Role of post-transcriptional regulators in the establishment and maintenance of cell identity

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Afzal Pasha Syed
aus Indien

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Mihaela Zavolan
fakultät verantwortliche und Dissertationsleiterin

Prof. Gerhard Christofori
Korreferent

Basel , den 19.05.2015

Prof. Dr. Jörg Schibler
Dekan

Abstract

Regulation of gene expression, which is essential for the unfolding of all processes taking place in multicellular organisms, is very complex. Gene expression is controlled at the level of transcription, RNA processing and localization, translation and protein modification and decay. Among the various post-transcriptional regulators of gene expression, microRNA (miRNA)s contribute to the maintenance of gene expression patterns among various cell types in an organism. miRNAs are small, evolutionarily conserved non-protein coding RNAs, whose biogenesis involves multiple steps in the nucleus and cytoplasm of the cell. So far 35,828 miRNAs have been reported from 233 species and in humans they are present as one of the abundant gene families comprising over 2500 miRNAs. Mature miRNAs are loaded into Argonaute (AGO) proteins to form RNA induced silencing complexes (RISC), which find their targets via nucleotide complementarity between sites mostly present in 3' untranslated region (3'UTR)s of mRNAs and miRNAs. The outcome is destabilization or translational repression of the miRNA targets. Although the components of miRNA biogenesis are relatively well characterized, the mechanisms through which miRNAs execute their functional activities remain less understood. In the first chapter of this thesis, we have addressed two important aspects of miRNA mediated gene regulation.

Differential expression analysis based on high-throughput data sets generated upon modulating the expression of a given miRNA in a given model has helped to identify miRNA targets. Many computational target prediction models have been proposed. They are typically trained on high-throughput data sets, and are based on few parameters such as seed complementarity of targets, evolutionary conservation etc. Validation of predicted miRNA targets remains non-trivial and we believe that one reason could be lack of methods that consider the miRNA activity at multiple levels. An aspect that has been largely ignored so far is the time scale on which miRNAs regulate their targets. In one study we have addressed the kinetics aspects of miRNA regulation, and proposed a model that takes these aspects into account. The parameters of this model were inferred from a variety of low and high-throughput experimental data sets and we found that the model well describes the time dependent changes in the level of mRNA, proteins and ribosome density levels upon miRNA transfection and induction. We also found that miRNAs may not generally act as fast regulators of gene expression due to two bottlenecks, one is the miRNA loading into Argonaute proteins and the other is the rate of protein decay. These influence the time-scale and magnitude of miRNA mediated gene regulation.

Several recent studies have indicated that the miRNA binding sites present in the coding region (CDS) of an mRNA are functional, but their implications remain unclear. Use of high-throughput approaches such as cross-linking and immunoprecipitation (CLIP) to isolate AGO bound target sites indicate that there are as many sites located in CDS as in 3'UTRs. The second study presented in this thesis concerns itself with the function of coding region-located miRNA binding sites. Using published high-throughput data sets of Argonaute CLIP and ribosome protected fragment profiles upon miRNA transfections, we have shown that miRNA binding sites that are located in CDS and 3'UTRs have co-evolved and have similar sequence and structure properties. We also found that the miRNA binding sites located in CDS are capable of inhibiting translation, while those located in 3'UTR are more efficient in triggering mRNA degradation. This particular observation was validated experimentally using an inducible miRNA expression cell line, and with a luciferase reporter system containing CDS located binding sites of the cognate miRNA. Our study therefore suggests that miRNAs may co-target CDS and 3'UTR to fine-tune the time scale and magnitude of the post-transcriptional regulatory effect imposed by them.

Recent studies reported that miRNAs from miR302/367 cluster enhance the somatic cell reprogramming induced with embryonic stem cell (ESC) specific transcription factors: OCT4, SOX2, KLF4 and c-MYC. Few other reports also claimed that miR-302/367 cluster alone is enough to reprogram somatic cells to induced pluripotent stem cell (iPSC)s. However, the mechanisms underlying the miRNA mediated reprogramming are not clear. We tried to establish the miR-302/367 mediated reprogramming of fibroblasts in the primary mouse embryonic fibroblast (MEF)s, but, similar to other labs, we were unable to reproduce the initial result. However, we have succeeded in enhancing the reprogramming efficiency in the secondary transgenic mouse embryonic fibroblasts (TNG)-MEFs that contained a pluripotency marker "Nanog" tagged with a green fluorescent protein, along with miR-302/367. Apart from miRNAs, other post-transcriptional regulators we have focused in this thesis are tissue specific splicing factors. Recent evidence indicated that knockdown of muscle blind like (MBNL) proteins enhance the reprogramming efficiency. Our own analysis of already published mRNA-seq data sets of iPSCs and their parental cells also showed that the tissue specific splicing factors are differentially expressed between iPSCs and their parental cells. Especially ESRP1, ESRP2 and MSI1 showed striking changes in their expression levels during the course of reprogramming. Based on this analysis we hypothesized that these factors may enhance the reprogramming efficiency. To investigate this hypothesis we again used secondary TNG-MEFs as a model and we have transduced them with both lentiviruses and retroviruses as carriers to deliver our candidate splicing factors. Our experiment indeed revealed an increase in the reprogramming efficiency of 1.4 fold to 2 fold, with ESRP2 showing highest enhancement. As a follow up of these experiments, we aim to

decipher the cascade of events through which miR-302/367 and splicing factors enhance the efficiency of reprogramming induced by the ESC set of transcription factors. Since the major changes involved in reprogramming occur during the early and late phases, we plan to perform an early and late time series of mRNA profiling upon the induction of reprogramming with miR-302/367 and splicing factors in secondary TNG-MEFs.

In conclusion this thesis presents two main contributions to the field of miRNA-based regulation of gene expression. The first is a mathematical model that describes the kinetics of miRNA dependent gene regulation and the second shows that the miRNA binding sites located in CDS sites are functional and are more effective in inhibiting translation than the sites present in 3'UTR. Besides these two studies, I have obtained evidence that tissue specific splicing factors, in particular ESRP1, ESRP2 and MSI1 are able to enhance reprogramming efficiency up to 2 fold. Experiments are under way to uncover the mechanisms involved in the enhancement of reprogramming efficiency by tissue specific factors.

Since relatively little is known about the function of alternative splice forms in iPSC generation, these preliminary studies could set the ground for future research in iPSC and also towards clinical research. Being able to obtain iPSCs with more efficient and safer methods will enable studies of various diseases at the clinical level. On the other side, as miRNAs are currently being considered for various therapeutic approaches, a deeper understanding of the underlying mechanisms by which miRNAs regulate gene expression would help in the better design of therapeutic compounds. The work presented in this thesis may thus be beneficial for both the miRNA and the iPSC fields.

Acknowledgements

I am grateful to Prof. Mihaela Zavolan for her incredible support and constant encouragement all through my PhD studies. I am always amazed by her availability and patience. I am also thankful to her for pushing me in some fruitful projects, which are presented in this thesis and allowing me to explore the projects with the way I wanted to.

My sincere thanks to Dr. Jean Hausser for sharing his ideas and projects those were eventually successful. It was always exciting to discuss science with you.

Heartfelt thanks to Drs. Shivendra Kishore and Lukasz jaskeiwicz for their help in the lab and also for always being around to troubleshoot any problem. You both inspired me in many ways. It was a great learning experience.

I am grateful to Dr. Keisuke Kaji for kindly allowing me to work in his lab and teaching me secondary reprogramming. I thank him for his continuous support both in person and by emails, for very useful suggestions and discussions regarding “tissue specific splicing factors and iPSC” project. I would also like to thank members of Kaji lab for helping me in various ways during my stay and for all those friendly chats.

I am thankful to Prof. Gerhard Christofori for readily agreeing to be in my thesis committee.

I am fortunate to be always surrounded by some great friends. Asif, Meraj, Moin, Harish, Kapil, Kumar, Ramesh, my cousins Shahed, Zahed, Javid, and Wajid, thank you to all you wonderful people. Life is always happy with you.

I would like to express my gratitude to Liliane and TMCF, at University of Basel for their help with MEFs.

I am fortunate to be a part of an energetic and enthusiastic group. I thank Yoana, Dominik, Arnau, Alex, Andreas J. Gruber, Aaron, Nitish, and Bea for their useful comments, discussions and making the lab a second home. Special thanks to Georges for his around-the-clock help, wonderful organization of the lab and for proof reading this thesis.

I can never thank my family enough for their unconditional support especially Ammi, Baba, Big B and Afshan, without whom, this could not have been possible. This one is for you all.

Table of contents

Chapter 1

1. Fundamentals of gene regulation

1.1 Introduction

Life on this planet has evolved tremendous variety, from unicellular to multicellular organisms. They span a wide range of sizes, from the smallest microorganisms like parasitic bacteria to the blue whale, which is considered to be one of the largest animals to have existed. Nevertheless, regardless of size, all organisms are composed of fundamental units, which are the cells. Based on morphology, metabolism, and sub-cellular organization, two types of cells are known, prokaryotic or eukaryotic. Most multicellular organisms are eukaryotic, and in general prokaryotes are unicellular, with few exceptions like myxobacteria [1]. There remain many unknowns about the evolutionary transitions from prokaryotic life to eukaryotic life and from uni-cellularity to multi-cellularity. However, it is likely that ability to adapt to changes in environment by expressing different sets of genes and thereby cellular phenotypes was an important factor in this process. The emergence of development is even more remarkable in this respect because it involves "running" coherent gene expression programs that are encoded in the genome, over quite long periods of time.

1.1.1 Gene regulation

The "expression" of the genetic material is what distinguishes different cell types within an individual. During embryonic stages, the pluripotent stem cells that constitute the embryo express the same set of specific transcription factors that are responsible for initiating the transcription of "pluripotency" genes, which maintain the embryonic state. At some point however, through mechanisms that are intensely studied currently, differentiation towards various lineages is initiated. The processes that underlie the change in gene expression at different stages of the development or in response to the intracellular or extracellular environments are denoted by the term "gene regulation" [2]. Essentially every step of gene expression is regulated, including transcription from the DNA template and translation of messenger RNA (mRNA)s into proteins. Cell growth, proliferation, apoptosis, metabolism, differentiation rely on specific regulators to establish the appropriate gene expression patterns. For instance, increased expression of genes such as the polo like kinase1 (*PLK1)* and BUB1 mitotic checkpoint serine/threonine kinase *(BUB1)* are associated with high proliferation, whereas cyclin (CCN)s *CCNE1*, *CCND1* and *CCNB1* take part in cell cycle regulation [3]. A remarkable example of how vital transcriptional gene regulation could be is the induction of pluripotency in somatic cells by

the ectopic expression of pluripotency-associated genes like *OCT4*, *SOX2*, *KLF4* and *c-MYC* [4]. Another classic example is the conversion of fibroblasts to myoblasts by the ectopic expression of *MyoD* [5]. These observations (as well as many others that have emerged since the landmark paper of Yamanaka and colleagues [4]) make the point that differentiation is a reversible process, where the appropriate manipulation of gene expression, via a remarkably small number of transcription factors, can lead to a desired cellular phenotype [4-6].

Not surprisingly, altered gene expression as a result, for instance, of mutations, can lead to pathological conditions [7]. For instance, spinal muscular atrophy [8] is caused by mutations in the SMN1 gene. Somatic mutations that cause misregulation of genes associated with cellular functions like cell growth, proliferation, division, migration and foreign tissue invasions have been observed in various types of cancers [9, 10].

1.1.2 Transcription

Characterizing the mechanisms by which gene expression is regulated has been a most active field of research ever since the central dogma of molecular biology was proposed [11]. As already mentioned, gene expression is regulated at virtually all major steps including transcription, mRNA processing, mRNA stability, transport to the cytoplasm and translation [12]. General transcription factors (GTFs) [13] such as TFIIA, TFIIB, TFIID TFIIE, TFIIF and TFIIH [14] act together with the DNA-dependent RNA polymerase II (RNAP II) to form a basal transcriptional machinery and maintain the basal transcription in a cell. Condition-specific transcription is regulated through the binding of sequence specific transcription factors to *cis*-acting regulatory elements that are typically located upstream of genes, in the core promoter region, but also distally, in enhancers, silencers, insulators and "locus control regions". T*rans*-acting transcription factors can either activate or repress expression of a gene [15, 16].

Sequence specific transcription factors along with RNAP II and GTFs form a transcription pre-initiation complex at the core promoter region to initiate transcription. Synthesis of protein coding mRNAs is carried out by RNAP II in the nucleus. RNAP II is made of 12 subunits, with Rbp1 being the largest [17]. A distinctive feature of RNAP II is the extended carboxyl terminal domain (CTD) residing on the Rbp1 subunit. The CTD coordinates pre-mRNA synthesis and processing. It recruits complexes responsible for chromatin remodeling, histone modification to initiate the transcription. Changes in the phosphorylation state of the serines or threonines of the CTD is a defining factor in the recruitment of complexes responsible for pre-mRNA processing such as capping, splicing and 3'-end formation [14, 18, 19]. The mature mRNA, along with various RNA binding

protein (RBP) complexes is exported to the cytoplasm for translation by ribosomes [20, 21].

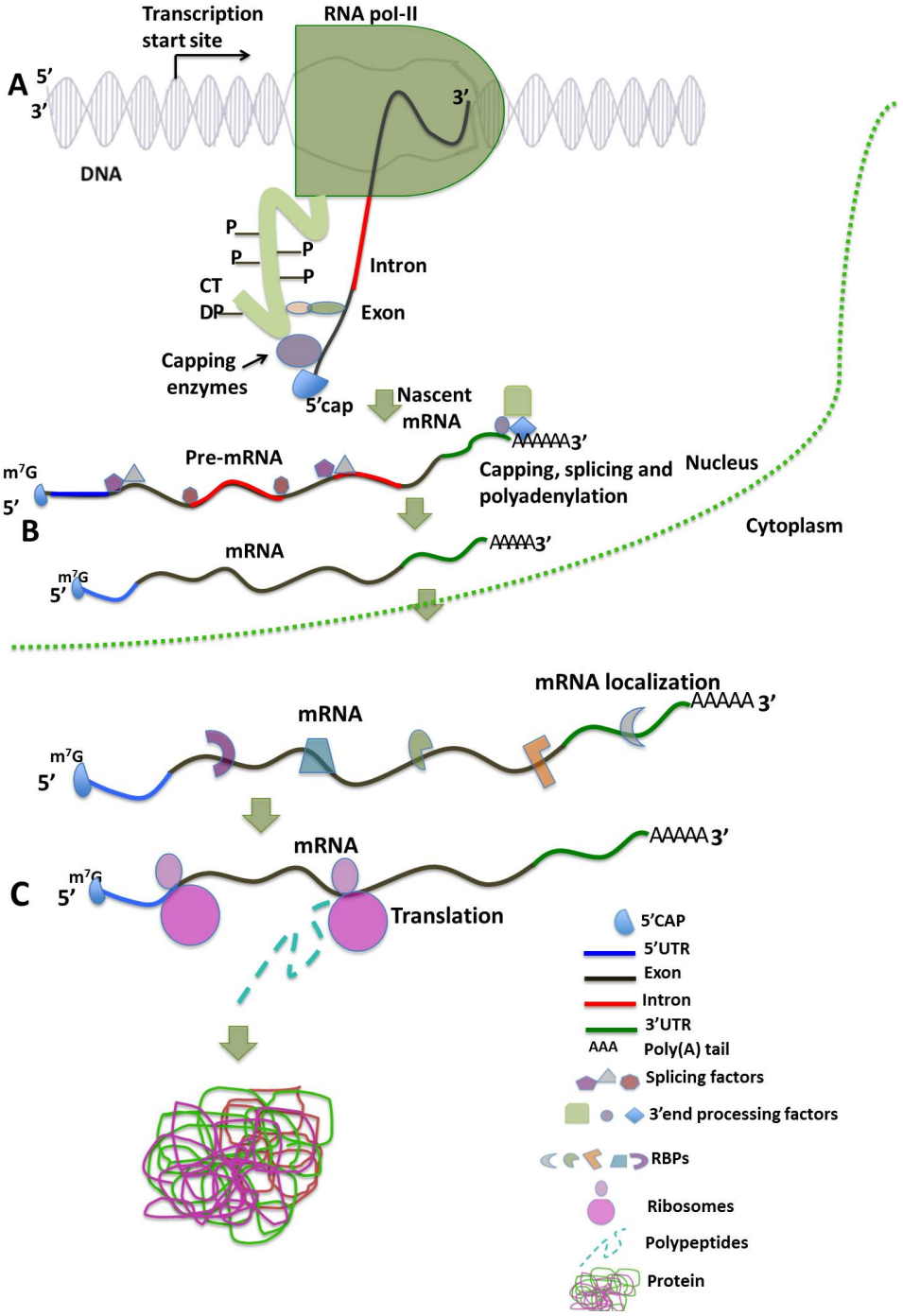


Figure 1: Main steps of gene expression. (A) All protein-coding genes are transcribed by RNA polymerase II (RNAP II), which has a characteristic carboxyl terminal domain residing on the largest subunit of RNAP II. The phosphorylation status of the serines or threonines of CTD direct the recruitment of variety of complexes responsible for the processing of pre-mRNAs. (B) Pre-mRNAs transcribed in the nucleus undergo co/post-transcriptional processes such as capping, splicing, polyadenylation to form a mature mRNA which will

further exported to the cytoplasm where they are subjected to translation (C) by the ribosomal machinery to form proteins.

Translation is also subject to regulation, as will be described later in this thesis. The main steps of gene expression and the corresponding regulatory processes are briefly summarized in Figure 1.

1.1.3 Co- and post-transcriptional mRNA processing and its regulation

1.1.3.1 Five-prime capping

The 5' cap is the first modification to be acquired by a nascent mRNA. The cap is a 7-methylguanosine (m^7G) structure, whose function is primarily to protect the pre-mRNA from 5'-3' exonucleases. The formation of the cap structure involves a series of enzymatic reactions. First a triphosphatase enzyme removes a phosphate molecule from the 5' terminus of the nascent mRNA, and then a guanosine monophosphate (GMP) is added by the enzyme guanylyltransferase to form the G(5')ppp(5')N cap. The cap is eventually methylated at the 7^{th} position of guanosine by an RNA methyltransferase enzyme to give m^7G cap [14, 22]. The first two capping enzymes are directly associated with activated CTD of RNAP II, and act on the nascent mRNA. Capping is important for the completion of transcript elongation by RNAP II [23]. Additionally, the cap binding protein (CBP) is recruited at the cap structure to form a complex that is important for further processing of the pre-mRNA in pre-mRNA splicing, pre-mRNA 3'end formation and for the nuclear export of RNA [24].

1.1.3.2 Polyadenylation

Protection of mRNA 3' ends from 3'-5' exonucleases is due to another modification, polyadenylation, which further assists in the nuclear export of mature mRNAs and in their efficient translation in the cytoplasm [24]. A complex of proteins such as the cleavage and polyadenylation specificity factor (CPSF) that binds to an AAUAAA element, the cleavage stimulatory factor (CstF) that binds to U/GU rich downstream elements and the cleavage factor Im (CFIm) that recognizes UGUA motifs lead to the recruitment of cleavage factor IIm (CFIIm) and poly (A) polymerase. The pre-mRNA is cleaved approximately 20 nucleotides downstream of the AAUAAA polyadenylation signal and a long stretch of adenines is added by the poly(A) polymerase [25]. The CTD of RNAP II facilitates the formation of the 3'-end processing complex [26-28].

1.1.3.3 Pre-mRNA splicing

Over 90% of nascent pre-mRNA is in fact noncoding sequence, in the form of introns. The removal of intronic sequences and joining of the remaining, largely coding sequences

called exons occurs by a co/post-transcriptional mechanism called splicing [14]. Most of the human genes undergo alternative splicing to give rise to different protein isoforms, which frequently differ in their biological activity [29]. For instance the Fas receptor undergoes alternative splicing to produce both soluble and membrane bound protein isoforms, with opposing effects on apoptosis [30]. In *Drosophila melanogaster*, the Down syndrome cell adhesion molecule (DSCAM) gene can theoretically give rise to more than 38,000 variants, more than the total number of genes [31-33]. Thus, splicing contributes to diversity in the proteome.

1.1.3.4 Types of alternative splicing

The locations where the pre-mRNA cleaved to remove the introns and join the exons are called splice sites [34]. Most exons in an mRNA are constitutive, they will be included in the majority of transcripts. Other exons, also known as “cassette exons” are included in some but not all of the transcripts [35]. Another common splice variation consists of small changes in the location of the splice site, which leads to variations in the length of exons. The 5’ terminal exons can further vary through an alternative choice of the promoter followed by alternative splicing, while the 3’ terminal exons can vary due to alternative splicing and alternative polyadenylation [36]. Retention of introns has also been described [33]. These major types of splicing are depicted schematically in Figure 2.

1.1.3.5 The splicing reaction

From the in vitro studies of radiolabelled pre-mRNA incubated in HeLa nuclear extract, it became evident that removal of introns and joining of exons is carried out in two steps [37-39]. In the first step, a trans-esterification reaction is triggered when the 2’hydroxyl group of adenosine at the branch site attacks the 3’-5’ phosphodiester bond of the 5’ splice site to cleave the 5’ intron from the exon. The 5’ guanosine end of the intron becomes attached covalently to the adenosine of the branch site through a 2’-5’ phosphodiester bond. This generates two intermediates, the cleaved off 5’exon and the intron attached to the 3’ exon in a branched circular configuration. The second step also involves a trans-esterification reaction in which a 3’ hydroxyl group of 5’ exon attacks the phosphodiester bond of the 3’ splice site releasing the branched intron and joining the exons [40, 41]. The outline of the splicing reaction is depicted in Figure 2H.

The 5’ splice sites are defined by a consensus sequence, GURAGU (where G: guanine U: uracil A: adenosine R: any purine). The 3’ splice site has three elements: a branch site with the consensus YNYURAY (where Y is any pyrimidine), a polypyrimidine tract, and a YAG splice acceptor site [42, 43]. Recognition of these elements by the spliceosome complex is

essential for gene expression particular in higher eukaryotes, in which the majority of transcripts consist of multiple exons.

The spliceosome is composed of small nuclear RNAs (snRNAs) and several auxiliary proteins. Five snRNAs, U1, U2, U4, U5, and U6, form the small nuclear ribonucleoprotein particles (snRNPs) that are involved in the assembly of spliceosome complex on the pre-mRNA. Spliceosome assembly on the splice sites begins by recognition of the 5' splice site by U1 snRNP, the binding of splicing factor (SF1) to the branch point, and the binding of U2 auxiliary factor (U2AF) to the polypyrimidine tract and 3' terminal AG [44]. This initial complex undergoes extensive conformational changes; SF1 is replaced by U2 snRNP at the branch point, and further recruitment of U4/U6-U5 tri-snRNP complex occurs to make a catalytically active spliceosome complex [45].

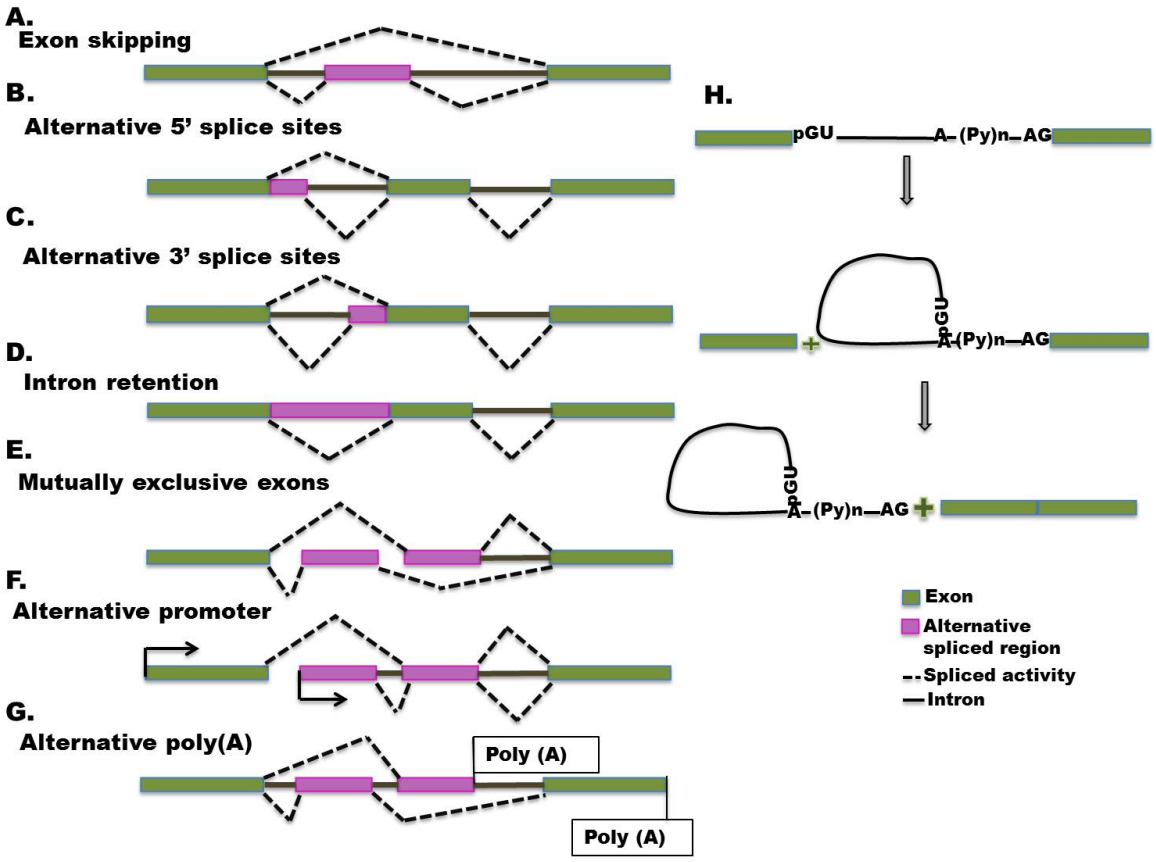


Figure 2: Types of alternative splicing and sketch of the splicing reaction. The simplest described splice variations are shown (combinations thereof can occur within a transcript). (A) A cassette exon can be included or excluded from the mRNA (exon skipping). (B, C) Changes in the selection of alternative 5' and 3' splice sites leads to isoform in which the affected exons have different lengths. (D) An intron can sometimes be retained. (E) Mutually exclusive exons are exons that not selected together in a transcript. (F, G) Selection of alternative promoters or polyadenylation sites can also lead to different choices of internal exons. (H) The splicing reaction involves two trans-esterification steps, wherein the first step the 5'exon is

detached and an intron/3'exon fragment forms a branched structure and in the second step two exons are ligated and branched structure is released.

1.1.3.6 Cis and trans-acting elements of splicing

Splice sites can be weak or strong, presumably depending on their affinity for splicing factors. These differences in the strength of splice sites open the possibility for regulating the binding of splicing factors by other RNA-binding proteins. This involves *cis*-acting RNA sequence elements and *trans*-acting protein regulators. *Cis*-acting elements include exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) [46], whereas trans-acting factors include proteins of the Ser/Arg (SR) family and heterogenous nuclear RNPs (hnRNPs) [47, 48].

Members of the SR protein family typically bind to ESEs. They recruit the U1 snRNP to 5' splice site and the U2AF and U2 snRNP to the 3' splice site [49, 50]. Studies showed that SR proteins also have RS (Arg-Ser) domains with which they bind to several regulatory factors such as transformer2 (TRA2), and SR-related nuclear matrix proteins (SSRM1 and SSRM2) to enhance splicing [51, 52]. Phosphorylated SR proteins act as sequence-dependent splicing activators [53], whereas dephosphorylated SR proteins bind to a regulatory protein, SRp38, to act as splicing repressors [54]. .

In contrast to ESEs, ISSs are generally known to bind to hnRNPs, a class of proteins known to be associated with the unspliced pre-mRNA (hnRNA- also called heterogenous nuclear RNA). hnRNPs have both RNA binding domains as well as protein-protein interaction domains [55]. Of the various hnRNPs, only few hnRNPs have well characterized roles in splicing. For example, hnRNPA1, hnRNPA2 and hnRNPH mostly share mechanisms to inhibit splicing, promote exon skipping, and participate in 5' splice site selection. A recent study showed that hnRNP L binds to an ESS to inhibit the pairing of U1 and U2 snRNPs and promote exon skipping in the *CD45* pre-mRNA [56]. In another report, a U-rich element present adjacent to the 5'splice site of the K-SAM exon of the *FGF2* transcript was shown to bind to the TIA-1 protein, this binding interferes with spliceosome assembly and induces U1 snRNP to bind to 5' splice site [57]. Like other steps of gene expression, splicing is also regulated in a combinatorial manner. For example hnRNP A1 and the SR proteins SF2 and SC35 have antagonistic effects on the splicing of β-tropomyosin exon 6B [58], whereas SR protein SFRS7 and the hnRNPs F and H compete to regulate the splicing of exon2 in α-tropomyosin [59]. HnRNPH and hnRNPF have also been reported to act positively on splicing as part of intronic splicing enhancer complexes with KSRP and PTB [60, 61].

1.1.3.7 Tissue-specific splicing

Alternative splicing is part of tissue-specific programs of gene expression and is based on the differential expression of splicing regulatory proteins between tissues. High-throughput technologies like micro-arrays and RNA sequencing have revealed the breadth of alternative splicing that occurs as a consequence of combinatorial effects of various splicing regulatory proteins [62]. The brain appears to have the highest occurrence of alternative spliced isoforms which is due to the expression of a number of brain-specific splicing factors such as neuro-oncological ventral antigen (*Nova)1*, *Nova2* and nuclear polypyrimidine tract-binding proteins (*nPTB)* [63]. *nPTB* is expressed in neural progenitor cells, but its expression is downregulated in the differentiated neurons [64], whereas *Nova1* and *Nova2* are differentially expressed in various regions of the nervous system, with *Nova1* being expressed in hindbrain and spinal cord, and *Nova2* being highly expressed in neocortex and hippocampus [65, 66]. Similarly, epithelial cells express the RBM35a (ESRP1) and RBM35b (ESRP2) splicing factors, which regulate the expression of epithelia-specific exons. Downregulation of ESRP1 leads to the loss of epithelial splicing during epithelial to mesenchymal transition (EMT) [67, 68].

1.1.4 microRNAs

A recently discovered class of post-transcriptional regulators is the small regulatory non-coding RNAs called microRNAs (miRNAs). As the first part of my PhD dealt with regulation of gene expression by miRNAs, here I describe in some detail miRNA biogenesis, their functions and mechanisms through which miRNAs regulate gene expression. The biogenesis and functions of miRNA are shown schematically in Figure 3.

1.1.4.1 Discovery

The first report of a miRNA (lin-4) came into light in 1993, when Victor Ambros and colleagues reported that lin-4 represses the heterochronic lin-14 gene, which plays a role in the transition of *Caenorhabditis elegans (C.elegans)* between larval stages [69]. However, it was not until Gary Ruvkun and his colleagues reported in 2000 that let-7 miRNA is a 21-nucleotide RNA that is complementary to elements in 3'UTRs of several heterochronic genes, whose expression it probably regulates [70], that the field started to expand dramatically. A series of studies that emerged in rapid succession catalogued miRNAs in fly, zebrafish, mouse and human [71-73]. To date, 35,828 miRNAs have been reported from 233 species, many being conserved over large evolutionary distances [74]. MiRNAs form one of the most abundant gene families, which comprises over 2500 miRNAs in humans [74]. Most human genes are in fact regulated by miRNAs and a protein-

coding mRNA typically contains binding sites for one or more miRNAs in its 3' untranslated region (UTR).

