of concept study of miravirsen, an oligonucleotide targeting miR-122, in treatment naive patients with genotype 1 chronic HCV infection. *Hepatology* **54,** (2011).
119. Trang, P., Wiggins, J. F., Daige, C. L., Cho, C., *et al.* Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice. *Mol Ther* **19,** 1116-1122 (2011).
120. Trang, P., Medina, P. P., Wiggins, J. F., Ruffino, L., *et al.* Regression of murine lung tumors by the let-7 microRNA. *Oncogene* **29,** 1580-1587 (2010).
121. Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435,** 834-838 (2005).
122. Volinia, S., Calin, G. A., Liu, C. -G., Ambs, S., *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. USA* **103,** 2257-2261 (2006).
123. Chen, Y., Liu, W., Chao, T., Zhang, Y., *et al.* MicroRNA-21 down-regulates the expression of tumor suppressor PDCD4 in human glioblastoma cell T98G. *Cancer letters* **272,** 197-205 (2008).
124. Selaru, F. M., Olaru, A. V., Kan, T., David, S., *et al.* MicroRNA-21 is overexpressed in human cholangiocarcinoma and regulates programmed cell death 4 and tissue inhibitor of metalloproteinase 3. *Hepatology (Baltimore, Md)* **49,** 1595-1601 (2009).
125. Hiyoshi, Y., Kamohara, H., Karashima, R., Sato, N., *et al.* MicroRNA-21 regulates the proliferation and invasion in esophageal squamous cell carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **15,** 1915-1922 (2009).
126. Liu, C., Yu, J., Yu, S., Lavker, R. M., *et al.* MicroRNA-21 acts as an oncomir through multiple targets in human hepatocellular carcinoma. *J. Hepatol.* **53,** 98-107 (2010).

127. Yao, Q., Xu, H., Zhang, Q. Q., Zhou, H. & Qu, L. H. MicroRNA-21 promotes cell proliferation and down-regulates the expression of programmed cell death 4 (PDCD4) in HeLa cervical carcinoma cells. *Biochem Biophys Res Commun* **388,** 539-542 (2009).
128. Schetter, A. J., Leung, S. Y., Sohn, J. J., Zanetti, K. A., *et al.* MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *JAMA : the journal of the American Medical Association* **299,** 425-436 (2008).
129. Markou, A., Tsaroucha, E. G., Kaklamanis, L., Fotinou, M., *et al.* Prognostic value of mature microRNA-21 and microRNA-205 overexpression in non-small cell lung cancer by quantitative real-time RT-PCR. *Clinical Chemistry* **54,** 1696-1704 (2008).
130. Dillhoff, M., Liu, J., Frankel, W., Croce, C. & Bloomston, M. MicroRNA-21 is overexpressed in pancreatic cancer and a potential predictor of survival. *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract* **12,** 2171-2176 (2008).
131. Qian, B., Katsaros, D., Lu, L., Preti, M., *et al.* High miR-21 expression in breast cancer associated with poor disease-free survival in early stage disease and high TGF-beta1. *Breast cancer research and treatment* **117,** 131-140 (2009).
132. Yan, L. X., Huang, X. F., Shao, Q., Huang, M. Y., *et al.* MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA (New York, NY)* **14,** 2348-2360 (2008).
133. Zhu, S., Wu, H., Wu, F., Nie, D., *et al.* MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Research* **18,** 350-359 (2008).
134. Asangani, I. A., Rasheed, S. A., Nikolova, D. A., Leupold, J. H., *et al.* MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* **27,** 2128-2136 (2008).
135. Chan, J. A., Krichevsky, A. M. & Kosik, K. S. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer research* **65,** 6029-6033 (2005).
136. Jazbutyte, V. & Thum, T. MicroRNA-21: from cancer to cardiovascular disease. *Current drug targets* **11,** 926-935 (2010).
137. Fujita, S., Ito, T., Mizutani, T., Minoguchi, S., *et al.* miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. *J Mol Biol* **378,** 492-504 (2008).
138. Hatley, M. E., Patrick, D. M., Garcia, M. R., Richardson, J. A., *et al.* Modulation of K-Ras-dependent lung tumorigenesis by MicroRNA-21. *Cancer Cell* **18,** 282-293 (2010).
