Novel Bioinformatics Approaches for MicroRNA Detection and Target Prediction

by

Subramanian Shankar Ajay

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Bioinformatics)
in The University of Michigan
2009

Doctoral committee:

Professor Brian D. Athey, Co-Chair
Research Investigator Inhan Lee, Co-Chair
Professor Daniel M. Burns Jr.
Assistant Professor Zhaohui Qin
Professor David J. States, University of Texas Health Sciences Center at Houston

To amma, appa and panni

Acknowledgements

I would like to show my immense gratitude to Prof. Brian Athey and Dr. Inhan Lee who took me under their wings when I first joined the Bioinformatics doctoral degree program. Their enthusiasm has been so infectious that I derived confidence from it every time I interacted with them. They have been a pillar of knowledge and support during the course of my graduate studies. Their ideas have helped shape this thesis to a large extent and I am grateful to them for the encouragement that they continued to provide along every step of the way.

I would like to thank Prof. David States for his invaluable inputs and discussions. The influence his pedagogical role has had in the first few years cannot be emphasized enough. Many thanks are also due to my doctoral committee members, Dr. Zhaohui Qin and Prof. Daniel Burns, for all their help when I turned to them. I have had the pleasure of learning bioinformatics methods and concepts under their tutelage. I really appreciate the fresh perspectives that Dr. Jeffrey DeWet lent when I had a tough time sorting out scientific problems, especially with respect to experiments. The extremely resourceful Bioinformatics graduate program staff, Alicia, Yuri, Julia, Sandy and Janet have helped make my transition and stay in the program facile.

Much of the work in this thesis would not have been possible without the support from laboratories of Dr. Arul Chinnaiyan from the Department of Pathology at the

University of Michigan, Dr. Haiming Chen from the Department of Psychiatry at the University of Michigan and Dr. JongIn Yook at Yonsei University. I am thankful to members of Dr. Chinnaiyan's group, Mohan, Bharathy and Anju who have taught me experimental techniques, from using a pipette to cell transfections.

Outside the scientific community some people have had an immeasurable impact in my life. The unconditional love and support (and money) from my family has seen me through highs and lows. The friends I've made through college and graduate school are ones I'll forever stand by - the spasmers from BITS to the ommalites in Ann Arbor, among whom Jaggi and Sweta deserve a special mention for everything they've helped me with. Work would not have been as much fun without k2 and uthra, and lunches would not have been as entertaining without the discussions on everything from politics to movies with my colleagues at Green Court. I thank them all.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

The draft human genome sequence assembly in 2001 was a landmark achievement in the field of genomics (Lander et al. 2001; Venter et al. 2001), paving the way for sequencing of nearly a hundred eukaryotic organisms (Liolios et al. 2008). Improvement in sequencing technology has steadily been lowering costs and turnaround times for access to vast amounts of data. The wealth of information that is being "mined" from this data through sophisticated tools has helped further our knowledge in biology and disease. Comparative genomics, benefiting from the data, has been making strides alongside, which has aided our understanding of genetic variation and evolution. Eventually, scientists aim to define features and gene expression control mechanisms that lend uniqueness not only at the tissue level but also at the level of an organism. Though many new regulatory elements have been characterized at both the DNA and RNA level, it is the harmonious action and interaction of all these elements that hold the key to understanding the emergent properties. The "–omics" revolution in different areas of biology is centered upon large-scale analysis of data mostly through computational approaches followed by experimental validation of models and predictions that are generated. Study of gene regulation has also seen such combinatorial approaches being adopted to elucidate elements and explain mechanisms involved. In this thesis we focus

on aspects related to one subset of eukaryotic gene regulation mechanisms, namely, post-transcriptional regulation mediated by microRNAs (miRNAs).

Broadly, the two main areas of contribution by this body of work relate to the detection of miRNAs using microarrays and to prediction of miRNA targets. Novel approaches for microarray probe design and for identification of regulatory elements specific to miRNAs followed by experimental validations for our predictions are described. In this chapter concepts related to these areas are introduced and a brief overview of challenges faced is presented.

1.1 Post-transcriptional gene regulation

Regulation of gene expression is a complex process that requires the coordination of multiple factors at several different steps (Orphanides and Reinberg 2002). Control begins with the organization of genetic material at the chromatin level which regulates access for DNA transcription (Richards and Elgin 2002). Transcriptional regulation is a highly regulated step that involves sequence and protein elements (Sperling 2007). Following RNA polymerase transcription of DNA, transcripts are further subjected to various processes like capping, editing, polyadenylation and regulation by non-coding RNAs before they are finally turned over (Moore 2005). Genome-wide studies of these post-transcriptional gene regulation mechanisms revealed their importance in cellular processes (Halbeisen et al. 2008).

One highly conserved mechanism of post-transcriptional gene regulation is RNA interference (RNAi), which causes gene silencing mediated by double-stranded RNA (dsRNA) (Cerutti and Casas-Mollano 2006). The term was first coined when repression

mediated by dsRNA was characterized in the nematode worm, *Caenorhabditis elegans* (*C. elegans*) (Fire et al. 1998). The most well-studied mechanism of mRNA degradation through a complex protein machinery is that by small-interfering RNA (siRNA) (Hannon and Rossi 2004). This species of RNA and the associated RNAi machinery has been investigated in several eukaryotic systems, making it an excellent tool for functional genetic studies (Elbashir et al. 2001) and also for therapeutics (Zimmermann et al. 2006). Though usually exogenous, many endogenous siRNAs like small scan-RNAs (scnRNAs), trans-acting siRNAs (tasi-RNAs) and repeat-associated siRNAs (rasiRNAs) have been discovered in other organisms (Kim 2005). In addition to siRNAs, other endogenous small RNAs like miRNAs and Piwi-interacting RNAs (piRNAs) share the RNAi machinery for transcriptional silencing, mRNA degradation or translational repression (Farazi et al. 2008).

1.2 Introduction to microRNAs

MicroRNAs (miRNAs) are 21-23 nucleotide (nt) long post-transcriptional regulators of gene expression many of which are conserved across metazoans (Bartel 2004). The first miRNA, *lin-4*, was discovered by genetics analysis in *C. elegans*, where it was found to negatively regulate protein levels of LIN-14 (Lee et al. 1993; Wightman et al. 1993). The term 'microRNA' was first coined in 2001 when tens of small RNAs with regulatory potential were discovered in *C. elegans* (Lau et al. 2001; Lee and Ambros 2001). A term search for 'microRNAs' on PubMed emphasizes the increasing interest and impact it has had (**Fig. 1.1**). At present there are over 8000 miRNAs that have been identified in over 30 different species through experimental and computational approaches (Griffiths-Jones

et al. 2008). Through diverse experimental approaches their prevalence has been brought to light not only in plants and animals but also in a unicellular alga (Zhao et al. 2007).

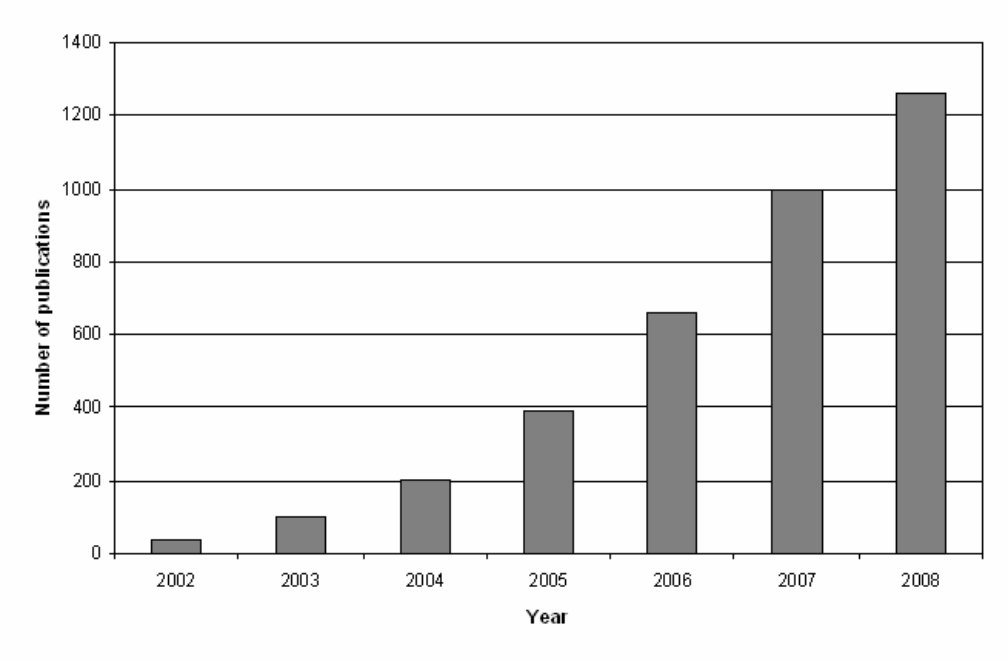


Figure 1.1 Trend showing increase in microRNA research since 2001

Their functions are diverse ranging from development control to apoptosis to involvement in disease like cancer (Kloosterman and Plasterk 2006; Bushati and Cohen 2007; Croce 2008). With an estimated 3% of human miRNAs each targeting hundreds of mRNAs their significance in post-transcriptional gene regulation is unmistakably large (Bartel 2004; Engels and Hutvagner 2006).

1.2.1 MicroRNA biogenesis

The mature form of a miRNA is generated through a multi-step process (Kim 2005). miRNA genes are first transcribed from genomic loci either as independent units or as

part of introns of other protein-coding genes (Du and Zamore 2005), and like mRNAs, they contain a 5′-cap structure and a poly-A tail (Bracht et al. 2004; Cai et al. 2004). This primary transcript (called pri-miRNA) exists in a hairpin conformation ~200nt long (**Fig. 1.2**), which is further processed by enzymes downstream.

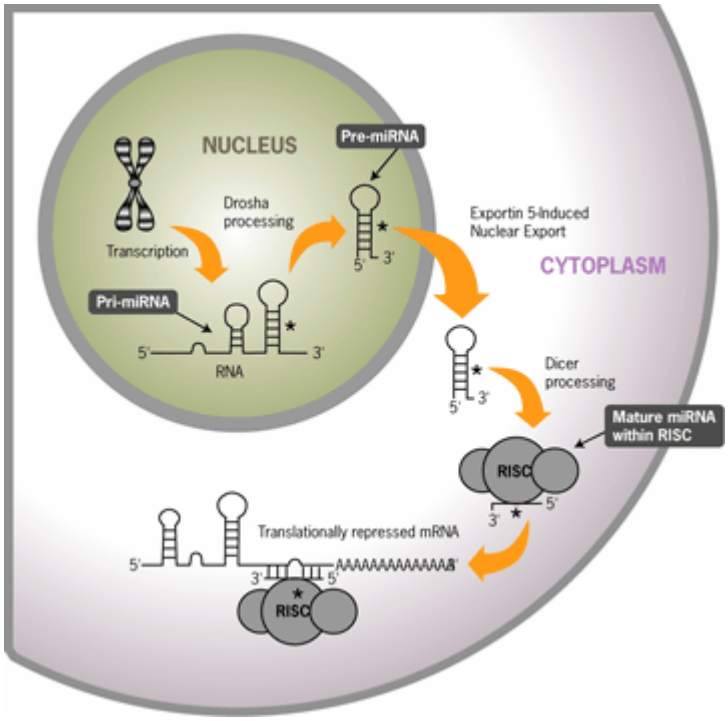


Figure 1.2 MicroRNA biogenesis (Source – http://www.ambion.com)

Drosha, an endoribonuclease belonging to the RNase III family, cleaves the pri-miRNA to form the precursor miRNA (pre-miRNA) molecule (Peters and Meister 2007). The enzyme Exportin-5 mediates the transport of pre-miRNA to the cytoplasm (Yi et al. 2003; Lund et al. 2004). Precursor miRNAs are then processed in the cytoplasm by the enzyme Dicer, which cleaves the hairpin loop to produce miRNA duplexes (Bernstein et al. 2001). In most cases only one of the strands in the duplex functions as a mature miRNA. This is decided by the thermodynamic stability of the ends of the duplex (Khvorova et al. 2003).

1.2.2 Mechanisms of miRNA-mediated repression

The mature strand of the miRNA is incorporated into a complex of ribonucleotide proteins (RNPs) to form the miRNP, also called the miRNA-induced silencing complex (miRISC). The primary proteins in this complex are members of the Argonaute (AGO) family, each of which possesses repressive capabilities. Mammals have four AGO proteins (AGO1-AGO4) of which only AGO2 has the potential to cleave target sequences due to its RNaseH-like domain (Peters and Meister 2007). The mature miRNA is used as a guide in the miRNP to recognize its target mRNA, to which it may be complementary with different degrees. In plants, miRNAs exhibit a near-perfect match to targets, thereby triggering an RNAi-like mechanism that results in cleavage of target mRNAs (**Fig. 1.3**), one of the modes of miRNA-mediated regulation (Jones-Rhoades et al. 2006). In animals, however, there is imperfect complementarity between a miRNA-target pair leading to several alternative mechanisms of repression (Filipowicz et al. 2008).

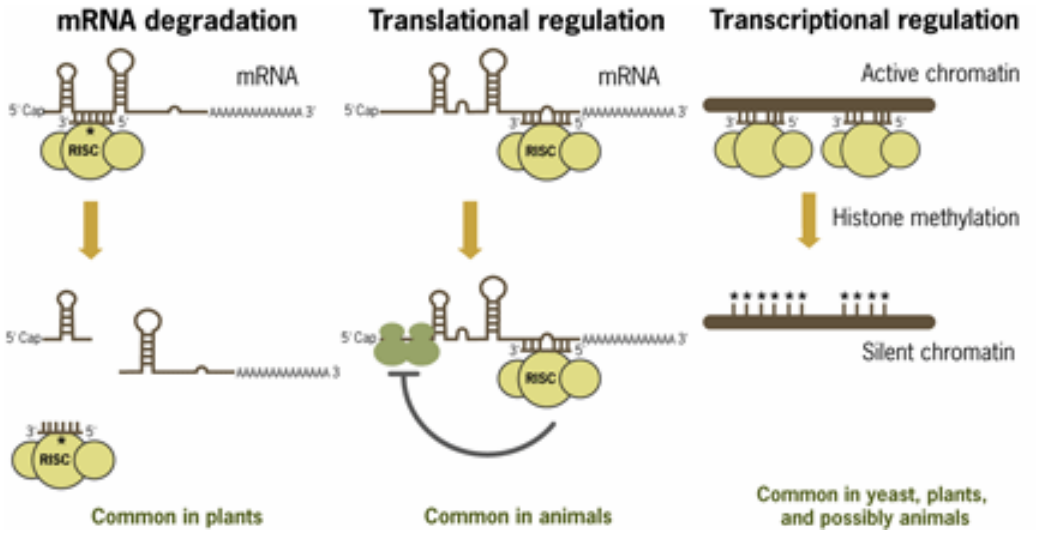


Figure 1.3 Mechanisms of miRNA-mediated post-transcriptional regulation (Source – www.ambion.com)

Majority of animal miRNA targets are regulated by repressing protein translation either at the initiation stage (Humphreys et al. 2005; Pillai et al. 2005) or during the elongation phase (Nottrott et al. 2006; Petersen et al. 2006). There is also recent evidence to show that mRNA destabilization occurs in certain cases (Bagga et al. 2005; Giraldez et al. 2006). Reconciliation of these mechanisms or at least understanding the context in which each mechanism predominates is an area of active work. One report shows promoter dependency for different translation repression mechanisms (Kong et al. 2008).

1.3 Nearest-neighbour thermodynamics

Much of the subject matter dealt in this thesis, especially microarray probe design tailored to miRNAs, employs a thermodynamic component based on the nearest-neighbour model. The approaches described in our work rely on nearest-neighbour thermodynamic quantities to assess the outcomes from our predictions, so a fair introduction to this topic will lay a foundation for rest of the chapters in this thesis.

Thermodynamic quantities have long been used to understand spontaneity of processes and one such is the change in Gibbs free energy. Conceptually, it is the amount of useful work that can be done by a system or the amount of work that must be done on a system for the process to take place (Haynie 2001). By this definition a negative change in free energy indicates spontaneity. Mathematically, the change in free energy at constant pressure and temperature is defined as

$$\Delta G = \Delta H - T\Delta S$$

where ΔH is the change in enthalpy, T is the absolute temperature and ΔS is the change in entropy.

An important application of free energy change is in the study of thermodynamics of nucleic acid base-pairing. Although thermodynamic measurements can be made for a sequence-pair of interest, it is not conceivable to record these values for every possible one comprised of Watson-Crick (WC) base-pairs, mismatches and other modified nucleotides. Clearly, a theoretical approximation that could predict these quantities for sequences that have not been studies would be very valuable. It is well understood that stability of base-pairing comes is correlated with the GC content of the nucleic-acid sequence. However, the largest contribution to helix stability comes from vertical stacking of bases in a sequence-dependent manner (Devoe and Tinoco 1962). These short-range interactions are useful to study the thermodynamic properties of a sequence as a function of its structure. The nearest-neighbour model assumes that the thermodynamic parameters of a given base-pair only depend on the adjacent pair and that the stability of helix formation may be approximated by pairwise addition of these nearest-neighbour parameters (Turner 1996; SantaLucia 1998).

Oligonucleotides and polymer duplex sequences have been used to estimate free energy and enthalpy changes by studying the 'melting' of these sequences assuming a two-state cooperative process (Borer et al. 1974). The procedure is repeated for many known sequences to derive the individual nearest-neighbour contributions to overall stability. Contributions by adjacent WC base-pairs have been studied and compiled by several groups for both the DNA (SantaLucia et al. 1996; SantaLucia 1998) and RNA backbones (Uhlenbeck et al. 1973; Borer et al. 1974; Breslauer et al. 1986; Freier et al.

1986). Apart from WC base-pairs, parameters for dangling ends, internal mismatches and loops have since been compiled (SantaLucia and Hicks 2004).

The change in free energy or in enthalpy for an unknown sequence may be calculated using the following equation,

$$\Delta G^{\circ}_{37\,(total)} = \Delta G^{\circ}_{37\,initiation} + \Delta G^{\circ}_{37\,symmetry} + \Sigma \Delta G^{\circ}_{37\,stack} + \Delta G^{\circ}_{37\,AT\,ter\,\min al}$$

where $\Delta G^{\circ}_{37\,initiation}$ is the parameter for duplex initiation, $\Delta G^{\circ}_{37\,symmetry}$ is the symmetry penalty for self-complementary sequences, $\Sigma \Delta G^{\circ}_{37\,stack}$ is the sum of individual nearest-neighbour contributions from tabulated data and $\Delta G^{\circ}_{37\,AT\,ter\,\min al}$ is the penalty for a terminal AT nearest-neighbour. The magnitude of negative free energy change thus calculated provides an indication of the strength of base-pairing. This formula is also applied to calculate the total changes in enthalpy and entropy. The melting temperature (T_m), defined as the temperature at which 50% of the oligonucleotide molecules are single stranded, is then given by

$$T_m = \Delta H^{\circ} \times 1000 / \left(\Delta S^{\circ} + R \times \ln\left(C_T / x\right)\right) - 273.15$$

where ΔH° (kcal/mol) and ΔS° (entropy units) are the changes in standard enthalpy and entropy, R is the gas constant (1.9872 cal/K-mol), C_T is the total molar strand concentration and x equals 4 for nonself-complementary duplexes and 1 for self-complementary duplexes.

Applications of nearest-neighbour thermodynamics to nucleic acid base-pairing exist in secondary structure prediction, primer design and microarray probe design among others. A quantum mechanical or a statistical mechanical treatment of thermodynamic quantities can provide an understanding of the molecular interactions involved when

base-pairing occurs, but the idea is to be able to predict if a given process can occur spontaneously, which the free energy change in the classical sense using the nearest-neighbour model adequately does.

A phenomenon that is common to miRNA expression detection and miRNA target recognition is nucleic-acid hybridization. In the case of miRNA detection, a nucleotide probe or primer is designed to efficiently pair with the miRNA sequence intended to be captured. Similarly, irrespective of the degree of complementarity between a miRNA and target sequence, this process is at least partly responsible for miRNA action. In this thesis, we use tools that employ nearest-neighbour parameters to assess the spontaneity of these processes.

1.4 Problem Statement

Understanding the mechanisms of miRNA-mediated post-transcriptional gene regulation is critical to learning how disease phenotypes are manifested because of dysfunctional regulation and being able to develop therapeutic solutions. Even before we can begin to understand mechanistic implications, the ability to accurately profile miRNA expression and the ability to predict which genes they target, be it in different tissues or in normal and disease states, will bring us closer to this goal.

We developed methods to generate high-confidence predictions in both these respects. We provide an overview of the current approaches and challenges for both miRNA expression profiling and target prediction, which are motivations for this thesis.

1.4.1 Detection of miRNA expression

Being able to determine expression patterns of all miRNAs in different tissues can help us better understand their roles in development and gene regulation. The first described method for miRNA expression detection used a northern blot procedure (Lee et al. 1993). Although this gel-based method can convincingly determine the length of the hybridizing sequence and is fairly sensitive the biggest disadvantage is the amount of time consumed by this technique. This makes it unsuitable for profiling the expression of hundreds of miRNAs simultaneously. Even though quantitative RT-PCR (qRT-PCR) offers a highly sensitive method (Fulci et al. 2007), the lack of parallelism is evident.

Expanding an earlier cloning and sequencing study (Lagos-Quintana et al. 2002) across tens of different tissues and various cell-lines has produced very comprehensive datasets that can be visualized to study miRNA expression (Landgraf et al. 2007); however it has its shortcomings too (Lim and Linsley 2007). A high-throughput alternative to the aforementioned techniques is a microarray. The use of microarray technology was pioneered by Schena et al. to study the expression of a set of *Arabidopsis thaliana (A. thaliana)* genes on a complementary DNA microarray (Schena et al. 1995) and the first global gene expression study followed two years later (DeRisi et al. 1997). Since then microarrays have been used for both genomic and transcriptomic analysis, and have also been adopted to study the expression of miRNAs (Nelson et al. 2004; Thomson et al. 2004). Not only have they been used to profile expression in different tissue types (Babak et al. 2004; Liu et al. 2004)but also in studying their impact on mRNA repression and to gain insight into target evolution (Farh et al. 2005).

Some of the challenges associated with detecting miRNAs using microarrays are associated with the inherent nature of miRNA sequences. It has been proven that the mature sequences are involved in hybridizing with the designed probes and not their hairpin precursors (Barad et al. 2004). This means that probes designed to hybridize with the mature miRNAs are, in essence, reverse complementary to the candidate (mature) miRNA. Since there is only one candidate sequence for a probe the challenges at least three-fold:

- *Cross-hybridization with closely-related sequences* – many miRNAs are grouped into families based on the similarity of sequences. One such family is the human *let-7* group of miRNAs (**Fig. 1.4**). It is evident that these miRNAs share extensive similarity and are different only by one or two nucleotides. We show that probes that are designed perfectly complementary to the intended sequence cross-hybridize with non-target sequences.
- *Non-uniform melting temperatures* – the base compositions of all mature sequences are fairly varied which results in a diverse range of melting temperatures. Finding a suitable experimental condition to assay for all miRNAs is, therefore, not a trivial task.

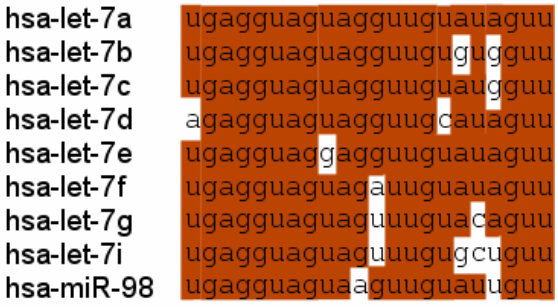


Figure 1.4. The human let-7 miRNA family consists of highly similar mature sequences

- *Probe secondary structure* – Probes may fold upon themselves to form secondary structures, rendering them unable to bind to target sequences being assayed for.

The approaches taken to increase specificity of hybridization or to balance melting temperatures of sequences either use modified nucleotides (Guo et al. 1997; Castoldi et al. 2006) or linker sequences (Wang et al. 2007). Using high-throughput sequencing technologies for RNA profiling (Nagalakshmi et al. 2008) has begun to alleviate specificity issues but current costs, errors rates and associated lack of standards for data analysis still pose an impediment.

We present an alternative probe design strategy that uses naturally occurring nucleotides wherein mismatches introduced in the probes eliminate the above unfavourable scenarios.

1.4.2 Computational miRNA target prediction in animals

Genetic approaches helped identify the first miRNA-target pair in the nematode worm, *C. elegans* (Lee et al. 1993; Wightman et al. 1993). Sequencing data from several different species further led to the discovery of many miRNAs, which in turn spurred the development of computational techniques to identify targets. The mechanisms behind miRNA action have not been revealed completely which pose a challenge in identifying true targets. Some of the earliest data in flies showed regulatory motifs on the 3′-UTRs of mRNAs that were complementary to the 5′-end of miRNAs (Lai 2002). It soon became clear that a short region (6-8 nt) on the 5′-end of the mature miRNA called the ‘seed’ was

the primary participant in Watson-Crick base-pairing with the 3′-UTR of mRNAs and contributed to efficient repression (Lewis et al. 2003). Several predictions emerged both in fruit-fly (Enright et al. 2003; Rajewsky and Socci 2004) and vertebrates that similarly reported the involvement of the 5′-end of the miRNA in target recognition (John et al. 2004; Kiriakidou et al. 2004). Results from these predictions show that animal target sites are only perfectly complementary to miRNAs.

Considering the short length (6-8 nt) of a match, this is a source for many false-positive predictions. To counter this problem most programs employ a combination of two or more of three major criteria to identify miRNA targets: 1) seed-match between miRNA and target 3′-UTR – while some programs require or prefer perfect seed-matches (Krek et al. 2005; Lewis et al. 2005) others allow imperfect base-pairing (Enright et al. 2003; Kiriakidou et al. 2004), 2) free energy of binding between the miRNA and target site, and 3) cross-species conservation of miRNAs and/or target sites – all programs use *a priori* conservation information across two or more species based on the idea that evolutionary constraint could signify function. Considerable variation in predictions from different algorithms coupled with the fact that only a small fraction of predictions are validated in each case leaves us with very little knowledge to make predictions with confidence.

The approaches mentioned above leave out possible target sites that may not be conserved yet are functional. It is plausible that these non-conserved sites are unique to certain miRNAs in a species – a source of variation in targeting, and hence in gene regulation, across different organisms. In chapter 3 we present work beginning with

formulation of a hypothesis based on this idea, leading to testing with sequence data on a whole-genome scale followed by validation in biological systems of interest.

1.5 Contributions

The obstacles outlined in the problem statement currently prevent us from harnessing the power of a high-throughput technology like microarrays to detect miRNAs using conventional DNA probes. Though sensitivity and specificity of probes can be improved with artificial probes they are less cost effective. Specificity with respect to target prediction was raised with current approaches that consider conservation across multiple species. Here, we detail the contributions made by this thesis to alleviate these challenges

1.5.1 Target-specific microarray probe design

In combination with the use of nearest-neighbor thermodynamics, we discuss strategies to generate optimal probes for the entire complement of mature human miRNAs. Unlike conventional perfectly-matched probes, we introduce base changes in a probe sequence that serve to eliminate cross-hybridization, reduce probes with high melting temperatures and/or secondary structural features. This strategy was adopted from a study that demonstrates the dependence of oligonucleotide melting temperatures on the natures of mismatches and their positions (Lee et al. 2004). Computationally, Probe Design Guru (ProDeG) was able to design probes for all miRNAs in the human dataset (Lee et al. 2008). We employed the services of LC Sciences, Inc. (Houston, TX) to validate probes designed for six of the *let-7* family of miRNAs using spiked-in samples cDNA and RNA samples. Cross-hybridization observed with perfectly-matched probes was eliminated

when ProDeG designed probes were used. As a collaborative effort with Dr. Haiming Chen (Department of Psychiatry, University of Michigan) total RNA from lymphoblastoid cell-line was used to assay *let-7* miRNAs from a true biological sample. We demonstrate the fidelity of ProDeG probes (microarray experiments at LC Sciences, Inc.) by reproducing expression patterns of the *let-7* miRNAs as determined by qRT-PCR (by Dr.Chen's group).

1.5.2 Discovery of endogenous 5′-UTR target sites

Majority of animal miRNA target prediction programs rely on conservation of sites in two or more species and only consider interaction of the seed region with these sites. A hypothesis was proposed by Dr. Inhan Lee suggesting that the 3′-end of the miRNAs may interact with regions on the 5′-UTR that are less conserved – a source of species-specific or gene-specific variation in targeting. We used data from a previously compiled genome-wide motif study (Xie et al. 2005) to examine the propensity of miRNAs, both 5′- and 3′-ends, to interact with 5′-UTR and 3′-UTR motifs of different degrees of conservation. We first show that motifs from the 5′-UTR with little or no conservation interact preferably with 3′-ends of miRNAs. Taken together with the seed-matches, we surmised that a miRNA may target both 5′-UTR and 3′-UTR of a gene simultaneously.

We collaborated with Dr. JongIn Yook (Dental School, Yonsei University, Seoul, South Korea) and Dr. Arul Chinnaiyan (Department of Pathology and Urology, University of Michigan) for *in vitro* verification of results from the computational study. Sequence analysis of two genes, human AXIN2 and the well studied *C. elegans* LIN-28, revealed conserved sites on the 3′-UTRs and non-conserved sites on the 5′-UTRs for *hsa-*

miR-34a and *cel-lin-4*. Reporter gene experiments revealed that the 5′-UTR site of AXIN2 was able to repress protein translation independent of the 3′-UTR site (work done by Dr. Yook's group). Inhibiting the endogenous expression of *hsa-miR-34a* produced a greater rescue from repression when both sites were present compared to when either site was present alone.

We also performed similar experiments to validate the lin4-lin28 pair using a designed siRNA that contains an intact seed-match but modified 3′-end with compensatory modifications in the 5′-UTR. This was done in Dr. Chinnaiyan's laboratory under the guidance of Dr. Saravana Dhanasekaran. We show that the 5′-UTR target site for modified siRNA was found to influence reporter gene product (luciferase) expression. The interaction was determined to be sequence-specific by mutating target sites on the UTRs.

Having established the ability of endogenous 5′-UTRs to interact with miRNAs, we examined a known regulatory element on the 5′-UTR, namely upstream AUGs (uAUG), for their potential to interact with miRNAs. Using uAUG sequences extracted from alignments of human and mouse 5′-UTRs (Churbanov et al. 2005) we demonstrate that these elements are probable target sites specific for miRNAs. We also show that the ability of uAUG motifs to confer cell-specific expression of the gene product correlates with the expression of miRNAs predicted to interact with the uAUG motifs in these cell-lines.

1.6 Thesis Outline

The following is a brief outline of work done to tackle the problems discussed above: Chapter 2 deals with the design of microarray probes tailored to miRNAs, providing methods for producing high-fidelity probes. In Chapter 3, we provide computational and experimental evidence to show that non-seed regions of miRNAs can target endogenous 5′-UTR sites. Chapter 4 provides a possible unified mechanism for the action of uAUGs, another post-transcriptional regulatory element, along with miRNAs by serving as binding sites. We conclude by discussing findings in this thesis and provide future directions in Chapter 5.

Chapter 2

Microarray probe design for miRNAs

2.1 Background

Many miRNAs are conserved across several species and are highly similar to other miRNAs in the genome. There is a great demand for accurate expression profiling of these miRNAs to better understand their tissue specificities (Babak et al. 2004; Barad et al. 2004; Liu et al. 2004; Chapman and Carrington 2007) and their role in development (Watanabe et al. 2005; Bushati and Cohen 2007; Moss 2007; Zhao and Srivastava 2007) and disease (Lu et al. 2005; Fulci et al. 2007; Jay et al. 2007; Soifer et al. 2007; van Rooij and Olson 2007).

Techniques for determining miRNA expression include Northern blot analyses (Valoczi et al. 2004), quantitative RT-PCR (Fulci et al. 2007), and microarrays (Thomson et al. 2004). Among these, the oligonucleotide microarray platform offers a simple and high-throughput experimental procedure for genome-wide miRNA profiling. Barad *et al.* have shown in expression profiling experiments that mature microRNA sequences, not their precursors, are responsible for fluorescence signals (Barad et al. 2004). By positioning short probes away from a solid support via an unrelated linker sequence, they have demonstrated efficient miRNA hybridization to the probes.