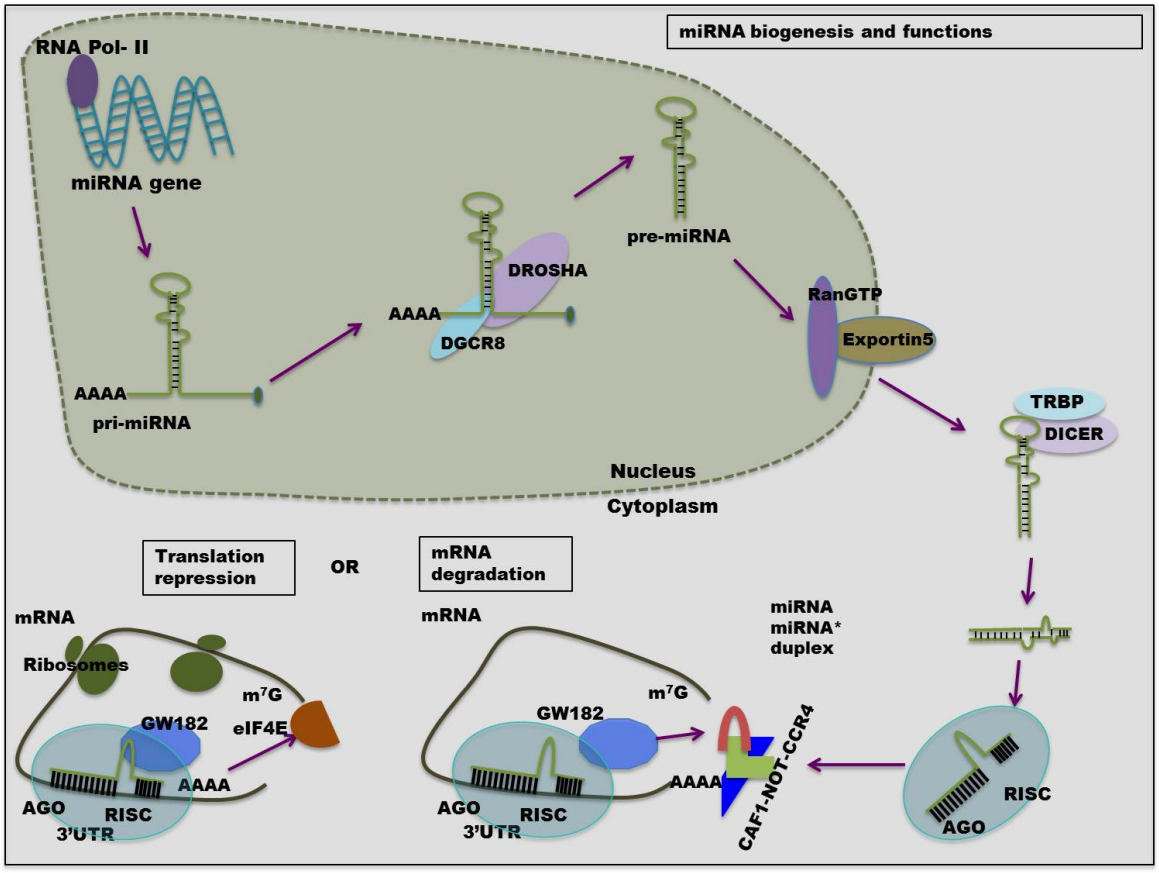


Figure 3: Biogenesis and functions of miRNAs: miRNA genes are transcribed by RNA polymerase II as primary-miRNA (pri-miRNA), which are processed by Drosha-DGCR8 complex to make precursor-miRNA (pre-miRNA). Pre-miRNAs are then exported to the cytoplasm via the RanGTP-Exportin5 complex, and further processed there by the Dicer-TRBP complex to yield a mature miRNA-miRNA* duplex. The mature miRNA is loaded into an AGO protein to form an RNA-induced silencing complex (RISC) and this complex is guided by miRNA to an mRNA target to exert its functions. The AGO-GW182 complex induced miRNA-dependent target degradation by recruits the CAF1-NOT1-CCR4 deadenylase complex to trigger deadenylation of target mRNAs. MiRNAs also repress translation of target mRNAs via AGO which competes with translation initiation factor eIF4E in binding to cap structure to block assembly of ribosomes.

1.1.4.2 Transcription and nuclear processing of miRNAs

MiRNAs are mostly located in the introns of coding and non-coding transcripts, and their expression is controlled by the promoters of the host transcripts. A few studies reported that miRNA have independent transcription start sites [75, 76]. Thus, for most miRNAs, promoters can be identified by collectively analyzing the data from mRNA sequencing (mRNA-seq), mapping of CpG islands and chromatin-immunopreciptation sequencing (ChIP-seq). For instance, association of RNAP II with the promoter of miR-*23a~27a~24*-2

cluster was revealed from ChIP data [77]. Some miRNA genes are organized in polycistronic transcription units, from which the miRNAs are co-transcribed [78]. Although some miRNAs like let-7 are deeply conserved in evolution [70], the targets seem to evolve rapidly [79]. The activities of individual miRNAs in a polycistronic miRNA cluster may differ. For instance, the most conserved miRNA cluster, mir~100~let-7~mir-125 plays a role in the development of invertebrates. In mammals it is only let-7 and no other miRNAs that is downregulated during early stages of development [80].

Although α-amanitin (an RNAP II inhibitor) sensitivity experiments showed that mammalian miRNAs are transcribed by RNAP II (and not by RNA polymerase-I (RNAP I) or RNA polymerase-III (RNAP III) [77, 81]), some viral miRNAs are transcribed by pol-III. For instance, miRNAs from a mouse virus, MHV68, are encoded in tRNA-like primary transcripts, which are transcribed by Pol-III [82]. In animals, transcribed primary-microRNAs (pri-miRNAs) are usually kilobases (kbs) long, have cap structures at the 5' end, miRNA-encoding stems of up to 35 base pairs (bp), stem terminal loop and a poly(A) tail [83]. Pri-miRNAs are processed in the nucleus by the RNase III enzyme Drosha together with its essential co-factor DGCR8. The two proteins form the so-called microprocessor complex [84]. Drosha is a ~160 kDa protein which has two RNase III domains (RIIIDs) and a double stranded RNA binding domain (dsRBD) at its carboxy-terminal end. Cleavage of the pri-miRNA by Drosha defines the 5' terminus of the so-called 5p miRNA, which is encoded in the 5' arm of the pre-miRNA hairpin structure. Because the sub-sequences located at the 5' end of the mature miRNA is important for function [85], Drosha can thereby determine the miRNA specificity. One RIIID of Drosha cuts the pri-miRNA at 11bp away from the junction of single and double stranded RNA and the second RIIID cuts the pri-miRNA at 22bp away from the junction linked to the terminal loop [83, 86]. This processing gives rise to precursor-miRNAs (pre-miRNAs), which are exported to the cytoplasm by a nuclear protein called exportin5 (EXP5), encoded by *XPO5* gene. EXP5 together with pre-miRNA and Ran-GTP (Ran-guanosine triphosphate) forms a transport complex that translocates the pre-miRNA. This complex is dissociated by the hydrolysis of GTP to release the pre-miRNA into the cytosol [87].

1.1.4.3 Cytoplasmic processing of miRNAs

In the cytoplasm, another type III RNase enzyme, Dicer, together with transactivation responsive RNA-binding protein (TRBP) cleaves off the terminal loop to release a small RNA duplex. Dicer recognizes a two-nucleotide 3' overhang, generated by Drosha, and cleaves the precursors at a distance of 21-25 nucleotides from the 3' end of the terminus. The levels of Drosha and Dicer are regulated by their binding partners, DGCR8 and TRBP, and vice versa [86, 88]. Dicer is a 200 kDa protein, which also has two RIIIDs which form a

catalytic center located as in Drosha, at its carboxyl terminus. At the amino terminus Dicer has a helicase domain, DUF283, which functions in the recognition of pre-miRNAs [89]. Dicer releases a small double stranded duplex, which contains both mature miRNA or guide strand and its complementary strand known miRNA* or the passenger strand. Only one of the two strands is generally loaded onto an Argonaute (AGO) protein to form a pre-RNA induced silencing complex called RISC [90]. For some miRNAs, both strands of the duplex can be loaded with comparable efficiency into AGO, and in these cases, the mature miRNAs are named based on the miRNA gene name, to which the 5p or 3p suffix is appended [86].

1.1.4.4 Strand selection

Dicer and its counterpart TRBP not only cleave the pre-miRNA, but also assist in the loading of the small RNA onto the AGO protein [91]. The strand that is loaded into AGO, also called the "guide strand" is selected by AGO protein based on the thermodynamically stability of the two ends of the duplex. That is, the end that is least stable gives the small RNA whose 5'end will be anchored in the AGO protein to form the RNA-induced silencing complex (RISC) [92, 93]. After the guide strand selection, in an ATP-independent biochemical reaction, the passenger strand is released. Release of the passenger strand from pre-RISC is also assisted by Dicer, which takes it to the degradation machinery [94]. The passenger strand can also be involved in gene silencing, although typically it is less active compared to the guide strand [95, 96].

1.4.5 The Argonaute proteins

Dicer, AGO and TRBP form a RISC loading complex (RLC) or pre-RISC, which aids in the loading of miRNAs onto AGO protein. After guide strand selection, the miRNA guides RISC to its complementary mRNA targets. RISC is the major processing machinery of the RNA interference (RNAi) mechanism through which silencing of genes takes place [97]. RNA silencing is an evolutionary conserved phenomenon, wherein miRNAs together with other RBPs degrade their other RNA targets or repress translation. AGO proteins form three subclasses, namely AGO, PIWI and worm-specific AGO proteins (WAGOs) [98]. In humans there are four Argonaute proteins encoded by four Ago genes. In humans, all AGO proteins have the ability to induce translation repression and mRNA degradation, but only AGO2 has the "slicing" activity that allows it to directly cleave miRNA targets [99]. This however, requires that the guide miRNA is perfectly complementary to its target mRNAs. AGO2 is also involved in gene silencing as AGO1, AGO3 and AGO4 [100]. AGO proteins have four domains: the N-terminal, PAZ, MID and PIWI [101]. The PAZ domain has a characteristic feature to recognize the two-nucleotide 3' overhang, with which it binds to the guide RNA of the small RNA duplex. The MID domain contacts the miRNA 5' end,

which frequently starts with the U nucleotide. Within MID there also lays a MC motif that is known to be involved in the translation control of target mRNAs [102]. Whereas the PIWI domain has a structure similar to RNase H, which contains active site that bears slicing activity, the N-domain assists in the unwinding of the duplex RNA [103].

1.1.4.6 The miRNA "seed" sequence

The most important part of the miRNA for its efficiency in repression is the region of nucleotides 2-7 at the 5' end. This has been called the miRNA "seed" sequence [104]. The binding of the miRNA to the target mRNA is nucleated by the seed sequence and the additional binding of the miRNA 3'end to the target may also play a role in enhancing the target repression [105]. MiRNA binding sites that are efficient in host transcript degradation are located mostly in the 3'UTRs of the mRNAs [106]. Other than the seed sequence, factors that influence the binding of the miRNAs to targets and the efficiency of repression are the accessibility and the position of the binding site in the 3'UTR, as well as the nucleotide composition around miRNA binding sites [107]. RBPs also compete with miRNAs in binding or even enhance the binding of the miRNA [108].

1.1.4.7 MiRNA functions and mechanisms

Several molecular mechanisms for miRNA-induced gene silencing have been proposed. The fact that these mechanisms emerged from experiments conducted in different models with different techniques, could be an underlying factor in their diversity. The mechanisms with the most substantial supporting evidence are discussed below.

From the initial studies in C.*elegans*, wherein the first discover miRNA lin-4 repressed the synthesis of lin-14 protein [69], it was thought that miRNAs mainly function as translational repressors. In contrast, in plants miRNAs mostly act by cleaving and degrading their target mRNAs [109]. Translation of mRNAs occurs in three steps, initiation, elongation and termination. Most data indicates that miRNA repress translation at the level of translation initiation. The eukaryotic translation initiation factor (eIF) eIF4F contains the eIF4E subunit, which is responsible for the identification of cap structure of the mRNA to initiate translation. Another subunit, eIF4G, conjoins eIF4E and the poly(A) binding protein (PABP) to circularize mRNAs. This is thought to stimulate translation re-initiation through the recycling of ribosomes. The interaction of PABP with translation initiation factors in turn increases the efficiency of translation [14, 110].

It has been observed that the MC motif of the MID domain of AGO has sequence homology to the eIF4E [102] and this lead to the hypothesis that AGO competes with eIF4E for cap binding, thereby preventing the initiation of translation. This interference of AGO also leads to the inhibition of binding of 40S or 60S subunit to the repressed mRNA,

which subsequently represses translation [102, 111]. In addition, co-sedimentation experiments in human cells yielded polysomes that contained miRNAs, AGO proteins and their mRNA targets suggesting that miRNAs are also able to block translation post-initiation [112, 113].

Although the initial studies indicated that miRNAs only participate in translational repression, more recent studies found that miRNAs also destabilize their targets. Microarray-based measurements of transcript levels after perturbing the miRNA pathway or miRNA levels revealed significant changes in the levels of already validated targets [114-116]. General mRNA degradation is carried out either in the 3'-5' direction by exosomes or by decapping, followed by 5'-3' degradation by the endonuclease XRN1 [117]. Isolation of miRNA and their cognate targets from the processing bodies (P-body), which are cytoplasmic foci where mRNA degrading and translation repression components are found, also support the idea of miRNA mediated mRNA degradation. Whether this is a consequence of other modes of repression is still disputed [118]. The P-body protein GW182 (with 3 isoforms in human, TNRC6A-C) interacts with AGO proteins to recruit the deadenylase CAF1-CCR4-NOT complex. Deadenylation is followed by decapping through the DCP2 enzyme, which then promotes degradation of the target mRNAs by nucleases. The role of miRNAs in their target degradation was elucidated when accumulation of mRNAs was observed upon knockdown of the components of CAF1-CCR4-NOT complex [115, 116].

1.1.4.8 Methods to identify miRNA targets

Since miRNAs are known to be involved in the regulation of many fundamental biological processes such as development, differentiation, immune system and many more [119-121], it is important to establish methods to identify the expressed miRNAs and their targets in a given system. Classical methods of identifying miRNA targets include cloning of miRNA genes to elicit the functional response of their targets, with which first miRNAs lin-4 and let-7 [69, 70] were discovered. However, much of what we have learned about miRNA-dependent mRNA regulation came from computational analyses and analysis of high-throughput mRNA expression data generated by microarrays and later with sequencing. In an initial study of this type, tissue specific miRNAs, miR-1 and miR-124a, were transiently transfected in HeLa cells, and changes in mRNA expression were determined with microarray analysis [122]. Computational analysis revealed that the downregulated mRNAs carried in their 3' UTRs sequences complementary to the seed of the transfected miRNA. Synthetic oligos complementary to miRNAs of interest have also been used to antagonize the miRNA activity and increased the mRNA and protein levels of miRNA targets [123-125]. This type of studies yielded miRNA targets that were

differentially expressed in response to the ectopic expression of specific miRNAs, but also false positive candidates. Computational methods that exploited not only the seed sequence complementarity but also the degree of evolutionary conservation of putative target sites, such as TargetScan [104] and ELMMO [126] were reported to be effective in defining miRNA targets [104, 127-129]. Proteomics data was generated after miR-1 transfection using stable isotope labeling by amino acids in cell culture (SILAC) to identify both destabilized and translationally repressed targets [130]. Upon analysis of such data, 12 targets for miR-1 were identified, of which 6 were validated with luciferase assays. Following this, other groups also used slightly modified approaches to identify the targets for various other miRNAs [131, 132].

1.1.5 Next generation sequencing and CLIP

The application of sequencing technologies helped to crack some important questions in molecular biology. An important landmark was the shot-gun sequencing and assembly of the human genome in the year 2003 which was accomplished with the first generation method called Sanger sequencing invented by Edward Sanger [133]. After the successful completion of human genome sequencing, the second generation or next generation sequencing methods brought new methodologies to answer complex questions such as determining gene expression profiles transcriptome wide. mRNA sequencing is the most widely used method to profile all polyadenylated mRNAs in a given population of cells [134]. This method has been modified to infer alternative splice forms, alternative promoter sites and alternative poly(A) sites [135, 136]. High-throughput sequencing methods are employed to infer the binding sites of the RNA binding proteins, for instance with Argonaute crosslinking and immunoprecipitation (CLIP) [137, 138]. CLIP relies on the *in-vivo* crosslinking using ultraviolet (UV) C light (254nm) of RNA binding proteins to their target RNAs intact in the cells. These complexes are pulled down with a specific antibody and are further processed to yield RNA fragments that are eventually sequenced. The CLIP method has been used extensively to discover not only miRNA targets but also non-coding RNAs such as small nucleolar RNAs (snoRNAs) [139] associated with snoRNPs. A modified protocol of CLIP called photoactivatable ribonucleoside CLIP (PAR-CLIP) [140] was proposed by Tuschl group wherein photoactivatable ribonucleosides such as 4-thiouridine (4-SU) or 6-thiouridine (6-SU) are incorporated into RNAs before crosslinking with UV A (at 365nm) light. This induces mutations (T-to-C with 4-SU or G-to A with 6-SU) during the reverse transcription, which helps in the determination of binding sites at a nucleotide resolution level. These mutations are further used as diagnostic markers in distinguishing the signal from noise in CLIP and to determine high affinity binding sites of miRNAs. Based on transcriptome-wide Argonaute CLIP [138, 140] our group has succeeded in developing a computational method to infer an empirical model called

MIRZA that allows one to quantify the strength of miRNA-target interactions. This model then enables identification of the miRNAs that most likely guided the interaction of AGO with each of the CLIPed mRNA sites. Moreover, this model enables one to infer both canonical and noncanonical miRNA targets [141].

2. Timescales and bottlenecks in miRNA-dependent gene regulation

High-throughput approaches generate huge data sets, which need to be analyzed computationally. Importantly, these data are generated in a limited number of conditions. To generalize as well to model miRNA-dependent regulation in the context of other processes, computational target prediction models remain useful. Currently, such models are based on several important aspects of miRNA target regulation such as the complementarity of targets to the miRNA seed sequence, their evolutionary conservation etc. Importantly, the experimental validation of predicted targets may induce side effects that complicate the interpretation of the data. Moreover, as described above, the miRNA mechanisms continue to be debated, partially due to the lack of methods that allow a comprehensive characterization of the behavior of miRNA targets, simultaneously and at all levels. An aspect that has not been considered in analyzing miRNA regulation is the time scale on which miRNA regulation takes place. In one study, we have addressed these kinetics issues and showed that they need to be taken into account when designing and analyzing experiments aimed at characterizing the regulatory function of miRNAs by means of miRNA perturbation and subsequent measurements of the induced changes in mRNA and protein levels. In addition, we have analyzed a detailed model of miRNA action, investigating how different parameters such as the rate at which miRNAs load into AGO protein and their decay rate influence the time-scale and magnitude of miRNA-mediated gene regulation.

2.1 Statement of my work

In the study described below, I carried out the experimental work. I have constructed the inducible miRNA constructs and used them to establish cell lines that stably express miRNAs in an inducible manner. I carried out an intensive screening of clones with Northern blots to check the consistency and proper expression of miRNAs. After establishing stable inducible cell lines, I cloned putative targets of the miRNAs of interest in the reporter vector and validated them by luciferase assays in the inducible miRNA expression cell lines. I have also established a stable cell line with constitutive expression of the miRNA target and inducible expression of its cognate miRNA. I have tested this cell line for the consistency of the miRNA target downregulation by luciferase assays until

several days after miRNA induction. This cell line was used for the final experiments described in the study below.

Molecular Systems Biology 9; Article number 711; doi:10.1038/msb.2013.68
Citation: ***Molecular Systems Biology*** **9:711**
www.molecularsystemsbiology.com

molecular systems biology

Timescales and bottlenecks in miRNA-dependent gene regulation

Jean Hausser[1,3,*], Afzal Pasha Syed[1], Nathalie Selevsek[2], Erik van Nimwegen[1], Lukasz Jaskiewicz[1], Ruedi Aebersold[2] and Mihaela Zavolan[1,*]

[1] Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Basel, Switzerland and [2] Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland
[3]Present address: Department of Molecular Cell Biology, Weizmann Institute of Science, Herzl Street 234, 76100 Rehovot, Israel
* Corresponding authors. J Hausser or M Zavolan, Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Klingelbergstrasse 50/70, Basel, 4056, Switzerland. Tel.: +972 8 934 4447; Fax: +972 8 934 4125; E-mail: jean.hausser@weizmann.ac.il or Tel.: +41 61 267 1577; Fax: +41 61 267 1585; E-mail: mihaela.zavolan@unibas.ch

Received 9.5.13; accepted 30.10.13

MiRNAs are post-transcriptional regulators that contribute to the establishment and maintenance of gene expression patterns. Although their biogenesis and decay appear to be under complex control, the implications of miRNA expression dynamics for the processes that they regulate are not well understood. We derived a mathematical model of miRNA-mediated gene regulation, inferred its parameters from experimental data sets, and found that the model describes well time-dependent changes in mRNA, protein and ribosome density levels measured upon miRNA transfection and induction. The inferred parameters indicate that the timescale of miRNA-dependent regulation is slower than initially thought. Delays in miRNA loading into Argonaute proteins and the slow decay of proteins relative to mRNAs can explain the typically small changes in protein levels observed upon miRNA transfection. For miRNAs to regulate protein expression on the timescale of a day, as miRNAs involved in cell-cycle regulation do, accelerated miRNA turnover is necessary.
Molecular Systems Biology **9**: 711; published online 3 December 2013; doi:10.1038/msb.2013.68
Subject Categories: simulation and data analysis; RNA
Keywords: gene expression regulation; kinetics; miRNAs; modeling; protein turnover

Introduction

The cellular abundance of proteins appears to be controlled to a substantial extent at the level of translation (Schwanhäusser *et al*, 2011). Among post-transcriptional regulators, miRNAs are short, evolutionarily conserved, non-protein-coding RNAs that modulate the rates of both protein translation and mRNA decay (Bartel, 2009; Fabian *et al*, 2010; Huntzinger and Izaurralde, 2011). MiRNAs guide Argonaute (Ago) proteins to specific elements that are located mostly in the 3′ untranslated regions (UTRs) of mRNAs and are typically complementary to the miRNAs' 'seed' region (7–8 nucleotides at the 5′ end) (Bartel, 2009). They are involved in virtually every cellular process, from early development (Wightman *et al*, 1993; Kanellopoulou *et al*, 2005; Song *et al*, 2011) to organ function, and their perturbed expression has been associated with numerous human diseases, such as diabetes (Trajkovski *et al*, 2011), cancer (Lu *et al*, 2005) and viral infection (Pfeffer *et al*, 2004). Furthermore, miRNAs appear to be able to initiate, on their own, the complex process of reprogramming somatic cells into pluripotent stem cells (Anokye-Danso *et al*, 2011).

Evidence for context-specific modulation of both miRNA biogenesis (Heo *et al*, 2009) and decay (Chatterjee and Grosshans, 2009; Krol *et al*, 2010; Rissland *et al*, 2011) has started to emerge. Yet, the implications of a flexible modulation of miRNA expression dynamics for the dynamics of the controlled processes remain largely unexplored. Rather, from a kinetic stand point, miRNAs are generally viewed as fast regulators of gene expression, in contrast to transcription factors, whose expression additionally requires protein synthesis (Shimoni *et al*, 2007; Hobert, 2008). Here, we use a variety of low- and high-throughput data sets to uncover the kinetics of various steps in the miRNA-dependent regulatory cascade. Contrary to the expectation that miRNAs are fast regulators, we find that miRNAs may not be acting as rapidly as commonly assumed due to two bottlenecks, one at the level of miRNA loading into Ago and the other at the level of protein decay. The derived quantitative model enables us to predict the effect of manipulating miRNA expression on mRNA, protein and ribosome density levels, to uncover these effects from experimental data sets and to make predictions about the dynamics of miRNA-dependent gene regulation in various physiological scenarios.

Results

Inference of a kinetic model of miRNA-dependent regulation

A stumbling block for a quantitative description of miRNA-dependent gene regulation is that the parameters

of miRNA expression dynamics and of miRNA–target interaction are generally unknown. The fluorescence cross-correlation spectroscopy experiments of Ohrt *et al* (2008) offer a possible solution. This study measured the fraction of small-interfering RNA (siRNA) in complex with Ago as well as the fraction of Ago in complex with siRNAs, as a function of the time after siRNA micro-injection into cells. Because the siRNA and miRNA pathways share many components, these data allow us to obtain initial estimates of the dynamics of Ago loading, which we can then use to predict the dynamics of miRNA-induced changes in miRNA targets.

Figure 1A illustrates the structure of the Ago-loading model that we constructed. At time $t=0$, X_0 siRNAs are micro-injected into the cytoplasm from where they either decay at a rate d or associate with free Ago at a rate $\gamma = b\, f_0$, with b being the rate of binding and f_0 the fraction of free Argonaute proteins. Ago-complexed siRNAs dissociate at a rate u. With reasonable assumptions supported by experimental evidence (Khan *et al*, 2009), namely that the fraction of free Ago f_0 is small and that siRNAs load into Ago by competing out endogenous miRNAs, the fraction of Ago proteins loaded with the injected siRNA can be described by a bi-exponential function. This form facilitates the modeling of mRNA and protein dynamics of miRNA targets, as will be described shortly. We obtained maximum-likelihood estimates and confidence intervals on the model parameters γ, d, u, X_0 by Markov Chain Monte Carlo (Supplementary Figure S1) and found a good agreement between the data and the model (Figure 1B). Therefore, Ago loading upon small RNA micro-injection can be well approximated by a bi-exponential function.

Does this model describe well the dynamics of Ago loading upon miRNA transfection? The above model predicts that the peak in Ago loading would be about 10 h after the delivery of the small RNA (Figure 1B). However, the peak in mRNA repression in miRNA transfection experiments, which should occur at the peak of Ago loading with the miRNA, appears to be considerably delayed. For example, in the experiments of Grimson *et al* (2007) target mRNAs are typically more repressed at 24 h than at 12 h after miRNA transfection (Supplementary Figure S2A). This is likely due to a delayed delivery of transfected miRNAs to the Ago proteins (Broderick and Zamore, 2011; Stalder *et al*, 2013) compared with the situation in which the small RNAs were directly micro-injected into cells. To investigate the dynamics of miRNA loading into Ago under transfection conditions, we therefore used the fine-grained time series of mRNA expression of Wang and Wang (2006) who transfected miR-124 in HepG2 cells. However, because up to this point our model only describes the dynamics of Ago loading with a small RNA, we first need to derive the consequences of Ago-loading dynamics for the mRNA and protein level of the small RNA targets.

The abundance of transcripts and proteins associated with individual genes is frequently described in terms of the rates of mRNA transcription, decay and translation and the rate of protein decay (e.g., Hargrove and Schmidt, 1989; Schwanhäusser *et al*, 2011; and Khanin and Higham, 2009). A large body of evidence indicates that miRNAs modulate both the decay and translation rates of mRNAs. How these rates respond to changes in miRNA expression is not known. However, because the studies of Djuranovic *et al* (2012) and Béthune *et al* (2012) showed that once loaded in Ago, miRNAs find their targets relatively fast, we can assume that the relative changes in mRNA decay and translation are proportional to the fraction of miRNA-loaded Ago. The resulting model, describing the mRNA and protein abundance of a given miRNA target, is parametrized by two proportionality factors δ and λ that relate the fraction of loaded Ago to the relative increase in mRNA decay and decrease in translation (see Materials and methods). $\lambda > 1$ indicates that miRNAs predominantly affect the translation of the message, while λ between 0 and 1 corresponds to the case where miRNAs impact mostly the mRNA decay.

We first fitted the mRNA expression dynamics in response to miR-124 transfection measured by Wang and Wang (2006). To do so, we fixed the Ago-loading parameters γ, d, u to the values inferred from the fluorescence cross-correlation spectroscopy measurements of Ohrt *et al* (2008) (Figure 2A, in green). As a

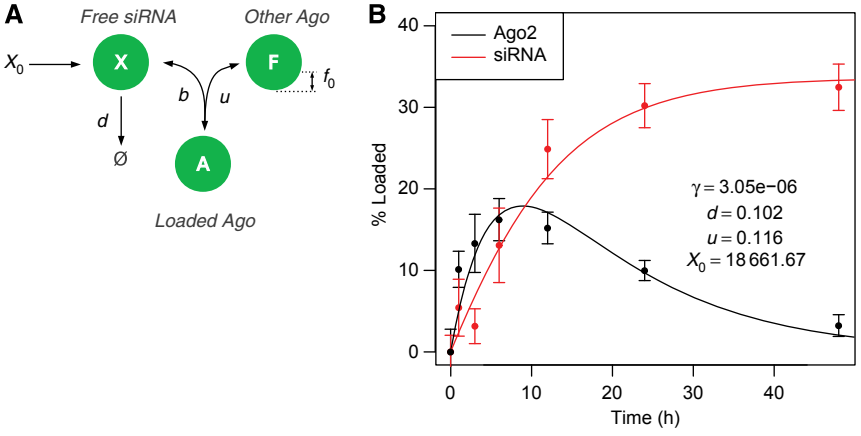


Figure 1 The amount of siRNA-loaded Ago following siRNA micro-injection can be modeled by a bi-exponential function. (**A**) Cartoon illustrating the model parameters: at time $t=0$, X_0 siRNAs are injected into the cell, after which the siRNAs X either decay with rate d or load into a free Ago f_0F with rate b to form siRNA-loaded Ago complexes A. Small RNAs dissociate from Ago at rate u. (**B**) Measured (dots) and fitted (lines) fractions of complexed Ago and complexed siRNAs from the data set of Ohrt *et al* (2008). Error bars represent the 95% confidence interval on the mean measured fraction of Ago and siRNA in complex. Maximum-likelihood parameter estimates also appear in the figure. $\gamma = bf_0$ is the Ago–siRNA association rate b normalized to the fraction of free Ago f_0. See also Supplementary Figure S1.