139. Cheng, Y., Liu, X., Zhang, S., Lin, Y., *et al.* MicroRNA-21 protects against the H(2)O(2)-induced injury on cardiac myocytes via its target gene PDCD4. *Journal of molecular and cellular cardiology* **47,** 5-14 (2009).
140. Wickramasinghe, N. S., Manavalan, T. T., Dougherty, S. M., Riggs, K. A., *et al.* Estradiol downregulates miR-21 expression and increases miR-21 target gene expression in MCF-7 breast cancer cells. *Nucleic Acids Res* **37,** 2584-2595 (2009).
141. Davis, B. N., Hilyard, A. C., Lagna, G. & Hata, A. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454,** 56-61 (2008).
142. Du, J., Yang, S., An, D., Hu, F., *et al.* BMP-6 inhibits microRNA-21 expression in breast cancer through repressing deltaEF1 and AP-1. *Cell Research* **19,** 487-496 (2009).

143. Frezzetti, D., De Menna, M., Zoppoli, P., Guerra, C., *et al.* Upregulation of miR-21 by Ras in vivo and its role in tumor growth. *Oncogene* **30,** 275-286 (2011).
144. Medina, P. P., Nolde, M. & Slack, F. J. OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. *Nature* **467,** 86-90 (2010).
145. Zheng, D., Chen, C. Y. & Shyu, A. B. Unraveling regulation and new components of human P-bodies through a protein interaction framework and experimental validation. *RNA* (2011).
146. Bashkirov, V. I., Scherthan, H., Solinger, J. A., Buerstedde, J. M. & Heyer, W. D. A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J Cell Biol* **136,** 761-773 (1997).
147. van Dijk, E., Cougot, N., Meyer, S., Babajko, S., *et al.* Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J* **21,** 6915-6924 (2002).
148. Ingelfinger, D., Arndt-Jovin, D. J., Lührmann, R. & Achsel, T. The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. *Rna-a Publication of the Rna Society* **8,** 1489-1501 (2002).
149. Sheth, U. & Parker, R. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300,** 805-808 (2003).
150. Cougot, N., Babajko, S. & Séraphin, B. Cytoplasmic foci are sites of mRNA decay in human cells. *J Cell Biol* **165,** 31-40 (2004).
151. Jakymiw, A., Lian, S., Eystathioy, T., Li, S., *et al.* Disruption of GW bodies impairs mammalian RNA interference. *Nat Cell Biol* **7,** 1267-1274 (2005).
152. Liu, J., Valencia-Sanchez, M. A., Hannon, G. J. & Parker, R. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* **7,** 719-723 (2005).
153. Leung, A. K., Calabrese, J. M. & Sharp, P. A. Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 18125-18130 (2006).
154. Rehwinkel, J., Behm-Ansmant, I., Gatfield, D. & Izaurralde, E. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *Rna-a Publication of the Rna Society* **11,** 1640-1647 (2005).
155. Liu, J., Rivas, F. V., Wohlschlegel, J., Yates, J. R., *et al.* A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* **7,** 1261-1266 (2005).
156. Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., *et al.* Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J Cell Biol* **169,** 871-884 (2005).
157. Eulalio, A., Behm-Ansmant, I., Schweizer, D. & Izaurralde, E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol* **27,** 3970-3981 (2007).
158. Pauley, K. M., Eystathioy, T., Jakymiw, A., Hamel, J. C., *et al.* Formation of GW bodies is a consequence of microRNA genesis. *EMBO Rep* **7,** 904-910 (2006).
159. Brengues, M., Teixeira, D. & Parker, R. Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science* **310,** 486-489 (2005).
160. Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Stress-induced reversal of microRNA repression and mRNA P-body localization in human cells. *Cold Spring Harb Symp Quant Biol* **71,** 513-521 (2006).

161. Cougot, N., Cavalier, A., Thomas, D. & Gillet, R. The Dual Organization of P-bodies Revealed by Immunoelectron Microscopy and Electron Tomography. *J Mol Biol* **420,** 17-28 (2012).
162. Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., *et al.* Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr Biol* **13,** 161-167 (2003).