However, miRNA arrays pose several challenges. One is the ability of design strategies to distinguish many highly similar sequences that differ by only a few nucleotides. Another is the mere ~22 nt length of miRNA, which allows no choice for a probe sequence other than the miRNA itself. Given the diverse range of miRNA melting temperatures (T_m), it is almost impossible to find one experimental condition to satisfy all genomic miRNA hybridizations simultaneously. Currently there exist two major strategies for balancing T_m: 1) by incorporating chemically modified nucleotides with higher affinity (Castoldi et al. 2006) and 2) altering probe sizes (Wang et al. 2007). However, discriminating highly similar sequences, thus featuring similar T_m, remains a challenge. Such sequences will hybridize similarly to the probes and the signal will not be specific any more. Guo *et al*. have shown experimentally that the introduction of an artificial nucleotide (lacking hybridization ability) into the probes enhanced specificity and allowed discrimination of single nucleotide polymorphisms (Guo et al. 1997). However, the small data set and use of an artificial nucleotide limit genome-wide application, as no microarray could utilize this feature.

Conventionally, mismatched sequences have been used in assessing noise levels rather than signals because hybridization can disappear with single or double nucleotide mismatches. The problem arises that background signals produced by these mismatched probes can be as strong as those of the matched probes.

Here, we present Probe Design Guru, or ProDeG (pronounced *prodigy*), a highly specific microarray probe design algorithm that also ensures a narrower calculated T_m range. This is achieved by following a base-change strategy previously outlined (Lee et al. 2004). We applied ProDeG to miRNA sequences as a first step in validating our probes

based on the importance and feasibility. Since the probes do not include any modified nucleotides or change of lengths, our methods are easy to incorporate into any microarray platform. Applying this method to human mature miRNAs from miRBase version 9.1, we found specific probes for all members of the *let-7*.

2.2 Computational Methods

2.2.1 Base change strategy

In a previous study, we identified mismatched sequences and positions which induced minimal or maximal changes in oligonucleotide hybridization compared to perfectly matched sequences. In addition, we found T_m variance with two-point mismatches to be greater than twice that with one-point mismatches (Lee et al. 2004). By carefully introducing mismatches into a probe sequence, we can increase differences in stabilities of hybridization between target and non-target sequences sufficient to achieve discrimination, as shown in **Fig. 2.1**. This technique allows the reduction of probe-target hybridization melting temperatures (T_m) when they significantly exceed the T_m of most other probe-target pairs. Introducing mismatches in the probe sequence can also serve to eliminate secondary structures of probes.

2.2.1 ProDeG algorithm

ProDeG follows a series of steps in scrutinizing each of the probes before reporting them as specific to a targeted miRNA. The flow chart in **Fig. 2.2** details all the steps in processing before final reporting on probes. Initially, the sole candidate probe is the mature miRNA sequence. Following this, probes are evaluated in two broad stages, first

addressing probe quality in respect to a target and secondly checking non-targets. In the first stage, probes are assessed for their structural properties and for their hybridization with the target sequence. Undesirable stable hairpin formations in probes and uniform T_m are evaluated. Melting temperature as a measure of hybridization stability is calculated using the nearest neighbor thermodynamics model (SantaLucia and Hicks 2004) with licensed software Oligonucleotide Modeling Platform (OMP; http://www.dnasoftware.com). Observing that OMP calculations correlated with experimental T_m better than our in-house program containing publicly available parameters, we then calculated T_m variance dependency on mismatch positions (Lee et al. 2004). Interestingly, Pizhitkov *et al.* recently reported a similar mismatch position dependency (Pozhitkov et al. 2006) based on microarray signal intensity data, leading us to utilize OMP in the miRNA probe design.

In the second stage, we use BLAST to search for similarities among all the other human miRNA sequences, making sure that the DUST program is turned off using the –F option so all sequence stretches are considered. Candidates predicted to cross-hybridize with matches ≥ 14 nucleotides are retained for further processing. Predicted cross-hybridizations between probes and non-target sequences may, in fact, not occur due to unstable interactions. Such interactions then undergo a round of thermodynamic stability evaluations using OMP. Probes without any stable cross-hybridization are then reported as specific.

Next, imperfectly matched probes are used to identify a target sequence when candidates fail to satisfy the conditions set forth in the two prior stages of evaluation. If stable cross-hybridizations are present, we change bases in order to alter binding

stabilities enough to distinguish between target and non-target sequences (**Fig. 2.1**). If target-candidate hybridization T_m is above a set temperature (75 ºC in the current case) or the candidates have strong secondary structure, imperfectly matched candidates are generated to destabilize secondary structure and also reduce excessively high T_m of hybridization between a candidate and its target. To assess the probe characteristics after base changes have been introduced, each of these modified probes is made a new candidate for which evaluations are repeated *from the start*. When imperfectly matched probes satisfy all the set criteria, they can be reported as specific to the target.

If a single round of changes in the probe sequence fails to weaken secondary structure formation, reduce high T_m of hybridization with the target, or eliminate hybridization with non-target sequences, we subject the probe to a defined maximum number *n* of rounds of base changes (currently, $n = 2$) and evaluate its hybridization properties. In spite of having two sets of introduced mutations, miRNA probes still showing some cross-hybridizations are reported as such.

Mature human miRNAs have a very wide T_m range of about 36 ºC, the lowest and highest melting temperatures being 56 ºC and 92 ºC for *miR-620* and *miR-663*, respectively (1M salt concentration). Since there are numerous miRNAs with melting temperatures between 65 ºC and 75 ºC, we set the ceiling for the T_m range at 75 ºC. The discriminating ΔT for our program is based on data from experiments conducted by Thomson *et al*. which showed that sequences with one mismatch are distinguishable (Thomson et al. 2004). By calculating T_m for *miR-124a* and the reverse complements of the perfectly and imperfectly matched probe sequences used in their experiments (data not shown), we concluded that 5 ºC is sufficient. Even though absolute T_m is a function of

parameters in the nearest neighbor model, T_m difference is not. Since our criteria for lowest hybridization T_m between probe and target is 56 ºC, we set the maximum hybridization T_m between probe and non-target to be 51 ºC. Detailed calculation parameters are given in **Table 2.1**.

ProDeG is a bundle of programs for the UNIX programming environment written in PERL and C++, process flow being controlled by a PERL script which calls all other programs within it. ProDeG uses two external programs, BLAST and OMP. BLAST is available for download for several platforms; OMP is a licensed application available on several platforms as well and may be purchased from the vendor. By calculating T_m using the nearest neighbor model and published parameters (SantaLucia and Hicks 2004), licensed OMP may feasibly be replaced.

2.3 Computational Results

2.3.1 Variance of T_m by introducing mismatches

One of the most abundant and well-studied miRNAs is the *let-7* family, associated with most cancers (Johnson et al. 2005; Brueckner et al. 2007). The *let-7* family of sequences and their corresponding DNA hybridization T_m with perfectly complementary pairs are shown in **Table 2.2**. Each family member differs by only one or two nucleotides. Predicted cross-hybridizations with $T_m \geq 52$ ºC are also presented in **Table 2.2**. With perfectly matched probes, there is no way to prevent cross-hybridizations (Wang et al. 2007). Utilizing our finding that T_m variance with two-point mutations is greater than twice that with one-point mutations (Lee et al. 2004), discrimination is now possible. This synergetic effect is not limited to nearest neighbor two-point mutation sites. Rather,

most positions of an oligonucleotide show this, unless they are close to the chain end. The discrimination of *let-7e* and *let-7a* exemplifies the process diagram in **Fig. 2.1**. One nucleotide among these differs near the middle of the sequences. T_m of a perfectly-matched *let-7e* probe – target is 66 ºC; T_m of a perfectly-matched *let-7e* probe – non-target (*let-7a*) is 62 ºC (**Table 2.2**), so that $\Delta T_1 = T_{m1} - T_{m3} = 4$ ºC. When we change the 10^{th} position sequence *A* of the *let-7e* probe to *T*, T_m for target and non-target becomes 60 and 54 ºC, respectively ($\Delta T_2 = T_{m2} - T_{m4} = 6$ ºC). After incorporation of a base change, the T_m difference between target and non-target is increased ($\Delta T_2 > \Delta T_1$)**.** This technique, moreover, allows probe-target hybridization T_m's to be reduced when they significantly exceed the T_m of most other probe-target pairs. Introducing mismatches in the probe sequence can also serve to eliminate secondary structures.

2.3.2 ProDeG probes for human miRNA cDNAs

Taking advantage of the fact that T_m variance with two-point mutations is greater than twice that with one-point mutations, ProDeG processed mature human miRNAs to design microarray probes with the parameters in **Table 2.1** and predicted probes for all 470 of them. Calculations treat samples as reverse complementary DNA sequences to mature miRNA and DNA probes as equivalent to mature miRNA, in accordance with cDNA microarray experiments. These cDNA probes will validate that microarray signals produced by ProDeG from highly similar sequences are discriminated. Moreover, as miRNA amplification methods become more advanced, probes for miRNA cDNA may prove valuable. ProDeG probes for the *let-7* family are shown in **Table 2.3** along with predicted cross-hybridizations where $T_m \geq 52$ ºC. Following several mutation steps, all

cross-hybridizations predicted in **Table 2.2** have been eliminated. In addition, all the probes shown in **Table 2.3** have uniform melting temperatures (mostly 57 and 58 ºC). Note that *miR-98* did not undergo the mutation steps because our T_m ceiling criterion was set at 75 ºC.

2.3.3 Characteristics of ProDeG probes for cDNA of human miRNAs

Among the probes for the 470 mature miRNA sequences, those for 432 miRNAs are target specific, including imperfectly matched probes for 224 miRNAs, 160 of them due to eliminating cross-hybridizations of perfectly matched probes. Secondary structures were eliminated in probes for 27 miRNAs. High T_m was eliminated in probes for 76 miRNAs**.** We were able to overcome these obstacles (cross-hybridization, secondary structures, and high T_m) using imperfectly matched probe sequences. Designed probes for 38 mature miRNAs presented cross-hybridization with non-target miRNAs (mostly with one other); the detailed sequences and T_m are in **Supplementary Table 3 (new Appendix A.1)**. 20 out of 38 miRNAs were 100% identical to at least one other miRNA except for bases at either end of the sequences, 5 of the 20 being complete subsets of the other miRNAs. 10 other miRNAs contained one mismatch with other mature miRNA sequences at the second or third position from the 3′-end. The remaining eight miRNAs have one middle A which differs from G in another miRNA sample, leading to T (probe)-A (target sample) and T (probe)-G (non-target sample) discrimination tasks.

2.3.4 ProDeG probes for RNA samples of human miRNAs

Since most miRNA profiles use fractionated small RNAs from total RNA, we designed probes for RNA samples using hybridization parameters of DNA-RNA pairs. Again, all

470 probes for RNA samples were predicted. **Table 2.4** shows perfectly matched (control) let-7 probes while **Table 2.5** shows ProDeG-designed let-7 probes and their predicted cross-hybridizations using the same criteria of $T_m \geq 52$ ºC. Probes for RNA samples are predicted to present some cross-hybridization on let-7a probe with *let-7c* and *let-7e* samples and on *let-7c* probe with *let-7b*. T_m's for targets are less uniform and a bit higher than cDNA sample cases.

2.4 Experimental validation – Methods

2.4.1 Microarray platform

Microarray services were provided by LC Sciences Inc. (Houston, TX), which made the detection probes by *in situ* synthesis using photogenerated reagent chemistry on a microfluidic chip. We augmented their microarray layout with custom probes to experimentally validate our probe design strategy. The whole block of probe sets is repeated six times in a microarray. Custom probes include DNA sequences to the *let-7* family in **Table 2.2** (as a control) and the ProDeG *let-7* family probes in **Table 2.3** for the cDNA spiked-in experiments. Custom probes also include the reverse complementary sequences of the *let-7* family (as a control; **Table 2.4**), as well as ProDeG-designed probes for RNA samples **(Table 2.5)** of both spiked-in and total RNA from the lymphoblastoid cell lines.

2.4.2 let-7 family spiked-in experiments

DNA and RNA oligonucleotides with fluorescence dye attached to their 5′-end were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). DNA

sequences are reverse complementary to the mature *let-7* member sequences, while RNA sequences are the same as the mature *let-7* family. For pairing, *let-7a*, *let-7c*, and *let-7f* were labeled with Cy-5 and *let-7b*, *let-7d*, *let-7e* with Cy-3. LC Sciences performed custom microarray fabrication, hybridization, and signal reading. All hybridization was performed for one hour in the presence of hybridization buffer (25% formamide, 6 × SSPE, pH 6.8) on a μParaflo microfluidic chip using a micro-circulation pump (Atactic Technologies, Inc.; Houston, TX). The signal intensities of each pair (*let-7a*/*7d*, *let-7b*/*7c*, and *let-7e*/*7f*) were recorded at seven temperature conditions (25 °C to 55 °C) for both cDNA and RNA cases. Because the microarray platform is microfluidic, the hybridization solution contains formamide, which reduces hybridization temperature (Hutton 1977) to minimize bubble formation in the chamber. Internal controls were used to compare multiple experiments. Hybridization images were collected using a laser scanner (GenePix 4000B, Molecular Devices, Inc; Sunnyvale, CA) and digitized using Array-Pro image analysis software (Media Cybernetics, Inc). Data were analyzed by first subtracting the background and then normalizing the signals using a LOWESS filter (Locally-weighted Regression) to compensate for the intensity difference between Cy5 and Cy3.

2.4.3 Hybridization experiment using lymphoblastoid cell-line small RNA

Lymphoblastoid cell lines were prepared from blood draws of six human subjects using established methods (Neitzel 1986). Briefly, peripheral blood mononuclear cells were isolated from whole blood with Histopaque reagent (Sigma). For each blood sample, 10 ml of Histopaque was added to a 50 ml sterile conical tube. In another 50 ml conical

tube, 10 ml of well-inverted blood was mixed with 10 ml of RPMI 1640 medium (Invitrogen). We then gently layered the blood and RPMI mixture on top of the Histopaque, and centrifuged at 1500-1700 rpm for thirty minutes.

In a bar-coded T_{25} flask, we added 0.15 ml of phytohemagglutinin reagent and 6 ml of 30% FBS complete medium. When blood centrifugation was complete, we aspirated off the top layer and transferred the white cloudy middle layer into a new 50 ml conical tube to wash the PBM cells with RPMI 1640 medium. We then re-suspended the cell pellet in 2 ml of RMPI 1640 medium. In the T_{25} flask prepared as described above, we added 2 ml of filtered EBV and the suspended pellet. We then filled the flask with 30% FBS complete medium up to 10 ml of total volume. The cells were placed in a CO_2 incubator for 6-8 weeks. At the half-way point (about 3 weeks), we fed the cells with 10% FBS complete medium. When the culture grew to a confluency of 10^6 cells/ml, we collected the cells and made stocks with freezing medium, storing the cell stocks in freezers at –140 °C.

Total RNA from each human lymphoblastoid cell line was isolated with Trizol reagent (Invitrogen) according to the manufacturer's protocol (Invitrogen Cat No. 15596). Following the recommendation of LC Sciences, Inc., we used 1.5 ml of isopropyl alcohol per 1 ml of Trizol Reagent for the initial homogenization. We incubated samples at -20°C overnight and centrifuged them at no more than 12,000 × g for 10 minutes at 4°C. These modifications were necessary for the recovery of small RNAs from our cell line samples (based on preliminary study), which would be lost otherwise.

Microarray assay was performed using a service provider (LC Sciences). The assay started from 2 to 5 µg total RNA sample, which was size fractionated using a YM-

100 Microcon centrifugal filter (from Millipore) and the small RNAs (< 300 nt) isolated were 3′-extended with a poly(A) tail using poly(A) polymerase. An oligonucleotide tag was then ligated to the poly(A) tail for later fluorescent dye Cy-3 staining. Hybridization took place at 34 °C. Wash temperatures for control and ProDeG probes were 53 and 47 °C, respectively.

2.4.5 Quantitative RT-PCR

We purchased TaqMan® 2X Universal PCR Master mix and primers of the *let-7* family and a control from Applied Biosystems Inc. and followed the supplied protocol. Briefly, for each 15 μL RT reaction, 7 μL of RT master mix was combined with 5 μL total RNA (5 ng) in a tube and gently mixed. 3 μL of RT primer was added to each reaction tube, gently mixed and placed on ice. The tubes of a mixture were loaded into a thermal cycler and reverse transcription was performed. For each 20 μL PCR reaction, 10 μL of Master Mix were mixed with 7.67 μL nuclease-free water. Once the mixture was added to the PCR reaction tube, 1 μL of 20X TaqMan MicroRNA Assay mix and 1.33 μL of the RT product were transferred and gently mixed. The PCR reaction plate was prepared with 20 μL of the complete PCR master mix in each well. We used three replicates per RT reaction. Applied Biosystems 7900HT Fast Real-Time PCR System detected the fluorescence intensity during PCR amplification. We used SDS2.1 software (Applied Biosystems, CA) for quantification analysis in conjunction with the comparative Ct method.

2.5 Experimental validation – Results

2.5.1 Verification of ProDeG cDNA probe specificity using let-7 spike-in experiments

Spiked-in experiments were performed to verify designed probe specificity within the *let-7* family. Since there is no significant cross-hybridization for *let-7g*, *7i*, or *miR-98*, we used designed probes for *let-7a* to *7f* (**Table 2.3**). Based on the T_m calculations (**Table 2.2**), we paired *let-7a/7d*, *let-7b/7c*, and *let-7e/7f* for two-color hybridization experiments. Average fluorescent signals from six adjacent spots of perfectly matched probes (controls) and of our probes are shown in **Fig. 2.3a** and **2.3b**. Each control or probe signal value is chosen for its optimal discriminating temperature (35 and 30 ºC, respectively, with formamide addition (Hutton 1977)) from 55 to 25 ºC data and normalized with the highest signal value from the respective control set or probe set. Two clear advantages over the controls become apparent. First, probe-target signal intensities align except in the case of *let-7e* probes, yielding much more homogeneous fluorescence signals, as predicted. Second, cross-hybridizing signals appearing in the control sets are mostly removed. In addition, the highest signal intensity value from the control set is nearly 4 times greater than that of the probe set. The minimal cross-hybridization signals in **Fig. 2.3b** are practically non-existent. The question arises whether the signals from our probes are strong enough for use in an application.

When we performed RNA spike-in experiments with these probes over 7 temperature points from 25 to 55 ºC, we found that the hybridizations were more stable than those in the case of DNA-DNA. Since some signals of the mismatched probes were much stronger than with cDNA, we prepared the normalized signal graph at 40 degrees

for both control and ProDeG probes in **Fig. 2.4a** and **Fig 2.4b**. The T_m calculations are basically held in the signal intensities except for the *let-7b* probe **(Fig. 2.4b)**. If we set aside the *let-7b* probe signal, the specificity of ProDeG probes were dramatically superior to the control probes **(Fig. 2.4a)**, with only mild cross-hybridization of *let-7c* on *let-7a* probes. Please note that the overall cross-hybridization of control probes was also much more prevalent compared to the case of cDNA. The normalized intensity of control probes is about three times higher than that of ProDeG probes.

2.5.2 Expression signals of ProDeG let-7 probes from human lymphoblastoid cell lines

We prepared total RNA of lymphoblastoid cell lines from a human subject to obtain miRNA profiles. In addition to LC Sciences probes, we incorporated custom probes containing controls (perfectly matched sequences) and ProDeG probes to compare signals among them. Since the hybridization temperature was 34 ºC, optimized for the company's probes, gentle wash condition (47 ºC) was performed to detect ProDeG signals compensating weaker signals in addition to the normal wash condition (53 ºC). Each microarray contained probe blocks repeated six times. The relative signal intensities compared to the *let-7a* signal are shown in **Fig. 2.5a** and **Fig 2.5b** for control and ProDeG probes, respectively. Interestingly, *let-7b* signal from ProDeG probes was detectable, in spite of the unusually low *let-7b* spike-in signal in **Fig. 2.4b**. Rather, the *let-7b* signal in control probes was minimal. On the other hand, the *let-7c* signals from the control probe were significant, while those of the ProDeG probe were non-existent.

In order to verify the presence of each *let-7* family, we performed qRT-PCR on the same total RNA. The relative amount compared to *let-7a* quantity is shown in **Fig 2.5c**. The relative amount pattern strikingly resembles the ProDeG probe signal intensity: practically non-existent *let-7c* and *let-7e*, while *let-7a* amount is the largest followed by *let-7f* amount. We therefore conclude that the *let-7c* signals from the conventional perfectly-matched probe were actually false signals from other let-7 family members (probably from cross-hybridization with *let-7a* based on **Fig. 2.4a**). ProDeG probes are highly reproducible using qRT-PCR and proved to be specific in our study.

2.6 Discussion

Which miRNAs need to be discriminated? Even though we definitely removed most cross-hybridizations, at least in computational terms, several remain **(Supplementary Table 2, new Appendix A.1)**. Eliminating these involves discriminating one nucleotide difference near or at the end of the miRNA and discriminating T-A and T-G pairs. We reported that mutation in the first or last three bases of a sequence produces minimal T_m changes. Moreover, the interaction energy between T-A and T-G are similar, indicating limited discrimination by mismatched probes.

This limitation would be overcome when discriminating one nucleotide difference near or at the end of the miRNA by simulating an internal mismatch which may be obtained by padding two or three nucleotides during sample preparation. This concept has already been implemented by other researchers (Wang et al. 2007). ProDeG can then be applied to mismatch probe design. However, among the miRNAs listed in

Supplementary Table 2, new Appendix A.1, some are only predicted, without experimental confirmation. Also, a nuclease might have cut one or more sequences in the process of miRNA maturation. We do not feel compelled to go further in discriminating end sequence differences.

Discrimination of T-A and T-G pairs can be addressed using reverse complementary sequences as probes and mature miRNAs as samples. T (probe)-A (target sample) and T (probe)-G (non-target sample) pairs in the original set become A (probe)-T (target sample) and A (probe)-C (non-target sample) pairs in this reverse set. There should be no miRNAs in common in the G-U wobble category of predicted cross-hybridizations. Therefore, two sets of experiments, one using probes with mature miRNAs and the other using their reverse complements, will ultimately discriminate T-A and T-G pairs.

ProDeG probes for cDNA samples are of significant value both in terms of T_m calculations **(Table 2.3)** and spike-in experiments **(Fig. 2.3b)**. One intrinsic concern, however, is that signal intensities from the ProDeG probes are relatively weak compared to the perfectly matched probes, thus raising a question regarding signal sensitivity in real applications. The next step is to optimize hybridization conditions and to find a balance between specificity and sensitivity. However, once techniques to obtain cDNAs of small RNAs are further developed and PCR amplification is routinely achievable, increased specificity to a target sequence using the ProDeG algorithm will be of some value.

RNA samples produced stronger signals and more cross-hybridization **(Table 2.4**, **Table 2.5** and **Fig. 2.4)** than cDNA samples. Since signals from the ProDeG probes were strong enough, we could use the same hybridization temperature for both control and

ProDeG probes in total RNA profiling experiments. During the revision process, the Sanger Institute miRBase updated its miRNA sequence database to version 10. Since cDNA samples established the correspondence between microarray signals and our calculations, cDNA data are meaningful by themselves. However, profiling total RNA involves endogenous miRNA, which needs to be updated based on the new information. In terms of the *let-7* family, however, only one nucleotide was added at the 3′-end position for *let-7d*, *e*, *g*, and *i*, whose influence is probably not significant. We added a corresponding sequence A to each *let-7* ProDeG probe, as calculated with version 9.1 and used for probes for RNA samples.

Comparing T_m calculation **(Table 2.5)** and spike-in experimental data **(Fig. 2.4b)**, either the *let-7b* T_m calculation was wrong or *let-7b* RNA synthesis was not desirable or both. Since there was no *let-7e* cross-hybridization to the *let-7a* ProDeG probe (different from the Table 2.5 prediction), thermodynamic parameters of RNA-DNA pairs might be less accurate than those of DNA-DNA pairs. Improved thermodynamic parameters will increase the quality of designed probes. On the other hand, considering the *let-7b* signal detection using total RNA samples **(Fig 2.5b)**, there might not be a high-purity yield of *let-7b* RNA, as the company warned, due to the difficulty of incorporating Cy-5 into RNA oligonucleotides. Despite these limitations, to our surprise, the relative signal intensity of total RNA using ProDeG probes matched the qRT-PCR data excellently, demonstrating the utility of ProDeG probes.

The presence of *let-7c* signal in the control emphasizes the false positive signal in miRNA microarray data which is prone to generate incorrect inferences in terms of miRNA expression. Another miRNA, miR-99a, is transcribed right next to *let-7c*

transcription site in the same intron of Chromosome 21 open reading frame 34. The expressions of these two miRNAs were reported to be correlated (Landgraf et al. 2007). In our total RNA sample, the *miR-99a* signal was absent from the microarray data. However, significant false signaling of *let-7c* in the control probes **(Fig. 2.5a)** would not yield such a correlation. With our probes, we can report both *let-7c* and *miR-99a* are probably absent from the transcription stage.

The ProDeG strategy is simple, powerful, cost-efficient and fully compatible with current profiling techniques, moreover considering only naturally occurring nucleotide hybridization. The use of mismatched sequences with natural nucleotides (less toxic than artificial ones) to enhance target specificity (minimal off-target effects) will allow safer *in vivo* applications. Like other hybridization calculations, ours lacks surface effects, which may have led to a lower than predicted *let-7e* signal in **Fig. 2.3b** (note that *let-7d* and *let-7e* are one nucleotide shorter than other members according to v9.1 of miRBase). To our surprise, however, the overall calculation predicted microarray intensity very well. All experimental data point to the validity of our computational algorithm.

Table 2.1 Parameters used in ProDeG microarray probes for mature human miRNA

Parameters	
Assay Temperature[1]	53 ºC
Maximum hybridization T_m	75 ºC
Maximum monomer folding T_m[1] (secondary structure measurement)	65 ºC
Minimum hybridization T_m between probe and target	56 ºC
Maximum hybridization T_m between probe and non-target	51 ºC
Na+ concentration	1 M
K+ concentration	0 M
Probe concentration	100 nM
Target concentration	100 nM
BLAST word size	7

[1]Specific parameters in the OMP software

Table 2.2 Mature human *let-7* family sequences in DNA and their hybridization T_m with perfectly complementary pairs

Name	Sequence[1]	T_m (ºC)[2]								
		let-7a	*let-7b*	*let-7c*	*let-7d*	*let-7e*	*let-7f*	*let-7g*	*let-7i*	*miR-98*
control *let-7a*	TGAGGTAGTAGGTTGTATAGTT	**64**	58	59	58	57	59	52	51	51
control *let-7b*	TGAGGTAGTAGGTTGT***G***T***G***GTT	58	**70**	65	51			51	54	
control *let-7c*	TGAGGTAGTAGGTTGTAT***G***GTT	59	64	**67**	51	51	52	52	51	51
control *let-7d*	***A***GAGGTAGTAGGTTG***C***ATAGT	55	51	51	**66**					
control *let-7e*	TGAGGTAG***G***AGGTTGTATAGT	62	55	57	56	**66**	55			
control *let-7f*	TGAGGTAGTAG***A***TTGTATAGTT	57					**62**			
control *let-7g*	TGAGGTAGTAG***T***TTGTA***C***AGT							**64**	55	
control *let-7i*	TGAGGTAGTAG***T***TTGT***GCT***GT							55	**68**	
control *miR-98*	TGAGGTAGTA***A***GTTGTAT***T***GTT									**63**

[1]Mismatch sequences compared to *let-7a* are shown in bold italics.
[2]Hybridization T_m 51 ºC is shown for reference but not expected to produce signals with our design criteria.

Table 2.3 ProDeG-designed probe sequences for cDNA of mature human *let-7* family and their hybridization T_m with targets and non-targets

Name	Sequence[1]	T_m (ºC)[2]								
		let-7a	*let-7b*	*let-7c*	*let-7d*	*let-7e*	*let-7f*	*let-7g*	*let-7i*	*miR-98*
probe *let-7a*	TGAG**a**TAGTAGGTTGTATAGTT	**57**			51					
probe *let-7b*	TG**t**GGTAGTAGG**c**TGTGTGGTT		**57**							
probe *let-7c*	TG**t**GG**c**AGTAGGTTGTATGGTT			**57**						
probe *let-7d*	AGAGGTAGTA**a**GTTGCATAGT				**58**					
probe *let-7e*	TGA**c**GTAGGAGGTTGTATAGT	51				**57**				
probe *let-7f*	TG**c**GGTAGTAGATTGTATAGTT	51					**57**			
probe *let-7g*	TGAGGTA**a**TAGTTTGTACAGT							**56**		
probe *let-7i*	TGAGGTAGTA**c**TTTGTGCTGT								**58**	
probe *miR-98*	TGAGGTAGTAAGTTGTATTGTT									**63**

[1]Mismatch sequences compared to the original are shown in bold lower case.
[2]Hybridization T_m 51 ºC is shown for reference but not expected to produce signals with our design criteria.

Table 2.4 Hybridization T_m for mature human let-7 family of microRNAs between DNA probes and RNA targets and off-targets before ProDeG run

Name	Sequence	T_m (ºC) [1]								
		let-7a	*let-7b*	*let-7c*	*let-7d*	*let-7e*	*let-7f*	*let-7g*	*let-7i*	*miR-98*
control *let-7a*	AACTATACAACCTACTACCTCA	**64**	59	62	59	63	57	*51*		*51*
control *let-7b*	AACCACACAACCTACTACCTCA	58	**70**	64	56	55			52	
control *let-7c*	AACCATACAACCTACTACCTCA	60	65	**67**	56	57				
control *let-7d*	AACTATGCAACCTACTACCTCT	61	55	58	**67**	59	52			
control *let-7e*	AACTATACAACCTCCTACCTCA	58	*51*	55	*51*	**67**				
control *let-7f*	AACTATACAATCTACTACCTCA	59	52	56	52	57	**62**	*51*		
control *let-7g*	AACTGTACAAACTACTACCTCA	57	52	55		54	54	**65**	58	
control *let-7i*	AACAGCACAAACTACTACCTCA	52	57	52				56	**68**	
control *miR-98*	AACAATACAACTTACTACCTCA	53		54						**63**

[1]Hybridization T_m 51 ºC is shown for reference but is not expected to produce signals when considering our calculation criteria.

Table 2.5 ProDeG-designed probe sequences for mature human *let-7* family (RNA as sample) and their respective hybridization T_m with targets and non-targets

Name	Sequence[1]	T_m (ºC)[2]								
		let-7a	*let-7b*	*let-7c*	*let-7d*	*let-7e*	*let-7f*	*let-7g*	*let-7i*	*miR-98*
probe *let-7a*	AACTATACAACCTACTA**t**CTCA	**59**	52	56	52	57				
probe *let-7b*	AACCACACAAC**t**TACTACC**a**CA		**60**	*51*						
probe *let-7c*	AACCATACAACCTA**t**TACCT**t**A		55	**58**						
probe *let-7d*	A**c**CTATGC**c**ACCTACTACCTCT				**59**					
probe *let-7e*	A**t**CTATA**a**AACCTCCTACCTCA					**58**				
probe *let-7f*	**t**ACTATACA**g**TCTACTACCTCA	53					**57**			
probe *let-7g*	AACTGTAC**t**AACTACTACCTCA	*51*						**60**	52	
probe *let-7i*	AACAGCAC**c**AACTACTACCTCA								**63**	
probe *miR-98*	AACAAT**t**CAACTTACTACCTCA									**58**

[1]Mismatch sequences compared to the original are shown in bold lower case.
[2]Hybridization T_m 51 ºC is shown for reference but not expected to produce signals with our design criteria.