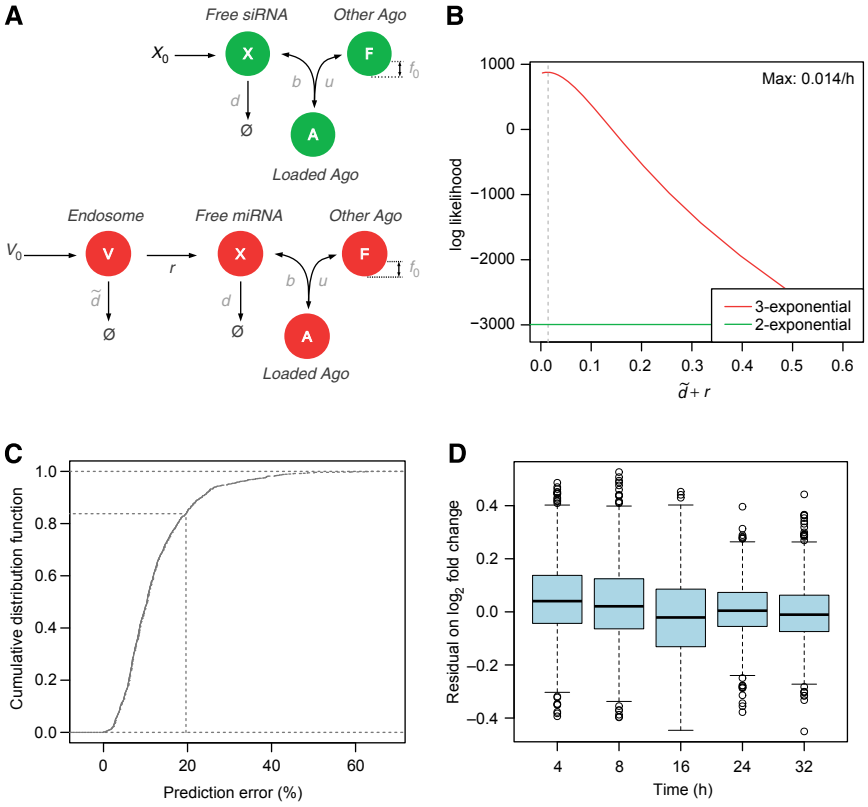


Figure 2 Changes in mRNA levels in the miR-124 transfection time course of Wang and Wang (2006) can be modeled assuming a tri-exponential Ago-loading function. (**A**) Cartoon illustrating two models of miRNA transfection experiments and their parameters. Free, fitted parameters appear in black, fixed parameters from Figure 1 in gray. The bi-exponential model (in green) is the same as in Figure 1A. Also shown is a tri-exponential model of Ago loading (in red), which is identical to the bi-exponential model, except for the addition of an extra compartment (*V*) in which V_0 miRNAs are loaded at time $t=0$, and two additional rates: rate of miRNA decay in this compartment ($\tilde{d}$) and rate of transfer to the Ago-accessible environment (r). (**B**) Log-likelihood profile of the clearance parameter $\tilde{d}+r$ given the mRNA profiling time-course data. The log-likelihood of the tri-exponential model (red line) is compared with that of the bi-exponential model (green line). (**C**) Cumulative distribution of the per-gene relative error between the model and the time-course data. The *x*-axis represents the per-gene relative error between the model prediction and the measurements. For any chosen cutoff on the relative error, the fraction of genes whose regulation following miRNA transfection could be predicted at the chosen error cutoff or less can be read on the *y*-axis. The dotted line marks a 20% error on the fold change typically observed in miRNA transfection experiments. (**D**) Boxplots of the model residual on $\log_2$ fold changes for genes that fit the measured mRNA fold changes with less than a 20% error. Boxes span the interquartal range and whiskers extending up to 1.5 times the interquartal range. See also Supplementary Figure S2.

result, the fraction of loaded Ago can be described by a bi-exponential function of time. One can show that mRNA expression dynamics are driven by a single free parameter $X_0\delta$: the product of the initial concentration of the small RNA X_0 and the maximum change in mRNA decay rate that can be induced by the small RNA δ. Alternatively, to account for the presumed delays in Ago loading upon miRNA transfection, we modified the Ago-loading model to include an additional compartment (presumably the endosomes) in which V_0 miRNAs are loaded at the time of transfection (Figure 2A, in red). From this compartment, miRNAs are either degraded or translocate to the cytoplasm where they can associate with Ago. These assumptions lead to a tri-exponential Ago-loading function. In this alternative model, describing the miRNA-induced changes in mRNA abundance requires two free parameters: $\tilde{d}+r$, the experiment-specific rate with which miRNAs are cleared from endosomes through degradation or transfer to the cytoplasm, and $rV_0\delta$, the gene-specific influence of the transfected miRNAs on mRNA degradation.

We compared the goodness of fit of both models by fitting their free parameters—$X_0\delta$ for the bi-exponential model, or $\tilde{d}+r$ and $rV_0\delta$ for the tri-exponential model i—to 1098 genes that had a miR-124 seed match in the 3′ UTR and were repressed on average 4, 8, 16, 24 and 32 h post transfection. This represents a total of 5490 data points. With a maximum log-likelihood of 877.3, the tri-exponential Ago-loading model fitted better the mRNA profiling data than the bi-exponential model (log-likelihood of -2994.6; Figure 2B; Supplementary Figure S2B). Replicate experiments would allow us to estimate an upper bound on the goodness of fit that can in principle be achieved, but unfortunately, the time series experiment did not include replicates. Nonetheless, we used a data set of six replicate experiments in which miR-124 was transfected into human embryonic kidney 293 (HEK293) cells and mRNA expression was subsequently profiled by microarrays (Karginov *et al*, 2007) to estimate this upper bound. The average standard deviation on the $\log_2$ mRNA fold change was $\sigma=0.26$ (Supplementary Figure S2C), corresponding to a 20% uncertainty on the fold change. At this cutoff, the tri-exponential model fitted 83.8% of the genes (Figure 2C), and the prediction error was uniform across time points (Figure 2D; Supplementary Figure S2D–F). Therefore,

accounting for delays in Ago loading is necessary to model changes in gene expression following miRNA transfection.

Validation of the derived models of mRNA and protein abundance changes in response to miRNA transfection and induction

To test whether the models derived above can describe the change in abundances of both mRNAs and proteins following miRNA perturbation, we selected a 'prototypical' miRNA, miR-199a, that is not normally expressed in HEK293 cells and established a cell line in which the expression of miR-199a is inducible with doxycylin from a pRTS-1 episomal vector (Bornkamm *et al*, 2005). We further cloned the 3′ UTR of the kinectin (KTN1) gene, containing a miR-199a-3p binding site, downstream of the stop codon of the renilla luciferase in a psiCHECK-2 vector and stably integrated this construct in the genome of cells containing the miR-199 expression vector. We then investigated the response of the KTN1 gene at both the mRNA and protein level upon induction of miR-199a or transfection of miR-199-3p mimic. For Ago loading, we used the tri-exponential model for the transfection experiment, while for miRNA induction experiment, we changed the one-time micro-injection represented in Figure 1A to add constant miRNA synthesis into the Ago-accessible compartment *X*. The measured and predicted mRNA and protein-level dynamics for both miRNA transfection and induction are shown in Figure 3 and Supplementary Figure S3.

From the perspective of the measured changes, we found that the transfection experiment had higher variability, presumably because the transfection efficiency varied to some extent between the replicate experiments. The changes in mRNA expression, measured by qPCR, also showed higher variability compared with changes in protein expression, which were estimated based on the luciferase activity. Nonetheless, the models give a reasonable fit to the mean observed changes. Furthermore, the protein-level changes, which are the final readout of our models' dynamics and whose prediction depends on all of the assumptions that our models make, are remarkably well described by the models. These results indicate that our models predict well mRNA and protein-level changes in response to changes in miRNA expression. Furthermore, as data from other groups also suggest (Karginov *et al*, 2007), controlling variability in transfection experiments is difficult, and miRNA target identification could therefore be more accurate in experiments in which the miRNA is induced instead of transfected.

As additional validation of our model, we further tested its ability to explain changes in mRNA abundance and translation efficiency (estimated through Ribosome Protected Fragment (RPF) sequencing) measured at two time points (12 and 32 h) after miR-155 and miR-1 transfection by Guo *et al* (2010). Indeed, we found that the maximum-likelihood estimates of both $\tilde{d}+r$ and $rV_0\delta$ were in the range of those previously observed in Wang's miR-124 and our miR-199 transfection experiments (Supplementary Figure S4A). The proportion of genes whose mRNA-level and RPF dynamics was fitted at the expected accuracy was 67.5% and 48.3% in the miR-155 and miR-1 transfections, respectively (Figure 4A). These numbers are lower than those we obtained for the Wang and Wang (2006) data, which could be due to the lower reproducibility of translation efficiency measurements by RPF—30% variability on the $\log_2$ fold change according to Ingolia *et al* (2009)—compared with mRNA profiling. Thus, our models that link the dynamics of the miRNA to the dynamics of its target at the mRNA and protein level explain well kinetic data from reporter systems as well as from high-throughput measurements.

Implications of Ago-loading kinetics and protein turnover for miRNA-dependent gene regulation

To explore the implications of our models for miRNA-dependent gene regulation, we investigated currently open questions in the field. One intensely debated aspect is the

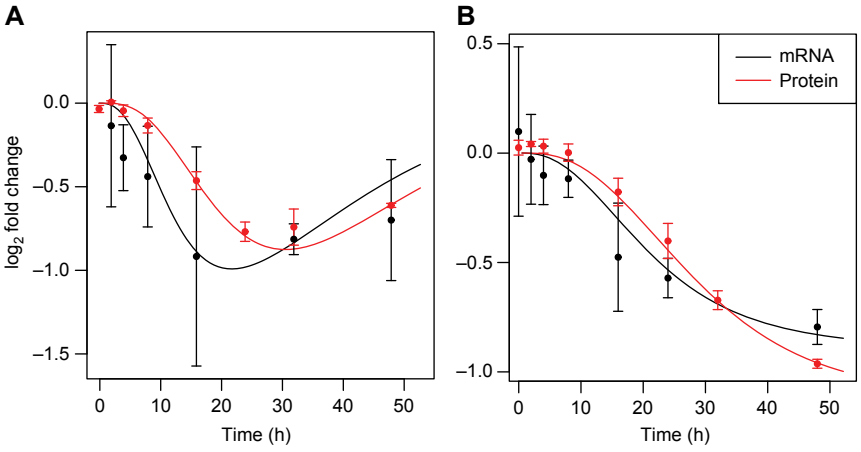


Figure 3 The kinetic model fits measured changes in mRNA and protein levels following miR-199a transfection (**A**) or induction (**B**) in HEK293 cells. The 3′ UTR of the KTN1 miR-199a target was cloned downstream of the stop codon of a luciferase reporter gene. Changes in mRNA expression following miR-199a transfection or induction were quantified by qPCR whereas changes in protein levels were determined by measuring luciferase activity. In the transfection experiment, changes in mRNA and protein levels were then fitted assuming the previously introduced three-exponential Ago-loading model. In the induction experiment, we assumed constant miRNA synthesis into an Ago-accessible environment. The best-fitted model appears as a continuous line and error bars represent 95% confidence intervals on the measured changes in mRNA and protein abundance. See also Supplementary Figure S3.

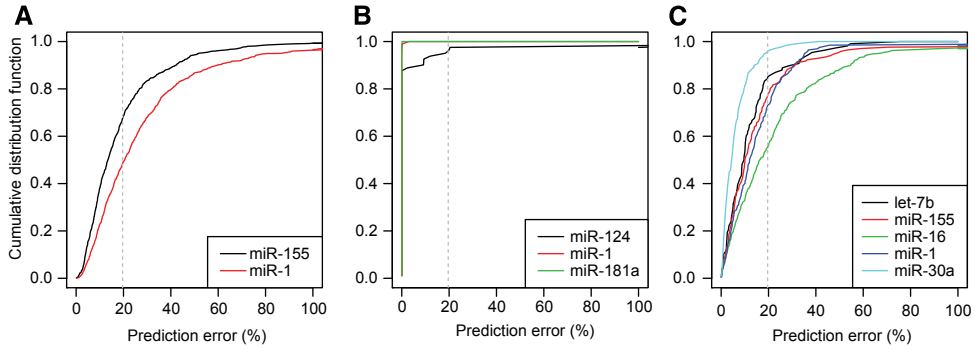


Figure 4 The kinetic model explains temporal changes in mRNA abundance, protein abundance and translation efficiency of most miRNA targets. Cumulative distributions of the relative error between the model prediction and the measurements in different transfection experiments. (**A**) Ribosome Protected Fragment (RPF) sequencing and mRNAseq experiments upon transfection of miR-155 and miR-1 by Guo *et al* (2010). (**B**) SILAC and microarray experiments upon miRNA transfection of miR-124, miR-1 and miR-181a by Baek *et al* (2008). (**C**) pSILAC and microarray experiments following the transfection of let-7b, miR-155, miR-16, miR-1 and miR-30a by Selbach *et al* (2008). The *x*-axis of each panel represents the per-gene relative error between the model prediction and the measurements. For any chosen cutoff on the relative error, the fraction of genes whose regulation following miRNA transfection could be predicted at the chosen error cutoff or less can be read on the *y*-axis. The dotted, vertical bars mark a 20% error cutoff on the fold change. This error level is typically observed in miRNA transfection experiments. See also Supplementary Figure S4.

extent to which miRNAs affect the decay as opposed to the translation rate of target mRNAs. To answer this question, a few studies obtained direct, high-throughput measurements of mRNA as well as protein-level changes upon miRNA transfection (Baek *et al*, 2008; Selbach *et al*, 2008). The general conclusion was that miRNAs predominantly affect the mRNA decay rate. If this were the case, then one would expect a strong correlation between the changes in levels of mRNA and in protein of individual genes, which is not what one typically observes. Rather, changes in protein abundance appear uncoupled from changes in mRNA abundance and miRNA targets typically change less at the protein compared with the mRNA level (see Section 3.4 of the Supplementary Material). By measuring protein decay rates by Selected Reaction Monitoring, we further found that fast-decaying proteins were preferentially detected as miRNA targets in these experiments (Section 3.4 of the Supplementary Material). Additionally, in the one experiment in which mRNA and protein-level changes were measured a week after the expression of a miRNA was abrogated yielded a much stronger correlation. All of these observations suggested that the mRNA and protein-level changes were measured far from steady state, and we re-analyzed these data in the framework of our kinetic model.

The three Stable Isotope Labeling by Amino acids in Cell culture (SILAC) experiments of Baek *et al* (2008) fitted the model very well. The dynamics of $>80\%$ of genes with a seed match to the transfected miRNA and downregulated 24 h post transfection was perfectly described by the model (Figure 4B; Supplementary Figure S4B). This included genes for which the mRNA was more strongly regulated than the protein (Supplementary Figure S4D). In addition, the values of the fitted parameters were consistent with the values obtained from the other data sets (Supplementary Figure S4). Application of the model to the pulsed SILAC (pSILAC) data from the five transfection experiments of Selbach *et al* (2008) revealed a similar picture, with 50–90% of downregulated mRNAs with seed match to the transfected miRNA fitting the model (Figure 4C; Supplementary Figure S4C), including genes for which protein levels were less affected than mRNA levels (Supplementary Figure S4E). Thus, our model explains to a large extent the measured effects of miRNA on mRNAs and protein levels. Dissecting the relative contribution of translational repression and mRNA decay to the miRNA-mediated repression of individual targets (Supplementary Figure S4A–C) we found that on average $\approx 25\%$ of the putative targets on a miRNA undergo predominantly translation repression. Thus, our analysis supports the previous assertions that miRNAs have a stronger impact on mRNA decay compared with translation (Baek *et al*, 2008, Guo *et al*, 2010). It further indicates that the discrepancy between the measured mRNA and protein-level responses of individual miRNA targets was due to the expression dynamics of the transfected miRNAs and to the confounding effect of protein decay rates on protein levels.

The second question that we sought to address relates to the observation that the expression of some miRNAs changes rapidly as a function of cellular state. In particular, Krol *et al* (2010) showed that the expression of miRNAs from the miR-183/96/182 cluster (but not of other miRNAs) changes rapidly in response to illumination and that these changes are important for the regulation of gene expression in the retina. Circadian changes in miRNA expression levels in the liver have also been described (Gatfield *et al*, 2009) and, examining the relative abundance of miRNAs in Ago2-immunoprecipitation samples prepared from HeLa cells in M phase and unsynchronized cells (Kishore *et al*, 2013) we also found that miRNAs of the let-7 family are two-fold upregulated during the M phase while miR-21 is 66% downregulated (Supplementary Table S1). In all of these circumstances the targets of the miRNAs whose expression is modulated should respond rapidly, on the timescale of about a day. We therefore asked under what conditions would the protein-level dynamics of miRNA targets reflect the rapid dynamics of the miRNA itself.

We used the model for miRNA induction experiments that was introduced in the previous section to compute the amplitude in the oscillation of a target protein level as a function of the protein decay rate and miRNA kinetics. To emulate the qPCR time series of Krol *et al* (2010), we explored a

regulatory scenario in which miRNA synthesis switched between two regimes during a 24-h cycle (Supplementary Figure S5A): the first 12 h of low miRNA expression (synthesis rate: half of its maximum level) and the next 12 h of high expression (maximal synthesis rate). Although it does not come as a surprise that proteins with median half-life (48 h; Schwanhäusser *et al*, 2011) undergo oscillations of $<1\%$ in amplitude (Figure 5A), increasing the protein turnover rate by two orders of magnitudes still only leads to a modest increase in the amplitude of protein oscillations ($<10\%$; Figure 5D). Responsible for these small changes are the slow kinetics of Ago loading and miRNA decay. It is only when fast protein turnover is combined with fast miRNA kinetics (rates of Ago association, dissociation, and of miRNA decay) that substantial oscillations in protein levels (60% of the baseline level assuming no translation repression and up to 2.5-fold with translation repression; Figure 5B) ensue.

Another highly relevant regulatory scenario is when the synthesis of a miRNA is abrogated at the transition between cell types or states. For example, rapid synthesis and subsequent clearance of the dre-miR-430 miRNA is important for proper zebrafish development (Chen *et al*, 2005; Giraldez *et al*, 2005). We thus investigated how fast target proteins are expected to change in response to a sudden block in miRNA synthesis (Supplementary Figure S5B). We again used the model introduced in the previous section for miRNA induction experiments and computed the time necessary for proteins to reach 90% of their steady-state level in the absence of the miRNA—the *recovery time*—as a function of the protein decay rate and of the miRNA kinetics. As expected, simulations revealed that the main factor limiting the recovery time is the protein decay rate (Figure 5C and D). However, they additionally allowed us to estimate the timescale of the protein response. Namely, proteins with a median half-life of 48 h (Schwanhäusser *et al*, 2011) would have a recovery times larger than 140 h (6 days). This implies that miRNAs are a slow means of regulating the levels of typical proteins. It may thus be expected that miRNAs preferentially target proteins with rapid turnover. We tested this hypothesis by comparing the decay rates of top compared with weakest predicted miRNA targets (highest and lowest, respectively, prediction scores according to the ElMMo algorithm (Gaidatzis *et al*, 2007) that estimates the strength of evolutionary selection of the miRNA-complementary sites) measured by Cambridge *et al* (2011). Indeed, we found that the highest scoring targets had significantly higher protein decay rates than low-scoring targets. This indicates that miRNAs preferentially target fast-decaying proteins, on whose expression they have a rapid regulatory impact (Supplementary Figure S5C). Alternatively, mechanisms such as increased protein turnover

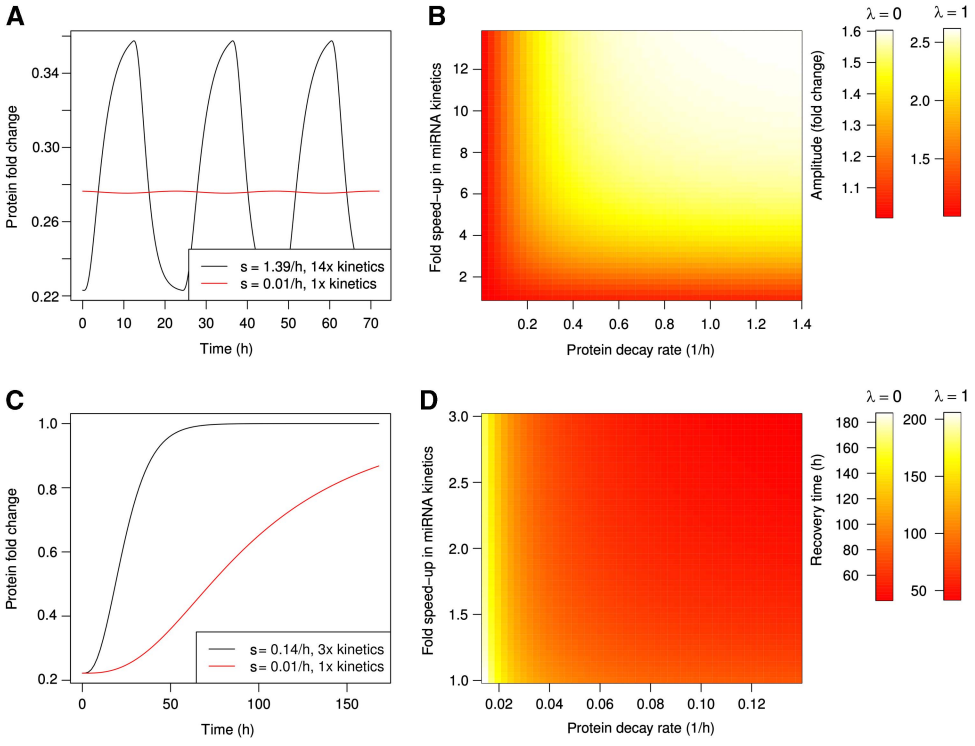


Figure 5 Parameter ranges that are compatible with a specific dynamic of protein targets. (**A**) Changes in protein levels induced by a miRNA whose synthesis switches between half-induction and full induction in 24 h cycles. Simulations were performed assuming the default kinetic parameter (48 h protein half-life, miRNA loading and decay rates estimated from biophysics data, red line) or faster kinetics (30 min protein half-life, 14-fold speed-up in miRNA loading and decay, black line). (**B**) Amplitude (fold change) of the oscillations in protein abundance as a function of protein decay and miRNA kinetics. The color bars correspond to the case where miRNA only affect mRNA decay ($\lambda = 0$) or equally regulate mRNA decay and translation ($\lambda = 1$). (**C**) Changes in protein levels following a sudden drop in miRNA synthesis given default kinetic parameter (48 h protein half-life, miRNA loading and decay rates from biophysics data, red line) or faster kinetics (5 h protein half-life, three-fold speed-up in miRNA loading and decay, black line). (**D**) Protein recovery time as a function of protein decay and miRNA kinetics. The color bars correspond to the case where miRNA only affect mRNA decay ($\lambda = 0$) or equally regulate mRNA decay and translation ($\lambda = 1$). See also Supplementary Figure S5.

or cell division would need to be recruited to diminish the recovery time of miRNA targets. Similar considerations apply to the situation when a miRNA is strongly induced to repress its targets.

Discussion

By contrast to transcription factors, miRNAs are often viewed as rapid means of regulating gene expression. Although synthesizing a miRNA likely requires less time than synthesizing a protein, miRNAs only function as part of complexes with Ago proteins. Some studies (Khan *et al*, 2009) already suggested that Ago levels within cells are limiting, implying that for a miRNA to exert its function, it has to displace other miRNAs from Ago. The models we introduced here abstract much of the molecular complexity of the miRNA pathway, but their parameters can be fitted from presently available data sets. As a result, it remains challenging to pinpoint the precise steps (binding the double-stranded miRNA to Ago, expelling the passenger/miR* strand, etc.) that are most responsible for the inferred bottlenecks. Nevertheless, our analysis indicates that it is the overall Ago-loading process that determines the timescale of miRNA-dependent gene regulation. At an estimated Ago-dissociation rate of $u = 0.116\,\mathrm{h}^{-1}$, the half-life of Ago residency is ~ 6 h, which sets a lower bound on the time that miRNAs need to regulate their target.

Re-analysis of data from experiments in which both mRNA and protein levels were measured after miRNA transfection shows that mRNAs underwent more pronounced changes than the proteins. The underlying explanation appears to be different in different experimental settings. In the SILAC experiments, it is the slow rate of protein decay that does not allow the full extent of protein downregulation to be reached before the miRNA is degraded. In contrast, in the pSILAC experiments the change of labeling medium, done 8 h post transfection, occurred before the full Ago loading with the transfected miRNA was reached (20–30 h post transfection). Thus, the proteins were labeled before the miRNA fully exerted its inhibitory effects and the confounding effect of protein decay rate was only partially circumvented. This suggests that changing the labeling medium at a later time point would better capture the effect of miRNAs of protein synthesis by pSILAC.

Simulating changes in mRNA and protein levels for two biological realistic scenarios, namely circadian oscillation of miRNA expression and a sudden change in miRNA synthesis, we found that protein decay rates critically limit the efficiency of miRNA regulation. Importantly, we found that oscillations of large amplitude in protein levels in response to an oscillating miRNA are predicted to be achievable only when the miRNA undergoes fast turnover and when the miRNA–Ago loading kinetics is fast, much faster than micro-injected siRNAs were found to load and unload from Ago. This is not a reflection of differences between siRNA and miRNA dynamics, because the siRNA–Ago loading model allows a very good fit to mRNA and protein-level data from a variety of experiments in which the miRNA expression was manipulated. Rather, our results suggest that active mechanisms increase miRNA turnover in specific *in vivo* conditions. The rate of Ago loading with a specific miRNA can be increased by activating the miRNA's transcription. Exonucleases such as Xrn2 (Chatterjee and Grosshans, 2009), RRP41 (Bail *et al*, 2010), PNPase(old-35) (Das *et al*, 2010) and Xrn1 (Chatterjee *et al*, 2011), which have been implicated in active miRNA degradation as well as destabilization by a complementary miRNA (Chen *et al*, 2011; De *et al*, 2013) can destabilize the miRNA, thereby critically regulating miRNA residency in Ago.

Large oscillations in protein levels in response to an oscillating miRNA require a fast protein turnover rate. This situation is probably exemplified by the miR-16 family of miRNAs that regulates the cell cycle (Linsley *et al*, 2007), directly targeting Cyclin D3, Cyclin E1 (Liu *et al*, 2008) and likely also Cyclin D2, the latter being one of the top predicted ElMMo (Gaidatzis *et al*, 2007) targets of these miRNAs with three conserved binding sites. Indeed, cell-cycle proteins do undergo fast decay (Schwanhäusser *et al*, 2011), and these cyclins have been shown to undergo proteolysis at different points of the cell cycle (Russell *et al*, 1999; Strohmaier *et al*, 2001; Kida *et al*, 2007). Thus, miRNAs should be able to meaningfully modulate the expression of these proteins if they can rapidly load and unload from the Ago protein and decay. A natural timescale for the decay of proteins is set by cell division. In cells that undergo substantial clonal expansion, the miRNA-induced repression can take hold on the timescale of protein dilution by cell division. In non-dividing retinal neurons (Krol *et al*, 2010) however, miRNA-based regulation of protein levels would be slow unless the miRNA targets have a high turnover rate.

In conclusion, we constructed a unifying mathematical framework of the kinetics of miRNA-mediated gene regulation starting with Ago loading, going through miRNA-dependent change in mRNA abundance, translation efficiency and protein abundance. We inferred the model parameters from various types of experiments and showed that it accurately describes regulatory dynamics in miRNA micro-injection, induction and transfection experiments. This framework will help in further designing miRNA target identification experiments and in characterizing miRNA function away from the steady state such as in development, cell cycle and circadian rhythms.

Materials and methods

Modeling the kinetics of Ago loading

We briefly outline the models describing Ago loading with small RNAs that we use in our study. Full derivations can be found in Section 3.2 of the Supplementary Material.

Modeling Ago loading following siRNA micro-injection

In a micro-injection experiment, X_0 siRNAs enter the cell at time $t = 0$ h. A fraction f_0 of the R Ago molecules is not bound to endogenous miRNAs (Khan *et al*, 2009) and therefore available for binding to siRNAs. If we assume that the fraction of free Ago f_0 remains fairly constant during the experiment—that is, if siRNAs load mostly by displacing miRNAs from Ago—and that the amount of free endogenous miRNAs is not significantly affected by the small RNAs micro-injection, that siRNAs only decay if they are not bound to Ago and that Ago–siRNA association and dissociation rates happen on a fast timescale compared with processes such as synthesis and degradation,

then we can write the dynamics of loaded Ago A and free siRNAs X as

$$\begin{cases} \frac{dA}{dt} &= bf_0RX - uA \\ \frac{dX}{dt} &= -dX - bf_0RX + uA \end{cases} \quad (1)$$

where free siRNAs X decay at rate d, associate with a free Ago at rate b and unload from a loaded Ago A at rate u. The equation for A specifies that the production rate of loaded Ago depends on the amount of free X, on the amount of free Ago f_0R and on the binding rate b. Loaded Ago A is then lost by dissociation that happens at rate u. In the equation that defines the dynamics of free siRNA X, the first term accounts for first-order decay (at rate d), the second for the association with a free Ago and the last for the release from a loaded Ago. For convenience when fitting the biophysical data of Ohrt *et al* (2008), we can rewrite these equations in terms of the fraction of loaded Ago $a = \frac{A}{R}$ and substitute $\gamma = bf_0$ which gives

$$\begin{cases} \frac{da}{dt} = \gamma X - ua \\ \frac{dX}{dt} = -(d+\gamma R)X + uRa \end{cases} \quad (2)$$

The analytical solution for the Ago-loading function $a(t)$ can be shown to be bi-exponential where each exponential accounts for a different phase of the Ago-loading dynamics. The initial phase corresponds to the displacement of endogenous miRNAs by the siRNAs, and the second phase corresponds to the slow clearance of siRNAs from cells.

Modeling the loading of Ago with miRNAs upon miRNA transfection

In miRNA transfection experiments, miRNAs are delivered with liposomes as vectors as opposed to being directly injected into cells. This likely introduces an additional delay as miRNAs need to traffic from liposomes to endosomes and cytoplasm (Stalder *et al*, 2013). To account for that delay, we introduce an additional compartment V to our model. V_0 miRNAs are delivered to V at time $t = 0$ h, and from there they can either decay with rate $\tilde{d}$ or transfer to the cytoplasm X with rate r. The ordinary differential equation (ODE) system describing these dynamics can be written as

$$\begin{cases} \frac{dV}{dt} = -\left(\tilde{d}+r\right)V \\ \frac{dX}{dt} = rV - (d+\gamma R)X + uRa \\ \frac{da}{dt} = \gamma X - ua \end{cases} \quad (3)$$

The analytical solution for the Ago-loading function is tri-exponential,

$$a(t) = \tilde{k}_1 e^{-\beta_1 t} + \tilde{k}_2 e^{-\beta_2 t} - (\tilde{k}_1 + \tilde{k}_2) e^{-(\tilde{d}+r)t}$$

whose parameters $\tilde{k}_1, \tilde{k}_2, \beta_1, \beta_2$ can be computed from the biophysical parameters illustrated in Figure 2A (see below).

Modeling Ago loading upon miRNA induction.

When the miRNA is induced (by doxycyclin in the case of our experiment), there is no need to account for the endosomal compartment. We therefore start from Equation (2) and slightly alter the model by assuming that miRNA synthesis $c(t)$ is inactive before 0 h and follows 0th order kinetics upon doxycyclin induction, that is

$$\begin{cases} \frac{dX}{dt} = c(t) - (d+\gamma R)X + uRa \\ \frac{da}{dt} = \gamma X - ua \end{cases} \quad (4)$$

with

$$c(t) = \begin{cases} 0 & \text{if } t < 0 \\ X_0 d & \text{if } t \geq 0 \end{cases}$$

The fraction of loaded Ago $a(t)$ is then coupled to changes in mRNA decay and translation as will be described below (Equation 8).

Modeling miRNA-induced changes in mRNA and protein abundance

We modeled mRNA and protein dynamics using an ODE model along the lines of those previously introduced by Hargrove and Schmidt (1989) and Khanin and Higham (2009)

$$\begin{cases} \frac{dm}{dt} &= c - (d_0 + \Delta d(t))m \\ \frac{dp}{dt} &= (l_0 + \Delta l(t))m - sp \end{cases} \quad (5)$$

where m is the mRNA abundance, p the protein abundance, c the transcription rate, d_0 and l_0 the mRNA decay and translation rates in the absence of the miRNA, s the protein decay rate, and $\Delta d(t)$ and $\Delta l(t)$ represent the time-dependent changes in mRNA decay and translation rates induced by the miRNA.

Assuming that mRNA and protein abundances are at steady state before miRNA transfection, that is, $m(0) = \frac{c}{d_0}$ and $p(0) = \frac{c}{d_0}\frac{l_0}{s}$, Equation (5) can be rewritten in terms of the fold change in mRNA and protein abundances f_m and f_p, that do not depend on the transcription rate c:

$$\begin{cases} \frac{df_m}{dt} &= d_0\left[1 - \left(1 + \frac{\Delta d(t)}{d_0}\right)f_m\right] \\ \frac{df_p}{dt} &= s\left[\left(1 + \frac{\Delta l(t)}{l_0}\right)f_m - f_p\right] \end{cases} \quad (6)$$

If after transfection the miRNA was to persist at a constant level until a new steady state was reached, then the change in protein abundance relative to the initial state would be given by $f_p = \left(1 + \frac{\Delta l(t)}{l_0}\right)f_m$, that is proportional to the change in mRNA abundance, the proportionality factor being the change in translation rate of the mRNA. The timing with which the protein abundance changes and how closely it follows the mRNA abundance will depend on the decay rate of the protein, s; fast-decaying proteins will reach the steady state faster in response to the miRNA compared with slow-decaying proteins.