163. Shav-Tal, Y., Darzacq, X., Shenoy, S. M., Fusco, D., *et al.* Dynamics of single mRNPs in nuclei of living cells. *Science* **304,** 1797-1800 (2004).
164. Grünwald, D. & Singer, R. H. In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. *Nature* **467,** 604-607 (2010).
165. Aizer, A., Brody, Y., Ler, L. W., Sonenberg, N., *et al.* The dynamics of mammalian P body transport, assembly, and disassembly in vivo. *Molecular Biology of the Cell* **19,** 4154-4166 (2008).
166. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115,** 787-798 (2003).
167. Mallanna, S. K. & Rizzino, A. Emerging roles of microRNAs in the control of embryonic stem cells and the generation of induced pluripotent stem cells. *Dev. Biol.* **344,** 16-25 (2010).
168. Vecchione, A. & Croce, C. M. Apoptomirs: small molecules have gained the license to kill. *Endocr. Rel. Cancer* **17,** F37-F50 (2010).
169. Zhao, Y. & Srivastava, D. A developmental view of microRNA function. *Trends Biochem. Sci.* **32,** 189-197 (2007).
170. Hendrickson, D. G., Hogan, D. J., McCullough, H. L., Myers, J. W., *et al.* Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biology* **7,** e1000238 (2009).
171. Kim, J., Krichevsky, A., Grad, Y., Hayes, G. D., *et al.* Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc Natl Acad Sci U S A* **101,** 360-365 (2004).
172. Lu, M., Zhang, Q., Deng, M., Miao, J., *et al.* An analysis of human microRNA and disease associations. *PloS one* **3,** e3420 (2008).
173. Mayr, C., Hemann, M. T. & Bartel, D. P. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* **315,** 1576-1579 (2007).
174. Thum, T., Gross, C., Fiedler, J., Fischer, T., *et al.* MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts. *Nature* **456,** 980-984 (2008).
175. Lu, T. X., Hartner, J., Lim, E. J., Fabry, V., *et al.* MicroRNA-21 limits in vivo immune response-mediated activation of the IL-12/IFN-gamma pathway, Th1 polarization, and the severity of delayed-type hypersensitivity. *J Immunol* **187,** 3362-3373 (2011).
176. Thomas, P. D., Kejariwal, A., Campbell, M. J., Mi, H., *et al.* PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research* **31,** 334-341 (2003).
177. Thomas, P. D., Kejariwal, A., Guo, N., Mi, H., *et al.* Applications for protein sequence–function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research* **34,** W645-W650 (2006).

178. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res* **31,** 3429-3431 (2003).
179. Kruger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* **34,** W451-W454 (2006).
180. Jannot, G., Vasquez-Rifo, A. & Simard, M. J. Argonaute pull-down and RISC analysis using 2'-O-methylated oligonucleotides affinity matrices. *Methods Mol Biol* **725,** 233-249 (2011).
181. Pall, G. S. & Hamilton, A. J. Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3,** 1077-1084 (2008).
182. van Dongen, S., Abreu-Goodger, C. & Enright, A. J. Detecting microRNA binding and siRNA off-target effects from expression data. *Nature Methods* **5,** 1023-1025 (2008).
183. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34,** D140-D144 (2006).
184. Cheloufi, S., Dos Santos, C. O., Chong, M. M. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465,** 584-589 (2010).
185. Burchard, J., Zhang, C., Liu, A. M., Poon, R. T., *et al.* microRNA-122 as a regulator of mitochondrial metabolic gene network in hepatocellular carcinoma. *Mol Syst Biol* **6,** 402 (2010).
186. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30,** 207-210 (2002).
187. Bissels, U., Wild, S., Tomiuk, S., Holste, A., *et al.* Absolute quantification of microRNAs by using a universal reference. *RNA* **15,** 2375-2384 (2009).
188. Filipowicz, W. & Grosshans, H. The liver-specific microRNA miR-122: biology and therapeutic potential. *Prog Drug Res* **67,** 221-238 (2011).
189. Burns, D. M., D'Ambrogio, A., Nottrott, S., Richter, J. D., *et al.* CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation. *Nature* **473,** 105 (2011).