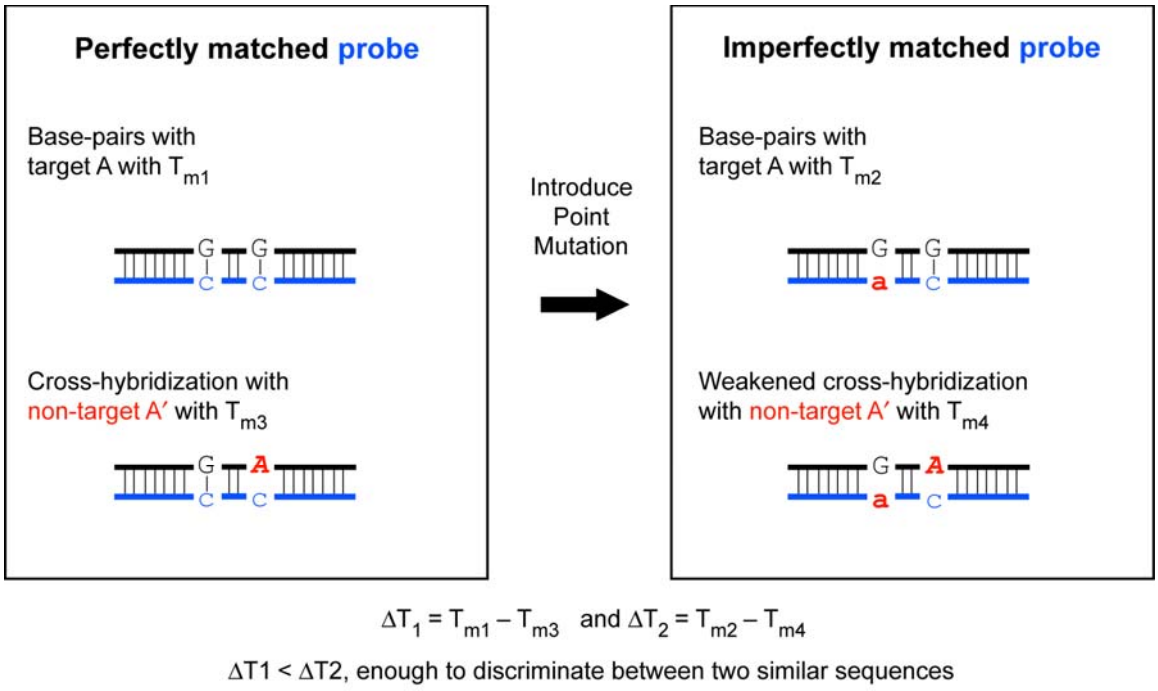


Figure 2.1 Schematic example of including an imperfectly matched probe to increase specificity. Probe strand is shown in blue and in lower case characters. Target A and non-target A' differ by one base at the position shown as sequence *A*. After incorporating a base change (sequence a), the difference in T_m between probe-target and probe-non-target pairs (right, $\Delta T2 = T_{m2} - T_{m4}$) is sufficient to discriminate similar sequences as compared to the difference in T_m before point mutation (left, $\Delta T1 = T_{m1} - T_{m3}$).

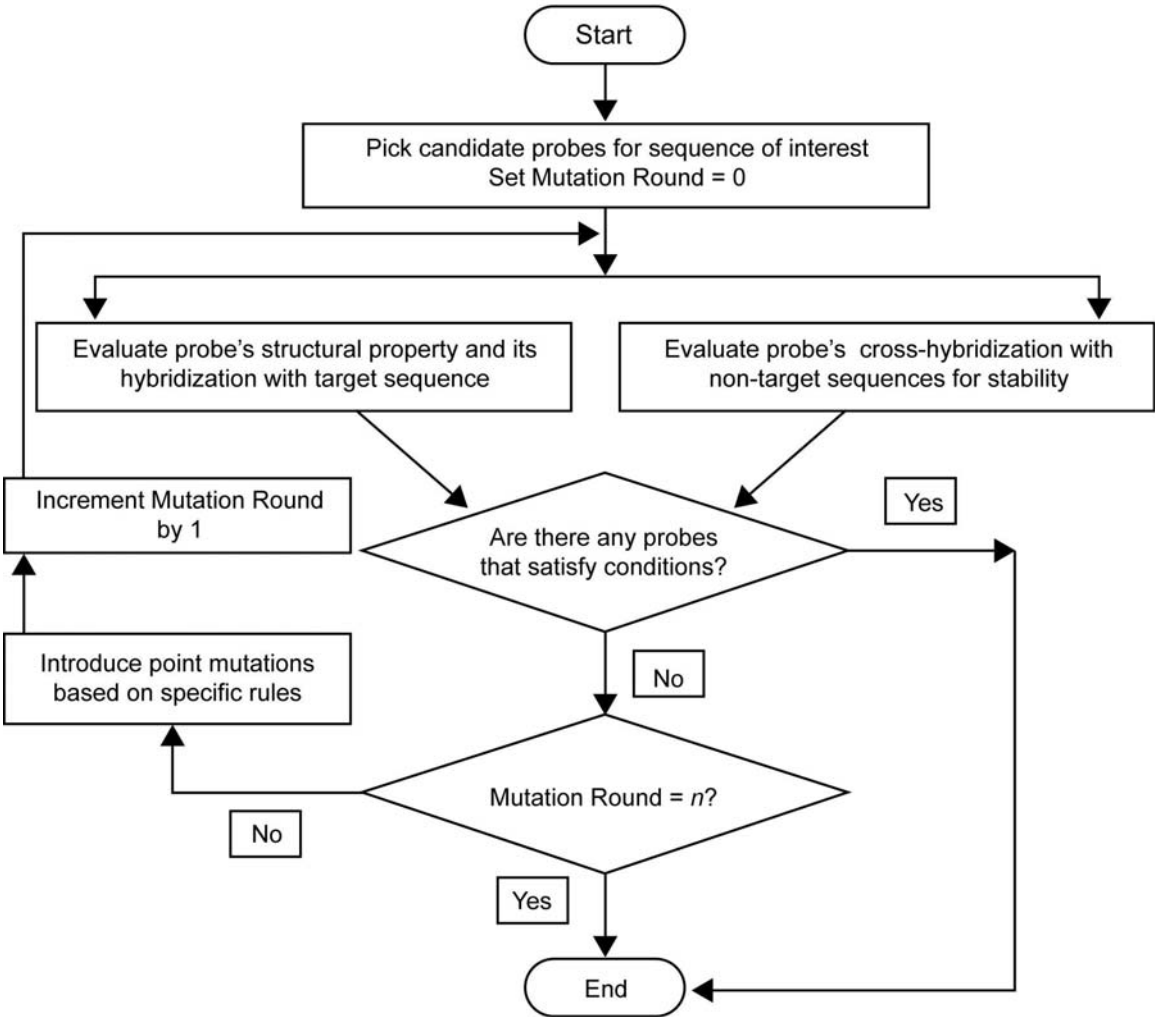


Figure 2.2 The ProDeG flowchart.

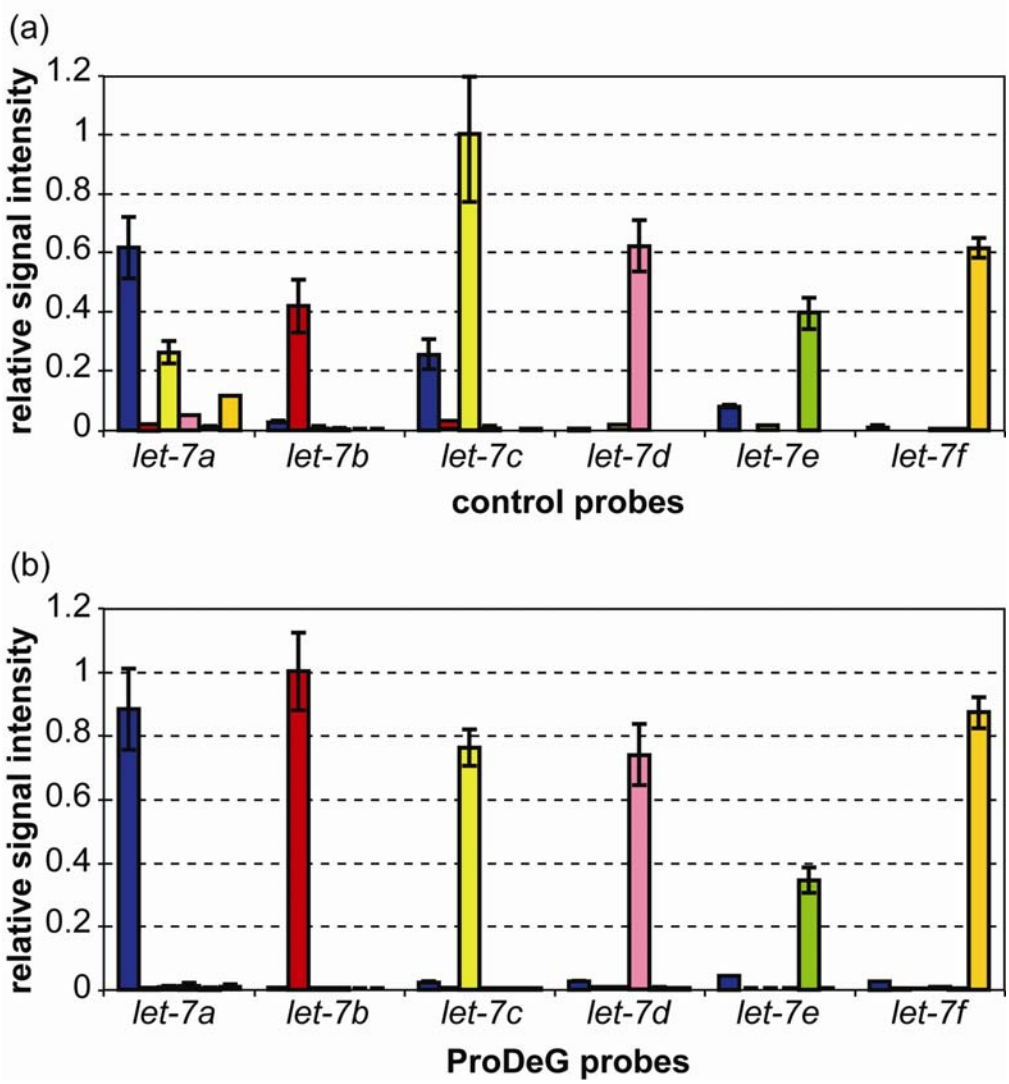


Figure 2.3 Relative signal intensities of the *let-7* family with spiked-in cDNA sequences. These are shown using control (perfectly matched) sequences as probes at 35 ºC **(a)** and ProDeG designed probes at 30 ºC **(b)**. Each x-axis category indicates probes used while the corresponding series shows the relative probe intensities normalized with highest intensity value. Spiked-in sample notations are as follows: blue bars, let-7a; red bars, let-7b; yellow bars, let-7c; pink bars, let-7d; lime bars, let-7e and orange bars, let-7f.

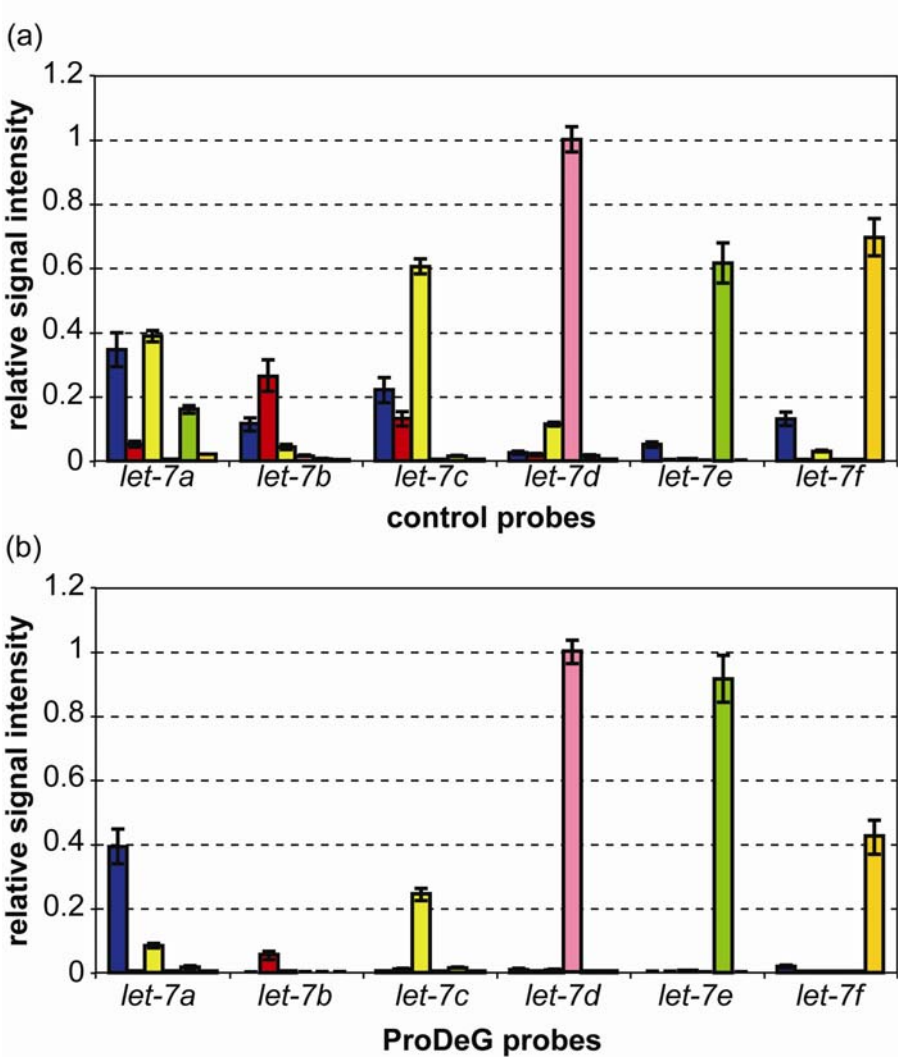


Figure 2.4 Relative signal intensities of the *let-7* family with spiked-in RNA sequences. These are shown using control (perfectly matched) sequences as probes **(a)** and ProDeG designed probes **(b)** at 40 ºC. Each x-axis category indicates probes used while the corresponding series shows the relative probe intensities normalized with highest intensity value. Spiked-in sample notations are as follows: blue bars, let-7a; red bars, let-7b; yellow bars, let-7c; pink bars, let-7d; lime bars, let-7e and orange bars, let-7f.

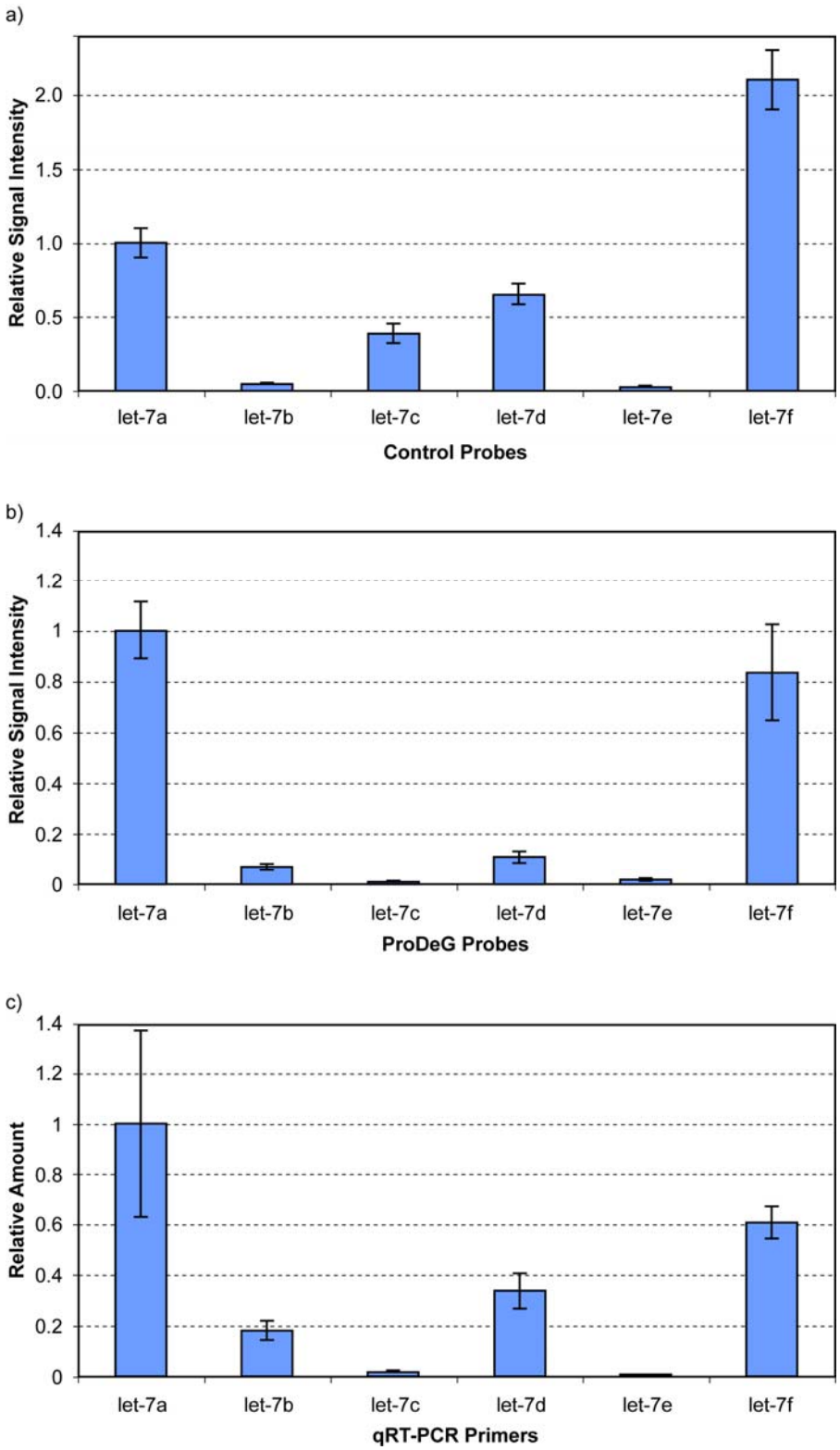


Figure 2.5 Total RNA sample data from a lymphoblastoid cell-line. All data are normalized against let-7a data. Relative signal intensity of the let-7 family control **(a)** and ProDeG **(b)** probes are shown after Cy-3 labeled total RNA from a lymphoblastoid cell-line was hybridized at 34 ºC. The relative amount of each let-7 family miRNA was quantified using TaqMan® qRT-PCR assay **(c)**.

Chapter 3

Discovery of endogenous 5′-UTR miRNA target sites

3.1 Background

miRNA target sites are known to primarily lie in 3′-UTR in animals. The first discovered miRNA, *lin-4* in *C. elegans*, was found to regulate developmental timing by targeting multiple sites in 3′-UTR of *lin-14* (Wightman et al. 1993). Since then several miRNA-target prediction programs have been developed (Rajewsky 2006; Sethupathy et al. 2006) that stress the importance of seed-match between the 5′-end of mature miRNA and 3′-UTR of the target mRNA, while some others show that the 3′-end of a miRNA may either complement a seed match or compensate for an imperfect one (Doench and Sharp 2004; Kiriakidou et al. 2004; Kloosterman et al. 2004; Grimson et al. 2007). The significance of this 5′-end of miRNA targeting 3′-UTR sites was recently confirmed by a proteomics study which showed superior protein inhibition capacity for the 3′-UTR sites over those in the coding region (Baek et al. 2008). It has been suggested a miRNA may target about 200 mRNAs (Krek et al. 2005), with varying degrees of protein repression (Baek et al. 2008; Selbach et al. 2008), though the number of predicted targets can range in the thousands. A major reason for these false positives lies in the partially complementary matches between miRNA and its targets.

To improve specificity of target prediction, many programs utilize sequence data to assess conservation of predicted target sites. This has also been done on a genome-

wide scale using motif data to uncover probable target sites (Xie et al. 2005). This study only discovered target sites on the 3′-UTR that interact with the seed region of miRNAs. Focusing on conserved target sites has the intrinsic limitation of not identifying species-specific sequences, including non-conserved miRNAs. Also, most miRNA target verification experiments have only used 3′-UTR interaction sites since many studies had shown miRNA effects with only portions of 3′-UTR. Even the first *lin-4* and *lin-14* experiments used not the whole mRNA, but rather the 3′-UTR together with coding region (Wightman et al. 1993). On the other hand, a few experiments have indicated possible target sites in the 5′-UTR (Jopling et al. 2005; Lytle et al. 2007; Orom et al. 2008).

We report here, based on both hybridization energy and sequence matches, many endogenous motifs within human 5′-UTRs specific to the 3′-ends of miRNAs. Rather than suggesting possible miRNA interactions with other regions of mRNA, we report combinatory interactions between a single miRNA and both end regions of an mRNA, based on our finding that many miRNAs contain significant interaction sites with mRNA 5′-UTR and 3′-UTR motifs through their 3′- and 5′-end sequences, respectively. As a model system, we experimentally verified that *hsa-miR-34a* function depends on both UTR sites of *AXIN2*. Additionally, we show that both UTRs of the *C. elegans lin28* gene is targeted by a modified *cel-lin-4*. We propose a new miRNA target class containing simultaneous 5′- and 3′-UTR interaction sites. This class can serve as an efficient screening tool for identifying real targets, especially in the case of non-conserved miRNAs or target sites.

3.2 Results

3.2.1 Presence of miRNA interaction sites in human 5′ UTR

We checked for genome-wide miRNA interaction motifs in human 5′-UTR and 3′-UTR. Xie *et al.* have reported conserved miRNA motifs in the 3′-UTR but not in the 5′-UTR or in coding sequences (Xie et al. 2005). We used the same UTR motif dataset sent to us by the authors but defined new conservation classes C=0 (non-conserved but human-enriched), 1 (minimally conserved and human-enriched), and ≥10 (highly conserved). To determine seed and non-seed region effects, all mature miRNAs were downloaded from miRBase (Release 11.0) (Griffiths-Jones et al. 2008) and split into their respective 5′- and 3′-ends, making miRNA halves. Following thermodynamic searches for half miRNA-UTR motif interaction using RNAhybrid (Kruger and Rehmsmeier 2006), we treated only consecutively-matched sequences as signals. To calculate significance, total numbers of pairwise interactions between half-miRNAs and UTR motifs were compared with the numbers of interactions with shuffled UTR motifs generated 1,000 times.

In these analyses, we identified 5′-UTR motifs (5U) interact significantly with the miRNA 3′-end (3P) in all conservation categories (5U3P's in **Fig. 3.1a**), most significantly in the case of C=0. 3′-UTR motifs (3U), on the other hand, show significant interactions with miRNA only in the case of highly conserved 8-mers (C≥10), which is consistent with previous reports (Conserved 10: 3U5P and 3U3P in **Fig. 3.1a**). Besides the most significant and well-known interaction of 3U5P, our identification of 3U3P interaction is in accordance with previous findings that the 3′-end of a miRNA may either complement a seed match or compensate for an imperfect one (Doench and Sharp 2004; Kiriakidou et al. 2004; Kloosterman et al. 2004; Grimson et al. 2007). Our new finding of

5U3P interaction was also observed with human-enriched 5′-UTR motifs when we followed Xie *et al.*'s conservation score (Methods and **Fig. 3.2**).

Viewed in terms of conserved and non-conserved miRNAs, interactions with conserved miRNAs show a trend similar to the one above, differing only in the levels of significance (**Fig. 3.1b**). Interestingly, 5U3P interactions with non-conserved miRNAs lack significance for C=0 motifs (**Fig. 3.1c**), the 5U3P signal in C=0 in **Fig. 3.1a** coming from that of conserved miRNAs. We also observed significant interactions between highly conserved 5′-UTR motifs and the 5′-end (5U5P) of non-conserved miRNAs **(Fig. 3.1c)**.

In conjunction with the significant interaction between the seed region of a miRNA and the 3′-UTR, the preferential 5′-UTR interaction with the 3′-end of miRNA raises the question whether a common miRNA may target both UTRs of an mRNA by interacting with different ends of the miRNA. Based on the significance data in **Fig. 3.1a**, 37 common miRNAs identified between 5U3P (C=0 and 1: total 250 miRNAs) and 3U5P (C=10: total 116 miRNAs) cases are listed in **Appendix B.1**. When these kinds of motifs exist in a single gene, will they be regulated by a single miRNA?

3.2.2 *hsa-miR-34a* targets *AXIN2* through both UTRs

A highly-conserved human miRNA, *hsa-miR-34a*, has such interaction sites in *AXIN2* **(Fig. 3.3a)**. Even though *miR-34a* is not one of the miRNAs in **Appendix B.1**, the 5′-end was predicted to interact with three highly-conserved (and one non-conserved) *AXIN* 3′-UTR sites, and the 3′-end with two overlapping 5′-UTR sites **(Fig. 3.3a)** present only in human and mouse but enriched in human 5′-UTRs **(Appendix B.2)**. We used *hsa-miR-*

34a and *AXIN2* as a model system to verify simultaneous UTR interactions. Since interactions between miRNA and 3′-UTR are well-established, we focused on the 5′-UTR interaction sites, using only minimal interaction sequences of 36-mer in the experimental constructs. As shown in **Fig. 3.3a**, the *hsa-miR-34a* effects on this 36-mer should mostly come from the 3′-end. Note that the full 3′-UTR inserted in the construct is 1,408 nucleotides long.

Reporter gene assay of MCF-7 cells revealed that *miR-34a* downregulated constructs containing either the 5′-UTR (5ULuc) or 3′-UTR (Luc3U) alone. When both *Axin2* UTR sites were present (5ULuc3U), luciferase expression was further repressed by *miR-34a* **(Fig. 3.3c)**. In order to identify endogenous miRNA effects in addition to those exogenously induced, we blocked endogenous *miR-34a* using inhibitor antisense RNA oligo. 5ULuc3U expression was greater than 5ULuc or Luc3U, suggesting that the 5′-UTR of *AXIN2* together with the 3′-UTR are functional target sites for *miR-34a* in the cells **(Fig. 3.3d)**. In addition, the fold change of 5ULuc3U 1.88 is greater than with the addition of 5ULuc and Luc3U 1.61. Considering the many interaction sites in the 3′-UTR, the synergetic 5′-UTR effect on endogenous miRNA function is remarkable. These data suggest that in conjunction with the 3′-UTR, the 5′-UTR of *AXIN2* plays a role in miRNA-mediated repression in human cells beyond fine-tuning. In order to confirm the sequence specificity of 5′-UTR effects, we created a construct with sites mutated (5UmutLuc3U). Separate luciferase experiments inducing *hsa-miR-34a* showed rescue of repression when the 5U interaction sites are mutated **(Fig. 3.3e)**.

3.2.3 Modified *cel-lin-4* targets both *lin28* UTRs

As an additional model system to verify simultaneous interaction of a miRNA with both UTRs, we chose the *C. elegans* lin-4 and lin-28 pair. The 3′-UTR of lin28 contains a single canonical target site conserved in the lin28 homologs of human, mouse and chimpanzee whereas the single 5′-UTR site predicted to bind with the 3′-end of the miRNA is lacking in all of the homologs. Expecting fewer cellular responses, we decided to use human cell lines. Due to the lower physiological temperature of *C. elegans*, we increased 5′-UTR and 3′-end interaction by changing GU pairs into GC pairs. The resulting construct-miRNA pair consisted of lin-4-like artificial miRNA (lin4msiRNA) and lin-28-like 5′-UTR sequences for 5U3P interaction; lin-4 and lin-28 sequences were for 3U5P interactions **(Fig. 3.4a)**. Constructs with mismatched sequences were also prepared to study interaction-site specificity.

Reporter gene assay of HEK293 cells showed that lin4msiRNA repressed luciferase expressions more consistently when the 5′-UTR site is intact (Wilcoxon rank-sum test $p < 0.005$ for 5UpmLuc3Upm and 5UmmLuc3Upm; $p = 0.005$ for 5UpmLuc3Umm and 5UmmLuc3Umm). It is clear that mismatches in the 5′-UTR corresponding to the 3′-end of lin4msiRNA disrupt interaction. We recognize that the increased luciferase due to 3′-UTR mutation is much greater than that due to 5′-UTR mutation. Possibilities are 1) there is an additional endogenous effect for the 3′-UTR site due to the site's existence in the human homologue LIN28 (*hsa-miR-125a-5p and hsa-miR-125b* have same seed region compared to *cel-lin-4*), 2) 5′-UTR effects may require more overall structural context in addition to short oligonucleotide sequences, 3) exogenously-induced vector and small

RNA may not be ideal for observing endogenous cellular effects, and 4) 5′-UTR effects may reflect species-specific fine-tuning in this case.

3.3 Methods

3.3.1 Bioinformatics and statistical analysis

Mature human miRNA sequences were downloaded from miRBase, version 11.0. These were separated into two categories, conserved and non-conserved. We define a conserved miRNA as one that has a similarly-named counterpart in at least one other species regardless of the percentage identity. For example, *miR-34a* exists in humans as well as mouse and many others whereas *miR-1178*, a non-conserved miRNA by our definition, exists only in humans. Following this, miRNAs were split into their respective 5-prime and 3-prime end halves.

Xie *et al*. kindly provided us with data on conservation of all possible 8-mer sequences from aligned 5′-UTRs and 3′-UTRs among human, mouse, rat and dog (Xie et al. 2005). Each 8-mer was listed along with the number of occurrences conserved in all four species (C), the number of occurrences in the human sequence (N), and the conservation rate (R) given by the ratio C/N, where $0 \leq R \leq 1$. We created five motif conservation categories: 1) C=0, non-conserved 8-mers ordered on decreasing N, 2) C=1, 8-mers with exactly one conserved occurrence, ordered on decreasing N, 3) C≥10, 8-mers with at least 10 conserved occurrences ordered on decreasing C and decreasing R, 4) positive MCS, and 5) negative MCS described below. Briefly, the motif conservation score (MCS, from Xie *et al.*) is reported as a Z-score calculated using binomial probability, $MCS = \left(C - \left(Np_0\right)\right) / \sqrt{Np_0\left(1 - p_0\right)}$, where C is the number of conserved

instances, N the number of occurrences in human and p0 the estimated rate of conservation. We calculated p_0 as the average conservation rate of all 65,536 8-mers. The top 540 highest scoring 5′-UTR and 3′-UTR 8-mers from each category above were then used for further analysis. RNAhybrid (Kruger and Rehmsmeier 2006) was used to search for potential interactions between the UTR motifs and each miRNA. Doench *et al.* having demonstrated the correlation between binding energy and fold repression (Doench and Sharp 2004), we set an energy threshold of -14 kcal/mol based on the RNAhybrid binding energy prediction for the *CXCR4* siRNA seed region and the corresponding target site used in Doench *et al.*'s paper. The results were then filtered for consecutive 8-mer matches with GU wobbles between the 8-mers and miRNA ends.

Shuffled 8-mers derived from the corresponding conservation category were used as controls to assess the significance of the number of interactions between motifs and miRNAs. The control datasets were generated 1000 times and the number of interactions calculated as an average over these iterations. We assumed the distribution of number of interactions to be normal and calculated p-value using the Z-test.

3.3.2 Experimental validation – *AXIN2 and hsa-miR-34a*

Luciferase coding sequences were amplified from pGL3 vector (Promega) and inserted between HindIII and BamHI sites of pcDNA3.1-Hyg(+) mammalian expression vector (Invitrogen) to generate luciferase expression construct. To make 3′-UTR constructs, the 3′-UTR of *AXIN2* (NM_004655; +1 ~ +1059) was amplified from genomic DNA of MCF-7 cells and cloned into the BamHI and NotI sites. The synthetic oligonucleotide containing 5′-UTR sequences targeted by *miR-34a* of *AXIN2* (5′-GCC

CGG GGG AGT CGG CTG GAG CCG GCT GCG CTT TGA, corresponding to +44 ~ +79) was inserted into NheI and HindIII sites upstream of luciferase vectors. Each reporter construct (25 ng) was co-transfected with 20 pmol of negative control RNA oligo (Ambion, AM17110) or *miR-34a* precursor RNA oligo (Ambion, product ID PM11030) using Lipofectamine 2000 (Invitrogen) for 48 hrs. In experiments inhibiting endogenous *miR-34a*, 5 ng of each construct was co-transfected with 40 pmol of anti-miR-34a inhibitor (Ambion, product ID AM11030) or anti-miR negative control (Ambion, product ID AM17010). Fold change by *miR-34a* or *miR-34a* inhibitor was measured by a dual-luciferase assay kit (Promega), and the firefly luciferase activity normalized relative to a simultaneously transfected SV40-driven *Renilla* luciferase expression plasmid. Experiments were performed in two sets of triplicates simultaneously, one for reporter gene assay and one for qPCR analysis.