Of the two approaches to measure protein abundance changes upon miRNA transfection, SILAC (Baek *et al*, 2008) measurements can be described directly by Equation (6). pSILAC (Selbach *et al*, 2008), on the other hand, measures newly synthesized proteins. That is, at $t_c = 8$ h after transfection, the growth medium was changed and the amount of protein synthesized between 8 and 32 h after either miRNA or mock transfection was estimated. The difference between the amount of protein produced in miRNA- and mock-transfected cells between 8 and 32 h is given by

$$\frac{df_p}{dt} = \frac{s}{1 - e^{-s(t-t_c)}}\left[\left(1 + \frac{\Delta l(t)}{l_0}\right)f_m - f_p\right] \quad (7)$$

Compared with Equation (6), the magnitude of measured changes in protein abundance no longer depends on the protein decay rate s, but rather on the factor $\frac{s}{1-e^{-s(t-t_c)}}$, which is larger than s, particularly shortly after the medium change ($t = t_c$). Consequently, pSILAC amplifies changes in protein levels by a factor $\frac{s}{1-e^{-s(t-t_c)}}$ that decreases with time. In addition, at constant time t, the strongest amplification occurs when $1 - e^{-s(t-t_c)}$ is small, that is for stable proteins (small s). The change of a stable protein is therefore amplified more than that of an unstable protein, and therefore the regulatory impact of a miRNA will be revealed whether the protein is stable or unstable (Supplementary Figure S5D). As a consequence, for a miRNA target to appear as a 'non-responder' in a pSILAC experiment, its decay rate has to be extremely small, which we could verify experimentally (see Supplementary Material). The similarity in the fold-change trajectories of fast- and slow-decaying miRNA targets is of course a main rationale of the pSILAC approach (Schwanhäusser *et al*, 2009) which by reducing the confounding effect of protein decay rates leads to higher correlation between mRNA and protein fold changes (average of $r = 0.27$ and $r = 0.41$ for all genes and genes with seed match, respectively, compared to an average of $r = 0.18$, for both categories of genes in SILAC, see Supplementary Table S2).

Assuming that the relative increase in mRNA decay and the decrease in translation is proportional to the fraction of Ago molecules loaded with the miRNA $a(t)$ yields

$$\frac{d_0 + \Delta d(t)}{d_0} = 1 + \delta a(t) = 1 + \tilde{a}(t) \quad (8)$$

$$\frac{l_0}{l_0 + \Delta l(t)} = 1 + \lambda\delta a(t) = 1 + \lambda\tilde{a}(t). \quad (9)$$

where δ and λ are proportionality factors. Because miRNAs increase mRNA decay and repress translation, both λ and δ should be positive.

$\tilde{a}(t) := \delta a(t)$ represents the relative change in the mRNA decay rate as a function of time.

Combining Equation (8), that relates the amount of miRNA-loaded Ago to changes in the *rate* of mRNA decay with Equation (6), that describes changes in mRNA *level*, we obtain a relationship between the change in mRNA level f_m and the time-dependent change in the mRNA decay rate $\tilde{a}(t)$,

$$\frac{df_m}{dt} = d_0[1-(1+\tilde{a}(t))f_m]. \quad (10)$$

Friedel *et al* (2009); Schwanhäusser *et al* (2009), measured mRNA decay rates in high-throughput. Although we can use the per-gene measurements of mRNA decay rates from these studies, the data reveal a fairly tight distribution around a median of 0.12/h (5.8 h half-life). For simplicity, we set the mRNA decay rate d_0 to 0.12/h in the above relation to analyze the mRNA response to miRNA transfection.

Similarly, by substituting Equations (8) and (9), that relate the amount of miRNA-loaded Ago to changes in the rates of mRNA decay and translation, into Equation (6) which describes changes in protein levels, we obtain

$$\frac{df_p}{dt} = s\left(\frac{f_m}{1+\lambda\tilde{a}(t)} - f_p\right) \quad (11)$$

which describes the change in protein level that would be measured as a function of time in an SILAC experiment. A similar equation can be derived for pSILAC measurements by substituting Equation (8) into Equation (7) to yield

$$\frac{df_p}{dt} = \frac{s}{1-e^{-s(t-t_c)}}\left(\frac{f_m}{1+\lambda\tilde{a}(t)} - f_p\right). \quad (12)$$

Because the protein decay rates have a much wider distribution than mRNA decay rates, in both models we used the measured (in HeLa cells) decay rates s of individual proteins Cambridge *et al* (2011) rather than a median decay rate.

Changes in mRNA and protein abundance f_m are driven by the relative change in mRNA decay $\tilde{a}(t)$ which depends on how the miRNA is delivered. In the case of a transfection experiment, one can solve Equation (3) to show that $\tilde{a}(t)$ takes the form

$$\tilde{a}(t) = \tilde{k}_1 e^{-\beta_1 t} + \tilde{k}_2 e^{-\beta_2 t} - (\tilde{k}_1+\tilde{k}_2)e^{-(\tilde{d}+r)t} \quad (13)$$

with

$$\begin{cases} \beta_{1,2} &= \frac{d+\gamma R+u \pm \sqrt{(d+\gamma R+u)^2-4ud}}{2} \\ \tilde{k}_1 &= \frac{\beta_1+d+\gamma R}{uR}\frac{\delta r V_0}{\beta_2-\beta_1}\frac{\beta_2+d+\gamma R}{\beta_1+\tilde{d}+r} \\ \tilde{k}_2 &= -\frac{\beta_2+d+\gamma R}{uR}\frac{\delta r V_0}{\beta_2-\beta_1}\frac{\beta_1+d+\gamma R}{\beta_2+\tilde{d}+r} \end{cases}$$

The parameters d, γ, R, u specify the kinetics of Ago loading and can be fitted from the experiments of Ohrt *et al* (2008) (see below). As a result, mRNA dynamics following miRNA transfection are determined by two free parameters, $\delta r V_0$ and $\tilde{d}+r$. Modeling changes in protein abundance requires an additional parameter λ.

In the case of an induction experiment, solving Equation (4) shows that the function $\tilde{a}(t)$ takes the form of bi-exponential

$$\tilde{a}(t) = X_0\delta\frac{\gamma}{u}\left(1-\frac{\beta_2}{\beta_2-\beta_1}e^{-\beta_1 t}+\frac{\beta_1}{\beta_2-\beta_1}e^{-\beta_2 t}\right) \quad (14)$$

with

$$\beta_{1,2} = \frac{d+\gamma R+u \pm \sqrt{(d+\gamma R+u)^2-4ud}}{2}.$$

There too, we use the values of the Ago-loading parameters d, γ, R, u inferred from the experiments of Ohrt *et al* (2008) (see below). From Equation (14), we see that mRNA changes upon miRNA induction $f_m(t)$ are driven by one free parameter $X_0\delta$. As in the transfection case, modeling protein dynamics $f_p(t)$ require an additional parameter λ, which accounts for the effect of miRNAs on translation relative to mRNA decay.

Measuring changes in target protein and mRNA abundance upon miR-199 transfection and induction by luciferase assays and qPCR

Establishing a cell line stably expressing a miRNA target as well as inducibly expressing the cognate miRNA

We generated a stable HEK293T cell line that simultaneously expresses hsa-miR-199a-3p (miRBase accession MIMAT0000232) and its target 3′UTR (Kinectin 1, KTN1, RefSeqID NM_004986). KTN1 was chosen among the hsa-miR-199a-3p targets predicted by ElMMo (Gaidatzis *et al*, 2007) that had at least one binding site with high probability of being under selective pressure ($P>0.7$) and a 3′ UTR shorter than 1500 nts so that it could be cloned efficiently. The approach we took is illustrated in Supplementary Figure S6. We PCR amplified the primary hsa-miR-199a-3p and the target KTN1 3′ UTR and cloned them into pGEM-T Easy vector (Promega). The hsa-miR-199a-3p insert was sequenced and subsequently cloned into an inducible pRTS-1 vector (Bornkamm *et al*, 2005), replacing the luciferase gene at *SfiI* restriction sites. We then transfected HEK293T cells with the miR-199a-3p-pRTS-1 plasmid and selected colonies after 2 weeks of culture in the presence of hygromycin B (Calbiochem, 100 μg/ml). Selected colonies were subsequently propagated in the presence of hygromycin B.

The cell line was tested for hsa-miR-199a-3p expression after the addition of doxycycline (1 μg/ml) with northern blotting (Supplementary Figure S7). The small RNA Northern was performed as previously described (Pall and Hamilton, 2008), with a minor change: we used the conventional TBE buffer instead of an MOPS-NaOH buffer.

The KTN1 3′ UTR insert from the pGEM-T Easy vector was sequenced and subsequently cloned downstream of the open reading frame of renilla luciferase in a psi-CHECK2 vector (Promega). The miR-199a-3p-pRTS-1 cell line was then used for co-transfection of KTN1-psiCHECK-2 and pPUR (Clontech) plasmids in a 3:1 ratio. Stable colonies were isolated after 2 weeks of culture in the presence of hygromycin B and puromycin (Sigma-Aldrich, 0.75 μg/ml). Selected colonies were subsequently propagated in DMEM with 10% FCS, penicillin-streptomycin (Sigma-Aldrich), hygromycin B and puromycin. We confirmed the genomic integration of the KTN1 3′ UTR by PCR (see Supplementary Figure S8).

MiRNA transfections and luciferase assay

KTN1 cells were split in 12-well plates for both transfection and induction experiments. hsa-miR-199a-3p mimic (c-300536-07-0005) and miRNA mimic negative control (CN-002000-01-05, Thermo Fisher Scientific) were transfected with a final concentration of 25 nM using lipofectamine 2000 (Invitrogen) transfection reagent according to the manufacturer's protocol. At the same time in a different set of experiment, hsa-miR-199a-3p expression was induced with doxycycline (Sigma-Aldrich, 1 μg/ml). Cells were collected at different time points (0, 2, 4, 8, 16, 24, 32 and 48 h). Both luciferase and qRT-PCR were performed with cells from a single well. Renilla and firefly activities were measured with the dual luciferase reporter assay system (Promega) with a luminometer (Centro LB960, Berthold Technologies). Firefly luciferase was used as an internal control.

Quantitative real-time PCR

We extracted total RNA with the TRI reagent (Sigma) as per the manufacturer's protocol, and then applied DNase digestion with RQ1DNase (Promega) followed by phenol-chloroform purification and cDNA synthesis with SuperScriptIII (invitrogen) reverse transcriptase according to the manufacturer's protocol. We measured mRNA levels with the Step One Plus real-time PCR system (Applied Biosystems) employing Power SYBR Green PCR Master Mix (Applied Biosystems). Firefly expression was used as an internal control. The primers for KTN1 3′ UTR were forward—GGGG CTCGAG TGGG AAACTGTTCATTTGAGG, reverse—TATT GCGGCCGC TTGCTGACGCC ATTACAAAA. Primers for hsa-miR-199a-3p were miR-199a-3p_1_forward—AAAAGGCCTCACTGGCCCCTCCCCCACTCTTTAGGAT,

miR-199a-3p_1_reverse—AAAAGGCCTCACTGGCCGTGGGGATGGCA GACTGATA. Primers for KTN1 cell line were GAGCGCGTGCTGAA GAACGAG (forward) and TTGCTGACGCCATTACAAAA (reverse). qRT-PCR primers were for firefly luciferase: hFluc-RT-rev—CGGTAGAC CCAGAGCTGTTC, hFluc-RT-for—TGCAGAAGATCCTGAACGTG, and for renilla luciferase: hRluc-RT-rev—CTAACCTCGCCCTTCTCCTT and hRluc-RT-for—TCGTCCATGCTGAGAGTGTC.

In the qPCR data, the standard deviation on the $\log_2$ fold change was on average 0.31 (corresponding to a 23.8% uncertainty on the fold change), as typically observed in miRNA transfection experiments (Supplementary Figure S2).

Fitting kinetic parameters

We now describe the strategies we used to fit the kinetic parameters to the different data sets. Table I provides an overview of all parameters and documents how we determined their values.

Fitting the parameters of Ago loading from the fluorescence cross-correlation spectroscopy siRNA micro-injection time course

We obtained the fractions of cytoplasmic Ago in complex with the siRNA and the cytoplasmic fractions of siRNAs in complex with Ago 0, 1, 3, 6, 12, 24 and 48 h after siTK3 micro-injection together with the standard errors on these fractions (Ohrt *et al*, 2008). This experiment was the largest of those performed by Ohrt *et al* (2008). We subtracted the fraction of bound Ago and bound siRNA at 0 h from all the measurements so that the fraction of bound Ago and siRNA was 0% at 0 h.

To fit the parameters, we simulated the fraction of Ago in complex with the siRNA and the fraction of siRNA in complex with Ago using Equation (2). Initial parameter space exploration by gradient ascent methods on the likelihood assuming a Gaussian error model suggested initial parameter estimates: $\gamma = 10^{-5}$, $d = 0.12$, $u = 0.12$ and $X_0 = 16\,000$. We then refined these estimates and determined their uncertainty by Markov Chain Monte Carlo. The chain was initialized at 10% of our initial estimates. From there, we performed Gaussian moves with standard deviation set to 33% of our initial parameter estimates, with a uniform prior that only allowed positive parameter values. The chain ran for 10^6 sampling steps and visual inspection of the sampling traces suggested that sampling was at equilibrium (Supplementary Figure S1). Acceptance probability was 20.8%. The parameter set with highest likelihood was $\gamma = 3.05 \times 10^{-6}$/molecule/h, $d = 0.102$/h, $u = 0.116$/h and $X_0 = 18\,661.67$ molecules. Finally, we determined the posterior probability distributions of the parameters from the MCMC samples, excluding the first 50 000 samples to ensure that the chain was 'burnt-in' (Supplementary Figure S1).

The product of the miRNA–Ago association rate b and the fraction of free Ago f_0, $\gamma = bf_0$, was in the range of 10^{-6}/molecule/h suggesting that the rate of small RNA binding to Ago is small or that there is little free Ago at steady state. In contrast, small RNA dissociation from Ago (u) appears to be about as fast as their decay (d), of the order of 10^{-1}/h. mRNA decay rates are in a similar range (Friedel *et al*, 2009; Schwanhäusser *et al*, 2011). Finally, we estimated that $X_0 \simeq 19\,000$ molecules of small RNA were injected in the cell at time $t = 0$ h.

Fitting changes in mRNA, protein and RPF abundance upon miRNA transfection by previous studies, and in the present study through luciferase activity and qPCR of a reporter gene

We fixed the baseline mRNA decay rate d_0 to the average of previously estimated values. That is, Friedel *et al* (2009) estimated a median decay rate of 0.15/h in murine fibroblasts and 0.13/h in human B cells, while Schwanhäusser *et al* (2011) reported a median mRNA decay rate of 0.08/h in murine fibroblasts. We used the average of these values, 0.12/h for d_0, corresponding to a mRNA half-life of 5.8 h. For the dynamics of miRNA loading into Ago, we assumed that the rates of association (b)/dissociation (u) of the Ago–miRNA complex, the miRNA decay rate d, and the fraction of free Ago at steady state f_0 were the same as in the experiment of Ohrt *et al* (2008). Hence, we fixed the parameters $\gamma = bf_0$, u and d to the values mentioned in Figure 1B. The remaining parameters X_0, the initial amount of miRNA and δ, the factor that relates the fraction of loaded Ago to the relative change in the mRNA decay rate, most certainly vary between experiments. Under the bi-exponential Ago-loading model, δ and X_0 always occur as a product in the analytical expression of $\tilde{a}(t)$ (Equation 7 of the Supplementary Material). The model has hence one free parameter δX_0 which we estimated by fitting mRNA profiling time-series data. Similarly, the analytical expression of $\tilde{a}(t)$ under the tri-exponential Ago-loading model (Equation 13) depends on two parameters: $\tilde{d} + r$, the rate with which miRNAs are cleared from endosomes through degradation or by transferring to the cytoplasm, and $rV_0\delta$, that quantifies the influence of the transfected miRNAs on mRNA degradation.

For each transfection experiment (Wang and Wang, 2006; Baek *et al*, 2008; Selbach *et al*, 2008; Guo *et al*, 2010), we scanned 25 values of $\tilde{d} + r$ between 0.003/h and 0.6/h, equally spaced on a logarithmic scale. For each of these values and for each gene, we first obtained the $rV_0\delta$ parameter by maximum likelihood from the mRNA profiling data,

Table I Parameters were either obtained from previous measurements or fitted to the data

		Figure 1 Ohrt	Figure 2 Wang	Figure 3 Transfection	Figure 3 Induction	Figure 4 Guo	Figure 4 Baek	Figure 4 Selbach
γ	miRNA–Ago association	Fitted	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]
d	miRNA decay	Fitted	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]
u	miRNA–Ago dissociation	Fitted	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]	Fixed[a]
X_0	Amount of micro-injected miRNAs	Fitted						
d_0	mRNA decay		Fixed[b]	Fixed[b]	Fixed[b]	Fixed[b]	Fixed[b]	Fixed[b]
s	Protein decay			Fixed[c]	Fixed[c]		Fixed[d]	Fixed[d]
$\tilde{d} + r$	Clearance from endosomes		Fitted	Fitted		Fitted	Fitted	Fitted
$rV_0\delta$	miRNA-induced change in mRNA decay (transfection)		Fitted	Fitted		Fitted	Fitted	Fitted
$X_0\delta$	miRNA-induced change in mRNA decay by miRNAs (induction)				Fitted			
λ	Translation repression			Fitted	Fitted	Fitted	Fitted	Fitted

Rows correspond to parameters and columns to data sets. Parameters were either fitted, fixed to the indicated reference. Empty cells represent cases in which a given parameter is irrelevant to the data set. To give an overview, we briefly indicate the references used to set the parameters. The main and Supplementary Methods sections describe how we determined them.

[a]The Ago-loading parameters γ, d, u were fixed to the values inferred from the micro-injection and fluorescence cross-correlation spectroscopy experiments of Ohrt *et al* (2008).

[b]The mRNA decay rate d_0 was fixed to 0.12/h, a value determined from the measurements of Friedel *et al* (2009) and Schwanhäusser *et al* (2011).

[c]The decay rate s of the luciferase was set to 0.14/h, (https://tinyurl.com/promegaBatesLuciferase).

[d]Protein decay rates were obtained from the mass-spectrometry measurements of Cambridge *et al* (2011).

assuming a Gaussian error model on the measured $\log_2$ mRNA fold change. To set the standard deviation of the error model in the case of the data of Wang and Wang (2006); Baek *et al* (2008); Selbach *et al* (2008); Guo *et al* (2010) where biological replicates were not performed, we used the value 0.25 from the miRNA transfection data of Karginov *et al* (2007) (Supplementary Figure S2C). In the case of our luciferase and qPCR measurements, we used the standard error computed from triplicate measurements. We focused on genes that carried a seed match to the transfected miRNA and whose cognate mRNA levels were downregulated following miRNA transfection, averaging $\log_2$ fold changes when several time points were available. When analyzing the qPCR measurements of the present study, we also discarded data points for which changes in mRNA and protein levels could not be determined in triplicates because the machine failed to amplify the target. We maximized the log-likelihood with respect to $rV_0\delta$ by a combination of golden section searches and successive parabolic interpolations implemented in the `optimize()` function of the R software (R Development Core Team, 2006). Finally, by repeating this procedure for the different values of $\tilde{d}+r$, we determine the value of $\tilde{d}+r$ that maximized the likelihood of each transfection experiment, as well as of the gene-specific $rV_0\delta$ parameters. With only one time point of mRNA and protein measurement in the experiments of Baek *et al* (2008), the parameter estimation problem is under-determined. As a result, the experiments of Baek *et al* (2008) do not suggest a single, most-likely estimate for $\tilde{d}+r$ but rather a range of values that are consistent with the measurements (Supplementary Figure S4B). The values of $\tilde{d}+r$ that we inferred from Wang's miR-124 transfection, our miR-199 transfection, and Guo's miR-155 and miR-1 transfections were within these ranges. Fixing the value of this parameter to 0.014/h—a value inferred from Wang's miR-124 transfection (Figure 2)—we then estimated the $rV_0\delta$ and λ parameters for each gene, based on the measured change in mRNA level at 24 h and protein change 48 h post transfection.

In experiments that also featured protein quantification or RPF profiling data, we further obtained maximum-likelihood estimates of the λ parameter for each gene given the previously determined $\tilde{d}+r$ and $rV_0\delta$ parameters and assuming the same Gaussian error model on the $\log_2$ fold change as for the mRNA data. When fitting the SILAC data of Baek *et al* (2008) and luciferase activities in our reporter system, we used the protein dynamics defined by Equation (6) while Equation (7) was used with the pSILAC data of Selbach *et al* (2008). In total, we thereby analyzed 117, 154 and 106 genes with seed matches to miR-124, miR-1 and miR-181 in Baek *et al* (2008) and 139, 211, 319, 197 and 157 genes in the let-7, miR-155, miR-16, miR-1 and miR-30a transfections of Selbach *et al* (2008) for which protein turnover measurements were available (Cambridge *et al,* 2011). In the case of the RPF profiling experiments of Guo *et al* (2010), we first computed the $\log_2$ fold change in translation efficiency r_i–m_i from the $\log_2$ fold change in RPF r_i and mRNA abundance m_i. In the case of our qPCR and luciferase measurements, we set the decay rate s of the luciferase to 0.14/h (see https://tinyurl.com/promegaBatesLuciferase). We then fitted λ by comparing the measured $\log_2$ fold change in translation efficiency to the model prediction $\log_2\left(\frac{1}{1+\lambda\tilde{a}(t)}\right)$ assuming the previously described Gaussian error model, where $\tilde{a}(t)$ is the change in the mRNA decay rate (Equation 3).

Fitting changes in luciferase activity and mRNA abundance upon miRNA induction

We first fitted the $X_0\delta$ parameter from temporal changes in the mRNA level by maximum likelihood assuming a Gaussian error model on the measured $\log_2$ fold change. The standard deviation was set to the standard error computed from triplicate measurements. By the same method as in the previous paragraph, we first determined the maximum-likelihood estimate of $X_0\delta$ from the qPCR data based on the Ago-loading dynamics of Equations (14) and (10), and then estimated λ from the luciferase activities using Equation (11).

Simulations

We simulated the dynamics of miRNAs and their targets under two regulatory scenarios using the same model as in an induction experiment (Equations 14, 10 and 11), where a miRNA is synthesized into an Ago accessible compartment X. We used parameter estimates from the biophysics data of Ohrt *et al* (2008) for association γ, dissociation u and decay d. δ was set to 7.32 (see below for the justification). Protein decay rate s was initially set to 0.015/h (48 h half-life).

Simulating protein recovery after a sudden block in miRNA expression

We set $X(0)$ to 18 000 which in the context of the other parameters implied that 50% of Agos are loaded with the miRNA at steady state (Supplementary Figure S5). This is representative of the case of highly expressed miRNAs such as miR-430 in the zebrafish embryo (Chen *et al*, 2005) or miR-122 in liver (Landgraf *et al*, 2007). We used measurements made on the miR-223 knockout neutrophils (Baek *et al*, 2008) to determine δ; namely, from the observations that miR-223 represents 5% of all miRNAs in neutrophils (Baek *et al*, 2008) and that miR-223 targets are upregulated by 37% upon miR-223 knock-out ($\log_2$ value of 0.45 in Supplementary Table S2), we can use Equation (9) to infer that the miRNA can change the rate of mRNA decay by a factor $\delta=\frac{0.37}{0.05}\simeq 7.4$ when Ago is fully saturated with the miRNA. The protein decay rate s was initially set to 0.015/h (48 h half-life). At $t=0$ h, miRNA synthesis stops, that is $c(t)=0$. Initial conditions were computed under the scenario that all variables are at steady state at $t=0$ h. By solving Equations (10), (11) and (4) at steady state, we obtain $a(0)=\frac{\gamma X(0)}{u}$, $f_m(0)=\frac{1}{1+\tilde{a}(0)}$ and $f_p(0)=\frac{f_m(0)}{1+\lambda\tilde{a}(0)}$. We then simulated free miRNAs X, loaded Ago a, mRNA f_m and proteins f_p according to Equations (10), (11) and (4) (see Supplementary Figure S5). In the absence of the miRNA, the fold change in protein level is $f_p=1$ by definition, but the value 1 is reached only as t tends toward infinity. We defined the recovery time as the amount of time necessary for $f_p(t)$ to exceed 0.9. We then monitored the recovery time as a function of the protein decay rate which we varied between 0.014/h and 0.14/h (corresponding to half-lives ranging from 5 to 48 h), and general miRNA kinetics where we multiplied $\gamma=bf_0$, u and d by a common factor k ranging from 1 to 3. Finally, simulations were performed under two scenarios: no translation repression $\lambda=0$ and equal contribution of translation repression and mRNA decay $\lambda=1$.

Computing the amplitude of oscillations in protein abundance in response to a miRNA with cyclic expression pattern

In this simulation, we sought to reproduce the observations of Krol *et al* (2010) by implementing the scenario in which miRNA synthesis was half-activated during the first 12 h of a 24-h cycle, and then fully activated during the second half of the cycle. We thus considered the miRNA synthesis function

$$c(t)=\begin{cases}\frac{1}{2}X_0d & \text{if} \quad 0\le t \bmod 24<12\\ X_0d & \text{if} \quad 12\le t \bmod 24<24\end{cases}$$

We again set X_0 to 18 000 which in the context of the other parameters implies that 50% of Agos would be loaded with the miRNA if miRNA synthesis remained fully active long enough for steady state to be reached. Note that defining miRNA synthesis as a function of the miRNA decay rate d makes it possible to change the rate d without changing the amount of free miRNA and loaded Ago. Initial conditions were computed numerically by simulating free miRNAs X, loaded Ago a, mRNA fold change f_m and protein fold change f_p according to Equations (10), (11) and (4) for 100 cycles (2400 h). From these initial conditions, we then simulated another three cycles (72 h) for visualization purposes (see Supplementary Figure S5). We then defined the amplitude of the oscillations as the ratio $\frac{\max f_p(t)}{\min f_p(t)}$ between the highest and smallest protein levels achieved in these three cycles. We monitored the amplitude of protein oscillations as a function of the protein decay rate which we varied between 0.014/h and 1.39/h (corresponding to half-lives ranging from 30 min to 48 h). In addition, the measurements of Krol *et al* (2010) suggest that miRNAs responding to light-dark adaptation undergo fast turnover, with half-lives as short

as 30 min, 14 times faster than the observed decay in the fluorescence cross-correlation spectroscopy experiments of Ohrt *et al* (2008) and used in our model. We therefore varied the miRNA decay rate, Ago association rate and Ago dissociation rates between their original values up to a 14-fold increase. Thereby, we varied the miRNA decay rate from $d = 0.10$/h to 1.39/h, the Ago dissociation rate from $u = 0.12$/h to 1.61/h and the normalized association rate from $\gamma = 3.11 \times 10^{-6}$/molecule/h to 4.27×10^{-5}/molecule/h.

Supplementary information

Supplementary Information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Lukas Burger (Bioinformatics platform, FMI Basel), Julien Béthune (Filipowicz lab, FMI Basel), Annecke Brümmer and members of the Zavolan and van Nimwegen group for input and stimulating discussions. We also thank Nitish Mittal for valuable input in the design and analysis of the reporter experiments and Helge Grosshans (FMI Basel) for constructive comments on the manuscript. JH acknowledges the support of the Swiss National Science Foundation (PBBSP3_146961) and EMBO (ALTF 1160-2012). Work in the Zavolan laboratory was supported by the Swiss National Science foundation grant #31003A_127307. RA is supported by the European Research Council (ERC-2008-AdG 233226). MZ and RA were further supported by SystemsX.ch, the Swiss Initiative for Systems Biology.