190. Roush, S. & Slack, F. J. The let-7 family of microRNAs. *Trends in Cell Biology* **18,** 505-516 (2008).
191. Garcia, D. M., Baek, D., Shin, C., Bell, G. W., *et al.* Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol* (2011).
192. Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* **11,** R90 (2010).
193. Xiao, F., Zuo, Z., Cai, G., Kang, S., *et al.* miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research* **37,** D105-D110 (2009).
194. Akerfelt, M., Morimoto, R. I. & Sistonen, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nat Rev Mol Cell Biol* **11,** 545-555 (2010).
195. Yuan, C. X. & Gurley, W. B. Potential targets for HSF1 within the preinitiation complex. *Cell Stress Chaperones* **5,** 229-242 (2000).

196. Pham, J. W., Pellino, J. L., Lee, Y. S., Carthew, R. W. & Sontheimer, E. J. A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell* **117,** 83-94 (2004).
197. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nature protocols* **3,** 1101-1108 (2008).
198. Reynolds, A., Anderson, E. M., Vermeulen, A., Fedorov, Y., *et al.* Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* **12,** 988-993 (2006).
199. Manche, L., Green, S. R., Schmedt, C. & Mathews, M. B. Interactions between double-stranded RNA regulators and the protein kinase DAI. *Mol. Cell. Biol.* **12,** 5238 (1992).
200. Liu, J. Jidong Liu: probing P-bodies for the secrets of silencing. Interview by Ruth Williams. *The Journal of cell biology* **178,** 894-895 (2007).
201. Meng, F., Henson, R., Wehbe-Janek, H., Ghoshal, K., *et al.* MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology* **133,** 647-658 (2007).
202. Majlessi, M., Nelson, N. C. & Becker, M. M. Advantages of 2′-O-methyl oligoribonucleotide probes for detecting RNA targets. *Nucleic Acids Research* **26,** 2224-2229 (1998).
203. Lui, W. O., Pourmand, N., Patterson, B. K. & Fire, A. Patterns of known and novel small RNAs in human cervical cancer. *Cancer Res* **67,** 6031-6043 (2007).
204. Tay, Y., Kats, L., Salmena, L., Weiss, D., *et al.* Coding-Independent Regulation of the Tumor Suppressor PTEN by Competing Endogenous mRNAs. *Cell* **147,** 344-357 (2011).
205. Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147,** 358-369 (2011).
206. Sumazin, P., Yang, X., Chiu, H. S., Chung, W. J., *et al.* An Extensive MicroRNA-Mediated Network of RNA-RNA Interactions Regulates Established Oncogenic Pathways in Glioblastoma. *Cell* **147,** 370-381 (2011).
207. Poliseno, L., Salmena, L., Zhang, J., Carver, B., *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465,** 1033-1038 (2010).
208. Pink, R. C., Wicks, K., Caley, D. P., Punch, E. K., *et al.* Pseudogenes: pseudo-functional or key regulators in health and disease? *RNA* **17,** 792-798 (2011).
209. Harper, L. V., Hilton, A. C. & Jones, A. F. RT-PCR for the pseudogene-free amplification of the glyceraldehyde-3-phosphate dehydrogenase gene (gapd). *Mol Cell Probes* **17,** 261-265 (2003).
210. Munz, C., Psichari, E., Mandilis, D., Lavigne, A. C., *et al.* TAF7 (TAFII55) plays a role in the transcription activation by c-Jun. *J Biol Chem* **278,** 21510-21516 (2003).
211. Mehan, S., Meena, H., Sharma, D. & Sankhla, R. JNK: a stress-activated protein kinase therapeutic strategies and involvement in Alzheimer's and various neurodegenerative abnormalities. *J Mol Neurosci* **43,** 376-390 (2011).
212. Turner, N. A. Therapeutic regulation of cardiac fibroblast function: targeting stress-activated protein kinase pathways. *Future Cardiol* **7,** 673-691 (2011).

213. Sampson, J., Mathews, M. B., Osborn, M. & Borghetti, A. F. Hemoglobin messenger ribonucleic acid translation in cell-free systems from rat and mouse liver and Landschutz ascites cells. *Biochemistry* **11,** 3636-3640 (1972).