3.3.3 Experimental validation – LIN28 and *lin-4* siRNA

Custom-designed lin-28 UTR sequences **(Appendix B.3)** were purchased from Integrated DNA Technologies, Inc. The expression reporter vector, pMIR-REPORT™, was purchased from Ambion, Inc. (Cat. # AM5795). 5′-UTR sequences were cloned into the BamHI restriction site upstream of the luciferase coding sequence and the 3′-UTR sequences were cloned into the multiple cloning site using HindIII and SpeI. UTR sequences and their orientation in the constructs were confirmed by DNA sequencing (University of Michigan DNA sequencing core).

Strands that make up the lin4msiRNA duplex were purchased from Integrated DNA Technologies, Inc. The single stranded molecules were later annealed using the

manufacturer's protocol. We used hsa-miR-16 (Ambion, Inc., product ID PM10339) as a negative control since there is no interaction site predicted in lin-28 UTRs.

HEK293 cells were grown to 80% confluence in Dulbecco Modified Eagle Medium (DMEM) with 10% Fetal Bovine Serum. Cells were then trypsinized and plated in 12-well plates with about 250,000-300,000 cells per well. 500ng of each firefly reporter construct and 50ng of internal control Renilla reporter pRL-tk (Promega, Cat. # E2241) were co-transfected with either 37 pmol of control miRNA (hsa-miR-16) or 170 pmol of siRNA using Lipofectamine 2000 (Invitrogen). Owing to mismatches in the duplex we used, we increased the siRNA concentration to compensate for any inefficiency in annealing. Transfections were performed in quadruplicate two independent times. Cells were lysed 24 hours post-transfection and assayed for luciferase expression using the Dual-Luciferase Reporter Assay System (Promega, Cat. # E1910) and GloMax® 96 Microplate Luminometer w/Dual injectors (Promega, Cat. # E6521) according to the manufacturer's protocol.

For the lin-28 study, experiments were repeated two independent times in quadruplicate each time. *Renilla* normalized luciferase values were normalized using values from a non-specific miR-16 transfection. To determine if there was significant difference between the pairs (5UpmLuc3Upm and 5UmmLuc3Upm) and (5UpmLuc3Umm and 5UmmLuc3Umm) we used the Wilcoxon rank-sum test to calculate p-values from the normalized luciferase values for each pair of constructs chosen.

3.4 Discussion

Translation repression has been reported to occur when a 3′-UTR target site for endogenous *let-7a* in HeLa cells is moved to the 5′-UTR (Lytle et al. 2007). We now show there exist many endogenous target sites in 5′-UTR for endogenous miRNAs, so that these 5′-UTR sites can contribute to miRNA function. The data in **Fig. 3.1a** is intriguing in that 1) significant miRNA interactions in the 5′-UTR occur only with the 3′-end of miRNA (5U3P), and 2) such 5U3P significance seems to arise in highly-conserved 8-mers and spread into less-conserved but highly-human-present motifs (C=0 and 1). Non-conserved sites have been explored under the assumption that each species or genome might employ them to attribute specificity in some manner (Farh et al. 2005). Considering that the 3′-end of miRNA family members (intra-species) and those of some miRNAs across species differ, the 3′-end of miRNAs may contribute to gene- or species-specific target site recognition of the 5′-UTR. Dividing miRNAs into conserved and non-conserved ones, it seems that human-specific 5U motifs interact with pre-existing miRNAs **(Fig. 3.1b)** and that human-specific miRNAs interact with pre-existing 5U motifs **(Fig. 3.1c)**. The significant 5U5P presence in the highly conserved UTR motifs and non-conserved miRNAs **(Fig. 3.1c)** may reflect an emergent feature of human-specific miRNAs, wherein miRNA and 5′-UTR are actively evolving in response to each other.

We used 36-mer sequences for the *AXIN2* 5′-UTR construct, which interacts mostly with the 3′-end of *miR-34a*. In contrast to 3′-UTR sites, which are well-dispersed across 1,408 nucleotides, making additive miRNA effects possible, the two 5′-UTR sites overlap, leaving no opportunity for additive effects. We expect to see four times higher

3′-UTR effects than with 5′-UTR, assuming the 5′-end represses translation in the 3′-UTR just as the 3′-end does in the 5′-UTR. Therefore, the contribution of *AXIN2* 5′-UTR sites in protein repression by *hsa-miR-34a* induction is no less than that of each site in the 3′-UTR **(Fig. 3.3c)**. Of some interest are the endogenous miRNA effects on both UTRs in this pair **(Fig. 3.3d)**. Not only is the inserted 5′-UTR site effect similar to that of the whole 3′-UTR (about 40 times longer than the inserted 5′-UTR sequences), but the presence of both UTRs has a synergetic effect on miRNA function. Exogenous *hsa-miR-34a* effects on top of endogenous *hsa-miR-34a* function may lead to saturation of repression capacity with 5ULuc3U in **Fig. 3.3c**, while repression of Luc3U is more easily achieved with exogenous *miR-34a*.

In order to fully understand miRNA function, therefore, we advise the insertion of both 5′ and 3′-UTR sequences in miRNA functional experiments, which has rarely been done before. We may see more protein reduction with 5′-UTR inclusion where interaction sites exist as seen in **Fig. 3.3a and 3.4a**. Moreover, this new class of miRNAs and targets may fall into the class of translation blockers prior to the 40S ribosome reaching the translation start region, preventing 60S association (Wang et al. 2008), one possible miRNA mechanism of translation repression.

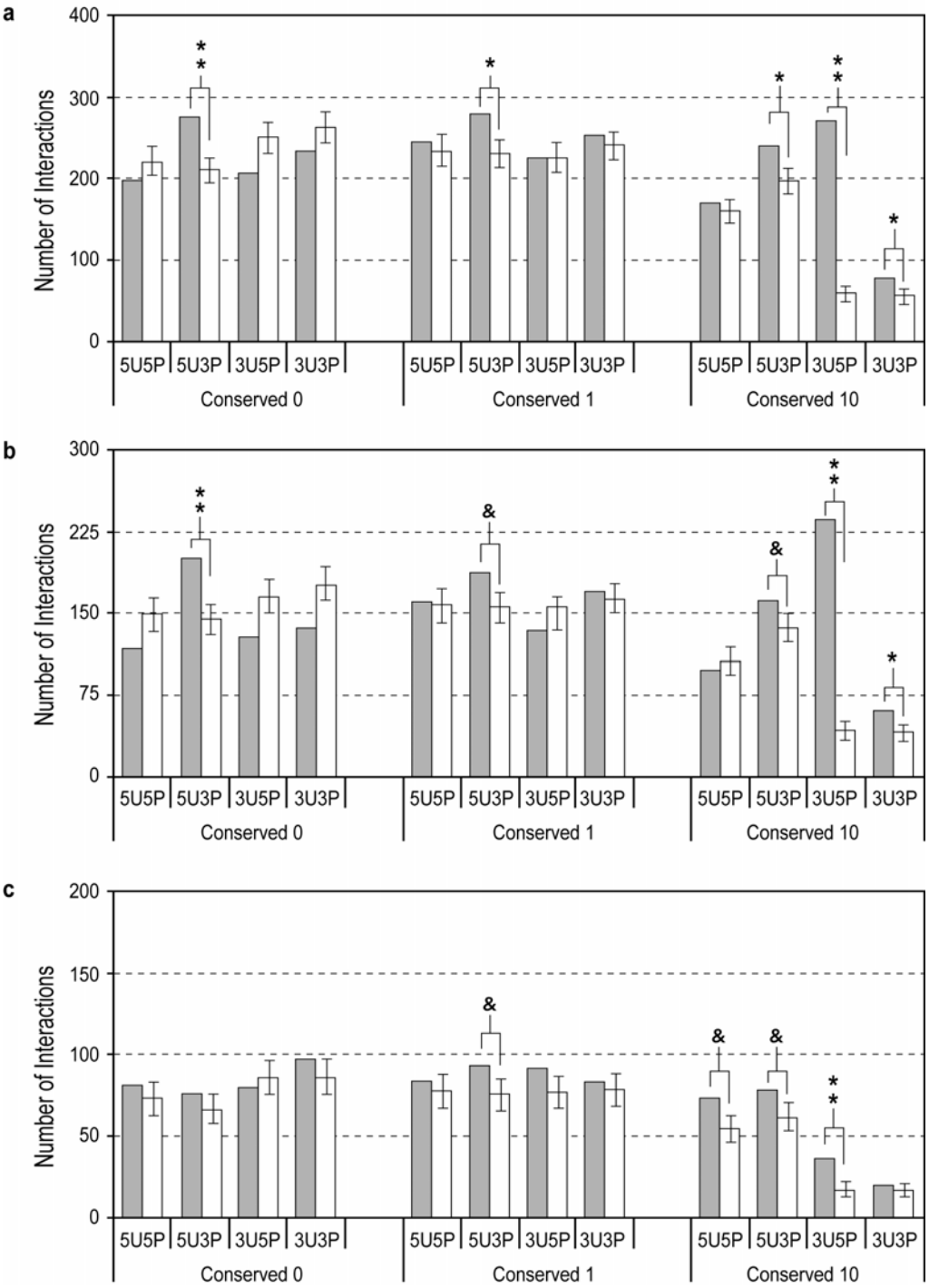


Figure 3.1 Analysis of predicted interactions between 8-mers from different conservation classes and miRNAs (based on number of occurrences). Closed bars indicate number of predicted interactions between 5′-UTR or 3′-UTR 8-mer sequences (indicated by 5U or 3U respectively) and 5′- or 3′- ends (indicated by 5P or 3P respectively) of a full set of mature miRNAs **(a)**, of conserved miRNAs **(b)**, and of non-conserved miRNAs **(c)**. Open bars correspond to mean number of interactions after 1000 shuffling iterations and error bars indicate standard deviations. Double asterisk indicates p<5e-05 and single asterisk p<5e-03

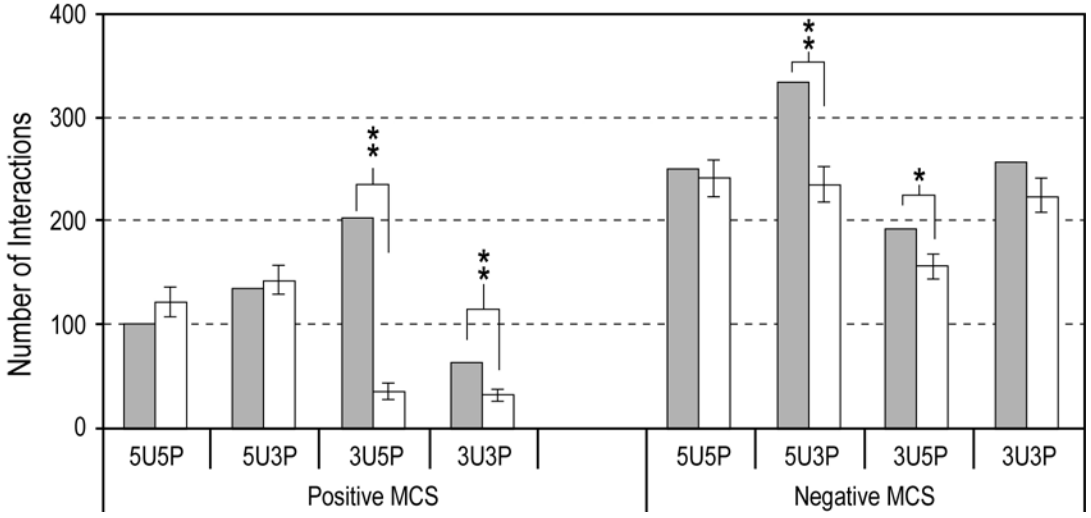


Figure 3.2 Analysis of predicted interactions between 8-mers from different conservation classes and miRNAs (based on conservation score). Closed bars indicate the number of predicted interactions between 5′-UTR or 3′-UTR 8-mers (5U and 3U respectively) and 5′- or 3′-ends of mature human miRNAs (5p and 3p respectively). Open bars indicate mean number of interactions using shuffled 8-mers after 1000 shuffling iterations. Double asterisk indicates p<5e-05 and single asterisk, p<5e-03.

Figure 3.3 Human miRNA *hsa-miR-34a* and target *AXIN2*. **(a)** Predicted interactions between hsa-miR-34a and Axin2 UTR sequences. Extended seed match between the 5′-end of miR-34a and one of the 3′-UTR binding sites is shown in bold red. All predicted 3′-UTR sites are marked in the Supplementary Information. Overlapping interactions between the 3′-end of miR-34a and the 5′-UTR inserted sequences are shown in bold blue. Energy was calculated using RNAhybrid. **(b)** Schematic showing vector constructs containing firefly luciferase reporter gene used in transfection experiments. The 5′-UTR and 3′-UTR inserts are indicated as 5U and 3U respectively. **(c)** Luciferase expression fold change with miR-34a (red bars) normalized with negative control RNA oligo (blue bars). Firefly luciferase protein expression was normalized with Renilla luciferase protein. **(d)** Reporter constructs were co-transfected with anti-miR-34a oligo (red bars, Ambion, product ID, AM11030) and normalized with negative control RNA oligo (blue bars). **(e)** Effect of mutations in the 5′-UTR site – luciferase protein levels when reporter constructs were co-transfected with miR-34a (red bars) or negative control (blue bars). Error bars in panels **(c)** to **(e)** represent standard deviation from triplicate experiments

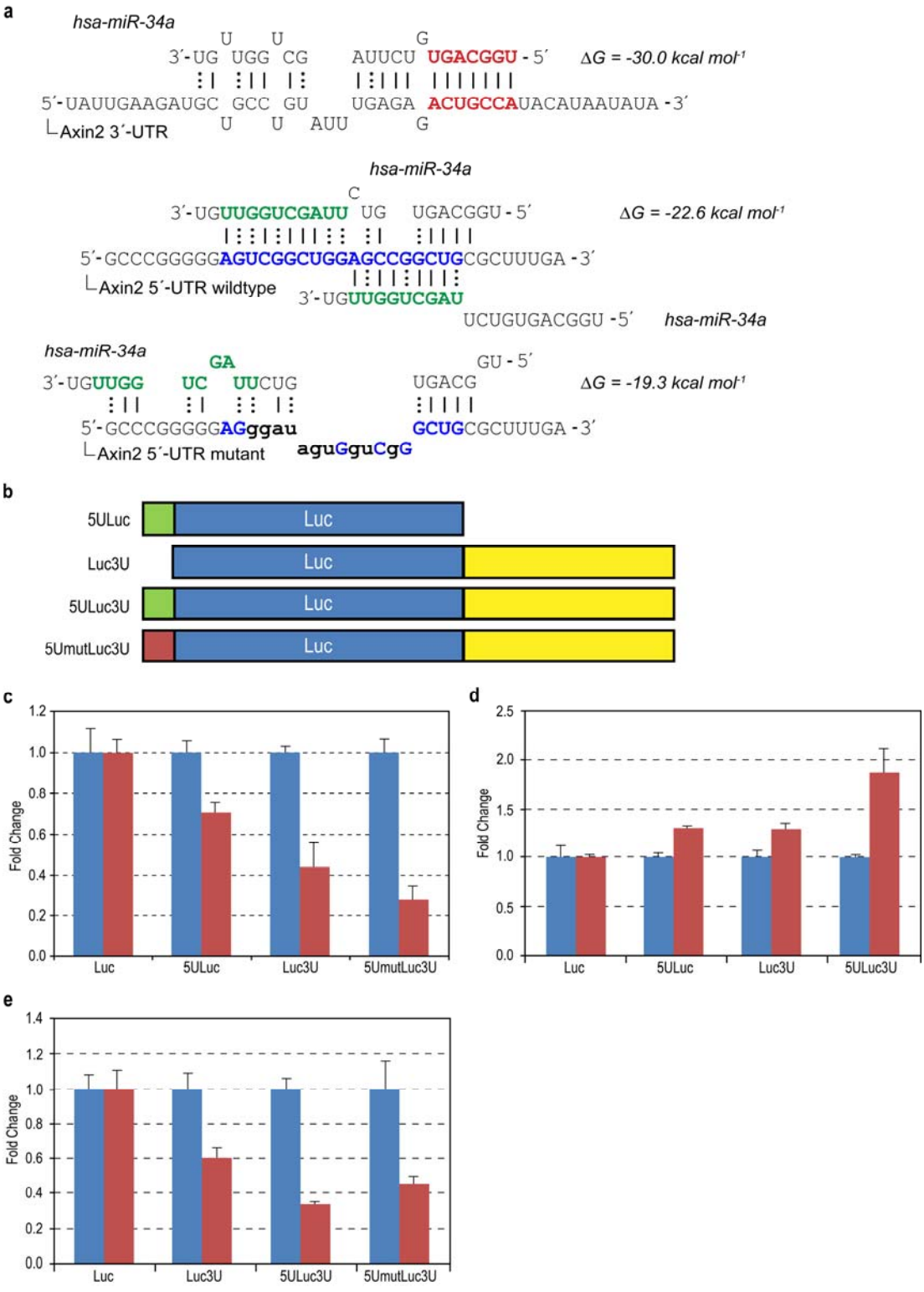
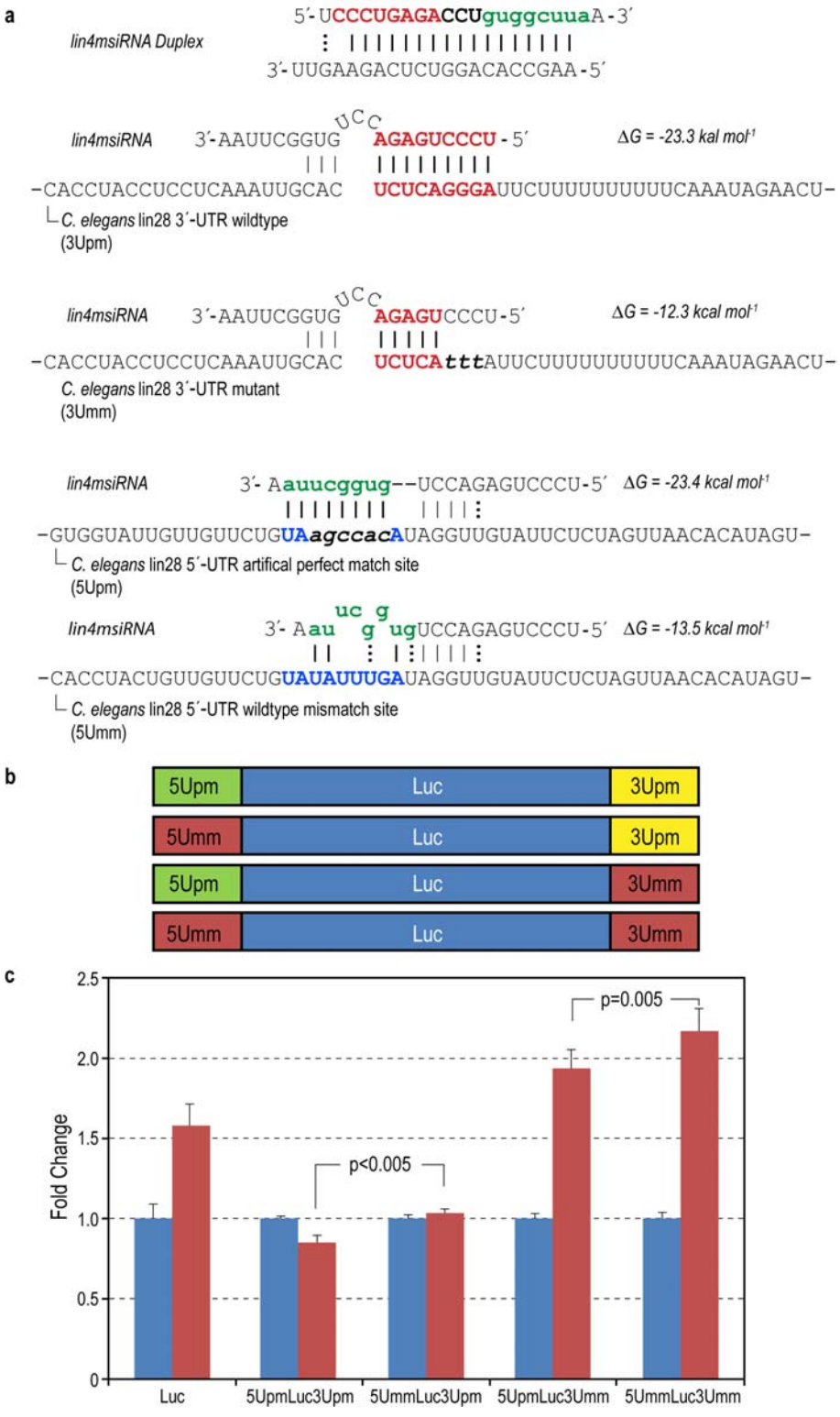

a
hsa-miR-34a
3′-UG UGG CG AUUCU UGACGGU - 5′ ΔG = -30.0 kcal mol-1
5′-UAUUGAAGAUGC GCC GU UGAGA ACUGCCAUACAUAAUAUA - 3′
Axin2 3′-UTR
hsa-miR-34a
3′-UGUUGGUCGAUU UG UGACGGU - 5′ ΔG = -22.6 kcal mol-1
5′-GCCCGGGGGAGUCGGCUGGAGCCGGCUGCGCUUUGA - 3′
Axin2 5′-UTR wildtype
3′-UGUUGGUCGAU
UCUGUGACGGU - 5′ hsa-miR-34a
hsa-miR-34a
3′-UGUUGG UC UUCUG UGACG GU - 5′ ΔG = -19.3 kcal mol-1
5′-GCCCGGGGGAGggau GCUGCGCUUUGA - 3′
Axin2 5′-UTR mutant aguGguCgG
b
5ULuc
Luc3U
5ULuc3U
5UmutLuc3U
Luc
c
Fold Change
Luc
5ULuc
Luc3U
5UmutLuc3U
d
Fold Change
Luc
5ULuc
Luc3U
5ULuc3U
e
Fold Change
Luc
Luc3U
5ULuc3U
5UmutLuc3U

Figure 3.4 Effect of 5′-UTR interaction site for lin4msiRNA on reporter expression levels. **(a)** Predicted interactions between lin4msiRNA and lin28 UTR sequences. The functional strand of the siRNA contains an intact cel-lin-4 seed region (red) while the 3′-end is modified (green). Extended seed match between the 5′-end of lin4msiRNA and the wild-type lin28 3′-UTR binding site (bold red) is disrupted by introducing mismatches (bold and italics) to create an imperfect match with the seed region. The 3′-end of lin4msiRNA is complementary to the artificial lin28 5′-UTR binding site created by introducing a few GC base-pairs (bold italics) to form a perfect match. The wild-type lin28 5′-UTR presents an imperfect match (bold blue). Structure and energy calculations were carried out using RNAhybrid. **(b)** Schematic showing vector constructs containing firefly luciferase reporter gene used in transfection experiments. Perfectly matched sites on the lin-28 5′-UTR and 3′-UTR segments are indicated as 5Upm and 3Upm respectively, and imperfectly matched sites are indicated as 5Umm and 3Umm respectively. **(c)** Fold changes of Renilla normalized firefly luciferase expression levels with respect to non-specific hsa-miR-16 (blue bars) upon treatment with lin4msiRNA (red bars). Error bars represent standard deviation recorded from 8 pooled replicates.


a
lin4msiRNA Duplex
5′-UCCCUGAGACCUguggcuuaA-3′
3′-UUGAAGACUCUGGACACCGAA-5′
lin4msiRNA
3′-AAUUCGGUG
AGAGUCCCU-5′
ΔG = -23.3 kal mol-1
-CACCUACCUCCUCAAAUUGCAC
UCUCAGGGAUUCUUUUUUUUUUUCAAAUAGAACU-
C. elegans lin28 3′-UTR wildtype
(3Upm)
lin4msiRNA
3′-AAUUCGGUG
AGAGUCCCU-5′
ΔG = -12.3 kcal mol-1
-CACCUACCUCCUCAAAUUGCAC
UCUCAtttAUUCUUUUUUUUUUUCAAAUAGAACU-
C. elegans lin28 3′-UTR mutant
(3Umm)
lin4msiRNA
3′- Aauucggug--UCCAGAGUCCCU-5′
ΔG = -23.4 kcal mol-1
-GUGGUAUUGUUGUUCUGUAagccacAUAGGUUGUAUUCUCUAGUUAACACAUAGU-
C. elegans lin28 5′-UTR artifical perfect match site
(5Upm)
lin4msiRNA
3′- Aau g ugUCCAGAGUCCCU-5′
ΔG = -13.5 kcal mol-1
-CACCUACUGUUGUUCUGUAUAUUUGAUAGGUUGUAUUCUCUAGUUAACACAUAGU-
C. elegans lin28 5′-UTR wildtype mismatch site
(5Umm)
b
5Upm
Luc
3Upm
5Umm
Luc
3Upm
5Upm
Luc
3Umm
5Umm
Luc
3Umm
c
Fold Change
2.5
2.0
1.5
1.0
0.5
0
p=0.005
p<0.005
Luc
5UpmLuc3Upm
5UmmLuc3Upm
5UpmLuc3Umm
5UmmLuc3Umm

Chapter 4

Post-transcriptional regulation by miRNA binding of uAUGs

4.1 Background

Translation initiation in eukaryotes is postulated to follow the ribosome scanning model (Kozak 2002), possibly constrained by multiple cis-elements on the 5′-UTR such as secondary structure (Kozak 1991b) and the 5′-terminal oligopyrimidine tracts (Avni et al. 1997) and upstream AUG (uAUG) nucleotides (Iacono et al. 2005). It is known that uAUGs cause a reduction in translational efficiency, therefore acting as a strong negative regulator of gene expression (Kozak 2002). Comparative genomic analysis has revealed that uAUGs are conserved in mammalian 5′-UTRs to a greater extent than in other segments of mRNAs, genes harboring them mainly coding for transcription factors (Churbanov et al. 2005). uAUGs may form alternative start sites forming upstream open reading frames (uORF), which are known to reduce efficiency of translation, possibly by translation of the uORF-encoded peptide (Morris and Geballe 2000). It has been noted that an uAUG/uORF can inhibit translation independent of a downstream secondary structure or its position relative to other uAUGs before the main ORF (Imataka et al. 1994; Jin et al. 2003).

Unlike the start codon of the main ORF, which in good initiation context is typically identified by the consensus Kozak sequence (Kozak 1991a), many of the uAUGs are in sub-optimal context for translation (Iacono et al. 2005). Some groups have been able to assay for *in vitro*-translated uORFs (Wang and Wessler 1998; Raney et al. 2000), which are not, however, readily detectable unless fused to a reporter gene (Kwon

et al. 2001; Song et al. 2007). One study showed that translation repression is not dependent on the encoded peptide sequence (Wang and Wessler 1998), which suggests that the peptide action may be non-specific. Further, Kwon *et al.* demonstrated that addition of a synthetic peptide encoded by an uORF did not alter translation of the protein-coding gene even though the uORF on the 5′-UTR was able to repress translation (Kwon et al. 2001).

Moreover, previous studies have reported that the uAUGs' effect on translation repression is specific to tissue type: though mRNAs containing uAUGs are expressed ubiquitously, the proteins are expressed only in specific tissues (Imataka et al. 1994; Nikolcheva et al. 2002). If indeed the translation of uORF limits downstream ORF translation, why does this repression occur only in certain cell-lines and tissues? There appears to be an additional mechanism of translation repression through uAUG other than upstream-encoded peptides.

In this report we identify certain miRNA interactions specific to the uAUG, preferentially through the 3'-end of the mature miRNA sequence. Based on our findings, we hypothesize that miRNAs expressed in one cell type but not in others may account for differences in protein expression in the cell types without changes in mRNA levels. Using miRNA expression data and results from prior work done with the *KLF9* gene in HeLa and N2A cells, we demonstrate the validity of our hypothesis. Our results suggest the role of miRNAs in cases where uAUG confers tissue-specific protein expression of the target mRNA.

.

4.2 Methods

4.2.1 uAUG and miRNA sequence data

Pairwise alignments between 5′-UTRs of mammalian human and mouse cDNAs were downloaded from the ftp site listed in Churbanov *et al.*(Churbanov et al. 2005). From each alignment we extracted 11-mer uAUG sequences in the human 5′-UTR beginning at position -4 and ending at position +7, with the 'A' being designated as +1 (e.g. NNNNAUGNNNN, where N is any nucleotide). When fewer than four nucleotides surround the uAUG, as in the case of a start of end of an alignment truncated n-mers ranging from 7 to 10 nucleotides in length were considered. Only uAUG sequences sharing 100% identity with the mouse homolog were categorized as conserved while others were considered as non-conserved uAUGs. Experimentally characterized uAUG sequences in Table 3 were obtained from the references listed in Table 2. For the *KLF* family of genes in Table 4, uAUG sequences were extracted from the 5′-UTR portions of the full RefSeq mRNA.

For the motif analysis, mature miRNA sequences were downloaded from miRBase (version 11.0) (Griffiths-Jones et al. 2008). miRNAs present in at least one other species (e.g. *hsa-let-7d* and *mmu-let-7d*) were categorized as conserved miRNAs (471 in total), and others as non-conserved miRNAs (206 in total). miRNAs were then split into their 5′- and 3′-halves to check for any preferential interaction with one end or the other.

4.2.2 Sequence complementarity search

A two-step strategy was employed in looking for matches between uAUG 11-mers and miRNA sequences. First, the thermodynamic search program RNAhybrid (Rehmsmeier et al. 2004) was used with –e option (ΔG) set to $\leq$ -14 kcal mol^{-1}**.** Next, hits with at least seven consecutive nucleotide matches were selected.

As control miRNAs were shuffled in order to keep the nucleotide composition of the sequences intact. The search strategy above was repeated over 1000 shuffling iterations and the average number of interactions was calculated. The resulting distribution of number of interactions was assumed to be normal and significance calculated using a Z-test.

4.2.3 miRNA expression data

For miRNAs from Landgraf *et al.*'s study (Landgraf et al. 2007), we used their web visualization tool to assess the presence or absence of miRNAs in a given cell-line. For data from Chen *et al.*'s study (Chen et al. 2008), we used a p-value cutoff of 0.01 to report the miRNA as expressed. We obtained expression evidence for miRNAs of interest in N2A cells from Hohjoh *et al.*'s (Hohjoh and Fukushima 2007) study through personal communication. Expression data from Lawrie *et al*.'s (Lawrie et al. 2008) and Takada *et al*.'s studies (Takada et al. 2006) were obtained directly from the manuscripts and supplementary information.

4.3 Results

4.3.1 uAUGs are potential miRNA target sites

An earlier study of excess conservation of uAUGs used a total of 1955 pairwise alignments of human and mouse 5′-UTR sequences (Churbanov et al. 2005). The authors generated the alignments after careful pre-processing steps to remove any coding sequences that may have been mis-annotated as leader sequences. We used this alignment data to compile sequences containing uAUGs from human 5′-UTRs (see Methods), generating a total of 4009 uAUG 11-mers. The number of uAUGs per 5′-UTR ranges from one to 20, with 68% of the 1955 human 5′-UTRs containing at most two **(Fig. 4.1A)**. In order to investigate conservation patterns of these n-mers we separated them into 2935 conserved and 1074 non-conserved sequences. The uAUG sequences appear to be highly conserved between both human and mouse UTRs, with all 7-mers having 100% identities and roughly 70% of 11-mers being conserved **(Fig. 4.1B)**.

Mature human miRNA sequences (miRBase, version 11.0) (Griffiths-Jones et al. 2008) were downloaded and categorized as conserved (471 sequences) or non-conserved (206 sequences) miRNAs (see Methods). To reveal preferential interaction with any portion of the miRNA we split each sequence into its 5′- and 3′-ends, the former containing the seed region. We then looked for sequence matches between miRNA ends and the uAUG-containing sequences generated. This was done in two steps: 1) a thermodynamics-based search using RNAhybrid (Rehmsmeier et al. 2004) with a ΔG cutoff $\leq$ -14 kcal mol^{-1} followed by 2) a filter step to look for 7 or more consecutive matches with zero or one GU wobbles. To control for spurious hits, the number of

interacting pairs was compared to the number obtained after shuffling the mature miRNAs sequences and repeating the search procedure.