Author contributions: NS and RA designed the mass spectrometry experiments and NS performed the SRM measurements with the help of LJ for sample preparation. NS and JH analyzed the mass spectrometry measurements to estimate protein decay rates. JH performed all other analyses. APS, LJ, JH and MZ designed the luciferase/qPCR time-series experiments following miRNA transfection/induction and APS performed the experiments. EvN inspired the use of the kinetic approach and designed the miRNA loading model. JH and MZ designed the research and wrote the paper with input from all authors.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, Zhang Y, Yang W, Gruber PJ, Epstein JA, Morrisey EE (2011) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Cell Stem Cell* **8:** 376–388

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

Bail S, Swerdel M, Liu H, Jiao X, Goff LA, Hart RP, Kiledjian M (2010) Differential regulation of microRNA stability. *RNA (New York, NY)* **16:** 1032–1039

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* **136:** 215–233

Béthune J, Artus-Revel CG, Filipowicz W (2012) Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* **13:** 1–8

Bornkamm GW, Berens C, Kuklik-Roos C, Bechet JM, Laux G, Bachl J, Korndoerfer M, Schlee M, Hölzel M, Malamoussi A, Chapman RD, Nimmerjahn F, Mautner J, Hillen W, Bujard H, Feuillard J (2005) Stringent doxycycline-dependent control of gene activities using an episomal one-vector system. *Nucleic Acids Res* **33:** e137

Broderick JA, Zamore PD (2011) MicroRNA therapeutics. *Gene Therapy* **18:** 1104–1110

Cambridge SB, Gnad F, Nguyen C, Bermejo JL, Krüger M, Mann M (2011) Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. *J Proteome Res* **10:** 5275–5284

Chatterjee S, Fasler M, Büssing I, Grosshans H (2011) Target-mediated protection of endogenous microRNAs in C. elegans. *Dev Cell* **20:** 388–396

Chatterjee S, Grosshans H (2009) Active turnover modulates mature microRNA activity in Caenorhabditis elegans. *Nature* **461:** 546–549

Chen PS, Su JL, Cha ST, Tarn WY, Wang MY, Hsu HC, Lin MT, Chu CY, Hua KT, Chen CN, Kuo TC, Chang KJ, Hsiao M, Chang YW, Chen JS, Yang PC, Kuo ML (2011) miR-107 promotes tumor progression by targeting the let-7 microRNA in mice and humans. *J Clin Invest* **121:** 3442–3455

Chen PY, Manninga H, Slanchev K, Chien M, Russo JJ, Ju J, Sheridan R, John B, Marks DS, Gaidatzis D, Sander C, Zavolan M, Tuschl T (2005) The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev* **19:** 1288–1293

Das SK, Sokhi UK, Bhutia SK, Azab B, Su ZZ, Sarkar D, Fisher PB (2010) Human polynucleotide phosphorylase selectively and preferentially degrades microRNA-221 in human melanoma cells. *Proc Natl Acad Sci USA* **107:** 11948–11953

De N, Young L, Lau PW, Meisner NC, Morrissey DV, Macrae IJ (2013) Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. *Mol Cell* **50:** 344–355

Djuranovic S, Nahvi A, Green R (2012) miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336:** 237–240

Fabian MR, Sonenberg N, Filipowicz W (2010) Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* **79:** 351–379

Friedel CC, Dölken L, Ruzsics Z, Koszinowski UH, Zimmer R (2009) Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res* **37:** e115

Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* **8:** 69

Gatfield D, Le Martelot G, Vejnar CE, Gerlach D, Schaad O, Fleury-Olela F, Ruskeepää AL, Oresic M, Esau CC, Zdobnov EM, Schibler U (2009) Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes Dev* **23:** 1313–1326

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308:** 833–838

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

Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466:** 835–840

Hargrove JL, Schmidt FH (1989) The role of mRNA and protein stability in gene expression. *FASEB J* **3:** 2360–2370

Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, Yeom KH, Han J, Kim VN (2009) TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138:** 696–708

Hobert O (2008) Gene regulation by transcription factors and microRNAs. *Science (New York, NY)* **319:** 1785–1786

Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* **12:** 99–110

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19:** 489–501

Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, Mandel G, Hannon GJ (2007) A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci USA* **104:** 19291–19296

Khan AA, Betel D, Miller ML, Sander C, Leslie CS, Marks DS (2009) Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* **27:** 549–555

Khanin R, Higham DJ (2009) Mathematical and computational modelling of post-transcriptional gene regulation by microRNAs. In *Handbook of Statistical Systems Biology* chap. 10

Kida A, Kakihana K, Kotani S, Kurosu T, Miura O (2007) Glycogen synthase kinase-3beta and p38 phosphorylate cyclin D2 on Thr280 to trigger its ubiquitin/proteasome-dependent degradation in hematopoietic cells. *Oncogene* **26:** 6630–6640

Kishore S, Gruber AR, Jedlinski DJ, Syed AP, Jorjani H, Zavolan M (2013) PAR-CLIP of snoRNA core proteins and small RNA-seq identify novel human snoRNA loci and give insights into snoRNA processing. *Genome Biol* **14:** R45

Krol J, Busskamp V, Markiewicz I, Stadler MB, Ribi S, Richter J, Duebel J, Bicker S, Fehling HJ, Schübeler D (2010) Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. *Cell* **141:** 618–631

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foà R, Schliwka J, Fuchs U, Novosel A, Müller RU *et al* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129:** 1401–1414

Linsley PPS, Schelter JM, Burchard J, Kibukawa M, Martin MMM, Bartz SRS, Johnson JJM, Cummins JMJ, Raymond CCK, Dai H, Chau N, Cleary M, Jackson AL, Carleton M, Lim L (2007) Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27:** 2240–2252

Liu Q, Fu H, Sun F, Zhang H, Tie Y, Zhu J, Xing R, Sun Z, Zheng X (2008) miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic Acids Res* **36:** 5391–5404

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert B, Mak R, Ferrando A, Downing J, Jacks T, Horvitz H, Golub T (2005) MicroRNA expression profiles classify human cancers. *Nature* **435:** 834–838

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

Pall GS, Hamilton AJ (2008) Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3:** 1077–1084

Pfeffer S, Zavolan M, Grässer FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks DS, Sander C, Tuschl T (2004) Identification of virus-encoded microRNAs. *Science* **304:** 734–736

R Development Core Team (2006) *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing. Vienna, Austria

Rissland OS, Hong Sj, Bartel DP (2011) Article microRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes. *Mol Cell* **43:** 993–1004

Russell A, Thompson MA, Hendley J, Trute L, Armes J, Germain D (1999) Cyclin D1 and D3 associate with the SCF complex and are coordinately elevated in breast cancer. *Oncogene* **18:** 1983–1991

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* **473:** 337–342

Schwanhäusser B, Gossen M, Dittmar G, Selbach M (2009) Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* **9:** 205–209

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

Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H (2007) Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol* **3:** 138

Song JL, Stoeckius M, Maaskola J, Friedländer M, Stepicheva N, Juliano C, Lebedeva S, Thompson W, Rajewsky N, Wessel GM (2011) Select microRNAs are essential for early development in the sea urchin. *Dev Biol* **362:** 104–113

Stalder L, Heusermann W, Sokol L, Trojer D, Wirz J, Hean J, Fritzsche A, Aeschimann F, Pfanzagl V, Basselet P, Weiler J, Hintersteiner M, Morrissey DV, Meisner-Kober NC (2013) The rough endoplasmatic reticulum is a central nucleation site of siRNA-mediated RNA silencing. *EMBO J* **32:** 1115–1127

Strohmaier H, Spruck CH, Kaiser P, Won KA, Sangfelt O, Reed SI (2001) Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. *Nature* **413:** 316–322

Trajkovski M, Hausser J, Soutschek J, Bhat B, Akin A, Zavolan M, Heim MH, Stoffel M (2011) MicroRNAs 103 and 107 regulate insulin sensitivity. *Nature* **474:** 649–653

Wang X, Wang X (2006) Systematic identification of microRNA functions by combining target prediction and expression profiling. *Nucleic Acids Res* **34:** 1646–1652

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

2.2 Extended results

2.2.1 Northern blot analysis of induced miRNA expression

We used the following experimental approach to validate the kinetic models of mRNA and protein level changes in response to miRNA transfections or induction. Two miRNAs which are not usually expressed in HEK cells, hsa-miR-199a-3p and hsa-219-5p, were either transfected or induced in the appropriate cell lines and then the kinetics of miRNA dependent gene regulation (mRNA level changes or luciferase activities were followed). Initially, to pin down *bona fide* targets for these miRNAs, stable cell lines with doxycycline (DOX) inducible miRNA expression were established. Primary hsa-miR-199a-3p, hsa-miR219-5p and hsa-miR-30a were cloned and sub-cloned in an inducible episomal pRTS-1 vector [142]. Stable clones were isolated based on hygromycin selection and tested for miRNA expression with Northern blot upon induction with DOX. MiR-30a clones were used as controls to compare the expression levels from the induction system with the endogenous miRNA levels from non-induced cells. To determine the dose dependency of the expression of inducible pRTS-1, stable cell lines of inducible miRNA-199a-3p and miRNA-219-5p were induced with different concentrations of DOX starting form 100ng/ml to 1µg/ml. 10µg of total RNA was isolated at different time intervals after induction with DOX. Expression of miRNA-199a-3p (Figure 4A&B) and miRNA-219-5p (Figure 4C) was observed to be dose dependent at the early time points. MiRNAs were expressed after 2 hours of induction. No leaky expression was observed in the Northern blots in the absence of DOX. The hsa-miR-30a expression was used to compare the expression of inducible system with the endogenous miRNA levels as shown in Figure 4D.

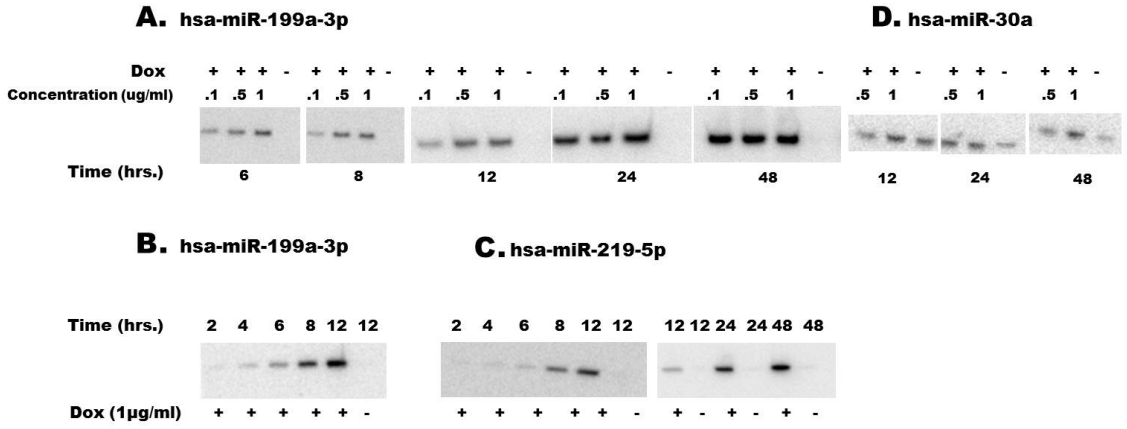


Figure 4: Northern blot analysis of miR-199a-3p and miR-219-5p. These miRNAs were cloned in pRTS-1 vectors and expressed upon induction with DOX. Expression from the pRTS-1 vector was induced with different concentrations of doxycycline and the expression of miRNAs was analyzed at different time intervals by Northern blot. 10µg of total RNA was isolated from stable HEK293T cells with inducible

expression of miRNAs for Northern blots. (A). Expression of miRNA-199a-3p was titrated with concentrations of doxycycline ranging from 100ng/ml to 1µg/ml. The data shows that expression is dose dependent at the early time points, eventually saturating after 48hrs of induction. Expression is so tightly regulated that in the absence of DOX no leaky expression is observed. (B) Expression of miR-199a-3p is visible on the blots 2hrs post induction with defined concentration of 1µg/ml. (C) Northern blot of miR-219-5p expression in presence and absence of DOX at different time intervals starting from 2 hrs. (D) In the cell line with inducible hsa-miR-30a expression, the expression upon induction with 1µg/ml of DOX was 1.5 fold higher compared to the expression from the endogenous locus, in absence of DOX induction. This indicates that the system was able to express miRNAs comparable to the endogenous levels, confirming the robustness of the system.

2.2.3 Validation of targets by luciferase assays

Targets for hsa-miR199a-3p and hsa-miR-219-5p were predicted using ElMMo [126]. We selected from predicted targets some that had one binding site in the 3'UTR with high probability of being under selective pressure (p>0.7) and with a length shorter than 1500 nts enabling efficient cloning. To validate the computationally predicted targets of miR-199a-3p and miR-219-5p, we performed luciferase assays. The 3'UTR of predicted target transcripts were cloned in the multiple cloning site of a dual luciferase reporter psiCHECK-2 vector. After transient transfections of these reporter vectors into the stable cell lines containing miR-219-5p or miR-199a-3p, the expression of these miRNAs was induced with 1µg/ml of doxycycline. 48 hours after induction, we performed the luciferase assay. None of the predicted targets of miR-219-5p showed a substantial downregulation upon miRNA induction (Figure 5B), whereas upon the induction of miR-199a-3p, some of its targets underwent 30-50% downregulation (Figure 5A). One target KTN1 for hsa-miR199a-3p was significantly downregulated and chosen for subsequent experiments. The list of primers used to clone pri-miRNAs in pRTS-1 vector, the 3'UTRs in psiCHECK-2 vector and the probes used for hybridization in miRNA Northerns are listed in Tables 1 to 4.

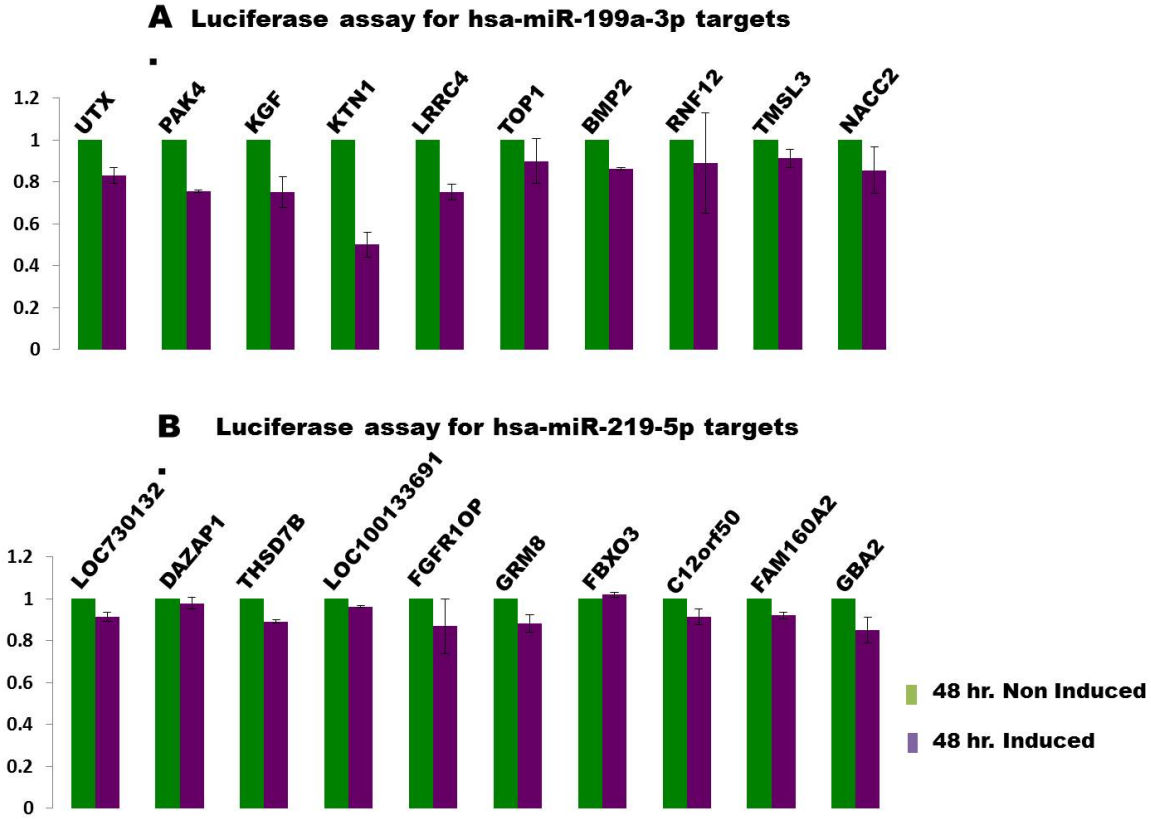


Figure 5: Validation of the predicted targets with luciferase assays. 3'UTRs with single binding sites for miRNAs were cloned downstream of Renilla luciferase. These vectors were transfected in the cell lines with inducible miRNA expression. After a 48-hour induction of miRNA expression, a luciferase assay was performed to measure the Renilla luciferase levels. Changes in the Renilla luciferase levels for hsa-miR199a-3p (A) and hsa-miR-219-5p (B) targets are shown here. The error bars represent ±SEM values and n=3.

2.2.4 Establishment of a cell line expressing a miRNA target as well as inducibly expressing the cognate miRNA

To establish a stable cell line constitutively expressing Renilla luciferase with KTN1 3'UTR and its cognate miRNA (referred hereafter as KTN1 cell line), I co-transfected the psiCHECK2-KTN1 vector and a mammalian selection vector carrying a puromycin gene pPUR (clontech) construct into the hsa-miR199a-3p stable cell line in a 3:1 ratio. Thus, isolated stable colonies were expanded in presence of 0.75µg/ml of puromycin (Sigma-Aldrich). A sketch of the experiment for constructing this cell line shown in Figure 7. The integration of psiCHECK2-KTN1 vector in the host genome was confirmed by PCR as shown in Figure 8. This cell line was further induced with DOX and the Renilla luciferase activity was measured every 24 hrs. for 8 days. The hsa-miR-199a-3p expression for 8 days was confirmed with Northern blot as shown in Figures 6A and B**.** The mRNA-level expression of Renilla luciferase upon DOX induction of miRNA expression was measured with quantitative real-time PCR (qRT-PCR) (Figure 6C) and the luciferase activity was assayed

for 8 days (Figure 6D). The expression of Renilla luciferase with KTN1 '3UTR was consistently downregulated for 8 days at both mRNA and protein levels as shown in Figures 6C and D.

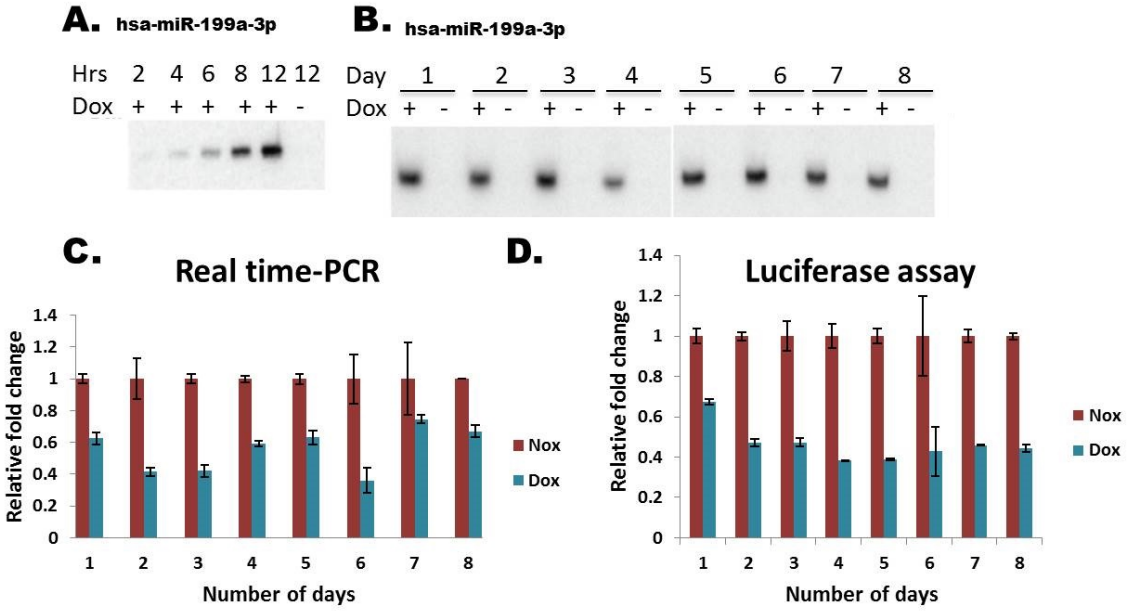


Figure 6: Characterization of the KTN1 cell line. (A) The expression of hsa-miR-199a-3p was detected already at 2 hrs. after DOX induction. Furthermore the expression was consistently detected until 8 days with Northern blot. The miRNA expression saturated after 24 hrs. Upon induction of miRNA expression, Renilla-KTN1 levels were measured at mRNA level by qRT-PCR (C) and protein level by luciferase assay for 8 days (D). The errors bars in (C & D) represent ± standard deviation of three technical replicates, n=3.

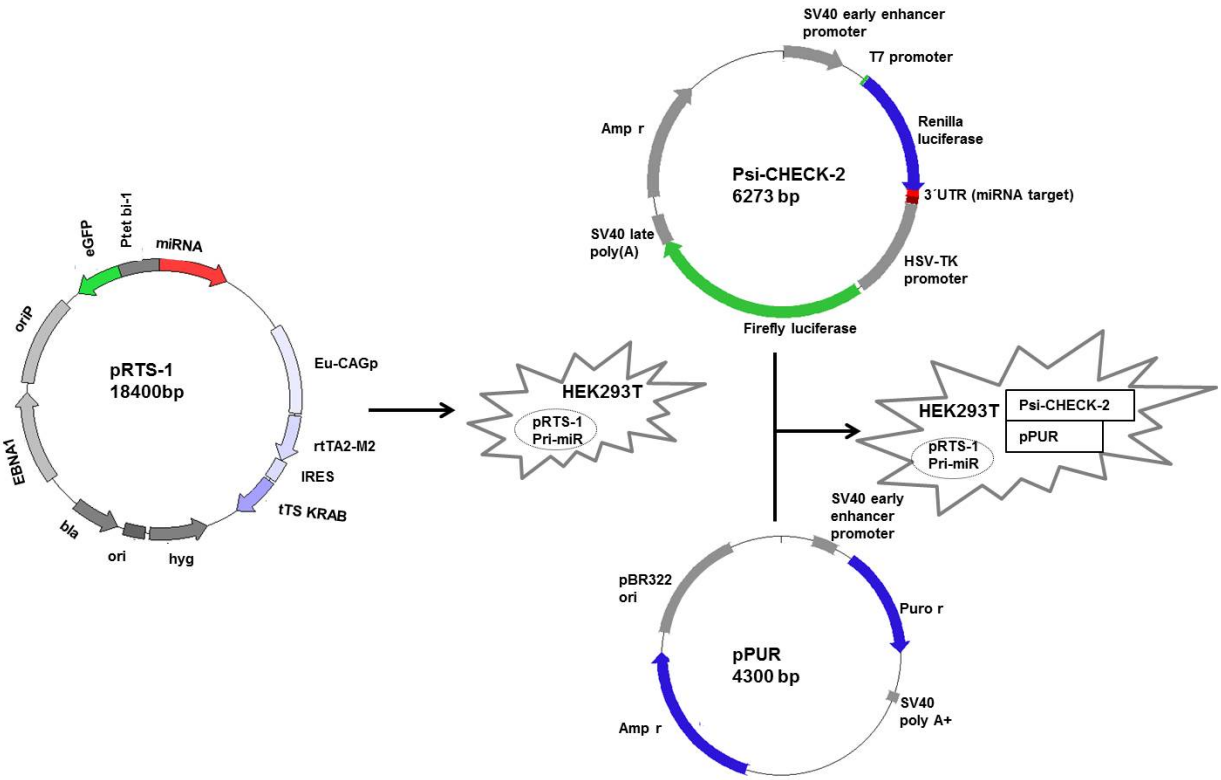


Figure7: Schematic representation of the experiment for the establishment of the KTN1 cell line. The 3' UTR of the kinectin 1 (KTN1) gene and the primary miR-199a-3p were PCR amplified and individually cloned into the multiple cloning site of a pGEM-T Easy vector. The 3' UTR was subcloned to a reporter psiCHECK-2 vector, whereas the primary miR-199a-3p was subcloned to an inducible episomal pRTS-1 vector through defined restriction sites. (A) First we constructed a stable cell line with inducible miRNA expression from a pRTS-1 vector. (B) Into the stable cell line with inducible miRNA expression, the reporter and pPUR vectors were co-transfected for stable integration.

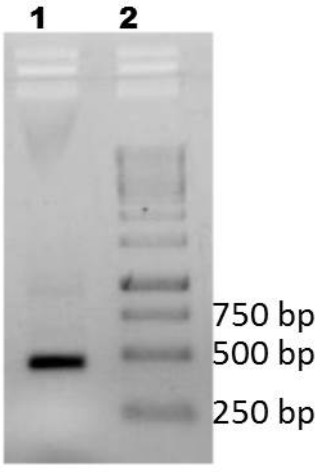


Figure 8: PCR confirming the genomic integration of KTN1 3' UTR-psiCHECK2 vector in the KTN1 cell line. Primers were designed to specifically amplify the 3'UTR of Renilla luciferase (KTN1 3'UTR) from the KTN1 cell line. Lane 1 shows the genomic PCR product of KTN1 3'UTR from the integrated psiCHECK-2 vector, lane 2 shows the standard DNA marker with the sizes of fragments labeled. Primers are listed in Table 4.

Table 1. List of targets of hsa-miR-199a-3p and their primers used to clone in the multiple cloning site of psiCHECK-2 vector.

hsa-miR-199a-3p		
Gene symbol	Forward primer (5'-3')	Reverse primer (5'-3')
UTX	GGGGCTCGAG TCCATGGACATTAAATGAGACCT	ATTTGCGGCCGC GGCAGCAATTGTTTATTGGTC
PAK4	CCCAGCGATCGC CCTTCCCCTCAACCAAAGAG	AAAAGCGGCCGC AACCTGGTTCTTCAGGCAGT
CITED2	GGCC CTCGAG CCGGCGAAAGAAATCAAAC	AATTGCGGCCGC TTGACATTGGCAAATTAAAATAGAA
LOC100133691	GCCCCTCGAG AGAACCAGTTCCAGTAGGGAGA	ATTTGCGGCCGC GCCATTATTATTATTATAGCAGCCATT
KTN1	GGGGCTCGAG TGGGAAACTGTTCATTTGAGG	TATTGCGGCCGC TTGCTGACGCCATTACAAAA
LRRC4C	GGGGCTCGAG ATTAAAAATGACACAAATGACTGG	ATATGCGGCCGC TCTGCTTTAGATCACAATAGAATTTTT
TOP1	GCCC CTCGAG CCAGTCTCAAGAGGCAGAGTT	TATAGCGGCCGC ACAGTTGATTAAAAGGGAATTCAT
BMP2	GGGG CTCGAG AAAAACCCCACCCCAGTTG	AAAAGCGGCCGC AGAATGAAATGCCAGAGATTTG
RNF12	CCCGCTCGAG AGCTGATATAGTGATGGGCAAA	TTATGCGGCCGC CATTTTTGAGGAGCAGACTTG
TMSL3	GCCG CTCGAG GCCGCCAATATGCACTGTA	TAAAGCGGCCGC AAAATTGTGCATTCCAATAATTAAA
NACC2	GGCC CTCGAG CCGAGTACTAGAGCTGCTTGC	TTATGCGGCCGC CATCAATTTTAAATACAAGCAGAATAA

Table 2. List of targets of hsa-miR-219-5p and their primers used to clone in the multiple cloning site of psiCHECK-2 vector.

hsa-miR-219-5p		
Gene symbol	Forward primer (5'-3')	Reverse primer (5'-3')
LOC730132	ATTTGCGATCGC TTGACTTTCCGTCTACCTTTTT	GTTAGCGGCCGC AAGGCAGAGTGAGGTGATGG
OTX2	CCCCCTCGAG AGAACCTCTTTTTGTGGGTGA	TTTTGCGGCCGC CTGCCAAATCCAGGAAGAAA
DAZAP1	GGGCCTCGAG CAGGTTACGGGCAGGACTT	AAATGCGGCCGC CCAGGGTCAAACTGGTCAAT
THSD7B	GGGGCTCGAG TCAACTGCCTTAACCGCTTT	TATTGCGGCCGC GCAGTCTCTTGAGCAGTTTATTTG
LOC100133691	GGGGCTCGAG AAGAACCAGTTCCAGTAGGGAGA	ATTTGCGGCCGC GCCATTATTATTATTATAGCAGCCATT
FGFR1OP	GGGGCTCGAG CACGAAGAAGGAAGTATTCTAATTAAC	TTTTGCGGCCGC CCTTAATGGTCTAACAAACCTTCC
GRM8	GGGGCTCGAG AGATCTTCCTCTACCAAGACAACA	ATAAGCGGCCGC AACATTTGCAAAACATTCCTTT
FBXO3	CGCCCTCGAG AAGCCTTCTGCTGATGGAAG	AATTGCGGCCGC CCAAAATTTGCTTTTAAAATGTG
C12orf50	CCCCCTCGAG ACCTGGAGGTTGAGAGAGAGAA	TTTAGCGGCCGC GTTTGGAGTAATGCACGGAAT
FAM160A2	GGGGCTCGAG TTTCTTCCATGGACAATCAGG	TATAGCGGCCGC ACATATTGCAACAAGTTCTCCA
GBA2	GGCGCTCGAG TCTGAACTGTGGGAGGGAAG	ATTAGCGGCCGC ACCCATGTGTCTGTCACACC

Table 3. Primers used to PCR-amplify primary miRNA sequences from HEK293 cells. The PCR products were sub-cloned into pRTS-1 vector.

Forward primer (5-3')	Reverse primer (5-3')
hsa-miR-199a-3p	
AAAAGGCCTCACTGGCCC CCCTTTTCCTGGTCCTAAA	AAAAGGCCTCACTGGCC CCTTCGGCAGTCTTTTCTCA
hsa-miR-219-5p	
AAAAGGCCTCACTGGCC CCACTTCCCACTCCAGACAT	AAAAGGCCTCACTGGCC AACACACCTAGGTCCCGAG
hsa-miR-30a	
AAAAGGCCTCACTGGCC GAAACTAGAAGCTCGGTGATGAA	AAAAGGCCTCACTGGCC ACAGAGCACCTCCTCAATGC

Table 4. List of DNA probes to hybridize the miRNAs on Northern blots and the genomic PCR primers for KTN1 cell line to confirm the integration the psiCHECK-2 in the miR-199a-3p cell line.

Probes	
hsa-miR-199a-3p	TAACCAATGTGCAGACTACTGT
hsa-miR-219-5p	AGAATTGCGTTTGGACAATCA
hsa-miR-30a	CTTCCAGTCGAGGATGTTTACA
KTN1 cell line primers	
Forward- GAGCGCGTGCTGAAGAACGAG	Reverse- TTGCTGACGCCATTACAAAA

3. Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation

Most research on miRNAs is confined to miRNA binding sites located in the 3'UTRs of the mRNAs. Recent studies, particularly with CLIP, revealed many miRNA target sites in the coding region (CDS), but the function of these sites in modulating gene expression remains largely unexplored. Some experimental studies showed that miRNAs do target the CDS regions and function in the modulation of gene expression, even in the absence of seed complementarity. These studies proposed that repression on these sites instead depends on the complementarity of the region downstream of the seed [121, 143]. In another study it was shown that miR-519 represses the translation of RBP HuR which in turn reduces the levels of downstream targets of HuR involved in cell division [144]. There are also reports showing that the plant miRNAs are able to induce translational repression via imperfect complementarity interactions with target sites in CDS and 3'UTRs [145, 146]. Similarly, evidence showing that functional miRNA target sites are located in CDS regions of mammalian transcripts has been provided [147-149], with miRNAs affecting both mRNA and protein levels. Reanalysis of the protein and mRNA level data generated by Baek et al. [132] found that there is strong effect in the regulation of the genes with target sites present in both CDS and 3'UTR compared to genes that only have miRNA binding sites in 3'UTR [150].

Application of high throughput approaches for isolating Argonaute- bound target sites indicates [138, 140] that CDS sites are as numerous as those located in 3' UTRs, although the density of Argonaute-bound sites is higher in 3' UTRs as compared to CDS. In the following study we analyzed putative miRNA target sites that are located in coding regions, with the aim of uncovering the function of such sites by comparison with sites that are located in 3' UTRs. The target sites that we used in our analysis were either predicted computationally or inferred on the basis of transcript- or protein-level changes following miRNA transfections. Our results indicate that CDS and 3' UTR target sites co-evolved, have similar sequences and structure properties, and may have similar efficiency in inducing translational repression of the transcripts in which they reside. On the other hand, sites located in 3'UTR are much more effective at inducing mRNA degradation, as reported [132, 140, 149, 150]. Analyzing recently published data of ribosome-protected fragment profiles upon miRNA transfection [151] from the perspective of the location of miRNA-complementary sites, we additionally found that sites located in the CDS are most

potent in inhibiting translation, while sites located in the 3' UTR are more efficient at triggering mRNA degradation. Our study suggests that miRNAs may combine targeting of CDS and 3' UTR to flexibly tune the time scale and magnitude of their post-transcriptional regulatory effects.

3.1 Statement of my work

My contribution to this study was to validate the observation that CDS-located miRNA binding sites were efficient in inhibiting translation. For this, I have established a stable inducible hsa-miR124-3p HEK cell line using an episomal vector pRTS-1 and I have cloned CDS regions with at least one miRNA binding site just upstream of the stop codon of Renilla luciferase in psiCHECK-2 vectors, by the method of overlap extension PCR. Psi-CHECK-2 vectors were transiently transfected in inducible miR-124 cell line. At certain time intervals, upon DOX induction of miRNA expression, changes in the mRNA and protein levels were measured by qRT-PCR and luciferase assays.

Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation

Jean Hausser,[1] Afzal Pasha Syed, Biter Bilen, and Mihaela Zavolan[1]

Biozentrum, University of Basel and Swiss Institute of Bioinformatics, 4056 Basel, Switzerland

Most of what is presently known about how miRNAs regulate gene expression comes from studies that characterized the regulatory effect of miRNA binding sites located in the 3′ untranslated regions (UTR) of mRNAs. In recent years, there has been increasing evidence that miRNAs also bind in the coding region (CDS), but the implication of these interactions remains obscure because they have a smaller impact on mRNA stability compared with miRNA-target interactions that involve 3′ UTRs. Here we show that miRNA-complementary sites that are located in both CDS and 3′-UTRs are under selection pressure and share the same sequence and structure properties. Analyzing recently published data of ribosome-protected fragment profiles upon miRNA transfection from the perspective of the location of miRNA-complementary sites, we find that sites located in the CDS are most potent in inhibiting translation, while sites located in the 3′ UTR are more efficient at triggering mRNA degradation. Our study suggests that miRNAs may combine targeting of CDS and 3′ UTR to flexibly tune the time scale and magnitude of their post-transcriptional regulatory effects.

[Supplemental material is available for this article.]

MicroRNAs (miRNAs) are ~21 nt (nucleotide)-long regulatory RNAs that are encoded in the genomes of species ranging from viruses to human. They form miRNA-induced silencing complexes (miRISCs) with Argonaute proteins which they guide, through hybridization, to target mRNAs whose expression is subsequently down-regulated (Bushati and Cohen 2007; Bartel 2009). In plants, miRNAs typically trigger the endonucleolytic cleavage of their targets through perfect or near-perfect complementarity interactions with transcript coding regions (CDS) (Jones-Rhoades et al. 2006). In contrast, in mammals they have been shown to interact predominantly through their "seed region" (nucleotides 2–8 from the 5′ end of the miRNA) with 3′ untranslated regions (3′ UTRs) of mRNAs (Lewis et al. 2005), inducing their destabilization and translational inhibition (Filipowicz et al. 2008). In recent years, the distinction between the mode of action of plant and animal miRNAs has become less clear. There is growing evidence that plant miRNAs can induce translational repression via imperfect complementarity interactions with target sites in CDS and 3′ UTRs (Brodersen et al. 2008; Lanet et al. 2009). Likewise, increasingly many miRNA target sites are discovered in coding regions of mammalian transcripts (Forman et al. 2008; Huang et al. 2010; Qin et al. 2010; Ott et al. 2011). Application of high-throughput approaches for isolating Argonaute-bound target sites indicates that CDS sites are as numerous as those located in 3′ UTRs (Chi et al. 2009; Hafner et al. 2010), though the density of Argonaute-bound sites is higher in 3′ UTR compared with CDS (Hafner et al. 2010).