214. Shav-Tal, Y. & Singer, R. H. RNA localization. *Journal of cell science* **118,** 4077-4081 (2005).
215. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79,** 351-379 (2010).
216. Alexiou, P., Maragkakis, M., Papadopoulos, G. L., Reczko, M. & Hatzigeorgiou, A. G. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* **25,** 3049-3055 (2009).
217. Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318,** 1931-1934 (2007).
218. Jacobsen, A., Wen, J., Marks, D. S. & Krogh, A. Signatures of RNA binding proteins globally coupled to effective microRNA target sites. *Genome Res* **20,** 1010-1019 (2010).
219. G. A. Held, G. G. Y. T. Relationship between gene expression and observed intensities in DNA microarrays—a modeling study. *Nucleic Acids Research* **34,** e70 (2006).
220. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev* **17,** 438-442 (2003).
221. Long, D., Lee, R., Williams, P., Chan, C. Y., *et al.* Potent effect of target structure on microRNA function. *Nat Struct Mol Biol* **14,** 287-294 (2007).
222. Hammell, M., Long, D., Zhang, L., Lee, A., *et al.* mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat Methods* **5,** 813-819 (2008).
223. Hausser, J., Landthaler, M., Jaskiewicz, L., Gaidatzis, D. & Zavolan, M. Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res* **19,** 2009-2020 (2009).
224. Ameres, S. L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* **130,** 101-112 (2007).
225. Kang, J. G., Majerciak, V., Uldrick, T. S., Wang, X., *et al.* Kaposi's sarcoma-associated herpesviral IL-6 and human IL-6 open reading frames contain miRNA binding sites and are subject to cellular miRNA regulation. *J Pathol* **225,** 378-389 (2011).
226. Moretti, F., Thermann, R. & Hentze, M. W. Mechanism of translational regulation by miR-2 from sites in the 5' untranslated region or the open reading frame. *RNA* **16,** 2493-2502 (2010).
227. Chi, S. W., Zang, J. B., Mele, A. & Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460,** 479-486 (2009).
228. Gu, S., Jin, L., Zhang, F., Sarnow, P. & Kay, M. A. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nature Structural & Molecular Biology* **16,** 144-150 (2009).
229. Shin, C., Nam, J. W., Farh, K. K., Chiang, H. R., *et al.* Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* **38,** 789-802 (2010).
230. Chi, S. W., Hannon, G. J. & Darnell, R. B. An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* **19,** 321-327 (2012).

231. Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., *et al.* The Pfam protein families database. *Nucleic Acids Res* **36,** D281-D288 (2008).
232. Kedde, M., Strasser, M. J., Boldajipour, B., Oude Vrielink, J. A., *et al.* RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131,** 1273-1286 (2007).
233. Galgano, A., Forrer, M., Jaskiewicz, L., Kanitz, A., *et al.* Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS One* **3,** e3164 (2008).
234. Meisner, N. C. & Filipowicz, W. Properties of the Regulatory RNA-Binding Protein HuR and its Role in Controlling miRNA Repression. *Adv Exp Med Biol* **700,** 106-123 (2011).
235. Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., *et al.* Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. *Mol Cell* **43,** 340-352 (2011).
236. Kim, H. H., Kuwano, Y., Srikantan, S., Lee, E. K., *et al.* HuR recruits let-7/RISC to repress c-Myc expression. *Genes Dev* **23,** 1743-1748 (2009).
237. Friend, K., Campbell, Z. T., Cooke, A., Kroll-Conner, P., *et al.* A conserved PUF-Ago-eEF1A complex attenuates translation elongation. *Nat Struct Mol Biol* **19,** 176-183 (2012).
238. Androsavich, J. R., Chau, B. N., Bhat, B., Linsley, P. S. & Walter, N. G. Disease-linked microRNA-21 exhibits drastically reduced mRNA binding and silencing activity in healthy mouse liver. *RNA* (2012).
239. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336,** 233-237 (2012).
240. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell* **146,** 353-358 (2011).
241. Wilson, G. M., Sutphen, K., Moutafis, M., Sinha, S. & Brewer, G. Structural remodeling of an A + U-rich RNA element by cation or AUF1 binding. *J Biol Chem* **276,** 38400-38409 (2001).