We observed many predicted interactions between uAUG sequences and the two miRNA ends, characterized by a dependency on conservation of miRNAs. Only conserved miRNAs showed a significant number of interactions while non-conserved miRNAs were no better than their shuffled cohorts **(Fig 4.2A and 4.2B)**. There were a number of 8-mer Watson-Crick complementary matches between the 5′-ends of conserved miRNAs and uAUG sequences **(Fig 4.2A)**. Interestingly, there seemed to be a greater number of such interactions at the 3′-ends **(Fig. 4.2A and Table 4.1)**, which suggests a preference for pairing between uAUGs and 3′-ends. A previous study also reported observations wherein 5′-UTR and coding regions participate in binding the 3′-end of the highly conserved miRNA, *let-7* (Forman et al. 2008). Further, when we included at most one GU wobble the only significant result that persisted was the interaction with the 3′-ends of conserved miRNAs **(Fig. 4.2B)**. We conducted a genome-wide motif study of 5′-UTRs and 3′-UTRs and observed a similar propensity for interaction between 5′-UTRs and 3′-ends of miRNAs (unpublished data). The preference for interaction with 3′-ends, perhaps, suggests the importance of non-seed region matches on the 5′-UTR. This may explain the fact that there are very few known endogenous targets on the 5′-UTR that exhibit seed-matches (Xie et al. 2005). We conducted a brief GO-term investigation into the nature of genes containing the uAUGs in **Table 4.1**. Out of a total 1071 genes that contained these uAUGs annotations were retrieved for 678 genes. Majority of these 678 were found to be involved in transcription factor activity (Supplementary data).

Considering that nearly 75% of the 11-mers were found to be conserved between human and mouse 5′-UTRs (2935 out of 4009) we investigated if the interactions with conserved miRNAs were a function of uAUG sequence conservation. Results showed no dependence on uAUG conservation when not allowing GU wobbles **(Fig. 4.2C)**. However, when allowing at most one GU wobble only conserved uAUGs exhibited significant interactions with 3′-ends of miRNAs **(Fig. 4.2D)**.

The above results indicate that uAUGs may participate in highly sequence-specific Watson-Crick base-pairing with miRNAs, particularly towards the 3′-ends. The fact that inclusion of a GU wobble still resulted in significant number of interactions between the 3′-ends and uAUGs probably suggests functionality.

4.3.2 Expressed miRNAs may bind endogenous uAUG sites

The analyses that follow are based on experiments with genes that contain uAUGs in their 5′-UTRs, drawing upon sequence data and results from previous experiments that attribute translational repression to the uAUGs. We also used miRNA expression evidence from several sources - these references are consolidated in the form of meta-data **(Table 4.2)**. We extracted 11-mer sequences containing uAUGs for these genes and looked for interactions with conserved miRNAs using the search strategy outlined above. Based on the observations in **Fig. 4.2A** and **4.2B**, we allowed one GU wobble for interactions with the 3′-end and none with the 5′-end. Many of genes contain multiple uAUGs/uORFs that have different inhibitory effects on translation. We assigned discrete values to these uAUGs that reflect their repressive capabilities on the downstream reporter. These were obtained by comparing the effect of uAUG on reporter expression

with a construct used as control. The values range from 1x to 6x, where 1x indicates that the uAUG is least repressive or does not show any effect.

We not only observed complementary matches with conserved miRNA sequences but also confirmed the presence of many of the predicted miRNAs in cell-lines where repression was observed **(Table 4.3)**. There also appears to be an association between repressive strength of uAUGs and miRNA target predictions. Two uAUGs that have little or no effect on repression, indicated by '1x' in **Table 4.3**, lack miRNA interaction sites. Conversely, uAUGs with strong repressive potential (2x-6x) are complementary to expressed miRNAs except in the case of the first uAUG in the ADH5/FDH gene where expressions of the predicted miRNAs have not been detected. Note that miRNAs can act in a combinatorial manner on uAUGs to produce a net repressive effect. These observations suggest that some of the uAUG sequences are miRNA-specific and functional target sites.

4.3.3 KLF genes are probable 5′-UTR miRNA targets

Kruppel-like factors (KLFs) are transcriptional regulators that contain a characteristic zinc-finger domain and are known to play a role in differentiation and other cellular events (Bieker 2001; Black et al. 2001). There are as many as 15 members in this family, seven of them containing at least one uAUG. Using the criteria set above we identified 7-mer matches between uAUG-containing sequences and miRNAs in all seven of these genes **(Table 4.4)**. Two of these, *KLF9* and *KLF13*, also called *BTEB1* and *RFLAT-1* respectively, are known to be translationally regulated by uAUGs in their 5′-UTRs (Imataka et al. 1994; Nikolcheva et al. 2002). The uAUGs in these two genes have been

implicated in cell-specific control of protein expression though their respective transcripts are present in many other tissues, suggesting a post-transcriptional mechanism of gene regulation (Imataka et al. 1994; Nikolcheva et al. 2002).

Specifically, protein expression of *KLF9*, whose 5′-UTR contains 10 uAUGs, is limited to brain tissue though its mRNA is expressed ubiquitously (Imataka et al. 1994). The 5′-UTR, particularly the portion containing uAUGs 6 and 7, suppressed reporter gene translation in HeLa cells but not in mouse neuroblastoma (N2A) cells (Imataka et al. 1994). This observation was even more intriguing because peptides from the two uORFs starting from uAUG6 and uAUG7 have not been detected (Imataka et al. 1994). Similarly, though *KLF13* mRNA is expressed in multiple tissues, protein expression was only detected in adult spleen and lung tissues (Song et al. 2000). While *KLF13* mRNA levels are constant throughout T-cell activation, KLF13 protein is only expressed later on in the activation process (Nikolcheva et al. 2002). The presence of several uAUGs in its 5′-UTR down-regulated translation of the reporter gene in Jurkat T-cells and, to a lesser degree, in HEK293 cells (Nikolcheva et al. 2002).

We decided to focus our analysis on *KLF9* uAUGs since the effects of wild-type and mutant constructs used to elucidate the roles of uAUGs were demonstrated in both cell-lines relevant to tissue specificity. We extracted uAUG 11-mers from the *KLF9* 5′-UTR sequence used in the experimental study (Imataka et al. 1994) and searched for interactions with both ends of conserved miRNAs. Since the 5′-UTR study for *KLF9* was also done in the mouse neuroblastoma (N2A) cell line, we used both mouse and human miRNAs in the analysis. All uAUGs except uAUG5 and uAUG8 interacted with at least one miRNA (Table 5). The ninth uAUG was predicted to interact with as many as five

miRNAs. Most of these predicted miRNAs are expressed in HeLa cells but not in N2A cells, including those that match uAUG6 and uAUG7. Only *mmu-miR-16* and *mmu-miR-543* were detected in N2A cells.

Regulatory roles of each uAUG/uORF may be studied by mutating one or more of the uAUGs to mitigate repression. In the case of *KLF9*, mutation of uAUG6 or 7 or both relieved translation repression (Imataka et al. 1994). However, uAUG6 inhibits translation to a greater extent compared to uAUG7, the translation efficiency of the uAUG6 mutant construct being 5 times that of the wild-type construct compared to a two-fold increase for the uAUG7 mutant, based on Imataka *et al*.'s figure 7 (Imataka et al. 1994). Interestingly, there are five human miRNAs that are predicted to interact with uAUG6, of which two are expressed in the HeLa cell lines and none in N2A cells **(Table 4.5 and Appendix C.3)**. Only one expressed miRNA, *hsa-miR-31*, is predicted to bind uAUG7. If these two uAUGs are indeed miRNA interaction sites, their mutation should presumably eliminate interactions with the miRNAs predicted in **Table 4.5**. To test this assumption we repeated the interaction-expression analysis using mutated uAUG sequences that had been shown to relieve translational repression. When mutated, uAUGs implicated in mediation of translation repression in *KLF9* showed fewer predicted interactions with miRNAs (**Table 4.5**, m6 and m7) compared to wild-type sequences. Moreover, there was little evidence for expression of miRNAs matching mutated uAUG sequences.

4.4 Discussion

Though uAUGs are known to act in post-transcriptional control of gene expression there is no clear account of the mechanism involved when differences in activity of uAUGs exist across cell or tissue types. While studying uAUGs and miRNAs independent of one another, researchers observed that uAUGs affect gene expression by changes to protein levels without a notable change in mRNA levels, a phenomenon that is also a characteristic of one of the mechanisms of miRNA-mediated gene regulation.

Target sites for miRNAs have conventionally been thought to reside on conserved regions of the 3′-UTR and are predicted to bind the seed-region of a miRNA (Lewis et al. 2005). Employing a combination of thermodynamic and sequence-based searches, we found many potential uAUG 5′-UTR sites that are predicted to preferentially bind to the 3′-ends of conserved miRNAs compared to seed regions (or 5′-ends), both ends showing a significant number of interactions. This is in sharp contrast to results which show a lack of an appreciable seed-matches on 5′-UTRs (Xie et al. 2005). Forman *et al.* have also shown *in silico* that a well-conserved miRNA, *let-7*, is predicted to base-pair with the 5′-UTRs through remainder of the miRNA apart from the seed portion (Forman et al. 2008). Based on these observations, we hypothesized that the overlap in function may arise from underlying sequence-specific interactions.

Examining many genes where uAUGs have regulatory properties, we demonstrate here the connection between uAUG-mediated repression and their likelihood as binding sites for conserved miRNAs. miRNA expression data support this link by confirming the presence of miRNAs in cell-lines where reporter translation is affected by uAUGs. Further, we predict that many uAUGs in the *KLF* family of genes are miRNA-binding

sites. Two uAUGs in the well-studied *KLF9* are proven down-regulators of protein expression with regulation observed only in HeLa cells. Many miRNAs likely to interact with these two sequences were found to be expressed in the HeLa and not in N2A cells where regulation was not observed.

Many genes that contain uAUGs are known to be transcription factors (Churbanov et al. 2005). In a very interesting recent report several miRNAs and transcription factors in *C. elegans* were shown to be involved in feedback circuits (Martinez et al. 2008). It is possible that the miRNAs in this study utilize seed matches on the 3′-UTRs while other miRNAs (not necessarily activated by the transcription factors) may bind the 5′-UTRs through either seed or non-seed region matches as means to achieve repression. For instance, we found several uAUGs on the 5′-UTRs of *LIN26* that were predicted to bind miRNAs other than the *miR-43* identified by the authors (data not shown).

Orom *et al.* showed that *miR-10a* binds sequences downstream of a 5′-oligopyrimidine tract (5′-TOP) on *RPS16*, a gene encoding a ribosomal protein, to regulate its translation (Orom et al. 2008). This exact binding site on the 5′-UTR was earlier shown to be responsible for conferring cell-specific translational regulation (Avni et al. 1997). Taken together with these findings, our results suggest that miRNAs can interact with uAUG sequences and confer tissue specificity. This would constitute a unifying mechanism of translation repression for miRNAs and uAUGs. We specifically propose that the interaction of miRNAs with uAUGs may impede the progress of the scanning 40S ribosome subunit. Interestingly, primer extension (toeprint) analysis reveals the presence of a 40S ribosomal subunit alone at the start codon on miRNA-repressed

mRNAs (Wang et al. 2008). The same technique also reveals stalling of ribosomes in the vicinity of uAUGs (Gaba et al. 2001; Kwon et al. 2001; Song et al. 2007). Furthermore, Ago2, a member of the Argonaute family of proteins (Peters and Meister 2007; Tolia and Joshua-Tor 2007) and a component of the functional micro-ribonucleoprotein (miRNP) complex, was found to co-sediment with 40S-containing complexes (Wang et al. 2008). These facts indicate that miRNAs associated with miRNPs may recognize uAUG sequences as target sites and prevent translation.

In this chapter we have presented observations that suggest a miRNA role in translational control by uAUG cis-elements on the 5′-UTR. Specifically, we identified many interactions between uAUG sequences and conserved miRNAs to suggest a sequence-specific binding mechanism between these post-transcriptional regulatory factors. We also presented evidence to show that miRNAs possibly to bind uAUGs that inhibit translation of downstream reporters in cells where the miRNAs are expressed, thus explaining differential control. This expands the range of probable miRNA targets to include many endogenous sites on the 5′-UTR.

Our current knowledge has limited us to think of miRNAs and uAUGs as distinct regulatory mechanisms. While distinct functions of miRNAs or uAUGs remain in other contexts, our study unifies them as a single translational repression phenomenon where uAUGs act as miRNA target sites and translation is hindered

* The 46 miRNAs represent conserved miRNAs
§ Only the portion of uAUG11-mer that interacts with the 3′-end of miRNAs without a GU wobble is presented

Table 4.1 MicroRNAs predicted to interact with uAUG-containing motifs

miRNA*	uAUG-containing motifs[§]
hsa-let-7d	AACUAUG, ACUAUGCAA, CUAUGCAAC
hsa-miR-130a/b	AUGCCCU
hsa-miR-132	GACCAUGGCU
hsa-miR-146a	ACCCAUGG, CCCAUGGAA
hsa-miR-146b-5p	GCCUAUGG, CCUAUGGAA
hsa-miR-194	CCACAUGGA, ACAUGGAG
hsa-miR-199a-3p	ACCAAUGUG
hsa-miR-202	UCCCAUGC, CCCAUGCC
hsa-miR-219-2-3p	ACAGAUGU, CAGAUGUCC, AGAUGUCCA
hsa-miR-297	GCACAUGC
hsa-miR-299-5p	AUGUAUGUGGG
hsa-miR-31	GCUAUGCCA, CUAUGCCAG
hsa-miR-324-5p	ACCAAUGCC, CAAUGCCC
hsa-miR-33a/b	GCAAUGCA, CAAUGCAA, AUGCAAC
hsa-miR-34b	AUGGCAG
hsa-miR-363	ACAGAUGGA, AGAUGGAU, CAGAUGGAU, GAUGGAU
hsa-miR-376b	AACAUGGAUU
hsa-miR-380	AAGAUGUGG, AGAUGUGGA, GAUGUGGA
hsa-miR-431	GCAUGACG, CAUGACGG
hsa-miR-432	CCCAAUGA, CCAAUGAC
hsa-miR-448	AUGGGAC
hsa-miR-450b-3p	AUGGAUGCA, GGAUGCAA
hsa-miR-455-3p	GUAUAUGC, AUAUGCC
hsa-miR-455-5p	CGAUGUAG, GAUGUAGU
hsa-miR-487a	CUGGAUGUC
hsa-miR-487b	GUGGAUGA, UGGAUGAC
hsa-miR-490-3p	CAGCAUGGAG, AGCAUGGAGU
hsa-miR-491-5p	CCUCAUGGAAG
hsa-miR-513b	AUAAAUGACA, AUGACAC
hsa-miR-556-3p	AAAGAUGAGC, AGAUGAGCU
hsa-miR-562	GCAAAUGGU
hsa-miR-580	CCUAAUGA, AUGAUUC
hsa-miR-583	UAAUGGGA, AAUGGGAC
hsa-miR-598	GACGAUGAC, ACGAUGACA
hsa-miR-609	AGAGAUGAG, GAGAUGAGA
hsa-miR-654-3p	GGUGAUGGU
hsa-miR-654-5p	GCACAUG, ACAUGUUCU
hsa-miR-767-3p	AACCAUGGG
hsa-miR-802	AAGGAUGAAU
hsa-miR-887	CGGGAUGG
hsa-miR-889	AAUGGUUG
hsa-miR-890	ACUGAUGC, CUGAUGCC
hsa-miR-942	CACAUGGCC, ACAUGGCCA
hsa-miR-944	UCCGAUG

Table 4.2 Genes used in uAUG-binding sequence analysis along with references

Gene	Evidence showing translational control by uAUG	miRNA expression evidence(s) used for analysis
KLF9/BTEB1	(Imataka et al. 1994)	(Hohjoh and Fukushima 2007; Landgraf et al. 2007; Chen et al. 2008)
KLF13/RFLAT-1	(Nikolcheva et al. 2002)	(Takada et al. 2006; Landgraf et al. 2007; Chen et al. 2008; Lawrie et al. 2008)
MOR	(Song et al. 2007)	(Landgraf et al. 2007; Chen et al. 2008)
CHOP	(Jousse et al. 2001)	(Landgraf et al. 2007; Chen et al. 2008)
MDM2	(Jin et al. 2003)	(Landgraf et al. 2007; Chen et al. 2008)
ADH5/FDH	(Kwon et al. 2001)	(Landgraf et al. 2007; Chen et al. 2008)

***** Evidence for expression of miRNAs in mouse N2A cells was acquired through personal communication with authors of (Hohjoh and Fukushima 2007).

† uAUGs shown in caps.
# uAUG not present in the GenBank entry but used in reporter constructs (Nikolcheva et al. 2002).
§ Numbers in parentheses indicate the miRNA end predicted to interact. miRNAs in italics indicate matches with one GU wobble.
‡ Reference for evidence of expression: 1 (Chen et al. 2008), 2 (Landgraf et al. 2007), 3 (Lawrie et al. 2008), and 4 (Takada et al. 2006).
*****Expression of the 5p arm of the precursor was detected, but that of 3p was not checked.

Table 4.3 Genes containing uAUGs predicted to interact with expressed miRNAs

Gene	uAUG[†]		Cell line used in experiments	miRNAs predicted to interact [§]	miRNA expression in cell-lines tested[‡]
MOR	1	gcccAUGcucc (1x)	HEK293	*hsa-miR-146a (3')* hsa-miR-202 (3′)	--- ---
	2	ggggAUGcuaa (2x)		hsa-miR-324-5p (5′) hsa-miR-517b (5′)	1 1
	3	aaggAUGcgcc (3x)		*hsa-miR-323-5p (3')* hsa-miR-324-5p (5′) *hsa-miR-450b-3p (3')*	--- 2 ---
CHOP	1	tatcAUGuuaa (1x)	HeLa	None	---
	2	aaagAUGagcg (6x)		hsa-miR-574-3p (5′) hsa-miR-556-3p (3′)	1,2 ---
	3	gcagAUGugcu (2x)		hsa-miR-219-2-3p (3′)	219-5p*
MDM2	1	aaagAUGgagc (3x)	HeLa	hsa-miR-363 (3′)	1
	2	tggaAUGaucc (1x)		None	---
ADH5/FDH	1	gcccAUGccuc (4x)	HeLa	*hsa-miR-146a (3')* hsa-miR-202 (3′)	--- ---
	2	ccggAUGucag (4x)		*hsa-miR-219-1-3p (3')* *hsa-miR-219-2-3p (3')* hsa-miR-487a (3′) hsa-miR-489 (5′)	219-5p* 219-5p* --- ---
KLF13	1	cacaAUGcgcg[#]	Jurkat	*hsa-miR-323-5p (3')* hsa-miR-103 (5′) hsa-miR-107 (5′) hsa-miR-33a (5′) hsa-miR-586 (5′)	--- 2 2 3,4 ---
	2	ccccAUGcgcu		hsa-miR-202 (3′)	---
	3	gcggAUGcgcg		*hsa-miR-450b-3p (3')* hsa-miR-324-5p (5′)	--- 2,3

Table 4.4 uAUGs from members of the KLF family predicted to interact with conserved miRNAs

KLF Gene[§]	uAUG[†]		miRNAs predicted to interact [‡]
KLF6 (NM_001300)	1	uugcAUGaaac	*hsa-miR-93 (3′)*
KLF7 (NM_003709)	1	cuggAUGccuc	hsa-miR-450b-3p (3′) hsa-miR-487a (3′)
	2	cuggAUGucug	*hsa-miR-450b-3p (3′)* hsa-miR-487a (3′)
KLF8 (NM_007250)	1	cucuAUGauuc	hsa-miR-376a (5′) hsa-miR-376b (5′) hsa-miR-376c (5′)
	2	cuuuAUGuuca	None
	3	gaggAUGggug	*hsa-miR-331-3p (3′)* *hsa-miR-363 (3′)* *hsa-miR-802 (3′)* *hsa-miR-99b (5′)*
	4	uuggAUGcuug	hsa-miR-450b-3p (3′)
	5	cgcuAUGucag	*hsa-miR-31 (3′)*
	6	cagaAUGgggc	*hsa-miR-448 (3′)* *hsa-miR-583 (3′)* *hsa-miR-136 (5′)*
	7	gagtAUGagcc	hsa-miR-767-3p (5′)
	8	cggcAUGaguu	hsa-miR-574-3p (5′)
KLF10 (NM_001032282, isoform a)	1	gauuAUGcaau	hsa-let-7d (3′) hsa-miR-153 (5′)
	2	agcaAUGgcuc	hsa-miR-160 (5′)
	3	caucAUGcauu	None
	4	aagaAUGuuuu	None
	5	uuuaAUGgaaa	None
KLF12 (NM_007249)	1	aucaAUGugac	hsa-miR-199a-3p (3′) hsa-miR-23a (5′) hsa-miR-23b (5′)
	2	acaaAUGgaug	hsa-miR-136 (5′)
	3	auggAUGaaug	hsa-miR-450b-3p (3′) hsa-miR-487b (3′) hsa-miR-802 (3′)
	4	augaAUGaaua	None

[§] KLF13 and KLF9 are presented along with miRNA expression data in Table 4.3 and 4.5, respectively.
[†] uAUGs are shown in caps.
[‡] Numbers in parentheses indicate the miRNA end predicted to interact. miRNAs in italics indicate matches with one GU wobble.

§ uAUG shown in caps, mutated sequences prefixed with letter ‘m’, and mutated positions shown in bold.
† Three letter species codes (hsa/mmu) are indicated only when one sequence interacts and omitted if both interact. Numbers in parentheses indicate the miRNA end predicted to interact. miRNAs in italics indicate matches with one GU wobble.
‡ Reference for evidence of expression: 1 (Chen et al. 2008), 2 (Landgraf et al. 2007), 3 (Hohjoh and Fukushima 2007)

Table 4.5 KLF9 uAUGs predicted to interact with miRNAs in HeLa cells

uAUG[§]		miRNAs predicted to interact [†]	miRNA expressed in cell-lines tested?[‡]	
			HeLa	N2A
1	cauaAUGgggu	hsa-miR-583 (3′) *hsa-miR-490-3p (3′)* *mmu-miR-490 (3′)*	1 --- ---	--- --- ---
2	aaagAUGuguc	miR-380 (3′) *hsa-miR-576-5p (3′)*	1 1	--- ---
3	gccaAUGccag	*miR-16 (3′)* hsa-miR-31 (3′) miR-324-5p (3′)	1,2 1,2 1,2	2,3 --- ---
4	aaagAUGuguc	miR-380 (3′) *hsa-miR-576-5p (3′)*	1 1	--- ---
5	uuaaAUGucag	None	---	---
6	cgugAUGggau	*miR-448 (3′)* *hsa-miR-583 (3′)* *hsa-miR-609 (3′)* miR-654-3p (3′) hsa-miR-605 (5′) *mmu-miR-325 (3′)*	--- 1 1 --- --- ---	--- --- --- --- --- ---
m6	cgugAAGggau	hsa-miR-491-3p (3′) miR-188-5p (5′) *hsa-miR-211 (3′)* *hsa-miR-520h (3′)*	--- 1 --- ---	--- --- --- ---
7	gagaAUGccgg	*hsa-miR-31 (3′)*	1,2	---
m7	gagaAAGccgg	None	---	---
8	gtgaAUGuccu	None	---	---
9	guggAUGcugc	hsa-miR-450b-3p (3′) hsa-miR-487b (3′) miR-103 (5′) miR-107 (5′) miR-338-3p (5′) mmu-miR-376b (3′) *mmu-miR-450a-3p (3′)*	--- --- 1 1 1,2 --- ---	--- --- 3 --- --- 3 ---
10	aaagAUGaggg	hsa-miR-556-3p (3′), hsa-miR-609 (3′)	--- 1	--- ---

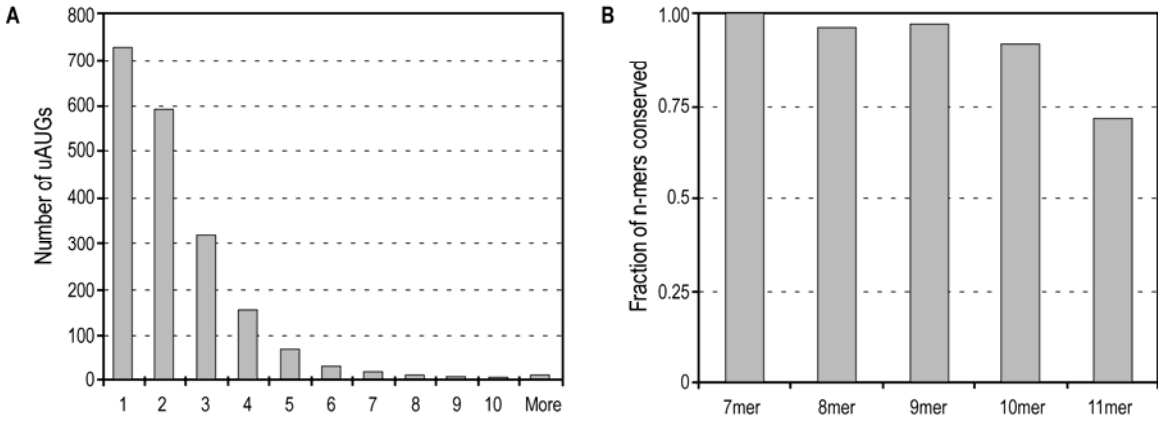


Figure 4.1 Number of uAUGs in 5′-UTRs and their conservation. **(A)** Distribution of uAUGs in human 5′-UTR sequences. **(B)** Fraction of uAUG-containing n-mer sequences conserved in human and mouse 5′-UTRs.

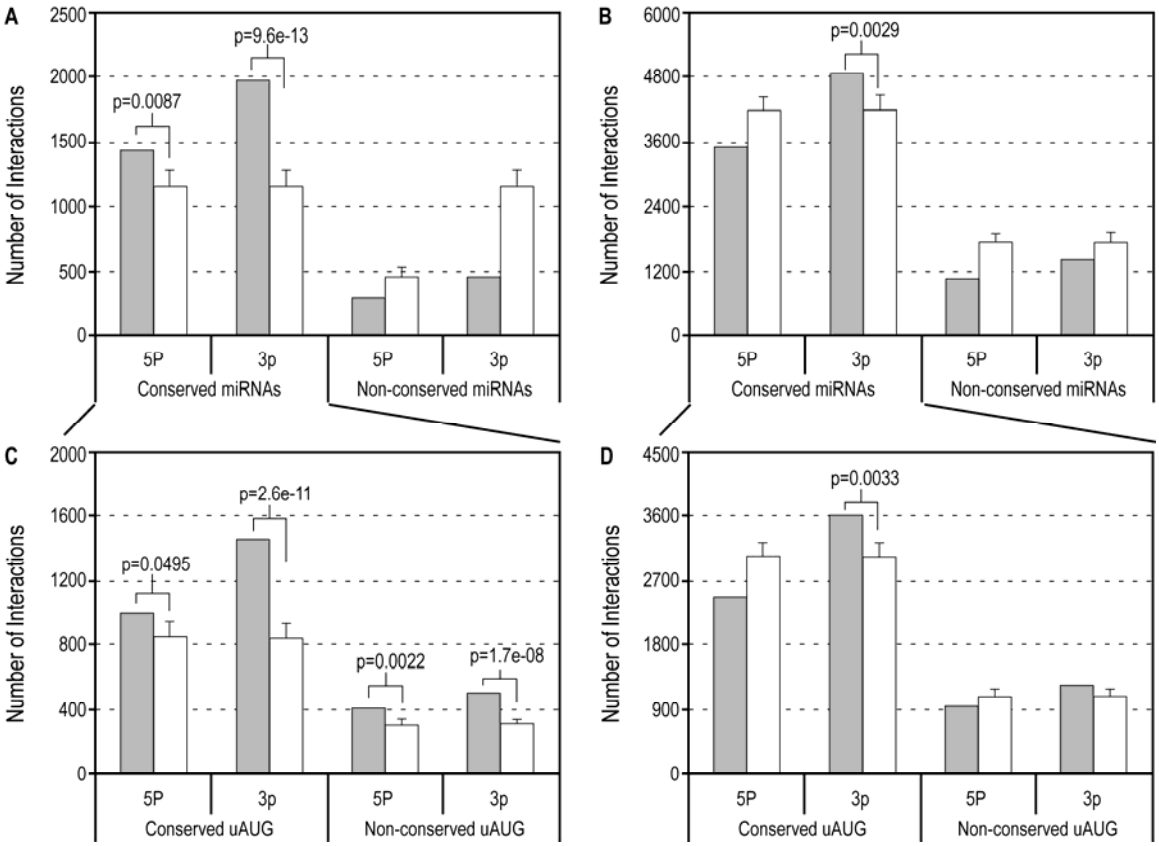


Figure 4.2 Interaction of miRNAs with uAUG sequences. Each predicted interaction is characterized by a 7-mer consecutive match between the indicated half of mature miRNA (5p and 3p for the 5′- and 3′-end respectively) and uAUG sequence with $\Delta G37 \leq -14$ kcal mol-1. Closed bars represent actual counts and open bars represent average number of counts over 1000 repetitions of miRNA shuffling. Error bars represent the standard deviations. Significant outcomes are indicated with the corresponding p-values **(A, B)** Number of interactions between uAUG sequences (4009 in total) and conserved and non-conserved miRNAs (471 and 206 in total respectively) without GU wobbles **(A)** and with at most one GU wobble **(B)**. **(C, D)** Number of interactions between conserved miRNAs and uAUG sequences (2935 conserved and 1074 non-conserved) without GU wobbles **(C)** and with at most one GU wobble **(D)**.

Chapter 5

Conclusions

5.1 Summary of work

In this thesis we have examined some outstanding challenges in the field of microRNA biology, provided techniques to overcome them and demonstrated the usefulness of these novel methods. Here, we provide a synopsis of the contributions by this thesis.

Many miRNAs have been grouped into families based on sequence homology, one being no different than another in the family but for one or two nucleotides. We used the example of the *let-7* family of miRNAs to demonstrate how conventional microarray probes designed to detect these sequences fail to distinguish them. Leveraging a novel base change strategy we provided a method to eliminate cross-hybridizations between similar sequences while creating a uniform melting temperature profile. The base-change strategy takes advantage of the changes in free energies of hybridization and in melting temperatures as dictated by the nearest-neighbour thermodynamics model. These changes are then utilized in the optimization process to generate one or more probes for a given mature miRNA.

Although conservation may signify a probable functional role the opposite does not necessarily hold true. Our work exemplifies this point in case with regards to miRNA targets. Using a combination of thermodynamic and sequence-based searches we showed

that the 3′-end participates in targeting 5′-UTR sites that are not conserved. We validated the hypothesis by using *AXIN2*, a gene part of the *Wnt* signaling pathway, containing a non-conserved 5′-UTR site and three conserved sites on the 3′-UTR, both of which are functional sites for *hsa-miR-34a*. Screening for targets that contain both UTR sites can reduce the set of probable biological targets.

Upstream AUGs are regulatory elements that are known to repress protein translation, with several mechanisms being proposed for their action. However, there is no explanation for how certain AUGs can confer tissue-specific expression of the protein. Utilizing data from sequence analysis and from miRNA expression in various cell types we provide evidence to show that the 3′-ends of miRNAs preferentially interact with the uAUGs and that the repressive action of uAUGs is correlated with the expression of miRNAs that bind them. We propose that miRNAs are responsible for cell-specific action of uAUGs by preventing the progress of a scanning ribosome.

5.2 Significance and impact

The methods developed in this thesis together with results generated can potentially influence the field of miRNA biology in a positive manner. It is evident that our strategy for designing miRNA probes is superior to probes conventionally designed with perfectly-matched sequences (Lee et al. 2008). Not only can ProDeG-designed probes be used for microarray experiments but also for *in situ* hybridization experiments to observe spatial expression of miRNAs, information which microarray experiments lose. Other methods developed for *in situ* probes involve using locked-nucleic acids for increasing specificity. However, this strategy requires adjustment of experimental conditions for

LNA-incorporated probes with different melting temperatures. In combination with the use of tetramethyl ammonium chloride (TMAC), as done in an earlier study (Deo et al. 2006) for highly specific wash conditions, our strategy might be able to generate more improved results.

The base-change strategy can also be used to design probes for DNA oligonucleotide microarrays that are used to detect gene expression. Lee *et al.* established specific guidelines to incorporate changes in probe sequences. These guidelines were based on the effects of base changes to melting temperatures as a function of length of probes, position and nature of mismatches. Using these rules one can reduce the search space for candidate probes and eventually distinguish similar sequences from a family of closely related genes. Being able to correctly detect expression of coding or non-coding transcripts in a particular cell-line, tissue or sample affects the inferences that are made downstream during miRNA-target analysis, especially with respect to correlation studies between miRNA and gene expression. Errors associated with expression data are propagated down the analysis pipeline that is used to predict possible targets that are regulated.