If CDS sites are as common as Argonaute cross-linking and immunoprecipitation (CLIP) studies indicate (Chi et al. 2009; Hafner et al. 2010), one wonders why there are relatively few reports on their involvement in gene regulation and why the studies that have been so far published suggest that CDS sites are much less effective in down-regulating mRNA levels upon miRNA transfection (Baek et al. 2008; Hafner et al. 2010; Fang and Rajewsky 2011; Schnall-Levin et al. 2011). A reason may be that CDS sites function in specific contexts, in which coding regions are accessible to the miRNA-loaded silencing complex, while under normal conditions the process of translation hinders miRNA binding to these sites (Bartel 2009; Gu et al. 2009). Alternatively, it may be that only a specific subset of miRNAs targets coding regions. For instance, it has recently been found that hsa-miR-181a targets multiple members of the C2H2 zinc finger domain family, through multiple CDS sites that occur precisely in the regions of the transcripts that encode the C2H2 domains (Schnall-Levin et al. 2011). However, the CDS sites that have been isolated in CLIP experiments do not seem to correspond to a restricted subset of miRNAs, and they also did not require that the cells were treated in some specific way to expose the use of CDS sites. Thus, none of the hypotheses mentioned above can explain the discrepancy between the apparent abundance of CDS sites and the paucity of reports about their function.

Evolutionary conservation has been successfully employed to predict regulatory elements, including binding sites for miRNAs in 3′ UTRs (Krek et al. 2005; Lewis et al. 2005; Gaidatzis et al. 2007). Hurst (2006) used a limited experimentally verified data set that was available at the time to demonstrate that miRNA target sites located in coding regions exhibit significantly low evolutionary rates in mammals. Evolutionary conservation-based approaches to predict miRNA target sites in coding regions followed. Forman et al. (2008) used alignments of CDS regions in 17 species to identify conserved miRNA-complementary sites in 700 human genes. Among the miRNAs with the most predicted CDS sites were hsa-let-7a-5p, hsa-miR-9-5p, hsa-miR-125a-5p, and hsa-miR-153. These authors further demonstrated experimentally that hsa-let-7b-5p down-regulates the miRNA-processing enzyme Dicer, whose transcript carries multiple complementarities to let-7 in its coding region. Schnall-Levin et al. (2010) allowed for the possibility that sites are not perfectly conserved among "all" genomes used in the inference to show that miRNA targeting in CDS is as common as in the 3′ UTRs in *Drosophila* species. They further predicted 26,000 and 14,000 sites in human 3′ UTR and CDS regions, respectively. Finally, Fang and Rajewsky (2011) found evidence that mRNAs that are simultaneously targeted in the CDS and in the 3′ UTR are slightly

[1]Corresponding authors
E-mail jean.hausser@unibas.ch
E-mail mihaela.zavolan@unibas.ch
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.139758.112. Freely available online through the *Genome Research* Open Access option.

23:604–615 © 2013, Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/13; www.genome.org

more destabilized than mRNA targeted only in the 3′ UTR, while Schnall-Levin et al. (2011) showed that mRNAs with several miRNA binding sites in the CDS can effectively be degraded.

In the present study we analyzed putative miRNA target sites that are located in coding regions, with the aim of uncovering the function of such sites by comparison with sites that are located in 3′ UTRs. The target sites that we used in our analysis were either predicted computationally or inferred on the basis of transcript- or protein-level changes following miRNA transfections. Our results indicate that CDS and 3′ UTR target sites co-evolved, have similar sequence and structure properties, and may have similar efficiency in inducing translational repression of the transcripts in which they reside. On the other hand, sites located in 3′ UTR are much more effective at inducing mRNA degradation, as reported before (Baek et al. 2008; Guo et al. 2010; Hafner et al. 2010; Fang and Rajewsky 2011).

Results

Both CDS and 3′ UTR sites are under evolutionary selection

The application of the ElMMo model to miRNA target prediction in 3′ UTRs was described before (Gaidatzis et al. 2007) and recent assessments indicate that ElMMo is among the most accurate miRNA target prediction methods available (Alexiou et al. 2009; Hafner et al. 2010; Sturm et al. 2010). ElMMo is parameter-free and, by building its background model from the appropriate type of regions, it can in principle predict miRNA target sites in transcript or genomic regions that do not correspond to 3′ UTRs. Here we applied ElMMo to predict miRNA target sites in coding regions. To determine whether the ElMMo-predicted CDS sites are functional we analyzed eight published data sets (listed in Table 1) that compared mRNA expression before and after transfections of individual miRNAs. In each of these experiments, we identified all mRNAs that carried precisely one 2–8 seed match to the transfected miRNA, and we selected 250 transcripts that were most down-regulated and 250 that responded least to transfection. We separated each of these sets in two subsets: transcripts in which the site was located in the CDS, and transcripts in which the site was located in the 3′ UTR. We then computed the *t*-values comparing the ElMMo scores of the sites that induced an effect with those that did not induce an effect and we computed the overall *t*-value over the data sets. As Figure 1A (and Supplemental Fig. 1) shows, both CDS and 3′ UTR sites that are located in down-regulated transcripts have a higher probability to be under selection compared with sites that are located in transcripts that do not respond to transfection. We further asked whether miRNA-complementary sites that are isolated in Argonaute CLIP experiments also have higher ElMMo scores compared with those that are not isolated (Fig. 1A). We used the Argonaute CLIP data of Kishore et al. (2011) (Table 1) and identified all matches to the 2–8 seed of the top 10 expressed miRNA families in transcripts from which CLIP reads were obtained. In principle, these transcripts were expressed and bound by Argonaute proteins, so the transfected miRNAs should have had access to all seed-matching sites in these transcripts. We separated the set of seed matches into those that were the most enriched in CLIP and those that were least enriched in CLIP and compared the distributions of ElMMo scores of the two subsets of sites. As shown in Figure 1A (and Supplemental Fig. 1), we found that CLIPed CDS sites had a significantly higher score compared with non-CLIPed sites. These results indicate that ElMMo predicts functional sites not only in 3′ UTRs but also in CDSs.

Table 1. The data sets that we used to identify properties of functional miRNA binding sites in the 3′ UTR and in the CDS

Data type	Source
mRNA expression profiling after miRNA transfection	Grimson et al. (2007), Karginov et al. (2007), Linsley et al. (2007), Baek et al. (2008), Selbach et al. (2008), Hausser et al. (2009), Guo et al. (2010)
EIF2C2 binding data from CLIP	Kishore et al. (2011)
Evolutionary selective pressure inferred with ElMMo	Gaidatzis et al. (2007)
Protein expression profiling upon miRNA transfection	Baek et al. (2008), Selbach et al. (2008)

ElMMo further estimates a miRNA-specific probability that a site that is complementary to the miRNA in the reference species (in this case human) is under selection in at least one other species taken into account in the miRNA target prediction. Here we used genome sequence data of the species *Pan troglodytes*, *Rhesus maccacus*, *Canis familiaris*, *Rattus norvegicus*, *Mus musculus*, *Bos taurus*, *Monodelphis domestica*, and *Gallus gallus* to predict sites that are under evolutionary selection. This probability is high when the miRNA-complementary motifs are strongly conserved and low when they are weakly conserved across species, relative to motifs that are not complementary to miRNAs. As shown in Figure 1B, the probabilities of miRNA-complementary sites to be under selection pressure are strongly correlated between CDS- and 3′ UTR-located sites of individual miRNAs ($r = 0.77$). For comparison, we estimated the average phastCons conservation score (Siepel et al. 2005) for CDS and 3′ UTR occurrences of individual motifs, and found a similar but weaker correlation (Supplemental Fig. 8). Further supporting the functionality of miRNA-complementary sites located in both CDS and 3′ UTR, the phastCons scores of these motifs are significantly higher compared with those of other 7mers in both types of regions (Supplemental Fig. 8). These results suggest that CDS and 3′ UTR sites evolve in parallel and that CDS sites must have a function that confers a selective advantage in evolution.

miRNAs that function in embryonic development co-target the 3′ UTR and the CDS

The proportion of miRNA-complementary sites estimated to be under evolutionary selection in CDS as opposed to 3′ UTRs differs between miRNAs (Fig. 2A). We found that this is due to a combination of factors. First, the sequence composition of CDS and 3′ UTR differ, with the result that some motifs are more abundant in CDS and others in 3′ UTR (Supplemental Fig. 9). Additionally, the relative lengths of CDS and 3′ UTR of the targeted transcripts and the relative selection pressure on miRNA-complementary motifs in the two types of regions (Supplemental Fig. 10) also contribute to preferential targeting of CDS vs. 3′ UTR. Interestingly, miRNAs that contain the AGCAGC motif at the 5′ end (hsa-miR-16-5p, hsa-miR-15a/b-5p, hsa-miR-195-5p, hsa-miR-103-3p, hsa-miR-107, hsa-miR-646, hsa-miR-424-5p, hsa-miR-497-5p) and that have been previously shown to regulate cell cycle (Linsley et al. 2007; Fulci et al. 2009; Forrest et al. 2010; Balatti et al. 2011) have many more complementary sites in the CDS compared with 3′ UTR (Fig. 2A). On the other hand, motifs that are complementary to miRNAs with embryonic expression—hsa-miR-302a-3p, hsa-miR-369-3p, hsa-miR-372, hsa-miR-373-3p, hsa-miR-374b/c-5p (Suh et al. 2004), the homologs of which have been shown to be involved in maternal

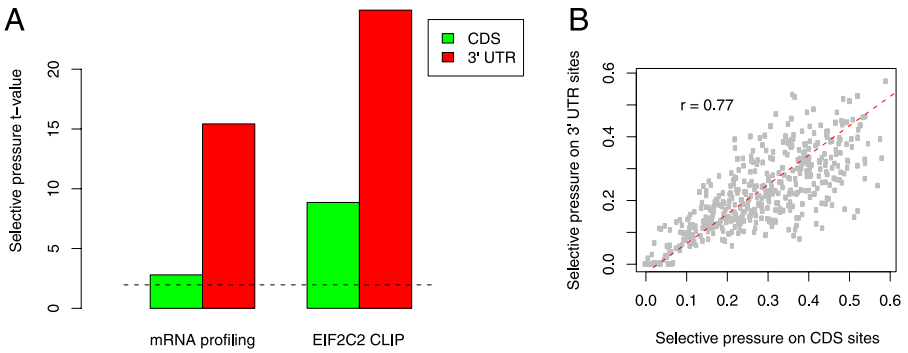


Figure 1. (*A*) CDS- and 3′ UTR-located miRNA binding sites that induce mRNA degradation or are isolated in EIF2C2 CLIP experiments are under stronger selection pressure compared with miRNA-complementary sites that were not functional in these experiments; 2274 CDS- and 3956 3′ UTR-located sites that induce mRNA degradation are compared with 3513 CDS- and 3268 3′ UTR-located sites that do not induce mRNA degradation; 751 CDS- and 786 3′ UTR-located sites isolated in EIF2C2 CLIP experiments are compared with 1059 CDS- and 956 3′ UTR-located sites that do not induce EIF2C2 binding. (*B*) Scatter plot of the inferred probabilities that CDS- and 3′ UTR-located sites complementary to individual miRNAs are under selection. (Dashed red line) First principal component of the scatter.

mRNA clearance in zebrafish (Giraldez et al. 2006)—or to the oncogenic miRNAs of the miR-17 family—hsa-miR-17-5p, hsa-miR-20a/b-5p, hsa-miR-93-5p, hsa-miR-106a/b-5p (He et al. 2005)—are more frequent in the 3′ UTRs (see also Supplemental Table 1).

Given that a large fraction of miRNAs appear to prefer targeting of either the CDS or the 3′ UTR, we asked whether some miRNAs tend to simultaneously target the 3′ UTR and the CDS of the same transcript. To answer this question, we determined the number of genes that contained both one of the 250 highest-scoring CDS binding sites and one of the 250 highest-scoring 3′ UTR sites for a given miRNA, as predicted by the ElMMo algorithm. We then determined the fold enrichments in the number of such co-targeted genes relative to what would be expected if miRNAs targeted the CDS and the 3′ UTR independently (see section "Enrichments in the number of genes co-targeted in the CDS and in the 3′ UTR" in Methods). We found significant evidence of CDS and 3′ UTR co-targeting for 62 human miRNAs ($P < 0.05$ at Fisher's exact test after Bonferroni correction; see Supplemental Table 2). Among these were many miRNAs with important functions: the epithelial cell-specific (Gregory et al. 2008) hsa-miR-141-3p/200a-3p ($P < 10^{-11}$), the embryonic stem cell-expressed (Suh et al. 2004; Stadler et al. 2010) hsa-miR-302/372/373-3p/520a-3p ($P < 10^{-8}$), the oncogenic component of the mir-17/92 cluster hsa-miR-19a-3p ($P < 10^{-6}$) (Ventura et al. 2008; Olive et al. 2009), the oncogenic hsa-miR-130b-3p/hsa-miR-301a-3p miRNAs ($P < 10^{-6}$) (Shi et al. 2011), and the hsa-miR-137 ($P < 10^{-6}$), that is involved in neural maturation (Smrt et al. 2010; Sun et al. 2011).

The tendency for these miRNAs to target both the CDS and the 3′ UTR of individual mRNAs cannot be explained by a general tendency of motif co-occurrence in the CDS and 3′ UTR of the same transcripts because co-targeting enrichments beyond threefold are very rarely observed for random motifs (Supplemental Fig. 11). Interestingly, the miRNAs for which we find evidence for CDS and 3′ UTR co-targeting are among those that preferentially target 3′ UTRs (*P*-value of the average CDS vs. 3′ UTR preference for these miRNAs is significantly >0, $P < 10^{-15}$; see "Statistical evidence that CDS and 3′ UTR co-targeting occurs mostly for miRNAs that preferentially target 3′ UTRs" in Methods). This is illustrated in Figure 2B. Each dot in the figure represents a miRNA. The *x*-axis shows the preference of the miRNA for CDS vs. 3′ UTR targeting defined as the signed distance to the line representing the scaling between the number of CDS and 3′ UTR sites (Fig. 2A) for the respective miRNA. The *y*-axis shows the co-targeting enrichment. miRNAs that target the CDS and 3′ UTR of the same transcripts significantly more often than expected by chance appear in red. Interestingly, overexpression of most of these CDS/3′ UTR–co-targeting miRNAs has been linked with tumorigenesis (Voorhoeve

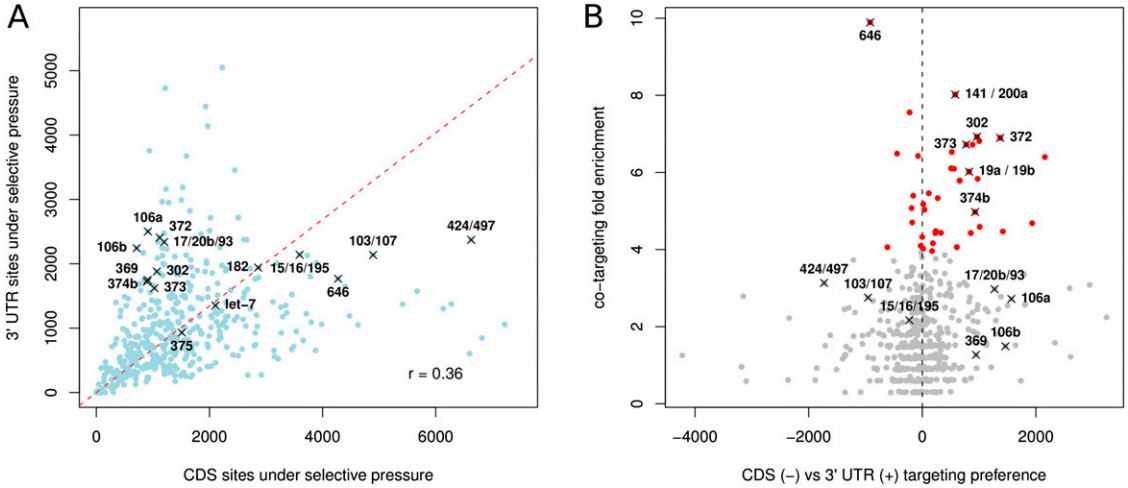


Figure 2. Individual miRNAs differ in their preference for targeting the CDS and the 3′ UTR as well as in their tendency to simultaneously target the CDS and the 3′ UTR of individual genes. (*A*) Number of complementary motifs, weighted by their respective ElMMo posteriors, of individual miRNAs in CDS vs. 3′ UTRs. (Red dashed line) Scaling between the number of CDS and 3′ UTR sites, defined as the line that goes through the origin and maximizes the projected variance. (*B*) Scatter plot of the CDS vs. 3′ UTR targeting preference of individual miRNAs against the fold enrichment in transcripts that are targeted in both the CDS and 3′ UTR relative to what would be expected if the sites were independently distributed. (Red dots) miRNAs with statistically significant co-targeting enrichment ($P < 0.05$ in Fisher's test after Bonferroni correction).

et al. 2006; Huang et al. 2008; Olive et al. 2009; Mateescu et al. 2011; Shi et al. 2011).

Functional CDS and 3′ UTR binding sites have similar sequence and structure properties

Previously, we found that miRNA binding sites that are located in 3′ UTRs and are effective in mRNA degradation have specific properties such as structural accessibility and a U-rich sequence context (Hausser et al. 2009). We further found that highly conserved miRNA target sites share these properties, suggesting that miRNA target sites in 3′ UTRs have been selected in evolution based on their ability to induce mRNA degradation. We therefore asked whether target sites that are located in the CDS and are functional according to the criteria we defined for 3′ UTR sites also share these properties. As before (Hausser et al. 2009), we took a systematic approach, making use of microarray, CLIP, pSILAC, and SILAC measurements as well as the predicted target sites that we obtained above, with their associated posterior probabilities (Table 1). Functional target sites were defined as those that had a high posterior probability of being under evolutionary selection or those whose associated transcripts or proteins responded in miRNA transfection experiments (see Methods). Nonfunctional target sites were defined as those that had a low posterior probability of being under evolutionary selection or whose associated transcripts or proteins did not respond in miRNA transfection experiments. For each target site, we computed 32 properties that quantify the structure and sequence context of miRNA binding sites (listed in section 1.2 of the Supplemental Material). We then calculated the *t*-value quantifying the difference between the mean values taken by the properties among the functional and nonfunctional target sites. A summary of the obtained *t*-values is shown in Figure 3 and indicates that CDS- and 3′ UTR-located miRNA binding sites have largely similar properties (see also Supplemental Figs. 1–7, 12). In particular, U-rich and structurally accessible environments characterize functional miRNA target sites, irrespective of their location in CDS or 3′ UTRs. The correlation is considerably smaller for sites obtained based on measurement of protein expression changes for reasons that are presently unknown (Hausser et al. 2009). These results again suggest that both types of sites experience similar selective constraints and should be equally functional.

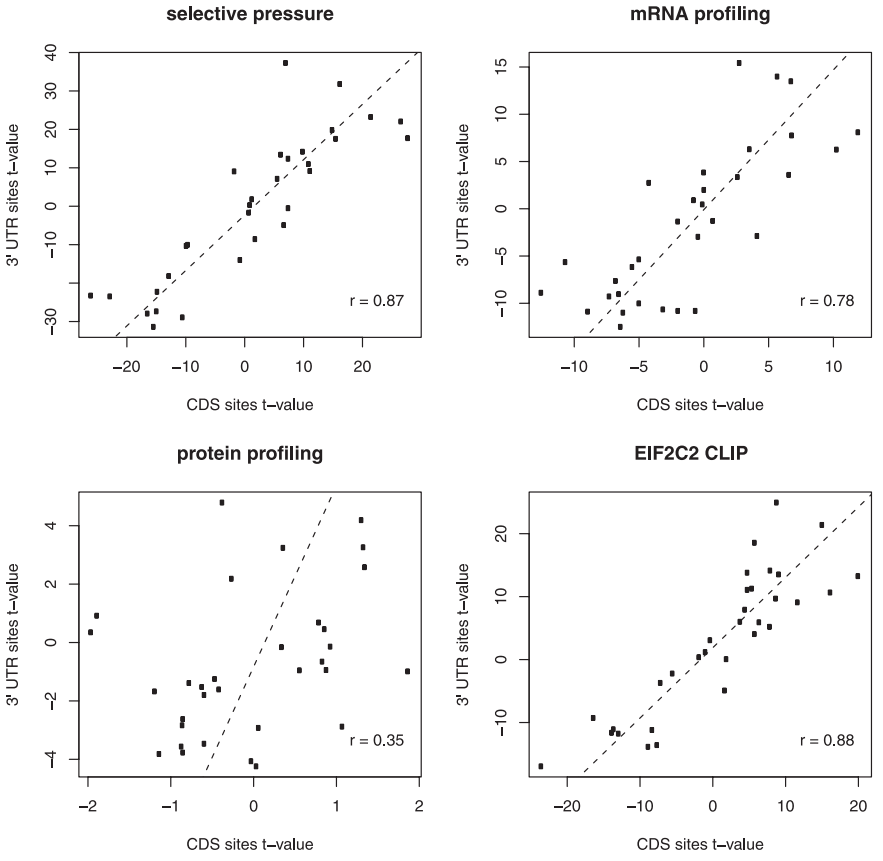


Figure 3. CDS and 3′ UTR sites share common sequence and structure properties. Sets of functional and nonfunctional binding sites were defined according to four different criteria—selective pressure, efficacy in mRNA degradation, efficacy in reducing protein levels, and binding to EIF2C2—each corresponding to a different panel. We then compared *t*-values obtained in comparing functional and nonfunctional sites from CDS (*x*-axis) and 3′ UTR (*y*-axis) regions. Each property is represented in each plot as a dot. Positive and negative values denote positive and negative predictors of functional miRNA binding sites, respectively.

mRNA destabilization occurs mainly when miRNAs bind to sites in 3′ UTRs

To investigate the possible function of CDS-located miRNA target sites, we first compiled, from the miRNA transfection experiments of Linsley et al. (2007) and Grimson et al. (2007), transcripts that satisfied one of the following four mutually exclusive constraints. They had either no seed match to the transfected miRNA in either CDS or 3′ UTR, or precisely one seed match in the CDS, or precisely one seed match located in the 3′ UTR, or precisely two seed matches, one in the CDS and the other in the 3′ UTR. A seed match was again defined as a match to positions 2–8 of the transfected miRNA. Figure 4 shows that compared with 3′ UTR sites, CDS sites have a smaller, though still significant effect in inducing the degradation of the host transcripts. At least in some data sets, transcripts that carry both types of sites are down-regulated to a significantly lower level compared with transcripts that have only one type of site (not shown), consistent with what Fang and Rajewsky (2011) reported based on the analysis of the Baek et al. (2008) and Selbach et al. (2008) data sets.

CDS binding sites can effectively inhibit translation

Although miRNAs significantly reduce the mRNA levels of their targets (Bagga et al. 2005; Lewis et al. 2005), the initial paradigm was that miRNAs inhibit translation without affecting mRNA levels (Lee et al. 1993; Wightman et al. 1993), and the relative contribution of these mechanisms is still a matter of debate (Eulalio et al. 2008; Filipowicz et al. 2008; Béthune et al. 2012; David 2012). Recently, as a way to assess the extent of translation inhibition, Guo et al. (2010) measured changes in ribosome occupancy upon miRNA transfection transcriptome-wide. They found that at 32 h post-

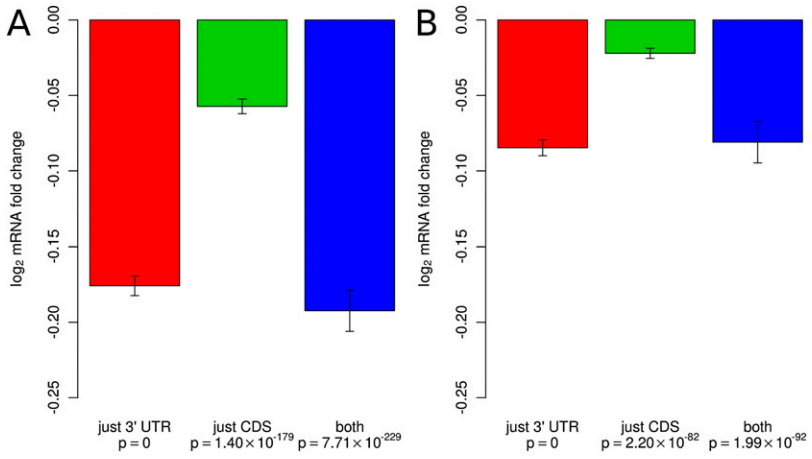


Figure 4. mRNA destabilization occurs mainly through sites located in the 3′ UTR. Shown are $\log_2$ fold changes in mRNA levels upon miRNA transfection in the experiments of (*A*) Linsley et al. (2007) and (*B*) Grimson et al. (2007). mRNAs with binding sites located in the CDS only, in the 3′ UTR only, and in both CDS and in 3′ UTR were analyzed separately. Fold changes were normalized to the average fold change of mRNAs that did not contain canonical binding sites to the transfected miRNA.

transfection, the reduction in ribosome occupancy of transcripts with miRNA binding sites in 3′ UTRs could be attributed in very large measure (84%) to the reduction in mRNA levels.

We asked whether this observation holds equally true for transcripts with miRNA-complementary sites that are located not in the 3′ UTR but rather in the CDS. We therefore analyzed the behavior of transcripts carrying a single miRNA seed match in the CDS and transcripts in which the single miRNA seed match occurred in the 3′ UTRs in the data of Guo et al. (2010). The data set included mRNA and ribosome occupancy levels at 12 and at 32 h post-transfection, which we analyzed as described in "Processing of quantitative proteomics, microarrays and deep sequencing data" in Methods. As shown in Figure 5 and previously reported by Guo et al. (2010), transcripts that are targeted in the 3′ UTR are strongly down-regulated at 32 h after transfection, and their ribosome occupancy at this time largely reflects the mRNA level. Transcripts in which the single miRNA seed match occurred in the CDS exhibit a similar behavior, though the reduction in their mRNA levels is more limited. Surprisingly, we found that transcripts with a single miRNA seed match in the CDS exhibited a significant reduction in ribosome occupancy at 12 h post-transfection that was not due to a corresponding reduction in mRNA levels ("translation, 12 h" bars in Fig. 5). In addition, the reduction in ribosome occupancy was increased for mRNAs with two miRNA seed matches in the CDS, corresponding to an estimated 18% reduction in the translation rate (Supplemental Fig. 13). This suggests that miRNA binding sites that are located in CDS are effective in inducing translational inhibition immediately after miRNA transfection while sites located in the 3′ UTR sites are effective in down-regulating the mRNA levels.

To assess the generality of these findings we revisited the data from a very recent study that investigated the miRNA-induced translational inhibition and mRNA degradation during maternal-zygotic transition in zebrafish (Bazzini et al. 2012). In contrast to Guo et al. (2010), who concluded that miRNAs have very limited effects on translation, Bazzini et al. (2012) identified a kinetic aspect of the miRNA-induced response. Namely, expression of the dre-miR-430 miRNAs caused an initial, transient, translational repression of the targets that was followed by the degradation of their corresponding mRNAs. Both CDS and 3′UTR sites were found to have a qualitatively similar effect, though CDS sites were much less effective than 3′UTR site (Bazzini et al. 2012). Reanalyzing the data of Bazzini et al. (2012) we found, as before, that CDS-located sites appear to induce translational repression at the early time points but are not effective in mRNA degradation (Supplemental Fig. 14). In contrast, 3′ UTR-located sites induce transient translational repression followed by mRNA degradation. We further analyzed transcriptomics and proteomics data that were obtained a week after the induction of miR-223 deletion in mouse (Baek et al. 2008). As shown in Supplemental Figure 15, the expression levels of mRNAs with CDS sites change very little upon knockout of miR-223 whereas their translation increases in the absence of the miRNA to a degree comparable to the translation of transcripts with binding sites in 3′ UTRs. Finally, we constructed reporters to measure the effect of a miRNA on the protein and mRNA level of miRNA targets with CDS-located target sites. First, we generated a HEK293T stable cell line containing an episomal pRTS-1 vector from which the expression of hsa-miR-124-3p can be induced with doxycyclin. We then selected two conserved CDS miR-124 target sites that were represented with a relatively high number of reads in the EIF2C2 (also known as Ago2) CLIP data of Hafner et al. (2010). These are likely to be functional hsa-miR-124-3p target sites. We inserted these sites in-frame, in their native sequence context, at the end of the Renilla-encoding coding region of the dual luciferase psiCHECK2 vector. We also constructed variants of these constructs that contained mutations in the miRNA seed-complementary region (see Supplemental Methods and Supplemental Fig. 16). We then estimated the protein ex-

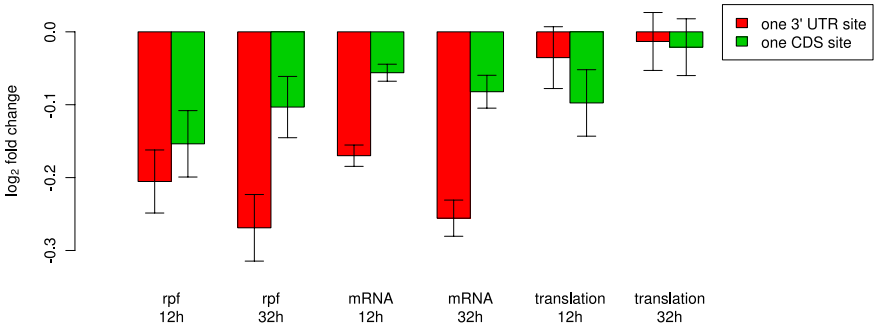


Figure 5. CDS-located binding sites transiently inhibit translation in miRNA transfection experiments. The figure shows $\log_2$ fold changes in mRNA levels (mRNA-seq), and ribosome-protected fragments (rpf) 12 and 32 h after hsa-miR-155-5p and hsa-miR-1 transfection. Changes in translation were estimated from the difference between changes in rpf and changes in mRNA levels. mRNAs with precisely one seed match to the transfected miRNA in the CDS and no seed match in the 3′ UTR were analyzed separately from mRNAs with precisely one seed match in the 3′ UTR and no seed match in the CDS. Fold changes were determined relative to the average fold change of mRNAs with no seed matches.

pression change (through luciferase assays) and the mRNA expression change (by qPCR) 24 h after the induction of hsa-miR-124 expression. As shown in Supplemental Figure 17, wild-type constructs exhibit a reduction in protein expression relative to the mutant constructs that cannot be explained by a corresponding decrease in mRNA levels. Thus, in human, mouse and zebrafish, CDS-located miRNA binding sites appear to be more effective in translational inhibition than in mRNA degradation.

Discussion

In spite of much research, the mechanisms of action and the function of miRNAs are insufficiently understood. An aspect that is still highly debated is whether miRNAs inhibit translation, induce mRNA degradation, or both. Recent high-throughput studies concluded that the reduction in protein levels upon miRNA expression is largely a consequence of the mRNA degradation induced by miRNAs (Baek et al. 2008; Selbach et al. 2008; Guo et al. 2010), though these initial experiments may have missed an important kinetic aspect (Bazzini et al. 2012). Another recurrent finding is that changes in target mRNA level upon miRNA transfection or overexpression are small. How these small effects can confer a selective advantage that is reflected in the strong evolutionary conservation of the target sites remains a puzzle. This applies especially to target sites located in coding regions, whose effects appear to be substantially smaller compared with those in 3′ UTRs (Gu et al. 2009; Hafner et al. 2010; Schnall-Levin et al. 2010, 2011; Fang and Rajewsky 2011). What then could the function of miRNA-complementary CDS sites be?