242. Kundu, P., Fabian, M. R., Sonenberg, N., Bhattacharyya, S. N. & Filipowicz, W. HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Res* **40,** 5088-5100 (2012).
243. Cisse, I. I., Kim, H. & Ha, T. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat Struct Mol Biol* **19,** 623-627 (2012).
244. Abdelmohsen, K. & Gorospe, M. Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdiscip Rev RNA* **1,** 214-229 (2010).
245. Kedde, M., van Kouwenhove, M., Zwart, W., Oude Vrielink, J. A., *et al.* A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat Cell Biol* **12,** 1014-1020 (2010).
246. Djuranovic, S., Nahvi, A. & Green, R. A parsimonious model for gene regulation by miRNAs. *Science* **331,** 550 (2011).
247. Pitchiaya, S., Androsavich, J. R. & Walter, N. G. Intracellular single molecule microscopy reveals two kinetically distinct pathways for microRNA assembly. *EMBO Rep* (2012).

248. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature Methods* **5,** 159-161 (2008).
249. Walter, N. G. Probing RNA structural dynamics and function by fluorescence resonance energy transfer (FRET). *Curr Protoc Nucleic Acid Chem* **Chapter 11,** Unit 11.10 (2003).
250. Querido, E. & Chartrand, P. Using fluorescent proteins to study mRNA trafficking in living cells. *Methods Cell Biol* **85,** 273-292 (2008).
251. Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods* **5,** 695-702 (2008).
252. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy* **224,** 213-232 (2006).
253. Cao, H., Xiong, Y., Wang, T., Chen, B., *et al.* A red cy3-based biarsenical fluorescent probe targeted to a complementary binding peptide. *Journal of the American Chemical Society* **129,** 8672-8673 (2007).
254. Yoda, M., Kawamata, T., Paroo, Z., Ye, X., *et al.* ATP-dependent human RISC assembly pathways. *Nature Structural & Molecular Biology* (2009).
255. Chen, C. Y., Zheng, D., Xia, Z. & Shyu, A. B. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. *Nat Struct Mol Biol* **16,** 1160-1166 (2009).
256. Arroyo, J. D., Chevillet, J. R., Kroh, E. M., Ruf, I. K., *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A* **108,** 5003-5008 (2011).
257. Sen, G. L. & Blau, H. M. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat Cell Biol* **7,** 633-636 (2005).
258. Zeitelhofer, M., Karra, D., Macchi, P., Tolino, M., *et al.* Dynamic interaction between P-bodies and transport ribonucleoprotein particles in dendrites of mature hippocampal neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28,** 7555-7562 (2008).
259. Nathans, R., Chu, C. Y., Serquina, A., Lu, C., *et al.* Cellular microRNA and P bodies modulate host-HIV-1 interactions. *Molecular Cell* **34,** 696-709 (2009).
260. Meister, G., Landthaler, M., Peters, L., Chen, P. Y., *et al.* Identification of Novel Argonaute-Associated Proteins. *Current Biology* **15,** 2149 - 2155 (2005).
261. Jagannath, A. & Wood, M. J. Localization of double-stranded small interfering RNA to cytoplasmic processing bodies is Ago2 dependent and results in up-regulation of GW182 and Argonaute-2. *Mol Biol Cell* **20,** 521-529 (2009).
262. Rüdel, S., Wang, Y., Lenobel, R., Körner, R., *et al.* Phosphorylation of human Argonaute proteins affects small RNA binding. *Nucleic Acids Res* **39,** 2330-2343 (2011).
263. Skube, S. B., Chaverri, J. M. & Goodson, H. V. Effect of GFP tags on the localization of EB1 and EB1 fragments in vivo. *Cytoskeleton (Hoboken)* **67,** 1-12 (2010).
264. Tsien, R. Y. The green fluorescent protein. *Annual review of biochemistry* **67,** 509-544 (1998).
265. Rodriguez, A., Shenoy, S. M., Singer, R. H. & Condeelis, J. Visualization of mRNA translation in living cells. *Journal of Cell Biology* **175,** 67-76 (2006).