In tackling the two-part problem, this thesis also studied aspects of miRNA targeting that, so far, have not been explored. Pan-genomic conservation has been used as a determinant in studying seed-match sites on the 3′-UTR. This strategy showed very few seed-matches on 5′-UTRs and coding regions of a miRNA. Employing conservation might be a convenient method to reduce search space of possible target sites complementary to a stretch of 6-7 nucleotides on the seed region of a miRNA. The drawback, however, is that it eliminates species-specific interactions between miRNAs

and target sites. In other words, are there sites that are specific to one species and not others?

Our studies revealed many endogenous sites on the 5′-UTR that possibly interact with 3′-ends of miRNAs. This was true of 5′-UTR sites that are conserved across multiple species. What is even more interesting, however, is the significant number of interactions with 3′-ends of miRNAs when we considered sites that are not conserved and are enriched in human sequences. In the absence of 3′-UTR, *in vitro* experiments in Chapter 3 showed that it is possible that sites on the 5′-UTR may contribute towards targeting by binding the 3′-end of the miRNA. If this is the case, it seems that current prediction strategies that use conserved 3′-UTR regions as perceived target sits might be grossly underestimating the number of genes that are under regulatory control by miRNAs.

When studying a small portion (~1%) of the human genome researchers observed that only 40% of constrained sequences mapped back to protein-coding loci and their associated UTR sequences (Birney et al. 2007). Relevant to this discussion some of the other intriguing results from their work are: 1) 5′-UTRs are more divergent, an observation that is thought to reflect positive selection and 2) a large fraction of experimentally characterized functional elements shows no evolutionary constraint at the sequence level. The authors hypothesize functional conservation of non-orthologous sequences. In keeping with the results obtained and the hypothesis, it is plausible that non-orthologous sites on the 5′-UTR are functional. It will be interesting to study the abundance of sites on both UTRs and understand under what circumstances they are functional.

As a special case, we identified many upstream AUGs that may be involved in binding the 3′-ends of miRNAs and cause cell- or tissue-specific control of translation (Chapter 4). This finding shares striking similarity with a result from the study by Orom *et al.* where *miR-10a* binds a region of the 5′-UTR of *RPS16* that was earlier found to confer cell-type specific regulation of translation (Avni et al. 1997; Orom et al. 2008). Using Orom *et al*'s strategy of crosslinking miRNAs by UV light will shed light on binding sites and will help determine if miRNA-uAUG interactions occur as hypothesized. Majority of the genes that harbour uAUGs on 5′-UTRs are transcription factors (Churbanov et al. 2005). A recent report shows a network of miRNAs and transcription factors in *C. elegans* involved in a negative feedback mechanism (Martinez et al. 2008). Though this study primarily considered 3′-UTR sites predicted by other programs, it is possible that a different set of miRNAs might be involved in binding the 5′-UTR uAUGs.

Finally, over-expression or inhibition of miRNAs followed by microarray analysis has revealed hundreds of genes that are whose mRNA levels are regulated (Krutzfeldt et al. 2005; Lim et al. 2005). However, not all these genes may be direct miRNA targets containing seed-match sites on the 3′-UTR. Some possible scenarios are: 1) direct targets of miRNAs (regulated at the protein or mRNA level) may regulate other genes downstream which, therefore, appear differentially expressed, and 2) sites on the 5′-UTR could act independently or in conjunction with 3′-UTR sites to cause differential mRNA expression. The contribution of 5′-UTR sites may be further explored, perhaps by using an unbiased tagging method (Orom et al. 2008), to understand if the attribution in the 2nd scenario is valid. Further experimentation of sites on the 5′-UTR might also shed light on

the mechanistic differences of miRNA-mediated repression that currently are not fully understood.

One salient feature of the methods that we have used involves considering thermodynamics of nucleic acid binding. In some cases, we also allow G:U wobbles in the hybridizations even though they are not the most optimal in terms of free energy changes unlike Watson-Crick pairs. Stability is usually related to structure which in turn is related to function. However, it may not be that stability needs be 'extremized' for function. For instance, single molecule experiments show that catalytic activity of DNA polymerase depends on the tension of DNA; any lower or higher than 6 pN affects the enzyme's activity (Wuite et al. 2000; Haynie 2001). Considering G:U wobbles may allow for flexibility that might be necessary for protein complexes associated with miRNA and mRNA duplexes to perform their functions.

Our contributions in this thesis will, no doubt, help further discoveries in the field of miRNAs and pave the way for their application in therapeutics, changing the landscape of medicine.

5.3 Future work

Though the first miRNA was discovered more than 20 years ago, bulk of the research in miRNA biology has taken place in just over 5 years. The interest that this field has garnered in such a short span of time indicates that its impact cannot be underestimated. This thesis has made important contributions to the field but our curiosity knows no bounds. Some of the areas that we wish to explore for future research include but are not limited to:

- Design and execution of experiments to test our hypothesis that unifies uAUGs and miRNAs. Validation of this hypothesis should bring us one step closer to understanding miRNA function.
- Develop machine-learning techniques to include *a priori* information and increase confidence in target predictions
- Understanding how miRNAs have evolved across species and how the process has had an effect on target site evolution.
- Use work done in this thesis and incorporate miRNA and gene expression data to build gene-miRNA regulatory networks. Incorporating high quality data from next-generation sequencing methods should produce better results.

Appendices

Appendix A

Table A.1 ProDeG probes for cDNA samples of mature miRNAs along with respective cross-hybridization(s)

Target microRNA	Probe sequence[1]	microRNA name	Target and Non-target sequence(s)[1]	T_m
Targets and non-targets differ by one or two bases at ends				
miR-17-5p	CAAAGT**a**CTTACAGT**t**CAGGTAGT	miR-17-5p	ACTACCTGCACTGTAAGCACTTTG	57
		miR-106a	**G**CTACCTGCACTGTAAGCACTTT**T**	54
miR-106a	AAAAGT**c**CTTACAGT**c**CAGGTAGC	miR-106a	GCTACCTGCACTGTAAGCACTTTT	58
		miR-17-5p	**A**CTACCTGCACTGTAAGCACTTT**G**	55
miR-449	TGGC**t**GTGTATTGTTA**c**CTGGT	miR-449	ACCAGCTAACAATACACTGCCA	57
		miR-449b	**G**CCAGCTAACAATACACTGCC**T**	55
miR-449b	AGGCA**a**T**c**TATTGTTAGCTGGC	miR-449b	GCCAGCTAACAATACACTGCCT	57
		miR-449	**A**CCAGCTAACAATACACTGCC**A**	55
miR-517a	ATCGT**t**CATCCCTTTAGA**a**TGTT	miR-517a	AACACTCTAAAGGGATGCACGAT	57
		miR-517b	AACACTCTAAAGGGATGCACGA	57
		miR-517b	AACACTCTAAAGGGATGCACGA	57

miR-517b[2]	TCGTG**a**ATCCCTTTA**t**AGTGTT	miR-517a	AACACTCTAAAGGGATGCACGA**T**	56
miR-128a	TCACAGTGAAC**aa**GTCTCTTTT	miR-128a miR-128b	AAAAGAGACCGGTTCACTGTGA **G**AAAGAGACCGGTTCACTGTGA	57 56
miR-128b	TCACAGTGAAC**at**GTCTCTTTC	miR-128b miR-128a	GAAAGAGACCGGTTCACTGTGA **A**AAAGAGACCGGTTCACTGTGA	57 56
miR-133a	TTG**c**TCCCCTTCAACCA**t**CTGT	miR-133a miR-133b	ACAGCTGGTTGAAGGGGACCAA **T**AGCTGGTTGAAGGGGACCAA	60 56
miR-133b	TTGGTCCCCTTCA**t**CC**c**GCTA	miR-133b miR-133a	**T**AGCTGGTTGAAGGGGACCAA **AC**AGCTGGTTGAAGGGGACCAA	59 60
miR-520c	A**t**AGTGCTTCCTTTTAGA**a**GGTT	miR-520c miR-520f miR-520b miR-526b*	AACCCTCTAAAAGGAAGCACTTT AACCCTCTAAAAGGAAGCACTT CCCTCTAAAAGGAAGCACTTT **G**CCTCTAAAAGGAAGCACTTT	60 60 58 58
miR-520f[2]	**t**AGTGCTTCCTTTTAGA**a**GGTT	miR-520f miR-520c miR-520b miR-526b*	AACCCTCTAAAAGGAAGCACTT AACCCTCTAAAAGGAAGCACTT**T** CCCTCTAAAAGGAAGCACTT**T** **G**CCTCTAAAAGGAAGCACTT**T**	60 60 58 58
		miR-520b	CCCTCTAAAAGGAAGCACTTT	60

miR-520b[2]	AAAGTGCTTCCTTTTA**a**AGGG	miR-520c	**AA**CCCTCTAAAAGGAAGCACTTT	61
		miR-526b*	**G**CCTCTAAAAGGAAGCACTTT	59
		miR-520f	**AA**CCCTCTAAAAGGAAGCACTT	61
		miR-520e	CCCTC**A**AAAAGGAAGCACTTT	56
miR-526b*	AAAGT**a**CTTCCTTTTAGAGGC	miR-526b*	GCCTCTAAAAGGAAGCACTTT	59
		miR-520c	**AAC**CCTCTAAAAGGAAGCACTTT	56
		miR-520b	**C**CCTCTAAAAGGAAGCACTTT	56
		miR-520f	**AAC**CCTCTAAAAGGAAGCACTT	55
miR-520h[2]	AC**g**AAGTGCTTCCC**c**TTAGAGT	miR-520h	ACTCTAAAGGGAAGCACTTTGT	59
		miR-520g	**AC**ACTCTAAAGGGAAGCACTTTGT	60
		miR-519d	**AC**ACTCTAAAGGGA**G**GCACTTTG	54
miR-520g	ACAAAGT**t**CTTCCCTTT**c**GAGTGT	miR-520g	ACACTCTAAAGGGAAGCACTTTGT	60
		miR-520h	ACTCTAAAGGGAAGCACTTTGT	56
		miR-519d	ACACTCTAAAGGGA**G**GCACTTTG	54
miR-518c	CAAAGCGC**c**TCTCTTT**c**GAGTG	miR-518c	CACTCTAAAGAGAAGCGCTTTG	57
		miR-518f	**TC**CTCTAAAGAGAAGCGCTTT	52
miR-518f	AAA**t**CGCTTCTCTTTAG**g**GGA	miR-518f	TCCTCTAAAGAGAAGCGCTTT	58
		miR-518c	**CA**CTCTAAAGAGAAGCGCTTT**G**	54

miR-518f*	CTCTAGA**c**GGAAGCA**g**TTTCTCT	miR-518f* miR-526a	AGAGAAAGTGCTTCCCTCTAGAG AGAAAGTGCTTCCCTCTAGAG	57 54
miR-526a[2]	CTCTAGAG**a**GAAGCACTTTCT	miR-526a miR-518f* miR-518c* miR-526b	AGAAAGTGCTTCCCTCTAGAG **AG**AGAAAGTGCTTCCCTCTAGAG **C**AGAAAGTGCTTCCCTC**C**AGAGA **AAC**AGAAAGTGCTTCCCTC**A**AGAG	59 60 55 55
Targets and non-targets differ by one or two bases at positions near chain end				
miR-103	AGCA**t**CATT**c**TACAGGGCTATGA	miR-103 miR-107	TCATAGCCCTGTACAATGCTGCT T**G**ATAGCCCTGTACAATGCTGCT	57 53
miR-107	AGCA**t**CATT**c**TACAGGGCTATCA	miR-107 miR-103	TGATAGCCCTGTACAATGCTGCT T**C**ATAGCCCTGTACAATGCTGCT	57 53
miR-215	ATGACCTATG**g**ATTGAC**g**GAC	miR-215 miR-192	GTCTGTCAATTCATAGGTCAT G**G**CTGTCAATTCATAGGTCA**G**	57 54
miR-376b	ATCATAG**g**GGAAAAT**g**CATGTT	miR-376b miR-376a	AACATGGATTTTCCTCTATGAT AC**G**TGGATTTTCCTCTATGAT	57 53
miR-520e	AAAGTGCTTCC**a**TT**g**TGAGGG	miR-520e miR-520c	CCCTCAAAAAGGAAGCACTTT **AA**CCCTC**T**AAAAGGAAGCACTTT	59 53

		miR-520f	**AA**CCCTC**T**AAAAGGAAGCACTT	*52*
miR-527	**a**TGCAAA**c**GGAAGCCCTTTCT	miR-527 miR-520d*	AGAAAGGGCTTCCCTTTGCAG **C**AGAAAGGGCTTCCCTTTG**T**AG**A**	62 57
miR-18a	TAAGGTG**g**ATCTAGTGC**g**GATA	miR-18a miR-18b	TATCTGCACTAGATGCACCTTA TA**A**CTGCACTAGATGCACCTTA	60 58
miR-18b	TAAGGT**c**CATCTAGTGCAGTT**t**	miR-18b miR-18a	TAACTGCACTAGATGCACCTTA TA**T**CTGCACTAGATGCACCTTA	60 57
miR-23b	ATCACA**g**TGCCAG**a**GATTACC	miR-23b miR-23a	GGTAATCCCTGGCAATGTGAT GG**A**AATCCCTGGCAATGTGAT	58 52
miR-27b	TTCACA**t**TGG**a**TAAGTTCTGC	miR-27b miR-27a	GCAGAACTTAGCCACTGTGAA GC**G**GAACTTAGCCACTGTGAA	57 53
	Non-target sequence and probe pair contains interior G-U wobble			
miR-20a	TAA**t**GTGCTTATAGTGCAGGTAG	miR-20a miR-20b miR-106a miR-17-5p	CTACCTGCACTATAAGCACTTTA CTACCTGCACTATGAGCACTTT**G** **G**CTACCTGCACT**G**TAAGCACTTT**T** **A**CTACCTGCACT**G**TAAGCACTTT**G**	64 61 63 63
miR-19a	TGTG**t**AAATCTATG**g**AAAACTGA	miR-19a	TCAGTTTTGCATAGATTTGCACA	59

		miR-19b	TCAGTTTTGCAT**G**GATTTGCACA	55
miR-135a	TATGGCTTTTTATT**a**CTA**g**GTGA	miR-135a	TCACATAGGAATAAAAAGCCATA	57
		miR-135b	CACATAGGAAT**G**AAAAGCCATA	53
miR-30a-5p	T**a**TAAACATCCTCGACT**a**GAAG	miR-30a-5p	CTTCCAGTCGAGGATGTTTACA	57
		miR-30d	CTTCCAGTCG**G**GGATGTTTACA	54
miR-30e-5p	**g**GTAAAC**c**TCCTTGACTGGA	miR-30e-5p	TCCAGTCAAGGATGTTTACA	57
		miR-30a-5p	**CT**TCCAGTC**G**AGGATGTTTACA	53
miR-517c	ATCGTGCATCCTTTT**t**GA**c**TGT	miR-517c	ACACTCTAAAAGGATGCACGAT	58
		miR-517a	**A**ACACTCTAAA**G**GGATGCACGAT	52
miR-519e	AAA**t**TGCCTCCTTTTAG**t**GTGT	miR-519e	ACACTCTAAAAGGAGGCACTTT	58
		miR-519d	ACACTCTAAA**G**GGAGGCACTTT**G**	52
miR-29c	T**t**GCACCATTTGAAATCGGT	miR-29c	ACCGATTTCAAATGGTGCTA	66
		miR-29a	**A**ACCGATTTCA**G**ATGGTGCTA	62
		miR-29b	**AAC**AC**T**GATTTCAAATGGTGCTA	60

[1] all sequences are shown from 5' to 3' direction.
[2] miRNAs that are shorter than and identical to another miRNA as shown.

Appendix B

Table B.1 Conserved miRNAs predicted to target both 5' and 3' UTR 8-mers.

miRNA	**5' UTR Motif[1]**	**3' UTR Motif[1]**
hsa-let-7i	GGCGGCAC (0, 48)	CTACCTCA (139, 263), ACTACCTC (63, 160), TTACCTCA (99, 262), GCTACCTC (48, 152), ATTACCTC (40, 156)
hsa-miR-16	GCCAGTGT (1, 53)	TGCTGCTA (124, 380)
hsa-miR-17	GCCTGCAC (0, 66), CCTGCACT (1, 63)	AGCACTTT (190, 601), GCACTTTG (119, 398), AAGCACTT (123, 483), GGCACTTT (64, 276)
hsa-miR-18b	GACTGCGC (0, 54)	GTGCCTTA (147, 271), GCACTTTA (193, 405), TGCACTTT (244, 596), GCACTTTG (119, 398), ATGCACTT (82, 302)
hsa-miR-20a	GCCTGCAC (0, 66), CCTGCACT (1, 63)	GCACTTTA (193, 405), AGCACTTT (190, 601), AAGCACTT (123, 483), GGCACTTT (64, 276)
hsa-miR-20b/93	GCCTGCAC (0, 66), CCTGCACT (1, 63)	AGCACTTT (190, 601), GCACTTTG (119, 398), GGCACTTT (64, 276)
hsa-miR-25	TCAGGCCG (0, 37), CGGGCTGA (1, 69), GACTGAGA (1, 57), GACCGGGG (1, 50)	AGTGCAAT (90, 273), GGTGCAAT (33, 116), AAGTGCAA (81, 306), GTGCAATG (46, 183)
hsa-miR-93	GCCTGCAC (0, 66),	AGCACTTT (190, 601), GCACTTTG (119, 398), GGCACTTT (64, 276)
hsa-miR-106a	GCCTGCAC (0, 66), CCTGCACT (1, 63)	AGCACTTT (190, 601), AAGCACTT (123, 483), GGCACTTT (64, 276)
hsa-miR-125a-3p	CTCCTAGG (0, 47)	GCCTTACT (36, 172)
hsa-miR-128	AGGGGACC (1, 64), GAGACCGG (1, 49)	CACTGTGA (132, 414), TACTGTGA (99, 354)
hsa-miR-133b	TGGCTGGT (1, 51)	GGGACCAA (44, 214)
hsa-miR-139-5p	TGGGGGCA (1, 63)	GTACTGTA (136, 338), TGTACTGT (129, 506), TGCACTGT (90, 418), GTACTGTG (55, 265)
hsa-miR-196a/b	CCCAGCGA (1, 57)	AACTACCT (43, 194)

hsa-miR-199a-3p	GCCAGTGT (1, 53)	ACTACTGT (59, 268)
hsa-miR-199b-3p	GCCAGTGT (1, 53)	ACTACTGT (59, 268)
hsa-miR-202	TCTGTGCC (1, 74)	TATACCTC (35, 161)
hsa-miR-330-5p	CTGAGACA (1, 60)	CTCAGGGA (129, 451)
hsa-miR-362-5p	CTTGCACC (0, 37)	TTCAGGGA (81, 395)
hsa-miR-373	GCCCTGAA (0, 56), CCCCAAGA (1, 49)	AGCACTTT (190, 601), AAGCACTT (123, 483), GGCACTTT (64, 276)
hsa-miR-503	TGCGGGAC (0, 47), AGAGCTGT (1, 54)	TGCTGCTA (124, 380), TTGCTGCT (145, 664)
hsa-miR-506	TGCTCGGG (0,45)	GTGCCTTA (147, 271), GTGCCTTG (133, 355), GGTGCCTT (82, 281), GGTGCTTT (85, 340)
hsa-miR-509-3p	CCCGCAGA (0, 55), ACTCACAG (0, 46), TCTACAGA (0, 38), CCCACAGG (1, 66)	GTGCCAAT (40, 142)
hsa-miR-517a/b	GCGCTCTG (1, 61)	GTGCATGA (38, 170)
hsa-miR-517c	GCGCTCTG (1, 61)	GTGCATGA (38, 170)
hsa-miR-519a	GCGCTCTG (1, 61)	TGCACTTT (244, 596), ATGCACTT (82, 302)
hsa-miR-520a-3p	AGTCCAGA (1, 52)	AAGCACTT (123, 483)
hsa-miR-520d-3p	ACCCACCA (0, 38), CTCGCCGA (0, 37), CGCCAGAG (1, 59)	AAGCACTT (123, 483)
hsa-miR-520g	GCGCTCTG (1, 61)	AGCACTTT (190, 601), GCACTTTG (119, 398), AAGCACTT (123, 483), GGCACTTT (64, 276)
hsa-miR-524-3p	CCGGAGGG (1, 90)	GTGCCTTT (179, 482), GTGCCTTC (107, 361), AGTGCCTT (227, 473), GGTGCCTT (82, 281) AAGTGCCT (196, 430),
hsa-miR-608	GCTGTCCT (0, 70), CGGGGTTG (0, 41)	ACTACCTC (63, 160), CTACCTCT (96, 284), GCTACCTC (48, 152)
hsa-miR-613	GGCGAAGG (0, 64), GGCAAGGG (1, 56)	ACATTCCT (85, 405)
hsa-miR-650	CTGAGGGT (0, 67), TCTGAGGG (1, 75)	GTTGCCTT (66, 319)
hsa-miR-652	GCGACCCT (0, 46), CGACCCTG (0, 44)	TGGTGCTA (116, 272), GGTGCTAT (59, 167), GTGCCATT (61, 277)
hsa-miR-661	TGCGGGCC (0, 59), CGCGTGGG (0, 43), CGTGGGCC (1, 61)	CTCAGGTA (39, 156), ACTCAGGT (37, 175)

hsa-miR-770-5p	GGTTCTGA (0, 44), CTTGGCGC (1, 52), CCTTGGCA (1, 52)	TGGTGCTG (118, 525)
hsa-miR-885-3p	GTCCGCTG (0, 38), TCCACTGC (1, 59), CCACTGCA (1, 58)	TTGCTGCT (145, 664)

[1]Numbers within parentheses indicate conserved instances and number of human occurrences

B.1 Sequences used in *hsa-miR-34a* and AXIN2 assay

AXIN2 5′-UTR (NM_004655). Sequence cloned upstream of the luciferase coding region is shown in italics with predicted overlapping binding sites for the 3′-end of miR-34a shown in bold blue italics.

cggctgtgattggcgcggcgggatcactggctccgcgagcctg*gcccgggggagtcggctggagc*
*cggctgcgctttga*taaggtcctggcaactcagtaacagcccgagagccgggaaataaaaataac
ccctcagagcgatggatttcggggccgcccggcggccgaggcgcccgccgaaggccctgctgtaa
aagagaggaggttcagatgagcccctgctgacttgagagagacagagagaccacgccgattgctg
agaggaactggaagaagaaaaattcccagactcagtgggaagagctccctcacc

AXIN2 3′-UTR (NM_004655). Seed match sites are shown in bold underlined red and

3′-end interaction sites in blue italics. Seed match site with no GU wobble is in caps.

gccctggggtctggctttggtgaactgttggagcccgaagctcttgtgaactgtcttggctgtga
gcaactgcgacaaaacattttgaaggaaaattaaaccaatgaagaagacaaagtctaaggaagaa
t*cggccagt*gggccttcgggagggcgggggaggttgattttcatgattcatgagctgggtactg
actgagataagaaaagcctgaactatttattaaaaacatgaccactcttggctattga**agatgct**
gcctgtatttgagag**ACTGCCA**tacataatatatgacttcctagggatctgaaatccataaacta
agagaaactgtgtatagcttacctgaacaggaatccttactgatatttatagaacagttgatttc
ccccatccccagtttatggat**atgctgct**ttaaacttggaaggggagacaggaagttttaattg
ttctgactaaacttaggagttgagctaggagtgcgttcatggtttcttcactaacagaggaatta
tgctttgcactacgtccctccaagtgaagacagactgttttagacagacttttaaaatggtgcc
ctaccattgacacatgcagaaattggtgcgttttgtttttttttttcctatgctgctctgttttg
tcttaaaggtcttgagggttgaccatgttgcgtcatcatcaacattttgggggttgtgttggatg
ggatgatctgttgcagagggagaggcagggaaccctgctccttcgggccccaggttgatcctgtg
actgaggctccccctcatgtagcctccccaggcccagggccctgaggcctgctagaat**cactgcc**
gctgtgctttcgtggaaatgacagttccttgttttttttgtttctgttttgttttacattagtc
attggacc*acagcca*ttcaggaactacccccctgccccacaaagaaatgaacagttgtagggagac
ccagcagcacctttcctccacacaccttcattttgatgttcgggtttttgtgttaagttaatctg
tacattctgtttgccattgttacttgtactatacatctgtatatagtgtacggcaaaagagtatt
aatccactatctctagtgcttgactttaaatcagtacagtacctgtacctgcacggtcacccgct
ccgtgtgtcgccctatattgagggctcaagctttcccttgtttttgaaaggggtttatgtataa
atatattttatgccttttttattacaagtcttgtactcaatgacttttgtcatgacattttgttct
acttatactgtaaattatgcattataaagagttcatttaaggaaaattacttggtacaataatta
ttgtaattaagagatgtagcctttattaaaattttatatttttcaaaa

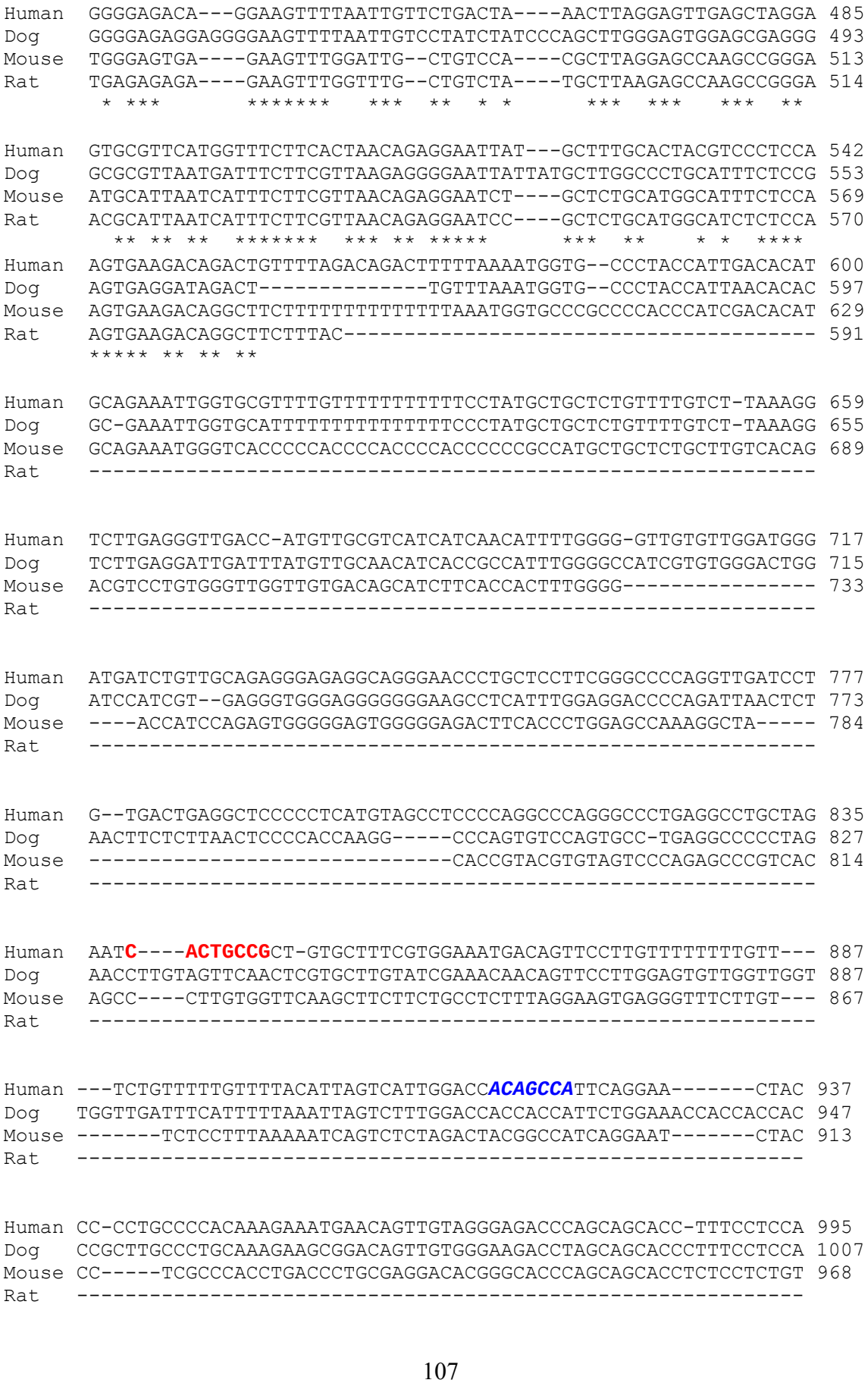
B.2 Multiple alignments of AXIN2 UTRs

B.2.1 Alignment of Axin2 5′-UTR sequences from 4 mammalian species (CLUSTAL 2.0.8 multiple sequence alignment). Predicted 5′-UTR target sites for the 3′-end of miR-34a are shown in bold blue italics.

```
Human  --------------------CGGCTGTGATTGGCGCGGCGGGATCACTGGCTCCGCGAGC 40
Dog    ------------------------------------------------------------
Mouse  GCGCGGCGGGATCACTGGCTCCCCGAGCCCGGCCCGGGGGAGTCGGCTGGAGCCGGCTGC 60
Rat    ------------------------------------------------------------

Human  CTGGCCCGGGGGAGTCGGCTGGAGCCGGCTGCGCTTTGATAAGGTCCTGGCAACTCAGTA 100
Dog    ------------------------------------------------------------
Mouse  GCTTTGATAAGGTCCTGGCAACTCAGTAACAGCCCAAGAACCGGGAAATAAAAATAAGCA 120
Rat    ------------------------------------------------------------

Human  ACAGCCCGAGAGCCGGGAAATAAAAATAACCCCTCAGAGCGATGGATTTCGGGGCCGCCC 160
Dog    ------------------------------------------------------------
Mouse  GCCGTTCGCGATGGATTTCGGGGCCACCCGGAGGCCGAGGCGTCCGCCTCCCCAAAGGAG 180
Rat    ------------------------------------------------------------

Human  GGCGGCCGAGGCGCCCGCCGAAGGCCCTGCTGTAAAAGAGAGGAGGTTCAGATGAGCCCC 220
Dog    ------------------------------------------------------------
Mouse  AGCTTTGCTGTAAAAGAGAGGAGGCTCACATGAGCCCCTGCTGACTTAAGAGAGACCAAG 240
Rat    ------------------------------------------------------------

Human  TGCTGACTTGAGAGAGACAGAGAGACCACGCCGATTGCTGAGAGGAACTGGAAGAAGAAA 280
Dog    ---------GAGAGAGAAAGAGAGACCACGCTGATTGCTGAGAGGAACTGGGGGAAGGGA 51
Mouse  CCGATTGCTGAGAGGAACTGGAAGAAGAAAAAGGAGGAGGAGGGAAAA-AAAGCAAAACA 299
Rat    --------TGAGAGGAACTGGAAGAAGAAAAAGGAAGAGGAAAAAAAA-AAAGCAAAACA 51
                *****  *  *  ***  *    *   *  **    **       **   *

Human  AATTCCCAGACTCAGTGGGAAGAGCTCCCTCACC 314
Dog    ACAAACAAACTCCGTCGCGAAGAACTCCCTCACC 85
Mouse  AAATCCAAACTCAGT-GAGACGCTCTCCCTCACC 332
Rat    AAACCCAAACTCAGT-GAGACGCTCTCCCTCACC 84
       *    * *        * ** *  **********
```