Through an appropriate choice of a background set of sequences, the ElMMo model that we previously developed (Gaidatzis et al. 2007) allows us to predict miRNA binding sites in regions other than 3′ UTRs. We thus predicted miRNA binding sites in CDS and used them to comparatively investigate the properties and effects of miRNA-complementary sites that are located in the coding domain or in the 3′ UTR. We found that the sites that are effective at various steps of the miRNA-induced cascade are under evolutionary selection pressure. Furthermore, properties that we previously found to characterize functional 3′ UTR sites also characterize the functional CDS sites. These findings suggest that CDS and 3′ UTR sites function through similar mechanisms and have a comparable impact on gene expression.

Simultaneous measurement of mRNA and protein levels is nontrivial for a variety of reasons. Obtaining a good coverage of the proteome is difficult. Furthermore, proteomics measurements require that the proteins are labeled and this imposes constraints on the timing of miRNA transfection (or induction of expression). Sequencing of ribosome-protected fragments (rpf) circumvents some of these problems. For the first time, our analysis of such data reveals an effect that CDS-located miRNA binding sites are capable of inducing more effectively than 3′ UTR-located sites, namely a rapid reduction in mRNA translation. In some of the data sets, the effect appears to be transient. Why on longer time scales the reduction in rpf can be largely explained by the reduction of the mRNA level is at the moment unclear. Some experiments involve miRNA transfection, which is transient by nature. However, the dre-miR-430 miRNA in zebrafish persists beyond 6 h post-fertilization (hpf) when the translation inhibition is apparently relieved. Thus, transient expression of the miRNA is probably not an explanation in this case, although it cannot be entirely ruled out because dre-miR-430 may be displaced from the RNA-induced silencing complex by other miRNAs at later time points.

Interestingly, in both the zebrafish system and our reporter system, some mRNAs with miRNA binding sites in coding regions appear to be up-regulated upon miRNA expression. This may reflect another unsuspected complexity of the kinetics of miRNA-dependent gene regulation (Bazzini et al. 2012; Béthune et al. 2012; Djuranovic et al. 2012). In zebrafish, one could hypothesize that the abundance of polyadenylated mRNAs, which are isolated for mRNA expression profiling, differs between Dicer-deficient and wild-type embryos. It is well known that polyadenylation is a common means of regulation of mRNA stability and translation (Telford et al. 1990; Audic et al. 1997; Henrion et al. 2000), with deadenylation being reported to be sufficient for mRNA degradation (Audic et al. 1997). In the Bazzini et al. (2012) data, however, what is observed is preferential stabilization of known dre-miR-430 targets, which suggests either a direct involvement of miRNA-dependent regulation or an indirect correlation caused by the overlap between the set of mRNAs that undergo polyadenylation changes in development and the set of dre-miR-430 targets. At this point it is difficult to imagine how miRNAs would be directly involved. miRNAs have been reported to induce deadenylation, particularly miR-430 during maternal-zygotic transition in zebrafish (Giraldez et al. 2006). However, at 2 hpf there should be no difference in the mRNA level of miR-430 targets in Dicer-deficient and wild-type cells, both of which lack miR-430. In Xenopus, it has been reported that siRNAs impair the accumulation of the embryonically expressed miR-427, restricting RNAi during early development (Lund et al. 2011). If this effect were present in zebrafish as well and the efficiency of the siRNAs were enhanced in Dicer-negative cells, it could contribute to the observed stabilization of miR-430 targets at 2 hpf. These considerations, however, do not apply to our reporter system. Because the effect appears to be reporter-specific, additional interplays with other regulators of mRNA stability and translation rate may be at work. A few examples of crosstalk between RNA-binding proteins and the miRNA pathway have been described (Bhattacharyya et al. 2006; Kedde et al. 2007; Kedde and Agami 2008; Kim et al. 2009) and more are likely to emerge in the future.

Finally, different families of miRNAs that are broadly expressed but are active in different contexts show distinct preferences for the CDS or 3′ UTR. On one side of the spectrum are miRNAs expressed in the embryonic cells and miRNAs of the miR-17-92 cluster, that target predominantly 3′ UTRs, but also show the strongest enrichment in CDS–3′ UTR co-targeting. This suggests that these miRNAs strongly and robustly down-regulate target genes, for example at developmental transitions. On the other side of the spectrum are miRNAs of the miR-16 family, which have been previously shown to regulate cell cycle (Linsley et al. 2007). These miRNAs appear to preferentially bind to CDS-located sites, which we found to be effective in rapid inhibition of translation. This type of response may be better suited on the time scale of the cell cycle. Future work will determine the magnitude and timing of gene repression induced by miRNAs binding to the coding regions to uncover new aspects of miRNA biology.

Methods

Estimation of the selection pressure on CDS and 3′ UTR target sites

In a previous study, we introduced the ElMMo model for inferring miRNA target sites based on comparative genomics data (Gaidatzis et al. 2007). ElMMo is parameter-free, requiring only a set of

miRNA-complementary, putative sites in a reference species and pairwise genome alignments between the reference species and other species. We only considered regions that were complementary to positions 1–7, 2–8, or 1–8 of the miRNA as putative miRNA binding sites. Thus, miRNAs which are identical at positions 1–8 defined a miRNA family whose members have indistinguishable target sites according to our model. ElMMo predicts miRNA target site by estimating the selection pressure on motifs that are complementary to specific miRNAs, relative to a "background" set of motifs that do not correspond to miRNAs. By simply changing the type of sequences in the input data set (coding regions or 3′ UTRs), ElMMo thus allows us to independently estimate the selection pressure on binding sites for individual miRNAs in CDS and 3′ UTRs, respectively. We used human as the reference species, and the RefSeq data set of human transcripts downloaded from the National Center for Biotechnology Information (NCBI) as the set of transcripts in which we predicted target sites.

Processing of quantitative proteomics, microarrays, and deep sequencing data

mRNA sequences, gene to mRNA mappings, mRNA to protein mappings, representative mRNA

We used the RefSeq mRNA database that we downloaded from NCBI (www.ncbi.nlm.nih.gov) on Jan 18th 2011 for all analyses described in this manuscript. The Entrez database provides us with mappings of mRNAs-to-genes and protein-to-genes. For each gene, we defined a representative mRNA as the longest mRNA in RefSeq featuring 5′ UTR, coding domain, and 3′ UTR annotation.

Computational analysis of quantitative proteomics and microarray data

We followed the methods previously described in Hausser et al. (2009), except for minor changes in the analysis pipeline and for the use of updated genomic, mRNA, and protein sequence databases. The corresponding methods are described in the Supplemental Material.

Computational analysis of ribosome protected fragment sequencing and mRNA profiling data from Guo et al. (2010)

Guo et al. (2010) analyzed mRNAs and ribosome-protected fragments (rpf) in HeLa cells that were mock-transfected or transfected with hsa-miR-1 or hsa-miR-155-5p. The deep sequencing data was downloaded from the Sequence Reads Archive (accession: GSE22004) and analyzed on the CLIPZ server (Khorshid et al. 2011). To obtain per-gene expression levels, we first computed the number of reads mapping to representative mRNAs (longest mRNA having annotated 5′ UTR, CDS, and 3′ UTR among those associated with the gene according to the Entrez Gene database of NCBI). The number of reads was subsequently normalized by the length of the CDS in the case of rpf samples, or the length of the representative mRNA in the case of the mRNA-seq samples. For the analysis, we only considered 10,222 genes with nonzero expression levels in all experimental conditions (mock, hsa-miR-155-5p, and hsa-miR-1 transfections/mRNA and rpf sequencing/0, 12, and 32 h after transfection).

To investigate the effect of the transfected miRNAs on mRNA stability and translation, we first computed the $\log_2$ fold change in rpf and mRNA expression upon miRNA transfection compared with mock transfection. mRNAs were then divided into four subsets: those with no binding sites to the transfected miRNA, those with exactly one binding site located in the 3′ UTR, those with exactly one binding site located in the CDS, and mRNAs with two binding sites in the CDS. Binding sites were defined as canonical seed matches—7mer-A1, 7mer-m8, 8mer (Bartel 2009). The overall effect of the transfected miRNA on rpf and mRNA levels was estimated by subtracting the $\log_2$ fold change of mRNAs without binding sites from that of mRNAs with binding sites (defined as above). Finally, the $\log_2$ fold change in translation induced by the transfected miRNAs was estimated from the difference between the $\log_2$ fold changes in rpf and mRNA levels.

EIF2C2 CLIP from Kishore et al. (2011)

We started from the raw CLIP and mRNA-seq reads (deposited in NCBI GEO under the accession GSE28865) of the 6 EIF2C2 CLIP experiments performed by Kishore et al. (2011). CLIP and mRNA-seq data were processed on the CLIPZ server (Khorshid et al. 2011). Only reads annotated as mRNA, mapping to a single genomic locus and to a single representative mRNA were used. CLIPed sites were defined as nonoverlapping 40-nt windows. For each mRNA, the first window was centered on the position of the mRNA that accumulated most reads. Additional windows were extracted similarly, under the constraint that they did not overlap with any previously extracted window. For each window, we computed the posterior probability that the number of CLIP reads associated to the window was larger than expected in a statistical model that only takes mRNA abundance into account (see "mRNA site extraction from CLIP reads" below). Finally, we determined which miRNA was most likely to be bound to the window by searching for 2–8 seed matches to the top 10 miRNA families expressed in HEK293 cells. In case several miRNAs mapped to a binding site, the site was assigned to most highly expressed miRNA. HEK293 miRNA profiles were determined from the two mild MNase digestion EIF2C2 CLIP samples of Kishore et al. (2011), using the methodology described in that paper.

mRNA site extraction from the EIF2C2 CLIP experiments of Kishore et al. (2011)

Outline of the statistical model

For each of the c sites yielded by the CLIP experiment, we compare the observed number of CLIP reads r_i with the number of reads expected under a background model in which the number of reads per CLIP site depends only on the abundance of the corresponding mRNA. We neglect biases related to sequence accessibility, sequence composition, etc. Instead, we will focus on how CLIP sites located on mRNAs of different abundance are sampled in a thought experiment in which all CLIP sites are equally prone to bind the RNA-binding protein of interest and generate reads.

Estimating mRNA frequencies from mRNAseq data

Let us assume there are m genes, expressing different mRNAs. Although a gene typically expresses multiple mRNA isoforms, we assume for simplicity that each gene is represented by a single mRNA, defined as the longest isoform with 5′ UTR, CDS, and 3′ UTR that we find in the RefSeq database of NCBI. Thus, having m genes represented by m mRNAs, we would like to compute the probability that mRNA m_i has relative abundance f_i from the number of reads n_i corresponding to mRNA m_i that we observe in a sample.

If n is the total number of mRNA-seq reads and q_i is the (unknown) probability that a read comes from mRNA m_i, then the number of mRNA-seq reads that map to mRNA m_i can be modeled to follow a binomial distribution:

$$P(n_i|q_i,n) = \frac{\Gamma(n+1)}{\Gamma(n-n_i+1)\Gamma(n_i+1)} q_i^{n_i}(1-q_i)^{n-n_i}.$$

Using Bayes's theorem, we can now compute the probability of q_i:

$$P(q_i \mid n_i, n) = \frac{P(n_i|q_i, n)P(q_i|n)}{\int_0^1 P(n_i|q, n)P(q|n)dq}.$$

Assuming a uniform prior on q_i, we obtain

$$P(q_i \mid n_i, n) = \frac{\frac{\Gamma(n+1)}{\Gamma(n-n_i+1)\Gamma(n_i+1)} q_i^{n_i}(1-q_i)^{n-n_i}}{\int_0^1 \frac{\Gamma(n+1)}{\Gamma(n-n_i+1)\Gamma(n_i+1)} q^{n_i}(1-q)^{n-n_i} dq}.$$

Setting $\alpha = n_i + 1$ and $\beta = n - n_i + 1$ and using the beta function to compute the denominator yields the distribution on the probability q_i for a read to come from mRNA m_i:

$$P(q_i \mid n_i, n) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} q_i^{\alpha-1}(1-q_i)^{\beta-1},$$

where n is total number of mRNA-seq reads and n_i is the number of reads mapping to mRNA m_i.

The probability q_i to obtain reads from mRNA m_i is proportional to both the relative abundance f_i of mRNA m_i, and the length l_i of mRNA m_i. Therefore,

$$q_i = \frac{1}{Z} f_i l_i$$

where Z is a normalizing constant $Z = \sum_{i=i}^{m} f_i l_i$. If the number of distinct mRNAs m is large and if the relative abundance of mRNAs is independent of their length, then on average, we can make the following approximation:

$$Z \simeq \frac{1}{m} \sum_{i=1}^{m} l_i.$$

What we are interested in is the relative abundance of mRNA m_i, $P(f_i)$ and to obtain it, we start from

$$\int_0^1 q_i^{\alpha-1}(1-q_i)^{\beta-1} dq_i = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

and perform a change of variable $q_i = \frac{l_i}{Z} f_i$, i.e., $dq_i = \frac{l_i}{Z} df_i$ which gives

$$\int_0^1 \left(\frac{l_i}{Z} f_i\right)^{\alpha-1} \left(1 - \frac{l_i}{Z} f_i\right)^{\beta-1} df_i = \frac{Z}{l_i} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

This further yields

$$P(f_i) = \frac{l_i}{Z} \frac{\Gamma(n+2)}{\Gamma(n_i+1)\Gamma(n-n_i+1)} \left(\frac{l_i}{Z} f_i\right)^{n_i} \left(1 - \frac{l_i}{Z} f_i\right)^{n-n_i}$$

Finally, one can compute the expected relative abundance $\langle f_i \rangle$ of mRNA m_i from the expected value of a beta distribution,

$$\left\langle \frac{l_i}{Z} f_i \right\rangle = \langle q_i \rangle = \frac{n_i + 1}{n + 2}$$

which gives

$$\langle f_i \rangle = \frac{Z}{l_i} \frac{n_i + 1}{n + 2}. \quad (1)$$

Similarly, the deviation around the expected value $\langle f_i \rangle$ can be obtained from the following fundamental property of the beta distribution:

$$\mathrm{var}(q_i) = \frac{(n_i+1)(n-n_i+1)}{(n+2)^2(n+3)}.$$

Since $\mathrm{var}(f_i) = \mathrm{var}\left(\frac{Z}{l_i} q_i\right) = \left(\frac{Z}{l_i}\right)^2 \mathrm{var}(q_i)$, the standard deviation around the expected value $\langle f_i \rangle$ becomes

$$\sigma_{f_i} = \frac{Z}{l_i(n+2)} \sqrt{\frac{(n_i+1)(n-n_i+1)}{n+3}}$$

or, in the limit of a large n,

$$\sigma_{f_i} \simeq \frac{Z}{l_i} \frac{\sqrt{n_i+1}}{n}. \quad (2)$$

The number n of reads mapping to mRNA is typically in the order of 10^6, with <1% of these reads mapping to the single most abundant mRNA. As a result, σ_{f_i} is $<10^{-4}$.

Computing the distribution of the number of reads that is expected to fall into CLIP sites based on the abundance of the mRNA in the mRNA–seq data

Let r be the total number of reads produced by the CLIP experiment, r_i of which map to CLIP site i, and let $\mu(i)$ be the mRNA on which CLIP site i is located. Let c be the number of CLIP sites observed in the experiment. Because we only sample a finite number of reads, CLIP sites compete with each other for yielding sequenced reads. In our thought experiment, all CLIP sites are equally prone to be bound by the RNA-binding protein of interest and generate reads. This implies that CLIP sites are sampled according to the abundance of the mRNAs on which they reside. In that case, the probability that a read maps to CLIP site i is $\frac{f_{\mu(i)}}{Q}$, where Q is the normalizing constant $\sum_{i=1}^{c} f_{\mu(i)}$.

And so, we can write the distribution of the number of reads r_i for site i as

$$P\left(r_i|r, f_{\mu(i)}\right) = \frac{\Gamma(r+1)}{\Gamma(r_i+1)\Gamma(r-r_i+1)} \left(\frac{f_{\mu(i)}}{Q}\right)^{r_i} \left(1 - \frac{f_{\mu(i)}}{Q}\right)^{r-r_i}.$$

However, we do not know the relative mRNA abundance $f_{\mu(i)}$, only its distribution $P(f_{\mu(i)})$. Ideally, we would like to integrate $f_{\mu(i)}$ out to compute

$$\begin{aligned} P(r_i|r, n_{\mu(i)}, l_{\mu(i)}) &= \int_0^1 P\left(r_i|r, f_{\mu(i)}\right) P\left(f_{\mu(i)}|n_{\mu(i)}, l_{\mu(i)}\right) df_{\mu(i)} \\ &= K \int_0^1 \left(f_{\mu(i)} l_{\mu(i)}\right)^{n_{\mu(i)}} \left(Z - f_{\mu(i)} l_{\mu(i)}\right)^{n-n_{\mu(i)}} f_{\mu(i)}^{r_i} \left(Q - f_{\mu(i)}\right)^{r-r_i} df_{\mu(i)} \end{aligned}$$

with

$$K = \frac{1}{Z^n Q^r} \frac{\Gamma(r+1)}{\Gamma(r_i+1)\Gamma(r-r_i+1)} \frac{\Gamma(n+2)}{\Gamma(n_i+1)\Gamma(n-n_i+1)}.$$

Since this integral cannot be solved analytically, we make the approximation that all the probability density is concentrated at $\langle f_{\mu(i)} \rangle$. Because mRNA-seq libraries are typically large ($n > 10^6$) compared with the number of genes ($m < 25{,}000$), and because the single most abundant mRNA usually represents <1% of the total mRNA pool, this is a reasonable approximation to make, as shown by Equation 2 and illustrated by Supplemental Figure 18. Approximating $P(f_{\mu(i)})$ by its expected value $\frac{Z}{l_{\mu(i)}} \frac{n_{\mu(i)}+1}{n+2}$ (see Equation 1) as opposed to the maximum likelihood estimate $\frac{Z}{l_{\mu(i)}} \frac{n_{\mu i}}{n}$ has the advantage that the relative abundance of any mRNA is always nonzero. This makes it possible to compute a probability of enrichment using a simple formula, even for sites located on mRNAs with no mRNAseq reads, as we will show now.

The expected value approximation of $P(f_{\mu(i)})$ yields

$$P(r_i|r) = \frac{\Gamma(r+1)}{\Gamma(r_i+1)\Gamma(r-r_i+1)}\left(\frac{f_{\mu(i)}}{Q}\right)^{r_i}\left(1-\frac{f_{\mu(i)}}{Q}\right)^{r-r_i}, \quad (3)$$

where $f_{\mu(i)} = \frac{Z}{l_{\mu(i)}}\frac{n_{\mu(i)}+1}{n+2}$ (see Equation 1) and $Q = \sum_{j=1}^{c} f_{\mu(j)} = \frac{Z}{n+2}\sum_{j=1}^{c}\frac{n_{\mu(j)}+1}{l_{\mu(j)}}$.

Finally, substituting $\frac{f_{\mu(i)}}{Q}$ back in Equation 3 leads to

$$P(r_i|r) = \frac{\Gamma(r+1)}{\Gamma(r_i+1)\Gamma(r-r_i+1)}p_i^{r_i}(1-p_i)^{r-r_i} \quad (4)$$

with

$$p_i = \frac{n_{\mu(i)}+1}{l_{\mu(i)}\sum_{j=1}^{c}\frac{n_{\mu(j)}+1}{l_{\mu(j)}}}.$$

Ranking and selecting significantly enriched CLIP sites

We use the probabilistic model of Equation 4 as the basis for testing whether site i is significantly enriched over the expected number of CLIP read given the abundance of mRNA $\mu(i)$.

Let ρ be the fraction of the r CLIP reads that map to site i. Note that this fraction ρ is unknown, but can be estimated from the number of CLIP reads r_i mapping to site i. Under the H_0 hypothesis that the site i is not enriched compared with what is expected given the abundance of the mRNA $\mu(i)$, we would have $\rho \le p_i$, p_i being the fraction of CLIP reads expected to map to site i given the abundance of mRNA $\mu(i)$. In contrast, under the H_1 hypothesis that the site i is significantly enriched over the number of reads expected from the abundance of mRNA $\mu(i)$ alone, we would have $\rho > p_i$. Note that $P(H_0) + P(H_1) = 1$.

Using Bayes's theorem, we can compute the posterior probability of H_1 from

$$P(H_1|r,r_i) = \frac{P(r_i|r,H_1)P(H_1)}{P(r_i)} \quad (5)$$

$$= \frac{P(r_i|r,H_1)P(H_1)}{P(r_i|r,H_0)P(H_0)+P(r_i|r,H_1)P(H_1)} \quad (6)$$

$$= \frac{P(r_i|r,H_1)}{P(r_i|r,H_0)+P(r_i|r,H_1)}, \quad (7)$$

where we assumed that both hypotheses H_0 and H_1 have equal prior probability $P(H_0) = P(H_1)$. $P(r_i|r, H_1)$ can be obtained from the likelihood function introduced in Equation 4:

$$P(r,r_i|H_1) = P(r_i|r,\rho > p_i) \quad (8)$$

$$= \int_{p_i}^{1} P(r_i|r,\rho)P(\rho)d\rho \quad (9)$$

$$= \int_{p_i}^{1} \frac{\Gamma(r+1)}{\Gamma(r_i+1)\Gamma(r-r_i+1)}\rho^{r_i}(1-\rho)^{r-r_i}d\rho \quad (10)$$

assuming a uniform prior on ρ ($P(\rho) = 1$, $\forall\rho \in [0, 1]$). By applying the same reasoning to $P(r_i|r, H_0)$, one can show the denominator of Equation 7 to be

$$P(r_i|r,H_0)+P(r_i|r,H_1) = \int_{0}^{1} \frac{\Gamma(r+1)}{\Gamma(r_i+1)\Gamma(r-r_i+1)}\rho^{r_i}(1-\rho)^{r-r_i}d\rho. \quad (11)$$

Finally, substituting Equations 10 and 11 into Equation 7 gives a simple expression for the posterior probability $P(H_1|r, r_i)$ that the number of CLIP reads r_i mapping to site i is larger than what would be expected given the abundance of the mRNA alone:

$$P(H_1|r,r_i) = \frac{\int_{p_i}^{1}\rho^{r_i}(1-\rho)^{r-r_i}d\rho}{\int_{0}^{1}\rho^{r_i}(1-\rho)^{r-r_i}d\rho} = \int_{p_i}^{1}\frac{\Gamma(r+2)}{\Gamma(r_i+1)\Gamma(r-r_i+1)}\rho^{r_i}(1-\rho)^{r-r_i}d\rho, \quad (12)$$

which is the reverse cumulative probability of a Beta distribution.

To obtain binding sites, 40-nt windows are ranked by decreasing posterior probability first, and in the case of ties, by decreasing enrichment ratios $\frac{r_i}{rp_i}$.

Extraction of functional and nonfunctional miRNA binding sites

Data sets used in the analysis

The microarray and deep-sequencing data for the following references were obtained from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under the accessions GSE6838 and GSE8501 for Linsley et al. (2007), GSE14537 for Hausser et al. (2009), GSE11968 for Baek et al. (2008), GDS1858 for Grimson et al. (2007), GSE22004 for Guo et al. (2010), and GSE28865 for Kishore et al. (2011). The microarray data from Selbach et al. (2008) were downloaded from http://psilac.mdc-berlin.de/download/. Protein profiling data from Selbach et al. (2008) and Baek et al. (2008) were retrieved from http://psilac.mdc-berlin.de/download/ and http://dx.doi.org/10.1038/nature07242, respectively. Finally, predictions of miRNA binding sites under selection pressure are available as flat files from http://www.mirz.unibas.ch/miRNAtargetPredictionBulk.php.

Extraction of functional and nonfunctional miRNA binding sites from combined miRNA perturbation and microarray experiments

Among the mRNA profiling data sets that we reanalyzed, the experiments performed by Grimson et al. (2007), Baek et al. (2008), Selbach et al. (2008), and Guo et al. (2010) did not feature biological replicates. For these data sets, we considered the top 250 down-regulated mRNAs that carried precisely one canonical seed match (7mer-A1, 7mer-m8, or 8mer [Bartel 2009]) to the transfected miRNA. After discarding mRNAs with seed matches located in the 5′ UTRs, we ended up with a set of "positive" (functional in mRNA down-regulation) seed matches. The negatives were obtained from the 250 mRNAs whose $\log_2$ expression fold changes were closest to 0 when comparing the miRNA-transfected samples to the mock-transfected samples. After discarding all seed matches located in the 5′ UTRs, we ended up with a set of "negative" seed matches.

For the experiments performed by Karginov et al. (2007), Linsley et al. (2007), and Hausser et al. (2009), which featured biological replicates, we applied a method that we designed for selecting transcripts that, with high probability, are affected in expression by the miRNA across all experiments in which the expression of the given miRNA was perturbed. The method was described previously (see the Supplemental Material of Hausser et al. 2009). Briefly, we first calculated, for each pairwise microarray comparison (referred to as contrast) k, the probability $P_k(f \mid -)$ that a transcript that is not a target will have a log fold change of f. To estimate the distributions $P_k(f \mid -)$ we assumed that they are Gaussian with means μ_k and standard deviation σ_k. Further assuming that transcripts that do not carry at least a heptameric seed-complementary site are unlikely to be real targets, we estimated μ_k and σ_k from the observed expression changes of transcripts with-

out such seed matches. We similarly calculated, for each contrast k, a distribution $P_k(f \mid +)$ that a transcript which is a true target of the miRNA will have a fold change f. Not knowing the distribution of the severity of the effect that miRNAs have on the expression of their targets we assumed as little as possible about the distribution $P_k(f \mid +)$, namely that a true target must change expression in the right direction, i.e., $f < 0$ for a miRNA overexpression experiment, and that expression changes are limited to a finite range over which the expression change has a "uniform" distribution. Finally, based on these distributions, we estimated the posterior probability that a transcript with fold change f was a functional target in a given experiment.

To obtain nonfunctional binding sites, we selected those transcripts with the smallest sum of squared $\log_2$ fold changes in the biological replicates. Finally, for the purpose of comparing the properties of functional and nonfunctional sites, we proceeded as with experiments where no replicates were performed: We selected 250 functional sites and 250 nonfunctional sites according to the criteria defined above and we discarded those cases in which the seed match was in the 5′ UTR.

Extraction of functional and nonfunctional sites from ElMMo predictions

From our predictions of miRNA target sites under evolutionary selection (Gaidatzis et al. 2007) and for each of the experimentally tested and evolutionarily conserved miRNAs (hsa-miR-30a-5p, hsa-let-7c, hsa-miR-155-5p, hsa-miR-1, hsa-miR-103-3p, hsa-miR-15a-5p, hsa-miR-16-5p, hsa-miR-106b-5p, hsa-miR-20a-5p, hsa-miR-141-3p, hsa-miR-200a-3p, hsa-miR-181a-5p, hsa-miR-124-3p and hsa-miR-17-5p), we selected the top 250 target sites in the order of their posterior probability of being under selection. We also selected an equal number of sites least likely to be under selection, i.e., that had the smallest posterior probability of being under selection. This procedure was applied to the 3′ UTR and CDS ElMMo miRNA target site predictions separately.

Extraction of functional and nonfunctional sites from CLIP experiments of Kishore et al. (2011)

The 40-nt windows were sorted by decreasing posterior probability, and, in the case of ties, by decreasing enrichment in CLIP reads (see "mRNA site extraction from EIF2C2 CLIP experiments" in Methods). We only kept windows with exactly one canonical seed match (7mer-U1, 7mer-m8, or 8mer [Bartel 2009]) to one of the top 10 expressed miRNA families in HEK293 cells. These top 10 families were determined from reads mapping to miRNAs in the EIF2C2 MNase CLIP experiments, as described in Kishore et al. (2011) and were hsa-let-7a-5p, hsa-miR-103-3p, hsa-miR-106a-5p, hsa-miR-10a-5p, hsa-miR-15a-5p, hsa-miR-19a-3p, hsa-miR-25-3p, hsa-miR-30a-5p, hsa-miR-320a, hsa-miR-7-5p. Functional 3′ UTR binding sites were obtained from the top 250 windows located in the 3′ UTR while nonfunctional miRNA binding sites were defined as the 250 windows with weakest enrichment. Finally, CDS binding sites were obtained in a similar fashion from windows located in coding domains.

Properties definition and computation

For data sets in which the effect could not be unambiguously attributed to a single site (all data sets other than those derived by CLIP and comparative genomic analysis), we only analyzed mRNAs that had precisely one canonical miRNA seed match (7mer-A1, 7mer-m8, or 8mer). In addition, we only considered sites that were at least 100 nt away from the 5′ and 3′ ends of the mRNA. For each individual putative target site we then computed 32 sequence and structure properties, described in detail in the Supplemental Methods and Hausser et al. (2009).

Enrichments in the number of genes co-targeted in the CDS and in the 3′ UTR

To determine whether individual miRNAs tend to target the same genes in the CDS and in the 3′ UTR more than expected by chance, we determined, for each miRNA, the number c_{11} of genes containing at least one of the top 250 CDS and at least one of the 250 top 3′ UTR binding sites predicted by ElMMo. These numbers are reported in the "genes targeted both in CDS and 3′ UTR" column of Supplemental Table 2. In addition, we also determined the number c_{10} of genes containing at least one of the top 250 CDS binding sites, but not any of the 250 predicted 3′ UTR binding sites. Similarly, we determined the number c_{01} of genes containing at least one of the top 250 predicted 3′ UTR binding sites but none of the top 250 predicted CDS binding sites.

Under the hypothesis that miRNAs target the CDS and 3′ UTR independently, we expect a fraction $\frac{c_{11}+c_{01}}{n}\frac{c_{11}+c_{10}}{n}$ of all mRNAs to be co-targeted in the CDS and in the 3′ UTR, where n = 18,430 is the number of representative mRNAs used in the analysis. We define the fold enrichment in the number of co-targeted mRNAs as the ratio between the observed fraction of co-targeted mRNAs $\frac{c_{11}}{n}$ and the expected fraction of co-targeted mRNAs under the model that miRNAs independently target the CDS and the 3′ UTR $\frac{c_{11}+c_{01}}{n}\frac{c_{11}+c_{10}}{n}$. The "co-targeting fold enrichment" reported in Supplemental Table 2 is defined as

$$\frac{nc_{11}}{(c_{11}+c_{10})(c_{11}+c_{01})}$$

and we use Fisher's test to assess whether the obtained ratio is significantly different from 1. The Bonferroni-corrected P-values are reported in the last column of Supplemental Table 2.

Statistical evidence that CDS and 3′ UTR co-targeting occurs mostly for miRNAs that preferentially target 3′ UTRs

To test for a statistical link between the tendency of individual miRNAs to co-target the CDS and 3′ UTRs of the same mRNAs and their preference for CDS vs. 3′ UTR targeting, we first selected n = 62 miRNAs that showed significant enrichment in the number of co-targeted mRNAs, as described in the previous section. These miRNAs had an average excess of μ = 585.7 predicted binding sites in the 3′ UTR. From the standard deviation σ = 589.5 between the individual 62 miRNAs, we can compute a Z-score for the tendency of these miRNAs to preferentially target 3′ UTRs:

$$z = \frac{\mu}{\sigma/\sqrt{n}} = 7.82,$$

which indicates a strong preference for 3′ UTR targeting. Further assuming that the preferences for CDS vs. 3′ UTR targeting are Gaussian distributed, we can compute the P-value ($P < 10^{-15}$) that miRNAs that co-target mRNAs in the CDS and in the 3′ UTR are mostly found among miRNAs that preferentially target 3′ UTRs.

Acknowledgments

We thank Nitish Mittal for valuable input in the design and analysis of the reporter experiments as well as members of the Zavolan and Filipowicz lab for comments and discussions. We also thank Antonio Giraldez and Miler Lee for sharing the processed mRNAseq and RPF zebrafish sequencing data. Work on this project was supported by the Swiss National Science Foundation Grant #31003A_127307 to M.Z. B.B. was supported by a Werner Siemens fellowship.

Author contributions: M.Z. and J.H. designed the research, J.H. and B.B. performed analyses, A.P.S. performed the experiments,

A.P.S. and J.H. analyzed the data, J.H. and M.Z. wrote the paper with the help of B.B. and A.P.S.