266. Griffin, B. A., Adams, S. R. & Tsien, R. Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **281,** 269-272 (1998).

267. Whittaker, V. P. An experimental investigation of the "ring hypothesis" of arsenical toxicity. *Biochem J* **41,** 56-62 (1947).
268. Adams, S. R., Campbell, R. E., Gross, L. A., Martin, B. R., *et al.* New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *Journal of the American Chemical Society* **124,** 6063-6076 (2002).
269. Madani, F., Lind, J., Damberg, P., Adams, S. R., *et al.* Hairpin structure of a biarsenical-tetracysteine motif determined by NMR spectroscopy. *Journal of the American Chemical Society* **131,** 4613-4615 (2009).
270. Rudner, L., Nydegger, S., Coren, L. V., Nagashima, K., *et al.* Dynamic fluorescent imaging of human immunodeficiency virus type 1 Gag in live cells by biarsenical labeling. *Journal of Virology* **79,** 4055-4065 (2005).
271. Ramdzan, Y. M., Nisbet, R. M., Miller, J., Finkbeiner, S., *et al.* Conformation sensors that distinguish monomeric proteins from oligomers in live cells. *Chem Biol* **17,** 371-379 (2010).
272. Park, H., Hanson, G. T., Duff, S. R. & Selvin, P. R. Nanometre localization of single ReAsH molecules. *Journal of Microscopy* **216,** 199-205 (2004).
273. Martin, B. R., Giepmans, B. N., Adams, S. R. & Tsien, R. Y. Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nature Biotechnology* **23,** 1308-1314 (2005).
274. Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J. & Waggoner, A. S. CYANINE DYE LABELING REAGENTS - SULFOINDOCYANINE SUCCINIMIDYL ESTERS. *Bioconjugate Chemistry* **4,** 105-111 (1993).
275. Walter, N. G., Huang, C. Y., Manzo, A. J. & Sobhy, M. A. Do-it-yourself guide: how to use the modern single-molecule toolkit. *Nature Methods* **5,** 475-489 (2008).
276. Wang, X., Li, Y., Xu, X. & Wang, Y. H. Toward a system-level understanding of microRNA pathway via mathematical modeling. *Biosystems* **100,** 31-38 (2010).
277. Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., *et al.* A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 7877-7882 (2002).
278. Li, S., Lian, S. L., Moser, J. J., Fritzler, M. L., *et al.* Identification of GW182 and its novel isoform TNGW1 as translational repressors in Ago2-mediated silencing. *Journal of cell science* **121,** 4134-4144 (2008).
279. Fujimura, K., Katahira, J., Kano, F., Yoneda, Y. & Murata, M. Selective Localization of PCBP2 to Cytoplasmic Processing Bodies. *Biochimica et biophysica acta* (2009).
280. Kedersha, N., Tisdale, S., Hickman, T. & Anderson, P. Real-time and quantitative imaging of mammalian stress granules and processing bodies. *Methods Enzymol* **448,** 521-552 (2008).
281. Cheezum, M. K., Walker, W. F. & Guilford, W. H. Quantitative comparison of algorithms for tracking single fluorescent particles. *Biophysical Journal* **81,** 2378-2388 (2001).
282. Manoharan, M. RNA interference and chemically modified small interfering RNAs. *Curr Opin Chem Biol* **8,** 570-579 (2004).

283. Parrish, S., Fleenor, J., Xu, S., Mello, C. & Fire, A. Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol Cell* **6,** 1077-1087 (2000).
284. Fahmy, K., Sakmar, T. P. & Siebert, F. Transducin-dependent protonation of glutamic acid 134 in rhodopsin. *Biochemistry* **39,** 10607-10612 (2000).
285. Stoica, C., Carmichael, J. B., Parker, H., Pare, J. & Hobman, T. C. Interactions between the RNA interference effector protein Ago1 and 14-3-3 proteins: consequences for cell cycle progression. *J Biol Chem* **281,** 37646-37651 (2006).
286. Stroffekova, K., Proenza, C. & Beam, K. G. The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins. *Pflügers Archiv : European journal of physiology* **442,** 859-866 (2001).