B.2.2 Alignment of Axin2 3′-UTR sequences from 4 mammalian species (CLUSTAL 2.0.8 multiple sequence alignment). Predicted seed match sites are shown in bold red and the 3'-end interaction sites in blue italics.

```
Human  --------------------GCCCTGGGGTCTGGCTTTGGTGAACTGTTGGAGCCCGAAG 40
Dog    --------------------GCCTCGGGGGTCGGGCCCCGGCGGACGC---GGGGCCACCG 38
Mouse  CCTTGGCCTCCTCGGCGTGCAACCTGGGCAAGCACCTCGGCGTGCACCATGGAGCCGAAG 60
Rat    CCTCGGCCTCCGCCGCGTGCA-CCTCCGCGTGCACCTCCGCGTGCACCACGGAGCCGGAG 59
                             *   *      *    * *  *     *   **   *

Human  CTC-TTGTGAACTGTCTTGGCTGTGAGCAACTGCGACAAAACATTTTGAAGGAAAATTAA 99
Dog    CCC-GCGCCACCAGCCTGGGCCATGACCGACCGCGACGAGACCTTTTGAAGGAAAACGAA 97
Mouse  CCCAGAGAC-CCTGTCTCAGGCCTACGCAACAGCCACGAAATATTCTGAAGGAAAATGAA 119
Rat    CCCAGAGAGACCTGTCTCAGGCCTACACAACAGCCATGAAATATTTTGAAGGAAAATGAA 119
       * *   *    * * **  *   *   * ** ** *  * *  ** **********  **

Human  ACCAATGAAGAAGACAAAGTCTAAGGAAGAATCGGCCAGTGGGCCTTCGGGA------GG 153
Dog    ACCAATGAAGAAGACAGAGTCTAGGGAAGACTTGGCCACTGGCCACGCGGGGAGGGCGGG 157
Mouse  ACCAATTAAGAAGACAAAGCCTAGGGAGGGACTGGCGCCTGGGCCTTCAGGA-------G 172
Rat    ACCAATTAAGAAGACAAAGCCTAGGGAGGGACTGGCCCCTGGGCCTTCAGGA-------G 172
       ****** ********* ** *** *** *    ***   *** *   * **        *

Human  GCGGGGGGAGGTTGATTTTCATGATTCATGAGC-TGGGTACTGACTGAGATAAGAAAAGC 212
Dog    GAGGGGGGAGGTTGGTTTTCATTATTCACGAGC-TGGGTACT----GAGATAAGAAAAGC 212
Mouse  GGCGGGGGTAGTTGATCTTCAGTCTCCAGGAGCCTGGGTACC----GAGATGAGAAAAGC 228
Rat    GGCGGGGGTGGTTGGTTTTCAATATCCACGAGC-TGGGTACT----GAGATCAGAAAAGC 227
       *  *****  **** * ****   * ** **** *******     ***** ********

Human  CTGAACTATTTATTAAAAACATGACCACTCTTGGCTATTGAAGATGCTGCCTGTATTTGA 272
Dog    CTGAACTATTTATTAAAAACATGACCACTCTTGGCTATTGATGATGCTGACTGTATTTGA 272
Mouse  CTGAACTATTTATTCAAAACATGACCACTCTGGGCTATAGAAGATGCTGAGTGTGTTTGA 288
Rat    CCGAACTATTTATTCAAAACATGACCACTCTGGGCTATAGAAGATGCTGAGTGC--TCGA 285
       * ************ *************** ****** ** *******  **   * **

Human  GAGACTGCCATACATAATATATGACTTCCTAGGGATCTGAAATCCATAAACTAAGAGAAA 332
Dog    GAGACTGCCATACATAATATATGACTGCCTAGGGATCTGAAATCCATAAACTAAGAGAAA 332
Mouse  GAGACTGACATACATAATAGATGACTTCCTAGGGTTCTGAAATTCATAGACTAAGAGAAA 348
Rat    GAGACTGCCATACATAATAGATGGCTTCCTAGGGATCTGAAATCCATAGACTAAGAGAAA 345
       ******* *********** *** ** ******* ******** **** ***********

Human  -CTGTGTATAGCTTACCTGAACAGGAATCCTTACTGATATTTATAGAACAGTTGATTTCC 391
Dog    -CTGTGTATAGCTTACCTGAACAGGAGTCCTTACTGCTATTTATTGAACAATTGATTCCC 391
Mouse  ACTGTGTATAGCTTGCCCGAACAGGAGTCCTTACTGATATTTATTGAACAGTCGATTCCC 408
Rat    ACTGTGTATAGCTTGCCCGAACAGGAGTCCTTACTGATATTTATTGAACAGTCGATTCCC 405
        ************* ** ******** ********* ******* ***** * **** **


Human  CC----------------CA-TCCCCAGTTTATGGAT-ATGCTGCTTT-AAACTTGGAAG 432
Dog    CC----------------CAGCCCCCAGTTTATGGAT-ATGCTGCTTT-AAACACGGAAG 433
Mouse  CTACCCGCC-----CTCCCTACCCCCAGCCCCCAGTTTATGCTGCTTTTAAACCTGGAAG 463
Rat    CTACCCCACTCCCACCCCCAACCCCCAACCCCCAGTTTATGCTGCTTT-AAACCTGGAAA 464
       *                  *   *****       * * ********** ****  ****
```


Human GGGGAGACA---GGAAGTTTTAATTGTTCTGACTA----AACTTAGGAGTTGAGCTAGGA 485
Dog GGGGAGAGGAGGGGAAGTTTTAATTGTCCTATCTATCCCAGCTTGGGAGTGGAGCGAGGG 493
Mouse TGGGAGTGA----GAAGTTTGGATTG--CTGTCCA----CGCTTAGGAGCCAAGCCGGGA 513
Rat TGAGAGAGA----GAAGTTTGGTTTG--CTGTCTA----TGCTTAAGAGCCAAGCCGGGA 514
Human GTGCGTTCATGGTTTCTTCACTAACAGAGGAATTAT---GCTTTGCACTACGTCCCTCCA 542
Dog GCGCGTTAATGATTTCTTCGTTAAGAGGGGAATTATTATGCTTGGCCCTGCATTTCTCCG 553
Mouse ATGCATTAATCATTTCTTCGTTAACAGAGGAATCT----GCTCTGCATGGCATTTCTCCA 569
Rat ACGCATTAATCATTTCTTCGTTAACAGAGGAATCC----GCTCTGCATGGCATCTCTCCA 570
Human AGTGAAGACAGACTGTTTTAGACAGACTTTTTAAAATGGTG--CCCTACCATTGACACAT 600
Dog AGTGAGGATAGACT--------------TGTTTAAATGGTG--CCCTACCATTAACACAC 597
Mouse AGTGAAGACAGGCTTCTTTTTTTTTTTTTTTAAATGGTGCCCGCCCCACCCATCGACACAT 629
Rat AGTGAAGACAGGCTTCTTTAC---------- 591
Human GCAGAAATTGGTGCGTTTTGTTTTTTTTTTTCCTATGCTGCTCTGTTTTGTCT-TAAAGG 659
Dog GC-GAAATTGGTGCATTTTTTTTTTTTTTTTCCCTATGCTGCTCTGTTTTGTCT-TAAAGG 655
Mouse GCAGAAATGGGTCACCCCCACCCCACCCCACCCCCCGCCATGCTGCTCTGCTTGTCACAG 689
Rat ----------
Human TCTTGAGGGTTGACC-ATGTTGCGTCATCATCAACATTTTGGGG-GTTGTGTTGGATGGG 717
Dog TCTTGAGGATTGATTTATGTTGCAACATCACCGCCATTTGGGGCCATCGTGTGGGACTGG 715
Mouse ACGTCCTGTGGGTTGGTTGTGACAGCATCTTCACCACTTTGGGG---------- 733
Rat ----------
Human ATGATCTGTTGCAGAGGGAGAGGCAGGGAACCCTGCTCCTTCGGGCCCCAGGTTGATCCT 777
Dog ATCCATCGT--GAGGGTGGGAGGGGGGGAAGCCTCATTTGGAGGACCCCAGATTAACTCT 773
Mouse ----ACCATCCAGAGTGGGGGAGTGGGGGAGACTTCACCCTGGAGCCAAAGGCTA----- 784
Rat ----------
Human G--TGACTGAGGCTCCCCCTCATGTAGCCTCCCCAGGCCCAGGGCCCTGAGGCCTGCTAG 835
Dog AACTTCTCTTAACTCCCCACCAAGG-----CCCAGTGTCCAGTGCC-TGAGGCCCCCTAG 827
Mouse ----------CACCGTACGTGTAGTCCCAGAGCCCGTCAC 814
Rat ----------
Human AATC----ACTGCCGCT-GTGCTTTCGTGGAAATGACAGTTCCTTGTTTTTTTTGTT--- 887
Dog AACCTTGTAGTTCAACTCGTGCTTGTATCGAAACAACAGTTCCTTGGAGTGTTGGTTGGT 887
Mouse AGCC----CTTGTGGTTCAAGCTTCTTCTGCCTCTTTAGGAAGTGAGGGTTTCTTGT--- 867
Rat ----------
Human ---TCTGTTTTTGTTTTACATTAGTCATTGGACCACAGCCATTCAGGAA-------CTAC 937
Dog TGGTTGATTTCATTTTTAAATTAGTCTTTGGACCACCACCATTCTGGAAACCACCACCAC 947
Mouse -------TCTCCTTTAAAAATCAGTCTCTAGACTACGGCCATCAGGAAT-------CTAC 913
Rat ----------
Human CC-CCTGCCCCACAAAGAAATGAACAGTTGTAGGGAGACCCAGCAGCACC-TTTCCTCCA 995
Dog CCGCTTGCCCTGCAAAGAAGCGGACAGTTGTGGGAAGACCTAGCAGCACCCTTTCCTCCA 1007
Mouse CC-----TCGCCCACCTGACCCTGCGAGGACACGGGCACCCAGCAGCACCTCTCCTCTGT 968
Rat ----------

```
Human CACACCTTCATTTTGATGTTCGGGTTTTTGTGTTAAGTTAATCTGTACATTCTGTTTGCC 1055
Dog   GACACCTTCATTTTGACGTTCGGGTTTTTGTGTTAAGTTAATCTGTACATTCTGTTTGCC 1067
Mouse TACCTTCCCCTTGGCGATCGCTCGGGTTTGTGTTAAGTTAATCTGTATGTCCTGTCTGCC 1028
Rat   ------------------------------------------------------------


Human ATTGTTACTTGTACTATACATCTGTATATAGTGTACGGCAAAAGAGTATTAATCCACTAT 1115
Dog   ATTGTTACCTGTACTATACGTCTGTATATATTGTACGACAGAAGAGTATTAATCCACTAT 1127
Mouse AGCGTTCCCTGTACTATAGGCCTGTGTATAGTGTAGGGCA---GAGCGTTGACCCACTGG 1085
Rat   ------------------------------------------------------------


Human CTCTAGTGCTTGACTTTA-AATCAGTACAGTACCTGTACCTGCACGGTCACCCGCTCCGT 1174
Dog   CTCTAGTGCTTGACTTGA-AATCAGTACAGTACCTGTACCTGCACGGCGCCCCGCTCCGT 1186
Mouse CT--AGTGCTTGACTTGGGAATCAGGACAGTACCTGTACAGGCACGGGGACCCGCTCCGT 1143
Rat   ------------------------------------------------------------


Human ----GTGTCGCCCTATATTGAGGGCTCAAGCTTTCCCTTGTTTTTTGAAAGGGGTTTATG 1230
Dog   ----GTGTCGCCCTATATTGAGGGCTCCAACTTTCCCTTGTTTTTTGAAAGGGGTTTATG 1242
Mouse CCGTGCGCCGCCCTATATTGAGGGCTCCAGCTCTCCCTTGGTTTTTGAAAGGGGTTTATG 1203
Rat   ------------------------------------------------------------

Human TATAAATATATTTTATGCCTTTTTATTACAAGTCTTGT-ACTCAATGACTTTTGTCATGA 1289
Dog   TATAAATATATTTTATGCCTTTTTATTACAAGTCTTGT-ACTCAATGACTTTTGCCATGG 1301
Mouse TATAAATATATTTTATGCCTTTTTATTACAAGTCTTGTTACTCAATGACTTTTGTCATGG 1263
Rat   ------------------------------------------------------------

Human CATTTTGTTCTACTT-ATACTGTAAATTATGCATTATAAAGAGTTCATTTAAGGAAAATT 1348
Dog   CATTTTGTTCTACTT-ATACTGTAAATTATGCATTATAAAGAGTTCATTTAAGGAAAATT 1360
Mouse CAGTTTGTTCTACTTTAGACTGTAAATTATGCATTATAAAGAGTTCATTTAAAGAAAACT 1323
Rat   ------------------------------------------------------------

Human ACTTGGTACAATAATTATTGTAATTAAGAGATGTAGCCTTTATTAAAATTTTATATTTTT 1408
Dog   ACTTGGTACAATAATTATTGTAATTAAGAGATGTAGCCTTTATTAAAATTTTATATTTTT 1420
Mouse ACTTGGTACAATAATTATTGTAATTAAGAGATGTAGCCTTTATTAAAATTTTATATTTTT 1383
Rat   ------------------------------------------------------------

Human CAAAA 1413
Dog   C---- 1421
Mouse C---- 1384
Rat   -----
```

B.3 Sequences used in lin28 assay

lin28, miRNA, and siRNA sequences used in reporter gene assays are shown below. Predicted binding sites and miRNA interaction sites are in bold; mutated sites are in bold small case.

lin28 3UTR with perfectly matched site (3Upm, wild-type sequence)

5′- CACCTACCTCCTCAAATTGCAC**TCTCAGGG**ATTCTTTTTTTTTTCAAATAGAACT- 3′

lin28 3UTR with imperfectly matched site (3Umm)

5′- CACCTACCTCCTCAAATTGCAC**TCTCAttt**ATTCTTTTTTTTTTCAAATAGAACT- 3′

lin28 5UTR with perfectly matched target site (5Upm)

5′- GTGGTATTGTTGTTCTG**TAagccacA**TAGGTTGTATTCTCTAGTTAACACATAGT- 3′

lin28 5UTR with mismatched site (5Umm, wild-type sequence)

5′- GTGGTATTGTTGTTCTG**TATATTTGA**TAGGTTGTATTCTCTAGTTAACACATAGT- 3′

cel-lin-4 miRNA	5′- **UCCCUGAGA**CCUCAAGUGUGA - 3′
lin4msiRNA	5′- **UCCCUGAGA**CC**UgugGcuUg**A - 3′ (functional strand) 5′- AAGCCACAGGUCUCAGAAGUU - 3′ (opposing strand)
hsa-miR-16	5′- UAGCAGCACGUAAAUAUUGGCG - 3′

Appendix C

C.1 GO-term analysis for genes containing targeted uAUGs

C.1.1 Genes containing uAUGs in Table 4.1 (Total 1071)

A1CF
ABCF2
ABCG1
ACCN1
ACTL6B
ACTR1A
ACVR2A
ADC
ADCK1
ADIPOR1
ADRA1B
ADRA2C
ADRM1
AFTPH
AGBL5
AHSA1
AJAP1
AKR1B10
ALDOA
ALX1
AMD1
AMFR
AMPH
ANKFY1
ANKH
ANKS1A
ANP32A
ANP32B
ANP32E
AP1G1
ARF1
DNAJB12
DNAJB5
ARF6
ARHGEF12
ARHGEF9
ARID4A
ARL15
ARL4C
ARL5A
ARL8A
ARNTL
ARPC4
ARPC5
ASB8
ASCC2
ASH1L
ASPH
ATAD2B
ATF4
ATF7
ATG5
ATG9A
ATOH8
ATP10A
ATP2A2
ATP2C1
ATP8A2
ATXN1
ATXN3
ATXN7
AZIN1
B3GALNT2
B3GALT2
DOC2A
DOC2B
B4GALT2
BACH2
BAI1
BAIAP2
BAIAP2L2
BANF1
BAP1
BARHL1
BARHL2
BAT1
BBC3
BBX
BCL11A
BCL11B
BCL3
BCL6
BCL7C
BCOR
BDNF
BDP1
BHLHB3
BHLHB5
BMP2
BMP2K
BMP6
BMPER
BRD2
BRD4
BTBD10
BTG1
C13orf7
DOK5
DOK6
C14orf100
C14orf147
C15orf41
C18orf1
C1orf119
C1orf164
C1orf25
C1orf27
C1orf76
C1QL2
C20orf24
C20orf67
C2orf25
C2orf33
C3orf10
C4orf18
C5orf41
C5orf5
CA10
CABIN1
CABYR
CACNA1E
CACNA1G
CACNA2D2
CACNG3
CADM1
CALB2
CALU
CAMK2B
CAMK2G
CAPRIN1
DOK7
DOLPP1
CASD1
CASK
CASP8AP2
CBLL1
CBX4
CBX6
CBX7
CBX8
CCDC109A
CCDC53
CCNB2
CCND1
CCNI
CCNJ
CD2AP
CD37
CDC2L1
CDH24
CDH8
CDK5RAP3
CDX1
CDYL2
CENTA2
CENTD1
CFL1
CHD2
CHD4
CHMP7
CHST11
CHSY1
CHUK
DOT1L
DPP4
CITED2
CKAP5
CLCF1
CLCN3
CLCN5
CLDN23
CLDND1
CLIC1
CMTM4
CNIH2
CNKSR2
CNOT4
CNOT7
CNTN4
CNTN6
COIL
COL1A1
COL3A1
COL4A1
COLQ
COPS3
COPS4
CPEB3
CPSF3
CRABP1
CRBN
CREB1
CREB3L2
CREBL2
CRK
CRKRS
DPYSL2
DRP2
CRYAB
CRYGC
CSDE1
CSK
CSMD3
CSNK1D
CSNK2A1
CSNK2A2
CTDSP1
CTDSPL2
CTNNA2
CTNNBIP1
CTNND1
CUL5
CUX1
CXorf6
CYBA
DAB1
DCTN2
DDEF2
DDX25
DGCR2
DGCR8
DGKI
DGKZ
DHCR24
DICER1
DLX1
DLX3
DLX5
DNAJA2
DSCAM
DSG1

DTNA
DTX3
DULLARD
DUSP15
DUSP16
DUSP6
DVL3
DYRK1A
DYRK1B
EBF1
EDA
EDC4
EEF1D
EEF1G
EFNA1
EFNA3
EFNA5
EFNB1
EFTUD2
EGLN2
EHBP1
EIF1
EIF1B
EIF2S2
EIF3B
EIF4A2
EIF4G2
EIF4G3
EIF4H
EIF5A
EIF5A2
ELAVL1
ELAVL2
ELK3
ELL2
ELOVL1
ELOVL6
EML5
EMX2
EN1
ENAH
ENTPD7
EPC1
EPHA3
EPHA4
EPHB2
ERF
ERGIC3
ERRFI1
ETF1
ETV1
ETV2
ETV5
EXOC1
EXOC5
EXT1
EYA1
FA2H
FAF1
FAM110B
FAM129A
FAM33A
FAM70A
FAM98A
FBXL3
FBXL4
FBXO42
FBXW11
FBXW2
FEM1B
FEN1
FEV
FEZF2
FGD1
FGD6
FGF14
FGF8
FGFR1
FIS1
FLII
FLRT3
FMO1
FMR1
FOXA1
FOXG1
FOXJ1
FOXJ3
FOXN3
FRS2
FRS3
FURIN
FYN
G3BP1
GABRB2
GABRG2
GADD45A
GALNT7
GATAD2B
GDF5
GGNBP2
GJB1
GLIS3
GNA11
GNAI2
GNAI3
GNAT1
GNAZ
GOLIM4
GPBP1
GPD1L
GPHN
GPR26
GPR61
GPR85
GPSM2
GRIA2
GRIA3
GRID1
GRIN2A
GRIPAP1
GRK1
GSC
GTF3C2
GTPBP1
HAND1
HCCA2
HCN2
HDAC2
HDAC4
HDAC5
HDGFRP3
HEATR3
HELZ
HERC4
HERPUD2
HES1
HEXIM1
HGS
HHIP
HIP1R
HIP2
HMBOX1
HMG2L1
HMGB1
HMGB3
HNRNPC
HNRNPR
HNRPAB
HOXA1
HOXA11
HOXA3
HOXA4
HOXA9
HOXB13
HOXB3
HOXB6
HOXC13
HOXC6
HOXC8
HOXD4
HOXD9
HPCA
HR
HRB
HS2ST1
HS6ST1
HS6ST3
HSD11B2
HSDL1
HSP90AB1
ICK
ID2
ID3
IHPK1
IKZF1
IL1RAPL1
IL7
ILK
ING3
INHBA
INHBB
INPP5A
IPO13
IRF2
IVNS1ABP
JAK1
JAKMIP2
JARID2
JAZF1
JDP2
JMJD1A
JMJD1C
JPH1
JPH4
JUN
KATNB1
KBTBD4
KBTBD8
KCNA4
KCNB1
KCND2
KCNH2
KCNIP1
KCNJ2
KCNJ8
KCNK3
KCNN3
KCNN4
KCNS2
KCNS3
KCTD10
KCTD15
KCTD17
KHDRBS3
KHK
KIAA0082
KIAA0427
KIAA0562
KIAA1219
KIAA1715
KIF3C
KIF5B
KIT
KITLG
KLF12
KLF13
KLF4
KLF7
KLF9
KLHL10
KLHL18
KLHL20
KLHL24
KLHL28
KPNA3
KPNA4
KRTAP4-5
LAD1
LASS6
LBX1
LCE1B
LCE1E
LCE3D
LCORL
LDB1
LDB2
LDLRAP1
LEPROTL1
LGI1
LHX5
LIN28
LMO1
LMO2
LMO4
LOXL1
LRFN2
LRP2
LRRC4
LRRC4C
LRRTM3
LRRTM4
LTBP1
LUC7L
LYAR
MAFB
MAGI3
Magmas
MAML3
MAP2K2
MAP3K11
MAP3K2
MAP3K7IP2
MAPK1
MAPK10
MAPK8IP3
MAPKAP1
MARCKS
MAST1
MATR3
MAX

MBD2
MBNL1
MBNL2
MBOAT2
MBTD1
MBTPS1
MCF2
MCRS1
MED31
MED7
MEF2C
MEIS1
MEMO1
MEN1
METTL3
MEX3C
MFSD2
MGAT2
MGAT3
MGAT4B
MGC4172
MIDN
MIER1
MINK1
MKRN1
MLF2
MLL5
MLLT3
MME
MNT
MORF4L1
MOSPD1
MOSPD3
MPP5
MRVI1
MSL3L1
MSX1
MTA1
MTCP1
MTMR14
MTPN
MYBL2
MYBPC1
MYC
MYL3
MYST2
MYST3
MYST4
MYT1
NAP1L1
NAT12
NBR1
NCK2
NCOA2
NDEL1
NDUFAB1
NEK11
NEK6
NELF
NEUROD6
NF1
NF2
NFAT5
NFATC3
NFATC4
NFE2L1
NFIA
NFIB
NFIX
NKD1
NKIRAS2
NKX2-8
NLGN3
NLK
NOC3L
NR1D1
NR2C2
NR4A2
NRBP1
NRIP1
NRN1
NRSN1
NRXN2
NRXN3
NTF5
NTNG1
NTRK3
NUDT3
NUMBL
NUS1
NUTF2
NXN
OAZ2
ODF2
OPA3
ORMDL2
OTUB1
PACS1
PAFAH1B1
PAK1
PAK3
PAPOLG
PARD6A
PARK2
PARP6
PARP8
PAX3
PAX6
PBRM1
PBX1
PBX3
PCBP2
PCDH7
PCGF1
PCGF2
PCSK1N
PCTK1
PCYT1B
PDAP1
PDCD10
PDE1B
PDE7B
PDGFA
PDGFC
PDPK1
PDXDC1
PDZD2
PELI2
PEX11B
PEX5
PFN1
PGM5
PGRMC1
PH-4
PHACTR1
PHACTR3
PHEX
PHF1
PHF10
PHF12
PHF2
PHF21A
PHF21B
PHF23
PHF3
PHOX2A
PHTF1
PI4K2A
PIAS1
PIAS3
PIAS4
PICALM
PIK3CG
PIK3R3
PITX1
PKIG
PLAG1
PLCB1
PLCH1
PLD5
PLEKHA2
PLK3
PLP1
PLSCR3
PNKD
PNRC2
POLR3F
PORCN
POU2F1
POU2F2
POU4F2
POU6F2
PPAP2A
PPAPDC3
PPARGC1A
PPFIA2
PPM1A
PPM1B
PPM1D
PPM1G
PPP1R10
PPP1R16A
PPP1R7
PPP2CA
PPP2CB
PPP2R2B
PPP2R5E
PPP3CA
PPP3CB
PPP4C
PPP4R1L
PRKCA
PRKCE
PRKD3
PRKG1
PROX1
PRPH2
PRRX1
PSCD2
PSKH1
PSMB3
PSMC4
PSME3
PTEN
PTGES3
PTP4A1
PTPLAD1
PTPN23
PTPRK
PUM2
PVRL4
PYGO2
RAB10
RAB13
RAB14
RAB1A
RAB2A
RAB31
RAB33A
RAB35
RAB39B
RAB5A
RAB6A
RAC1
RAC3
RAD23A
RAD50
RAI1
RALBP1
RALGPS2
RALYL
RANBP9
RAP2A
RARB
RASD1
RASGRP1
RAX
RBBP5
RBM10
RBM12
RBM12B
RBM18
RBM39
RCAN2
REEP1
REEP5
RELA
RELL2
REPS2
RER1
REV1
RFTN2
RGL2
RHOA
RHOB
RHOBTB2
RHOG
RIC8B
RICH2
RLBP1L1
RND3
RNF10
RNF126
RNF139
RNF144A
RNF41
RNPEPL1
ROD1
RORC
RPIA
RPL12
RPL13
RPL21
RPL29
RPLP0
RPP25
RPS2
RPS6KA1
RTN2
RUNX1T1

RUSC1
RXRA
RXRG
RYBP
SATB2
SBDS
SBF1
SCAMP1
SCARF2
SCN3A
SCN5A
SCYL1
SELI
SEMA4G
SEMA7A
SENP2
SERP1
SERTAD2
SF3B14
SFN
SFRS15
SFRS16
SFRS9
SH3GL1
SH3GLB2
SH3RF1
SH3RF2
SIAH2
SIRPA
SIX2
SIX3
SKP1
SLC20A1
SLC25A36
SLC25A5
SLC26A9
SLC35A4
SLC35E1
SLC39A13
SLC41A1
SLC43A2
SLC4A10
SLC4A3
SLC6A8
SLCO3A1
SMAD5
SMAD6
SMAD7
SMCR7L
SMURF1
SMURF2
SNCA
SNCAIP
SNCB
SND1
SNF1LK2
SNIP
SNN
SNRPB
SNRPD1
SNRPD2
SNRPD3
SNUPN
SNX24
SNX5
SOCS3
SORCS3
SOX21
SOX4
SP1
SP4
SPA17
SPAG7
SPI1
SRPK1
SRRM1
SSBP2
SSNA1
ST3GAL3
ST5
ST6GALNAC5
ST6GALNAC6
ST8SIA2
ST8SIA3
ST8SIA4
STAG1
STK36
STRN3
STX1B
SUMO2
SURF4
SUSD4
SUV420H1
SV2A
SYNCRIP
SYT13
TAF10
TAF5L
TAF9
TAGLN3
TAOK2
TBC1D15
TBL1XR1
TBPL1
TBR1
TBX2
TBX3
TCEB2
TCF7L2
TEAD1
TEAD2
TEK
TESK1
TFAP2E
TFE3
THAP1
THOC4
THRA
TIA1
TLE4
TLL1
TLOC1
TMCO6
TMEFF2
TMEM1
TMEM121
TMEM161B
TMEM49
TMOD4
TMSB10
TMTC2
TNFAIP1
TNFRSF11A
TNFSF8
TNMD
TNPO1
TNPO2
TNRC4
TOLLIP
TOX4
TPD52L2
TPM1
TPM2
TRA2A
TRAF4
TRAM1
TRIM11
TRIM2
TRIM3
TRIM33
TRIM37
TRIM46
TRIM62
TRIM66
TRIM8
TRIOBP
TRPM3
TRPS1
TSC1
TSC22D1
TSC22D3
TSC22D4
TSHZ2
TSPAN12
TSPAN17
TSPAN18
TSPAN5
TSPAN9
TSSK6
TTC9B
TTLL11
TTLL5
TULP4
TWIST1
TWIST2
TXNIP
TYRO3
UBA1
UBAC1
UBAP1
UBAP2L
UBE2E1
UBE2E2
UBE2M
UBE2Q1
UBE2R2
UBE4A
UBE4B
UBL7
UBOX5
UBP1
UBQLN2
UBTD1
UCHL3
UCK2
UGCG
UPF2
USF1
USF2
USP12
USP19
USP2
USP42
USP46
USP48
USP49
USP52
VANGL2
VAPA
VEGFA
VEZF1
VPS26B
VPS33B
VPS36
VPS4A
VSX2
VWC2
WAC
WAPAL
WASF2
WDR44
WNT10A
WNT11
WNT2
WNT9A
WRNIP1
WSB1
WSCD1
WWP1
XPO1
XPR1
YES1
YIPF4
YWHAE
YWHAQ
YY1
ZBTB2
ZBTB20
ZBTB33
ZBTB7B
ZC3H10
ZC3H15
ZC3H7B
ZCCHC14
ZCCHC17
ZCCHC6
ZDHHC14
ZDHHC3
ZDHHC5
ZFAND3
ZFAND5
ZFAND6
ZFHX3
ZFHX4
ZFP161
ZFR
ZFYVE27
ZIC1
ZIC3
ZIC5
ZMIZ1
ZMYND11
ZNF148
ZNF219
ZNF238
ZNF282
ZNF318
ZNF32
ZNF410
ZNF491
ZNF503
ZNF521
ZNF532
ZNF592
ZNF627
ZNF638
ZNF710
ZNF804A
ZNRF1
ZZZ3

C.1.2 GO-term analysis

Number of genes among 1071 (above) with available annotations = Number of genes in test set (X) = 678
Number of genes in reference set (N) = 8649
x indicates the number of genes in test set in the functional category listed
n indicates the number of genes in reference set in the functional category listed
p-values calculated using hypergeometric test.

GO-ID	p-value	corr p-value	x	n	X	N	Description
30528	1.1933E-17	7.6612E-15	145	923	678	8649	transcription regulator activity
3700	5.5821E-11	1.7919E-8	84	520	678	8649	transcription factor activity
16564	4.0031E-10	8.5667E-8	39	168	678	8649	transcription repressor activity
5488	3.7291E-9	5.9853E-7	527	5881	678	8649	binding
8134	1.9686E-8	2.5155E-6	59	356	678	8649	transcription factor binding
3676	2.3510E-8	2.5155E-6	138	1132	678	8649	nucleic acid binding
3702	4.2668E-7	3.9133E-5	36	189	678	8649	RNA polymerase II transcription factor activity
3677	6.9153E-7	5.5495E-5	104	836	678	8649	DNA binding
3924	1.5696E-6	1.1197E-4	27	128	678	8649	GTPase activity
5515	2.0281E-6	1.3020E-4	430	4753	678	8649	protein binding
4722	1.2510E-5	7.3011E-4	11	31	678	8649	protein serine/threonine phosphatase activity
16462	1.7399E-5	9.3085E-4	42	274	678	8649	pyrophosphatase activity
16817	2.0819E-5	9.5472E-4	42	276	678	8649	hydrolase activity, acting on acid anhydrides
16818	2.0819E-5	9.5472E-4	42	276	678	8649	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
17111	2.3087E-5	9.8813E-4	40	259	678	8649	nucleoside-triphosphatase activity
3712	3.3140E-5	1.3298E-3	40	263	678	8649	transcription cofactor activity
3714	4.9684E-5	1.8763E-3	20	97	678	8649	transcription corepressor activity
4674	1.3454E-3	4.7985E-2	30	215	678	8649	protein serine/threonine kinase activity

C.2 GO-term analysis for genes containing uAUGs and not targeted by miRNAs

C.2 Genes that do not exhibit Watson-Crick interactions with 3′-ends of conserved miRNAs (Total 716)

ABI2	*ARMC8*	*C12orf30*	*CDH2*	*CTR9*	*DYNLL2*	*FGF9*
ABLIM1	*ARMCX3*	*C14orf4*	*CDH7*	*CTTNBP2*	*E2F4*	*FGFR1OP*
ACSL4	*ARPC1B*	*C14orf43*	*CDK5*	*CTTNBP2NL*	*EED*	*FHIT*
ACTR2	*ARPC2*	*C17orf28*	*CDK8*	*CUGBP1*	*EFCBP2*	*FLI1*
ACTR3	*ASF1A*	*C1GALT1C1*	*CDKN2C*	*CUL1*	*EHD1*	*FLOT1*
ACTR6	*ASTN1*	*C1orf58*	*CEBPE*	*CUL3*	*EIF2C3*	*FLOT2*
ACVR1B	*ATAD1*	*C1orf77*	*CENTG2*	*CXCL14*	*EIF2S1*	*FOSB*
ADAMTS10	*ATN1*	*C20orf52*	*CFL2*	*CXXC5*	*ELF1*	*FOSL2*
AFF4	*ATP11B*	*C5orf13*	*CHCHD3*	*DAZAP2*	*ENC1*	*FOXB1*
AHCYL1	*ATP1B2*	*C6orf62*	*CHD6*	*DCUN1D1*	*ENOX2*	*FOXD2*
AIFM1	*ATP2A1*	*C7orf23*	*CHMP4B*	*DDEF1*	*ENSA*	*FOXM1*
AIP	*ATP6V1C1*	*C9orf126*	*CITED1*	*DDX3X*	*EPB41L4A*	*FOXN2*
AK3L1	*B3GAT1*	*CABP7*	*CKS1B*	*DDX6*	*EPB41L4B*	*FOXP2*
AKAP10	*B3GAT3*	*CACNA2D1*	*CLPTM1*	*DENND4A*	*EPC2*	*FRAP1*
AKT3	*B4GALT5*	*CACNB3*	*CNBP*	*DES*	*EPHA7*	*FRMD5*
ALDH9A1	*B4GALT7*	*CACNG2*	*CNN2*	*DHX30*	*ERH*	*G3BP2*
ALX4	*BAI2*	*CAMK1*	*CNOT10*	*DHX40*	*ETS1*	*GABRA1*
AMBN	*BCL2L1*	*CAMK4*	*CNOT2*	*DHX9*	*ETV3*	*GABRA4*
AMELX	*BCL2L11*	*CAMKK2*	*CNOT3*	*DIS3L*	*ETV6*	*GABRB1*
AMMECR1	*BCL7A*	*CAMTA2*	*CNOT6*	*DLG1*	*EWSR1*	*GALNT1*
ANKMY2	*BEX1*	*CAPN6*	*CNOT6L*	*DLG3*	*EXOC4*	*GALNT5*
ANKRD11	*BLCAP*	*CAPZA2*	*COL1A2*	*DMRT3*	*EXOSC1*	*GALNTL4*
ANXA1	*BMI1*	*CASC3*	*COL4A3BP*	*DMTF1*	*EZH2*	*GAP43*
AP1S2	*BMP7*	*CCBL2*	*COPG2*	*DNAJA4*	*FAM108B1*	*GARNL4*
AP2M1	*BNC2*	*CCDC130*	*COPZ2*	*DNAJB4*	*FAM126B*	*GBF1*
AP3M1	*BNIP2*	*CCDC55*	*CPEB4*	*DNAJC11*	*FAM134A*	*GBX2*
AP3S1	*BRD7*	*CCM2*	*CPS1*	*DNAJC6*	*FBLN5*	*GDI2*
APBB2	*BRDT*	*CCNG1*	*CREBBP*	*DOCK3*	*FBXL2*	*GGA1*
APBB3	*BTBD14A*	*CDC2L6*	*CRISPLD1*	*DPF2*	*FBXL5*	*GJC2*
APPL1	*BTF3*	*CDC42BPB*	*CRY1*	*DPH5*	*FBXO36*	*GLCE*
ARHGAP30	*BTG2*	*CDC73*	*CSNK1G2*	*DSCR3*	*FBXO46*	*GLRA2*
ARHGEF7	*BUB3*	*CDGAP*	*CSNK1G3*	*DTD1*	*FBXW7*	*GLTP*
ARL2	*BZW1*	*CDH11*	*CTCF*	*DYNC1LI1*	*FGF12*	*GNA13*
ARL8B	*C11orf73*	*CDH13*	*CTPS*	*DYNC1LI2*	*FGF17*	*GNAQ*

GNB2
GNB5
GNPDA2
GOLGA1
GPC3
GRASP
GRB2
GRHL2
GRIK5
GRIN3A
GRIP1
GRK5
GSK3B
GTPBP2
H1F0
H1FX
H2AFZ
HAS2
HBP1
HDGF
HDLBP
HESX1
HGD
HIPK1
HIPK3
HIST2H2AA3
HMGB2
HNRNPA2B1
HNRPF
HNRPH2
HNRPM
HNT
HOMER1
HOMER2
HOXA10
HOXA2
HOXA5
HOXA7
HOXB4
HOXB5
HOXB7
HOXB8
HOXB9
HOXC11
HOXC5
HOXC9
HOXD10
HOXD3
HSF2
HSP90AA1
HSPA4
IBSP
IBTK
ID4
IFRD1
IGF2BP1
IGF2BP2
IGSF9
ILF2
IMMP2L
IMP3
ING4
INSM1
INTS12
IRX2
IRX4
ISL1
ISL2
ITGB8
ITPR1
KBTBD7
KCNA6
KCNC1
KCNC4
KCNG3
KCNH5
KCNJ3
KCNK1
KCNMB2
KCTD4
KDELR1
KDELR2
KIAA0494
KIAA1267
KIF13A
KIF2A
KIFAP3
KLHDC3
KPNA1
KRT16
L3MBTL3
LARP5
LATS2
LCOR
LDLRAD3
LENG1
LHFPL2
LHX1
LHX2
LIM2
LINGO2
LLGL1
LMBR1L
LMBRD2
LRFN5
LRP1
LRP1B
LRRC42
LRRC59
LRRC8D
LRRTM1
LZIC
MAB21L1
MAB21L2
MACROD1
MAF1
MAFG
MAML2
MAN1A1
MAP2K4
MAP2K6
MAP4K3
MAPK14
MAPK6
MAPK8
MAPK8IP2
MAPRE3
MARCH5
MARCKSL1
MARK3
MBD3
MCTS1
MDH1
MDM1
MECP2
MEF2D
METT10D
MFSD5
MGAT5
MGAT5B
MIER3
MIR16
MLLT10
MOBKL1A
MOBKL3
MORF4L2
MOV10
MRLC2
MSI1
MTCH2
MTDH
MYCBP
MYF6
MYLIP
MYNN
MYOT
NACA2
NALCN
NARG1
NAV1
NCAM1
NCAM2
NCOA5
NDP
NEDD8
NEK9
NEUROD2
NFATC1
NIT1
NKX2-1
NKX2-2
NMNAT2
NMT2
N-PAC
NPTXR
NR2E1
NR2F1
NR2F2
NRAS
NUP93
OAZ1
OAZ3
ODF1
OGDH
OLIG3
ONECUT2
OPHN1
OTP
OTUB2
OTX1
OTX2
OXSR1
PAIP2
PALMD
PAX9
PCDHB8
PCDHGC3
PCGF3
PCMT1
PCMTD1
PDE6D
PDIK1L
PDP2
PFN2
PHC3
PHF14
PHF17
PHF20
PHIP
PHOX2B
PITPNA
PJA2
POLG
POLR2J2
POLS
POU2AF1
PPARG
PPARGC1B
PPL
PPM1E
PPM1H
PPP1CA
PPP1CB
PPP2R2C
PPP2R4
PPP5C
PPRC1
PPTC7
PRKAA2
PRKACA
PRKCB1
PRKCI
PRPSAP2
PSIP1
PSMA1
PSMA7
PSMC6
PSMD11
PSMD14
PSMD3
PTBP2
PTP4A2
PTPN11
PTPN12
PTPN4
PTPRG
PUM1
PURA
PURB
PVRL1
QKI
RAB11A
RAB1B
RAB2B
RAB30
RAB6B
RAB8A
RABL3
RAF1
RANBP2
RAP2B
RAP2C
RASIP1
RBBP6
RBBP7
RBM35B
RBM4B
RBM5
RBPMS
RDH10
REEP2
RFC1
RGS3
RLF
RNF122
RNF20
RNPS1
RPL10A
RPL15
RPL27A

RPL30
RPL31
RPN2
RPS8
RRAGC
SAP18
SAPS3
SAR1A
SARS
SAT1
SATB1
SEMA6A
SEPT5
SERBP1
SFRS10
SFRS12
SFRS6
SGIP1
SGMS2
SH3BGRL
SH3BGRL3
SHC1
SHOC2
SIRT1
SIX1
SLC10A7
SLC12A5
SLC15A2
SLC22A4
SLC22A5
SLC25A13
SLC25A14
SLC35B4
SLC39A10
SLIT3
SLITRK2
SLITRK5
SLK
SMAD9
SMAP1
SMAP1L
SMARCC2
SMARCE1
SMYD5
SNRPA1
SNRPN
SNX8
SORBS2
SOX17
SOX18
SOX2
SOX6
SOX9
SPARC
SPATA6
SPOP
SPRY2
SPTLC2
SRF
SRGAP1
SRP54
SRPK2
SRR
SSBP3
STAG2
STC1
STK32B
STK39
STK40
STMN2
STOML2
STRBP
STT3A
STUB1
STX5
STX8
STXBP6
SUMO1
SUMO3
SUZ12
SYN2
SYT14
SYT9
TACC2
TAF12
TAF5
TAOK1
TBC1D19
TBC1D22B
TCEB3
TCF20
TCF4
TFAP2B
TFAP2C
TFAP2D
TFDP2
TGFB3
THAP7
TIAL1
TLK1
TMED1
TMEM108
TMEM135
TMEM164
TMEM185A
TMEM39B
TMEM59
TMEM60
TMEM93
TMSB4X
TNFAIP3
TOP1
TOX
TPH2
TPM3
TRAIP
TRAPPC3
TRAPPC6B
TRDN
TRIB2
TRIP4
TRPM7
TSN
TSPAN15
TUB
TUBB6
TUBG2
TUSC2
TXNDC11
U2AF1L4
UBE2D3
UBE2L3
UBE2N
UBE2S
UBE2V2
UBE3A
UBN1
UBXD8
UGP2
USP25
USP28
USP5
USP8
UTX
VCPIP1
VIM
VKORC1L1
VPRBP
VPS25
VPS26A
VPS54
VPS72
WBP2
WDFY3
WDR1
WDR48
WDR67
WDR68
WNT10B
WNT3
WNT7A
WNT7B
WT1
WWP2
XKR6
XPNPEP1
XPO7
XYLT1
YBX1
YTHDF2
YTHDF3
YWHAG
ZBTB1
ZBTB10
ZBTB12
ZBTB16
ZBTB26
ZBTB37
ZBTB39
ZBTB7A
ZCCHC5
ZDHHC23
ZEB1
ZEB2
ZHX1
ZMAT2
ZMYM3
ZMYND8
ZNF24
ZNF275
ZNF281
ZNF384
ZNF428
ZNF496
ZNF512
ZNF513
ZNF706
ZNF76

C.2.2 GO-term analysis

Number of genes among 716 (above) with available annotations = Number of genes in test set (X) = 448
Number of genes in reference set (N) = 8648
x indicates the number of genes in test set in the functional category listed
n indicates the number of genes in reference set in the functional category listed
p-values calculated using hypergeometric test.

GO-ID	p-value	corr p-value	x	n	X	N	Description Genes in test set
30528	9.1528E-8	4.9059E-5	84	919	448	8648	transcription regulator activity
5488	2.2855E-7	6.1252E-5	352	5879	448	8648	binding
5515	3.5780E-7	6.3927E-5	297	4753	448	8648	protein binding
3676	2.2083E-6	2.9592E-4	93	1131	448	8648	nucleic acid binding
3700	1.2579E-5	1.3485E-3	50	519	448	8648	transcription factor activity
3677	5.0034E-5	4.4697E-3	69	835	448	8648	DNA binding
48027	1.3814E-4	1.0578E-2	3	3	448	8648	mRNA 5'-UTR binding
3723	4.3415E-4	2.9088E-2	29	288	448	8648	RNA binding
8134	7.5201E-4	4.0469E-2	33	355	448	8648	transcription factor binding
3702	7.5501E-4	4.0469E-2	21	189	448	8648	RNA polymerase II transcription factor activity

C.3 Predicted interactions between uAUG 6 and 7 of *KLF9* and conserved miRNAs

Shown below is a portion of the 5′-UTR of *KLF9* that contain uAUGs6 and 5. The interactions and the associated free energy of binding were predicted using 11-mers and miRNA 5′- or 3′-ends. Full miRNAs and sequences surrounding the uAUGs are shown for clarity.

uAUG6 *uAUG7*

641 5′–UUUGGUUUGUGACGUG**AUG**GGAUUCUGCGAAAUUGUUACUGAGCAAGAGA**AUG**CCGGAACGUGCGGAC–3′ 708

```
target 5' G       UCUGCGA 3'       (uAUG6)
          AUGGGAU
          ||||||:            ΔG: -14.2 kcal/mol
          UACCCUG
miRNA  3'         UAGGAUGUAUACGUU 5' hsa-miR-448
```

```
target 5' C          UCUG 3'       (uAUG6)
          GUGAUGGGAU
          ||:||||||:         ΔG: -16.5 kcal/mol
          CAUUACCCUG
miRNA  3'            GAAGGAGAAAC 5' hsa-miR-583
```

```
target 5' UUGUGACGUG        C 3'   (uAUG6)
                  AUGGGAUU
                  ||||||||  ΔG: -14.7 kcal/mol
                  UACCCUAA
miRNA  3' UCUUCCGUGG        AU 5' hsa-miR-605
```

```
target 5' C U       UUCUGCGA 3'   (uAUG6)
           G GAUGGGA
           | ||||:||       ΔG: -14.4 kcal/mol
           C CUACUCU
miRNA  3' U U       CUUUGUGGGA 5' hsa-miR-609
```

```
target 5'   C       GAUUCUGCGA 3'  (uAUG6)
             GUGAUGG
             |||||||       ΔG: -16.1 kcal/mol
             CACUACC
miRNA  3' UUC       AGUCGUCUGUAU 5'
                                   hsa-miR-654-3p
```

```
target 5' GAGA       AACGUGCG  3'  (uAUG7)
               AUGCCGG
               |||||:|      ΔG: -14.4 kcal/mol
               UACGGUC
miRNA  3' UCGA       GUAGAACGGA 5' hsa-miR-31
```

Bibliography

Avni, D., Biberman, Y., and Meyuhas, O. 1997. The 5' terminal oligopyrimidine tract confers translational control on TOP mRNAs in a cell type- and sequence context-dependent manner. *Nucleic Acids Res* **25**(5): 995-1001.

Babak, T., Zhang, W., Morris, Q., Blencowe, B.J., and Hughes, T.R. 2004. Probing microRNAs with microarrays: tissue specificity and functional inference. *Rna* **10**(11): 1813-1819.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. 2008. The impact of microRNAs on protein output. *Nature*.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* **122**(4): 553-563.

Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y. et al. 2004. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res* **14**(12): 2486-2494.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**(2): 281-297.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**(6818): 363-366.

Bieker, J.J. 2001. Kruppel-like factors: three fingers in many pies. *J Biol Chem* **276**(37): 34355-34358.

Birney, E. Stamatoyannopoulos, J.A. Dutta, A. Guigo, R. Gingeras, T.R. Margulies, E.H. Weng, Z. Snyder, M. Dermitzakis, E.T. Thurman, R.E. et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**(7146): 799-816.

Black, A.R., Black, J.D., and Azizkhan-Clifford, J. 2001. Sp1 and kruppel-like factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol* **188**(2): 143-160.

Borer, P.N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O.C. 1974. Stability of ribonucleic acid double-stranded helices. *J Mol Biol* **86**(4): 843-853.

Bracht, J., Hunter, S., Eachus, R., Weeks, P., and Pasquinelli, A.E. 2004. Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. *Rna* **10**(10): 1586-1594.

Breslauer, K.J., Frank, R., Blocker, H., and Marky, L.A. 1986. Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A* **83**(11): 3746-3750.

Brueckner, B., Stresemann, C., Kuner, R., Mund, C., Musch, T., Meister, M., Sultmann, H., and Lyko, F. 2007. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res* **67**(4): 1419-1423.

Bushati, N. and Cohen, S.M. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23**: 175-205.

Cai, X., Hagedorn, C.H., and Cullen, B.R. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**(12): 1957-1966.

Castoldi, M., Schmidt, S., Benes, V., Noerholm, M., Kulozik, A.E., Hentze, M.W., and Muckenthaler, M.U. 2006. A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). *Rna* **12**(5): 913-920.

Cerutti, H. and Casas-Mollano, J.A. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet* **50**(2): 81-99.

Chapman, E.J. and Carrington, J.C. 2007. Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* **8**(11): 884-896.

Chen, J., Lozach, J., Garcia, E.W., Barnes, B., Luo, S., Mikoulitch, I., Zhou, L., Schroth, G., and Fan, J.B. 2008. Highly sensitive and specific microRNA expression profiling using BeadArray technology. *Nucleic Acids Res* **36**(14): e87.

Churbanov, A., Rogozin, I.B., Babenko, V.N., Ali, H., and Koonin, E.V. 2005. Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes. *Nucleic Acids Res* **33**(17): 5512-5520.

Croce, C.M. 2008. MicroRNAs and lymphomas. *Ann Oncol* **19 Suppl 4**: iv39-40.

Deo, M., Yu, J.Y., Chung, K.H., Tippens, M., and Turner, D.L. 2006. Detection of mammalian microRNA expression by in situ hybridization with RNA oligonucleotides. *Dev Dyn* **235**(9): 2538-2548.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**(5338): 680-686.

Devoe, H. and Tinoco, I., Jr. 1962. The stability of helical polynucleotides: base contributions. *J Mol Biol* **4**: 500-517.

Doench, J.G. and Sharp, P.A. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev* **18**(5): 504-511.

Du, T. and Zamore, P.D. 2005. microPrimer: the biogenesis and function of microRNA. *Development* **132**(21): 4645-4652.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**(6836): 494-498.

Engels, B.M. and Hutvagner, G. 2006. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene* **25**(46): 6163-6169.

Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. 2003. MicroRNA targets in Drosophila. *Genome Biol* **5**(1): R1.

Farazi, T.A., Juranek, S.A., and Tuschl, T. 2008. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* **135**(7): 1201-1214.

Farh, K.K., Grimson, A., Jan, C., Lewis, B.P., Johnston, W.K., Lim, L.P., Burge, C.B., and Bartel, D.P. 2005. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310**(5755): 1817-1821.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **9**(2): 102-114.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669): 806-811.

Forman, J.J., Legesse-Miller, A., and Coller, H.A. 2008. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc Natl Acad Sci U S A* **105**(39): 14879-14884.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., and Turner, D.H. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *Proc Natl Acad Sci U S A* **83**(24): 9373-9377.

Fulci, V., Chiaretti, S., Goldoni, M., Azzalin, G., Carucci, N., Tavolaro, S., Castellano, L., Magrelli, A., Citarella, F., Messina, M. et al. 2007. Quantitative technologies establish a novel microRNA profile of chronic lymphocytic leukemia. *Blood* **109**(11): 4944-4951.

Gaba, A., Wang, Z., Krishnamoorthy, T., Hinnebusch, A.G., and Sachs, M.S. 2001. Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *Embo J* **20**(22): 6453-6463.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**(5770): 75-79.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**(Database issue): D154-158.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**(1): 91-105.

Guo, Z., Liu, Q., and Smith, L.M. 1997. Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nat Biotechnol* **15**(4): 331-335.

Halbeisen, R.E., Galgano, A., Scherrer, T., and Gerber, A.P. 2008. Post-transcriptional gene regulation: from genome-wide studies to principles. *Cell Mol Life Sci* **65**(5): 798-813.

Hannon, G.J. and Rossi, J.J. 2004. Unlocking the potential of the human genome with RNA interference. *Nature* **431**(7006): 371-378.

Haynie, D. 2001. *Biological Thermodynamics*. Cambridge University Press.

Hohjoh, H. and Fukushima, T. 2007. Marked change in microRNA expression during neuronal differentiation of human teratocarcinoma NTera2D1 and mouse embryonal carcinoma P19 cells. *Biochem Biophys Res Commun* **362**(2): 360-367.

Humphreys, D.T., Westman, B.J., Martin, D.I., and Preiss, T. 2005. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci U S A* **102**(47): 16961-16966.

Hutton, J.R. 1977. Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. *Nucleic Acids Res* **4**(10): 3537-3555.

Iacono, M., Mignone, F., and Pesole, G. 2005. uAUG and uORFs in human and rodent 5'untranslated mRNAs. *Gene* **349**: 97-105.

Imataka, H., Nakayama, K., Yasumoto, K., Mizuno, A., Fujii-Kuriyama, Y., and Hayami, M. 1994. Cell-specific translational control of transcription factor BTEB expression. The role of an upstream AUG in the 5'-untranslated region. *J Biol Chem* **269**(32): 20668-20673.

Jay, C., Nemunaitis, J., Chen, P., Fulgham, P., and Tong, A.W. 2007. miRNA profiling for diagnosis and prognosis of human cancer. *DNA Cell Biol* **26**(5): 293-300.

Jin, X., Turcott, E., Englehardt, S., Mize, G.J., and Morris, D.R. 2003. The two upstream open reading frames of oncogene mdm2 have different translational regulatory properties. *J Biol Chem* **278**(28): 25716-25721.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. 2004. Human MicroRNA targets. *PLoS Biol* **2**(11): e363.

Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., and Slack, F.J. 2005. RAS is regulated by the let-7 microRNA family. *Cell* **120**(5): 635-647.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. 2006. MicroRNAS and their regulatory roles in plants. *Annu Rev Plant Biol* **57**: 19-53.

Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., and Sarnow, P. 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309**(5740): 1577-1581.

Jousse, C., Bruhat, A., Carraro, V., Urano, F., Ferrara, M., Ron, D., and Fafournoux, P. 2001. Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the chop 5'UTR. *Nucleic Acids Res* **29**(21): 4341-4351.

Khvorova, A., Reynolds, A., and Jayasena, S.D. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**(2): 209-216.

Kim, V.N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* **6**(5): 376-385.

Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. 2004. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* **18**(10): 1165-1178.

Kloosterman, W.P. and Plasterk, R.H. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell* **11**(4): 441-450.

Kloosterman, W.P., Wienholds, E., Ketting, R.F., and Plasterk, R.H. 2004. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res* **32**(21): 6284-6291.

Kong, Y.W., Cannell, I.G., de Moor, C.H., Hill, K., Garside, P.G., Hamilton, T.L., Meijer, H.A., Dobbyn, H.C., Stoneley, M., Spriggs, K.A. et al. 2008. The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. *Proc Natl Acad Sci U S A* **105**(26): 8866-8871.

Kozak, M. 1991a. An analysis of vertebrate mRNA sequences: intimations of translational control. *J Cell Biol* **115**(4): 887-903.

-. 1991b. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* **266**(30): 19867-19870.

-. 2002. Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**(1-2): 1-34.

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. et al. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37**(5): 495-500.

Kruger, J. and Rehmsmeier, M. 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res* **34**(Web Server issue): W451-454.

Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**(7068): 685-689.

Kwon, H.S., Lee, D.K., Lee, J.J., Edenberg, H.J., Ahn, Y.H., and Hur, M.W. 2001. Posttranscriptional regulation of human ADH5/FDH and Myf6 gene expression by upstream AUG codons. *Arch Biochem Biophys* **386**(2): 163-171.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**(9): 735-739.

Lai, E.C. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**(4): 363-364.

Lander, E.S. Linton, L.M. Birren, B. Nusbaum, C. Zody, M.C. Baldwin, J. Devon, K. Dewar, K. Doyle, M. FitzHugh, W. et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M. et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**(7): 1401-1414.

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

Lawrie, C.H., Saunders, N.J., Soneji, S., Palazzo, S., Dunlop, H.M., Cooper, C.D., Brown, P.J., Troussard, X., Mossafa, H., Enver, T. et al. 2008. MicroRNA expression in lymphocyte development and malignancy. *Leukemia* **22**(7): 1440-1446.

Lee, I., Ajay, S.S., Chen, H., Maruyama, A., Wang, N., McInnis, M.G., and Athey, B.D. 2008. Discriminating single-base difference miRNA expressions using microarray Probe Design Guru (ProDeG). *Nucleic Acids Res* **36**(5): e27.

Lee, I., Dombkowski, A.A., and Athey, B.D. 2004. Guidelines for incorporating non-perfectly matched oligonucleotides into target-specific hybridization probes for a DNA microarray. *Nucleic Acids Res* **32**(2): 681-690.

Lee, R.C. and Ambros, V. 2001. An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**(5543): 862-864.

Lee, R.C., Feinbaum, R.L., and Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**(5): 843-854.

Lewis, B.P., Burge, C.B., and Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**(1): 15-20.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. 2003. Prediction of mammalian microRNA targets. *Cell* **115**(7): 787-798.

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. 2005. Microarray analysis shows that some

microRNAs downregulate large numbers of target mRNAs. *Nature* **433**(7027): 769-773.

Lim, L.P. and Linsley, P.S. 2007. Mustering the micromanagers. *Nat Biotechnol* **25**(9): 996-997.

Liolios, K., Mavromatis, K., Tavernarakis, N., and Kyrpides, N.C. 2008. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* **36**(Database issue): D475-479.

Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M. et al. 2004. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci U S A* **101**(26): 9740-9744.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* **435**(7043): 834-838.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. 2004. Nuclear export of microRNA precursors. *Science* **303**(5654): 95-98.

Lytle, J.R., Yario, T.A., and Steitz, J.A. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5' UTR as in the 3' UTR. *Proc Natl Acad Sci U S A* **104**(23): 9667-9672.

Martinez, N.J., Ow, M.C., Barrasa, M.I., Hammell, M., Sequerra, R., Doucette-Stamm, L., Roth, F.P., Ambros, V.R., and Walhout, A.J. 2008. A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev* **22**(18): 2535-2549.

Moore, M.J. 2005. From birth to death: the complex lives of eukaryotic mRNAs. *Science* **309**(5740): 1514-1518.

Morris, D.R. and Geballe, A.P. 2000. Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**(23): 8635-8642.

Moss, E.G. 2007. Heterochronic genes and the nature of developmental time. *Curr Biol* **17**(11): R425-434.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**(5881): 1344-1349.

Neitzel, H. 1986. A routine method for the establishment of permanent growing lymphoblastoid cell lines. *Hum Genet* **73**(4): 320-326.

Nelson, P.T., Baldwin, D.A., Scearce, L.M., Oberholtzer, J.C., Tobias, J.W., and Mourelatos, Z. 2004. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods* **1**(2): 155-161.

Nikolcheva, T., Pyronnet, S., Chou, S.Y., Sonenberg, N., Song, A., Clayberger, C., and Krensky, A.M. 2002. A translational rheostat for RFLAT-1 regulates RANTES expression in T lymphocytes. *J Clin Invest* **110**(1): 119-126.

Nottrott, S., Simard, M.J., and Richter, J.D. 2006. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol* **13**(12): 1108-1114.

Orom, U.A., Nielsen, F.C., and Lund, A.H. 2008. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* **30**(4): 460-471.

Orphanides, G. and Reinberg, D. 2002. A unified theory of gene expression. *Cell* **108**(4): 439-451.

Peters, L. and Meister, G. 2007. Argonaute proteins: mediators of RNA silencing. *Mol Cell* **26**(5): 611-623.

Petersen, C.P., Bordeleau, M.E., Pelletier, J., and Sharp, P.A. 2006. Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* **21**(4): 533-542.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**(5740): 1573-1576.

Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Staehler, P., Beier, M., and Tautz, D. 2006. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res* **34**(9): e66.

Rajewsky, N. 2006. microRNA target predictions in animals. *Nat Genet* **38 Suppl**: S8-13.

Rajewsky, N. and Socci, N.D. 2004. Computational identification of microRNA targets. *Dev Biol* **267**(2): 529-535.

Raney, A., Baron, A.C., Mize, G.J., Law, G.L., and Morris, D.R. 2000. In vitro translation of the upstream open reading frame in the mammalian mRNA encoding S-adenosylmethionine decarboxylase. *J Biol Chem* **275**(32): 24444-24450.

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. 2004. Fast and effective prediction of microRNA/target duplexes. *Rna* **10**(10): 1507-1517.

Richards, E.J. and Elgin, S.C. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**(4): 489-500.

SantaLucia, J., Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95**(4): 1460-1465.

SantaLucia, J., Jr., Allawi, H.T., and Seneviratne, P.A. 1996. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**(11): 3555-3562.

SantaLucia, J., Jr. and Hicks, D. 2004. The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* **33**: 415-440.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**(5235): 467-470.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**(7209): 58-63.

Sethupathy, P., Megraw, M., and Hatzigeorgiou, A.G. 2006. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat Methods* **3**(11): 881-886.

Soifer, H.S., Rossi, J.J., and Saetrom, P. 2007. MicroRNAs in Disease and Potential Therapeutic Applications. *Mol Ther*.

Song, A., Nikolcheva, T., and Krensky, A.M. 2000. Transcriptional regulation of RANTES expression in T lymphocytes. *Immunol Rev* **177**: 236-245.

Song, K.Y., Hwang, C.K., Kim, C.S., Choi, H.S., Law, P.Y., Wei, L.N., and Loh, H.H. 2007. Translational repression of mouse mu opioid receptor expression via leaky scanning. *Nucleic Acids Res* **35**(5): 1501-1513.

Sperling, S. 2007. Transcriptional regulation at a glance. *BMC Bioinformatics* **8 Suppl 6**: S2.

Takada, S., Berezikov, E., Yamashita, Y., Lagos-Quintana, M., Kloosterman, W.P., Enomoto, M., Hatanaka, H., Fujiwara, S., Watanabe, H., Soda, M. et al. 2006. Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Res* **34**(17): e115.

Thomson, J.M., Parker, J., Perou, C.M., and Hammond, S.M. 2004. A custom microarray platform for analysis of microRNA gene expression. *Nat Methods* **1**(1): 47-53.

Tolia, N.H. and Joshua-Tor, L. 2007. Slicer and the argonautes. *Nat Chem Biol* **3**(1): 36-43.

Turner, D.H. 1996. Thermodynamics of base pairing. *Curr Opin Struct Biol* **6**(3): 299-304.

Uhlenbeck, O.C., Borer, P.N., Dengler, B., and Tinoco, I., Jr. 1973. Stability of RNA hairpin loops: A 6 -C m -U 6. *J Mol Biol* **73**(4): 483-496.

Valoczi, A., Hornyik, C., Varga, N., Burgyan, J., Kauppinen, S., and Havelda, Z. 2004. Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. *Nucleic Acids Res* **32**(22): e175.

van Rooij, E. and Olson, E.N. 2007. MicroRNAs: powerful new regulators of heart disease and provocative therapeutic targets. *J Clin Invest* **117**(9): 2369-2376.

Venter, J.C. Adams, M.D. Myers, E.W. Li, P.W. Mural, R.J. Sutton, G.G. Smith, H.O. Yandell, M. Evans, C.A. Holt, R.A. et al. 2001. The sequence of the human genome. *Science* **291**(5507): 1304-1351.

Wang, B., Yanez, A., and Novina, C.D. 2008. MicroRNA-repressed mRNAs contain 40S but not 60S components. *Proc Natl Acad Sci U S A* **105**(14): 5343-5348.

Wang, H., Ach, R.A., and Curry, B. 2007. Direct and sensitive miRNA profiling from low-input total RNA. *Rna* **13**(1): 151-159.

Wang, L. and Wessler, S.R. 1998. Inefficient reinitiation is responsible for upstream open reading frame-mediated translational repression of the maize R gene. *Plant Cell* **10**(10): 1733-1746.

Watanabe, T., Takeda, A., Mise, K., Okuno, T., Suzuki, T., Minami, N., and Imai, H. 2005. Stage-specific expression of microRNAs during Xenopus development. *FEBS Lett* **579**(2): 318-324.

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

Wuite, G.J., Smith, S.B., Young, M., Keller, D., and Bustamante, C. 2000. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature* **404**(6773): 103-106.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**(7031): 338-345.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**(24): 3011-3016.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.J., and Qi, Y. 2007. A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. *Genes Dev* **21**(10): 1190-1203.

Zhao, Y. and Srivastava, D. 2007. A developmental view of microRNA function. *Trends Biochem Sci* **32**(4): 189-197.

Zimmermann, T.S., Lee, A.C., Akinc, A., Bramlage, B., Bumcrot, D., Fedoruk, M.N., Harborth, J., Heyes, J.A., Jeffs, L.B., John, M. et al. 2006. RNAi-mediated gene silencing in non-human primates. *Nature* **441**(7089): 111-114.