287. Kuszak, A. J., Pitchiaya, S., Anand, J. P., Mosberg, H. I., *et al.* Purification and functional reconstitution of monomeric mu-opioid receptors: allosteric modulation of agonist binding by Gi2. *J Biol Chem* **284,** 26732-26741 (2009).
288. Middleton, R. J. & Kellam, B. Fluorophore-tagged GPCR ligands. *Curr Opin Chem Biol* **9,** 517-525 (2005).
289. Ambros, V., Bartel, D. P., Burge, C. B., Carrington, J. C., *et al.* A uniform system for microRNA annotation. *Rna-a Publication of the Rna Society* **9,** 277-279 (2003).
290. Ma, X., Kumar, M., Choudhury, S. N., Becker Buscaglia, L. E., *et al.* Loss of the miR-21 allele elevates the expression of its target genes and reduces tumorigenesis. *Proc Natl Acad Sci U S A* (2011).
291. van Kouwenhove, M., Kedde, M. & Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat Rev Cancer* **11,** 644-656 (2011).
292. Weber, L. W., Boll, M. & Stampfl, A. Hepatotoxicity and mechanism of action of haloalkanes: carbon tetrachloride as a toxicological model. *Crit Rev Toxicol* **33,** 105-136 (2003).
293. Czech, B. & Hannon, G. J. Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* **12,** 19-31 (2011).
294. Ohrt, T., Mütze, J., Staroske, W., Weinmann, L., *et al.* Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Research* **36,** 6439-6449 (2008).
295. Grünwald, D., Singer, R. H. & Rout, M. Nuclear export dynamics of RNA-protein complexes. *Nature* **475,** 333-341 (2011).
296. Babcock, H. P., Chen, C. & Zhuang, X. W. Using single-particle tracking to study nuclear trafficking of viral genes. *Biophysical Journal* **87,** 2749-2758 (2004).
297. Itzkovitz, S. & van Oudenaarden, A. Validating transcripts with probes and imaging technology. *Nat Methods* **8,** S12-S19 (2011).
298. Walter, N. G. & Burke, J. M. Fluorescence assays to study structure, dynamics, and function of RNA and RNA-ligand complexes. *Methods Enzymol* **317,** 409-440 (2000).
299. Pinaud, F., Michalet, X., Iyer, G., Margeat, E., *et al.* Dynamic partitioning of a glycosyl-phosphatidylinositol-anchored protein in glycosphingolipid-rich microdomains imaged by single-quantum dot tracking. *Traffic* **10,** 691-712 (2009).
300. Saxton, M. J. & Jacobson, K. Single-particle tracking: applications to membrane dynamics. *Annual review of biophysics and biomolecular structure* **26,** 373-399 (1997).

301. Kusumi, A., Sako, Y. & Yamamoto, M. Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. *Biophys J* **65,** 2021-2040 (1993).
302. Ding, H., Wong, P. T., Lee, E. L., Gafni, A. & Steel, D. G. Determination of the oligomer size of amyloidogenic protein beta-amyloid(1-40) by single-molecule spectroscopy. *Biophys J* **97,** 912-921 (2009).
303. Chung, S. H. & Kennedy, R. A. Forward-backward non-linear filtering technique for extracting small biological signals from noise. *J Neurosci Methods* **40,** 71-86 (1991).
304. Karnani, N., Taylor, C. M., Malhotra, A. & Dutta, A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell* **21,** 393-404 (2010).
305. Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., *et al.* NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res* **39,** D1005-D1010 (2011).
306. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129,** 1401-1414 (2007).
307. Allawi, H. T., Dahlberg, J. E., Olson, S., Lund, E., *et al.* Quantitation of microRNAs using a modified Invader assay. *RNA* **10,** 1153-1161 (2004).
308. Neilson, J. R., Zheng, G. X., Burge, C. B. & Sharp, P. A. Dynamic regulation of miRNA expression in ordered stages of cellular development. *Genes Dev* **21,** 578-589 (2007).
309. Yeom, K. H., Heo, I., Lee, J., Hohng, S., *et al.* Single-molecule approach to immunoprecipitated protein complexes: insights into miRNA uridylation. *EMBO Rep* **12,** 690-696 (2011).