References

Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG. 2009. Lost in translation: An assessment and perspective for computational microRNA target identification. *Bioinformatics* **25:** 3049–3055.

Audic Y, Omilli F, Osborne HB. 1997. Postfertilization deadenylation of mRNAs in *Xenopus laevis* embryos is sufficient to cause their degradation at the blastula stage. *Mol Cell Biol* **17:** 209–218.

Baek D, Villén J, Shin C, Camargo F, Gygi S, Bartel DP. 2008. The impact of microRNAs on protein output. *Nature* **455:** 64–71.

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

Balatti V, Maniero S, Ferracin M, Veronese A, Negrini M, Ferrocci G, Martini F, Tognon MG. 2011. MicroRNAs dysregulation in human malignant pleural mesothelioma. *J Thorac Oncol* **6:** 844–851.

Bartel DP. 2009. MicroRNAs: Target recognition and regulatory functions. *Cell* **136:** 215–233.

Bazzini AA, Lee MT, Giraldez AJ. 2012. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336:** 233–237.

Béthune J, Artus-Revel CG, Filipowicz W. 2012. Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* **13:** 716–723.

Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125:** 1111–1124.

Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, Sieburth L, Voinnet O. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320:** 1185–1190.

Bushati N, Cohen SM. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23:** 175–205.

Chi SW, Zang JB, Mele A, Darnell RB. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460:** 479–486.

David R. 2012. miRNAs' strict schedule. *Nat Rev Mol Cell Biol* **13:** 340–341.

Djuranovic S, Nahvi A, Green R. 2012. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science* **336:** 237–240.

Eulalio A, Huntzinger E, Izaurralde E. 2008. Getting to the root of miRNA-mediated gene silencing. *Cell* **132:** 9–14.

Fang Z, Rajewsky N. 2011. The impact of miRNA target sites in coding sequences and in 3′UTRs. *PLoS ONE* **6:** e18067.

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

Forman JJ, Legesse-Miller A, Coller HA. 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* **105:** 14879–14884.

Forrest ARR, Kanamori-Katayama M, Tomaru Y, Lassmann T, Ninomiya N, Takahashi Y, de Hoon MJL, Kubosaki A, Kaiho A, Suzuki M, et al. 2010. Induction of microRNAs, mir-155, mir-222, mir-424 and mir-503, promotes monocytic differentiation through combinatorial regulation. *Leukemia* **24:** 460–466.

Fulci V, Colombo T, Chiaretti S, Messina M, Citarella F, Tavolaro S, Guarini A, Foà R, Macino G. 2009. Characterization of B- and T-lineage acute lymphoblastic leukemia by integrated analysis of MicroRNA and mRNA expression profiles. *Genes Chromosomes Cancer* **48:** 1069–1082.

Gaidatzis D, van Nimwegen E, Hausser J, Zavolan M. 2007. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* **8:** 69.

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish miR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312:** 75–79.

Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, Vadas MA, Khew-Goodall Y, Goodall GJ. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10:** 593–601.

Grimson A, Farh K, Johnston W, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Mol Cell* **27:** 91–105.

Gu S, Jin L, Zhang F, Sarnow P, Kay M. 2009. Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs. *Nat Struct Mol Biol* **16:** 144–150.

Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466:** 835–840.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-c, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129–141.

Hausser J, Landthaler M, Jaskiewicz L, Gaidatzis D, Zavolan M. 2009. Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets. *Genome Res* **19:** 2009–2020.

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, et al. 2005. A microRNA polycistron as a potential human oncogene. *Nature* **435:** 828–833.

Henrion G, Renard JP, Chesné P, Oudin JF, Maniey D, Brunet A, Osborne HB, Duranthon V. 2000. Differential regulation of the translation and the stability of two maternal transcripts in preimplantation rabbit embryos. *Mol Reprod Dev* **56:** 12–25.

Huang Q, Gumireddy K, Schrier M, le Sage C, Nagel R, Nair S, Egan DA, Li A, Huang G, Klein-Szanto AJ, et al. 2008. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat Cell Biol* **10:** 202–210.

Huang FWD, Qin J, Reidys CM, Stadler PF. 2010. Target prediction and a statistical sampling algorithm for RNA-RNA interaction. *Bioinformatics* **26:** 175–181.

Hurst LD. 2006. Preliminary assessment of the impact of microRNA-mediated regulation on coding sequence evolution in mammals. *J Mol Evol* **63:** 174–182.

Jones-Rhoades MW, Bartel DP, Bartel B. 2006. MicroRNAS and their regulatory roles in plants. *Annu Rev Plant Biol* **57:** 19–53.

Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, Mandel G, Hannon GJ. 2007. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci* **104:** 19291–19296.

Kedde M, Agami R. 2008. Interplay between microRNAs and RNA-binding proteins determines developmental processes. *Cell Cycle* **7:** 899–903.

Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JA, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, Ørom UA, et al. 2007. RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131:** 1273–1286.

Khorshid M, Rodak C, Zavolan M. 2011. CLIPZ: A database and analysis environment for experimentally determined binding sites of RNA-binding proteins. *Nucleic Acids Res* **39:** D245–D252.

Kim HD, Shay T, O'Shea EK, Regev A. 2009. Transcriptional regulatory circuits: Predicting numbers from alphabets. *Science* **325:** 429–432.

Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. 2011. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* **8:** 559–564.

Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, et al. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37:** 495–500.

Lanet E, Delannoy E, Sormani R, Floris M, Brodersen P, Crété P, Voinnet O, Robaglia C. 2009. Biochemical evidence for translational repression by *Arabidopsis* microRNAs. *Plant Cell* **21:** 1762–1768.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75:** 843–854.

Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120:** 15–20.

Linsley P, Schelter JM, Burchard J, Kibukawa M, Martin M, Bartz S, Johnson J, Cummins J, Raymond C, Dai H, et al. 2007. Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol Cell Biol* **27:** 2240–2252.

Lund E, Sheets MD, Imboden SB, Dahlberg JE. 2011. Limiting Ago protein restricts RNAi and microRNA biogenesis during early development in *Xenopus laevis*. *Genes Dev* **25:** 1121–1131.

Mateescu B, Batista L, Cardon M, Gruosso T, de Feraudy Y, Mariani O, Nicolas A, Meyniel J-P, Cottu P, Sastre-Garau X, et al. 2011. miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nat Med* **17:** 1627–1635.

Olive V, Bennett MJ, Walker JC, Ma C, Jiang I, Cordon-Cardo C, Li Q-J, Lowe SW, Hannon GJ, He L, et al. 2009. miR-19 is a key oncogenic component of mir-17-92. *Genes Dev* **23:** 2839–2849.

Ott CE, Grünhagen J, Jäger M, Horbelt D, Schwill S, Kallenbach K, Guo G, Manke T, Knaus P, Mundlos S, et al. 2011. MicroRNAs differentially expressed in postnatal aortic development downregulate elastin via 3′ UTR and coding-sequence binding sites. *PLoS ONE* **6:** e16250.

Qin W, Shi Y, Zhao B, Yao C, Jin L, Ma J, Jin Y. 2010. miR-24 regulates apoptosis by targeting the open reading frame (ORF) region of FAF1 in cancer cells. *PLoS ONE* **5:** e9429.

Schnall-Levin M, Zhao Y, Perrimon N, Berger B. 2010. Conserved microRNA targeting in *Drosophila* is as widespread in coding regions as in 3′ UTRs. *Proc Natl Acad Sci* **107:** 15751–15756.

Schnall-Levin M, Rissland OS, Johnston W, Perrimon N, Bartel DP, Berger B. 2011. Unusually effective microRNA targeting within repeat-rich coding regions of mammalian mRNAs. *Genome Res* **21:** 1395–1403.

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

Shi W, Gerster K, Alajez NM, Tsang J, Waldron L, Pintilie M, Hui AB, Sykes J, P'ng C, Miller N, et al. 2011. MicroRNA-301 mediates proliferation and invasion in human breast cancer. *Cancer Res* **71:** 2926–2937.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15:** 1034–1050.

Smrt RD, Szulwach KE, Pfeiffer RL, Li X, Guo W, Pathania M, Teng Z-Q, Luo Y, Peng J, Bordey A, et al. 2010. MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells* **28:** 1060–1070.

Stadler B, Ivanovska I, Mehta K, Song S, Nelson A, Tan Y, Mathieu J, Darby C, Blau CA, Ware C, et al. 2010. Characterization of microRNAs involved in embryonic stem cell states. *Stem Cells Dev* **19:** 935–950.

Sturm M, Hackenberg M, Langenberger D, Frishman D. 2010. TargetSpy: A supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics* **11:** 292.

Suh M-R, Lee Y, Kim JY, Kim S-K, Moon S-H, Lee JY, Cha K-Y, Chung HM, Yoon HS, Moon SY, et al. 2004. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol* **270:** 488–498.

Sun G, Ye P, Murai K, Lang M-F, Li S, Zhang H, Li W, Fu C, Yin J, Wang A, et al. 2011. miR-137 forms a regulatory loop with nuclear receptor TLX and LSD1 in neural stem cells. *Nat Commun* **2:** 529.

Telford NA, Watson AJ, Schultz GA. 1990. Transition from maternal to embryonic control in early mammalian development: A comparison of several species. *Mol Reprod Dev* **26:** 90–100.

Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J, Bronson RT, Crowley D, Stone JR, et al. 2008. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* **132:** 875–886.

Voorhoeve PM, le Sage C, Schrier M, Gillis AJM, Stoop H, Nagel R, Liu Y-P, van Duijse J, Drost J, Griekspoor A, et al. 2006. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* **124:** 1169–1181.

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

Received February 28, 2012; accepted in revised form January 16, 2013.

3.2 Extended results

To measure the effect of miRNA on mRNA and protein levels of its target with CDS-located binding site, a stable HEK293 cell line with inducible miR-124-3p expression was constructed using an episomal pRTS-1 vector. Two conserved CDS target sites (*LPCAT3 and PSEN1*) of miR-124, that showed a relatively high number of reads in AGO2-CLIP [140], were cloned in their native sequence context just before the stop codon of the Renilla luciferase in the psiCHECK-2 vector (WT) (Figure 9B & Figure 11A). As negative controls, constructs with mutations in the miRNA seed complementarity region were also generated (Mut). Mutations were generated such that the amino acid sequence was not changed (Figures 9C & D). Overlap extension PCR (Figure 11B) was used to clone these inserts at the end of the Renilla luciferase-encoding sequence. LPCAT3-psiCHEK-2 and PSEN1-psiCHEK-2 reporter vectors were transiently transfected in the miR124-3p stable cell line. miR-124-3p expression was induced with 1µg/mL DOX, and the expression was confirmed with small RNA Northern blots (Figure 9A). After 24hrs, mRNA and protein levels were measured by qRT-PCR and luciferase assay, respectively, and compared to those of non-induced cells (Figure 10). The mRNA levels of the Firefly luciferase were used an internal control in the qRT-PCR assay. The final readout of the protein levels is the activity of luciferase, which was downregulated in in the cells with WT constructs, but not in the cells with Mut constructs. Changes in the mRNA levels were not observed except for the WT LPCAT3 construct, which was upregulated two-fold. We computed the change in translation from the difference between the change in luciferase activity and the change in mRNA level upon miRNA induction. With both WT constructs translation was repressed, whereas with Mut constructs no translational repression was observed. The primers used for WT and Mut constructs, hsa-miR124-3p, DNA probe for hsa-miR124-3p, and qRT-PCR primers for Renilla and Firefly luciferase are listed in Tables 5 and 6.

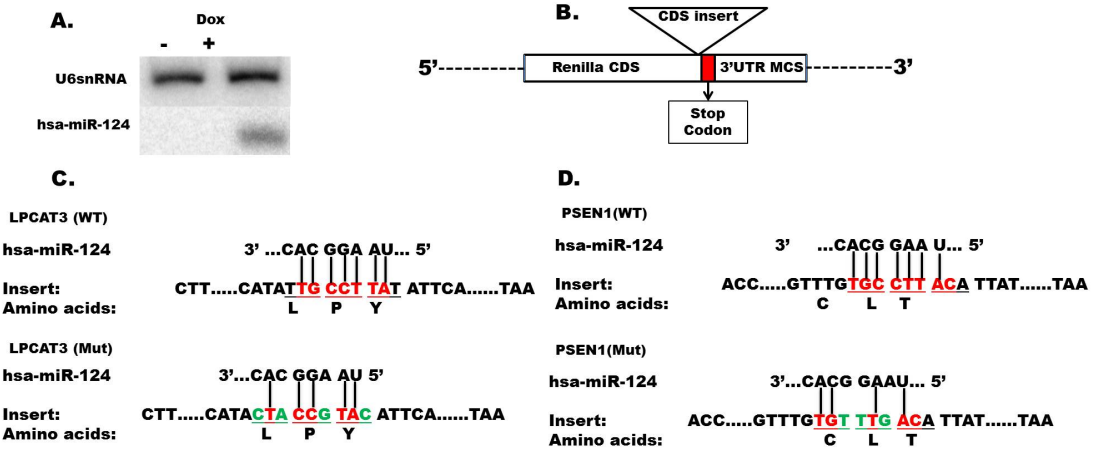


Figure 9: Design of the experiment to measure the effect of miR-124 on protein and mRNA levels. (A) Cells with an episomal hsa-miR-124 expression vector were split in a 24-well plate and induced with DOX

(1µg/mL). After 24 hrs., cells were collected and small RNA Northern blot was performed. The same blot was hybridized with U6 snRNA probe as an internal control. (B) Short inserts containing hsa-miR-124 binding sites from the LPCAT3 or PSEN1 genes were cloned upstream of the STOP codon of the Renilla luciferase. (C) Mutant constructs were generated from the wildtype (WT) sequence of the hsa-miR-124- complementary site in LPCAT3 by mutating 4 positions in the seed complementary region. The mutations were designed such that they did not change the encoded amino acid sequence. (D) Idem for PSEN1.

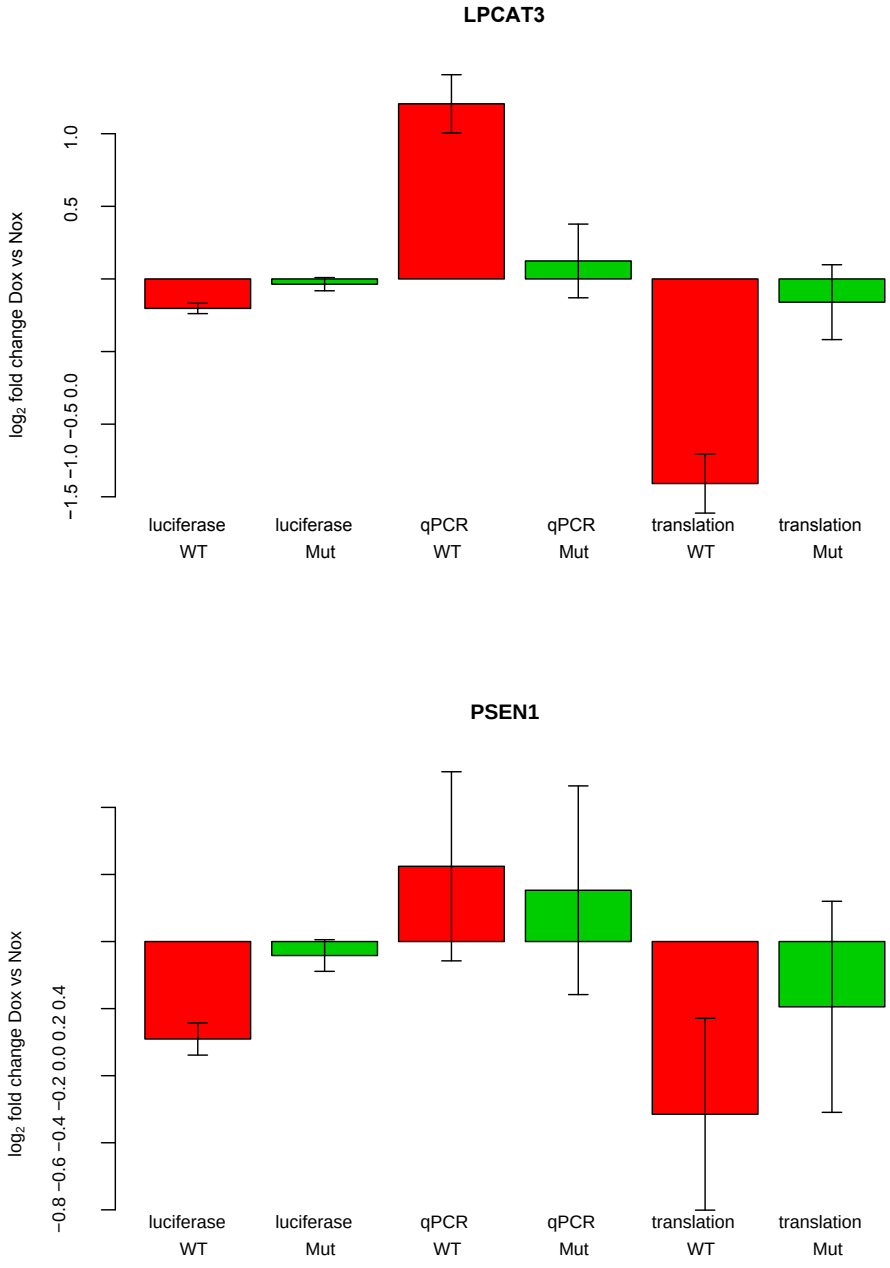


Figure 10: Translation of mRNAs with CDS-located binding sites is significantly repressed following induction of the cognate miRNA. We constructed a stable HEK293 cell line containing an episomal hsa-miR-124 expression vector that can be induced with doxycycline. We also constructed two reporter psiCHECK-2 plasmids in which wildtype (WT) miR-124 binding sites from the LPCAT3 or PSEN1 genes were inserted upstream of the STOP codon of the Renilla luciferase. As negative controls, we constructed two plasmids

with 3-4 point mutations in the binding site (Mut). At time t=0h, cells were transfected with one of the plasmids and doxycycline was added to the medium to induce hsa-miR-124 expression. The Figure shows log2 fold changes in luciferase activity measurements and mRNA levels (quantified by qRT-PCR) 24h post-transfection in DOX-induced cells vs non-induced cells. Changes in translation were estimated from the difference between changes in protein levels and changes in mRNA levels. Error bars represent Standard Errors of the Mean (SEM) from three independent experiments with three technical replicates each. The luciferase activities of cells transfected with WT plasmids were repressed for both constructs ($p < 10-8$, one-sided t-test) while cells transfected with Mut plasmids were not ($p > 0.18$). Furthermore, mRNA levels were unaffected ($p > 0.60$) except in the WT LPCAT3 construct whose expression was upregulated approximately two fold. In both WT constructs, translation appears to be repressed ($p < 10-11$ for LPCAT3, $p = 0.036$ for PSEN1) while repression is lost with both Mut plasmids ($p > 0.26$).

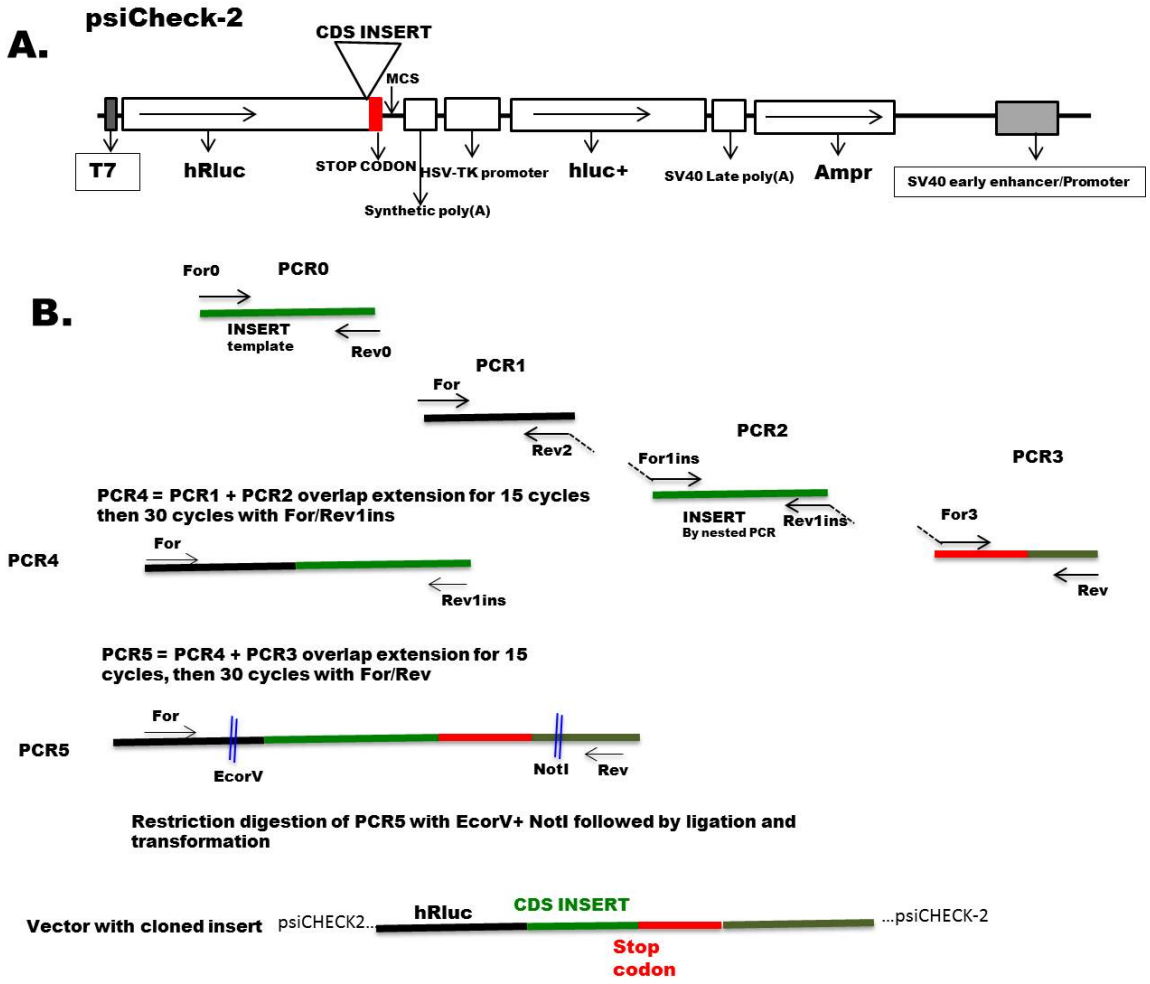


Figure 11: Methodology for cloning CDS-located miRNA binding sites in the Renilla luciferase gene just before stop codon. (A) psiCHECK-2 vector backbone showing its essential elements and the point where the CDS inserts were cloned. (B) Schematic representation of overlap extension PCR method with the primers used. 4 different PCRs were carried out to amplify various fragments from the vector and the CDS insert. Then, in the subsequent two PCRs (PCR4 and PCR5), all the fragments amplified from PCRs1 to 3 were joined, and restriction digestion of the final PCR product and psiCHECK-2 vector followed, ligation of the

digested fragments, transformation and screening of the clones to isolate the one with the insert at the right spot were carried out.

Table 5. List of primers used to amplify the primary-miR-124-3p sequence, which was eventually sub-cloned in pRTS-1 vector. DNA probe used for the hybridization of hsa-miR124-3p on the Northern blot. Primers used for qRT-PCR to detect Renilla luciferase (hRluc) and Firefly luciferase levels (hFluc).

hsa-miR124-3p PCR primers	
Forward (5'-3')	Reverse (5'-3')
AAAAGGCCTCACTGGCC GTGTGCTGTAAATGGCATGG	AAAAGGCCTCACTGGCC GTAGGAGCCTTTCCCTTTCG
hsa-miR124-3p Probe	
GGCATTCACCGCGTGCCTTA	
qRT-PCR primers	
hFluc-RT-for TGCAGAAGATCCTGAACGTG	hFluc-RT-rev CGGTAGACCCAGAGCTGTTC
hRluc-RT-for TCGTCCATGCTGAGAGTGTC	hRluc-RT-rev CTAACCTCGCCCTTCTCCTT

Table 6. The List of primers used in the construction of wild type and mutant reporters with CDS-located miRNA binding sites. A set of 8 primers were used for each construct and order of their use is depicted in Figure 11.

LPCAT3	
For	AGTCCTGGGACGAGTGGCCTGA
Rev	CCCCGAGCCGCCTCCGAATG
For0	TTCATGGGTTACTCCATGACT
Rev0	TGTGAAAAGGGAGACGAGTA
For1 ins	GCGTGCTGAAGAACGAGCAGCTTGGCCACATCTTCTTCCTG
Rev1 ins	GGCTCGAGCGATCGCCTAGAATTATTCCATCTTCTTTAACTTC
For3	GAAGTTAAAGAAGATGGAATAATTCTAGGCGATCGCTCGAGC
Rev2	CAGGAAGAAGATGTGGCCAAGCTGCTCGTTCTTCAGCACGC
LPCAT3 MUTANT	
Mutant for	AGCCTACTATTCATACTACCGTACATTCACAAAGCAATGGTGCC
Mutant Rev	GGCACCATTGCTTTGTGAATGTACGGTAGTATGAATAGTAGGCT
PSEN1	
For	AGTCCTGGGACGAGTGGCCTGA
Rev	CCCCGAGCCGCCTCCGAATG
For0	GTGTTCTGGTTGGTAAAGCC
Rev0	TTGTTAGATGTGGACACAGG
For1 ins	GCGTGCTGAAGAACGAGCAGACCATAGCCTGTTTCGTAGCCA
Rev1 ins	GGCTCGAGCGATCGCCTAGAACTAGATATAAAATTGATGGAA
For3	TTCCATCAATTTTATATCTAGTTCTAGGCGATCGCTCGAGC
Rev2	TGGCTACGAAACAGGCTATGGTCTGCTCGTTCTTCAGCACGC
PSEN1 MUTANT	
Mutant For	TAGCCATATTAATTGGT TTG TGT TTG ACATTATTACTCCTTGCCA
Mutant Rev	TGGCAAGGAGTAATAATGTCAAACACAAACCAATTAATATGGCTA

4. General Methods

The experiments in the above-mentioned studies mainly focused on DOX-inducible, stable miRNA-expressing HEK293T cell lines and luciferase reporter systems. The establishment of inducible miRNA cell lines was based on the pRTS-1 vector [142] (reverse transactivator silencer-1), which is an episomal vector with a size of 18.4kb that harbors a Tet regulatory system. It has a reverse tetracycline controlled transactivator (rtTA2^{S}-M2) and a Tet repressor (tTS^{KRAB}). The presence of both repressor and activator makes this vector more robust with low background and high inducibility with doxycycline. A bi-directional promoter P_{tet}bi-1 allows simultaneous expression of an enhanced green fluorescence protein (eGFP) and a luciferase gene. In the absence of doxycycline, tTS^{KRAB} repressor downregulates the expression of both genes by binding to P_{tet}bi-1, whereas doxycycline relieves this repression and facilitates the binding of rtTA2^{S}-M2, which in turn activates P_{tet}bi-1promoter, leading to the expression of the two genes flanking the promoter. pRTS-1 contains the hygromycinB gene as a mammalian selection marker and ampicillin as a bacterial resistance marker. SfiI restriction enzyme sites flank the luciferase gene, allowing generation of non-cohesive ends [142, 152]. In our pRTS-1 construct, the luciferase cassette was replaced by the pri-miRNA of interest through SfiI restriction sites. The pri-miRNA sequence was initially PCR amplified and cloned into the pGEMT-Easy cloning vector (Promega). The insert was sequenced and subsequently sub-cloned into the pRTS-1 expression vector. The HEK293T cell line was transfected with pRTS-1 construct using liopfectamine 2000 (Invitrogen). 24 hours post-transfection, 100µg/ml of hygromycinB (Calbiochem) was added to allow selective propagation of pRTS-1 positive clones. Several distinct foci of positive clones were grown separately in 24 well culture plates until confluent enough to be seeded in a 6 well plate. These were subsequently transferred to 10cm dishes and maintained in 50µg/ml of hygromycinB as required or cryopreserved. The whole process of establishing a stable cell line took 8-10 weeks. Clones were tested for the miRNA expression by Northern blots upon induction with doxycycline (Sigma-Aldrich) as shown in Figure 4.

On the other side, to study the effect of these miRNAs on its targets, a dual luciferase vector psiCHECK-2 (Promega) was used. psiCHECK-2 vector (Promega) expresses both Firefly and Renilla luciferase from two different loci. miRNA-regulated 3'UTRs of interest were cloned downstream of the Renilla luciferase, while the Firefly luciferase serves as a normalization control. To study the kinetics of miRNA mediated gene regulation, 3' UTR sequences with binding sites for miRNAs of interest were cloned in the multiple cloning site of the psiCHECK-2 reporter vector. The optimal amount of plasmid DNA to be used in the transfection for the luciferase assay was determined**.** Furthermore, a stable HEK293T

cell line has been generated that simultaneously expresses a miRNA and its target using these vector systems. Since the psiCHECK-2 vector lacks a mammalian selection marker, the psiCHECK-2 vector with cloned 3'UTRs was co-transfected with the mammalian selection vector pPUR (Clontech) with puromycin (0.75µg/ml) as a selection marker. This co-transfection was done in stable cell lines of pRTS-1 expressing miRNAs and allowed the psi-CHECK-2 reporter vectors to integrate stably within the HEK293T genome. Selection of clones stably hosting pRTS-1 and pSICHECK-2 was done exactly as described above.

To study the functional properties of CDS-located miRNA binding sites a stable HEK293 cell line with inducible hsa-miR-124-3p was established using the same methodology as described above. The cloning of CDS targets with miRNA binding sites just before the stop codon of Renilla luciferase in psiCHECK-2 vector was based on the overlap extension PCR method (Figure 11). All plasmid DNA transfections were performed with lipofectamine 2000 reagent as per the manufacturer's protocol, whereas the miRNA mimics were reverse transfected with lipofectamine RNAi max as per the manufacturer's protocol. Standard molecular cloning methods were followed [153]. The small RNA Northern was performed as previously described [154], with a minor change: we used the conventional TBE buffer instead of an MOPS- NaOH buffer.

5. General conclusions

In general, miRNAs are essential post-transcriptional regulators, which influence biological processes from cellular differentiation to the development of the various tissues [119-121]. However, the mechanisms through which miRNAs regulate gene expression remain insufficiently understood. So far, focus in the field has been on canonical targets, that match perfectly the miRNA 5' end, but recent technological advances revealed many functional targets with non-canonical sites, which typically show modest changes in their mRNA and protein levels when miRNA expression is perturbed, albeit more modest [155]. A relatively un-appreciated aspect has been that miRNA kinetics is important in the interpreting experimental results, and that the mode of miRNA delivery, transfection of miRNA mimics, microinjections of siRNAs and induction of miRNA expression plays an appreciable role. Proper experimental design is essential, because, as we have shown, the interpretation of the data is not trivial, and should take into account the time scale of miRNA kinetics. Such variations add another layer of complexity in understanding the miRNA mediated gene regulation. Compared to transcription factors, miRNAs were regarded as fast regulators of gene expression. This is because the activity of transcription factors requires additionally protein synthesis [156]. In the study " timescales and

bottlenecks in miRNA dependent gene regulation", using a variety of both high and low-throughput experiments, that involved transfection of miRNA mimics, microinjection of siRNAs and induction of miRNA expression from inducible stable cell lines, we have inferred several parameters of miRNA kinetics and formulated a mathematical model of miRNA dependent gene regulation. Based on the analyses derived from this model, we propose that miRNAs may not be acting so rapidly due to major bottlenecks of the miRNA pathway. From recent studies it appears that the kinetics of miRNA action is mainly dependent on the levels of AGO, which is a limiting factor in cells [157], and as miRNAs function as part of complexes with AGO proteins, displacement of miRNAs from AGO should occur for other miRNAs to come into action. Therefore our analysis indicates that it is the overall AGO loading process that plays a main role in setting the time scale of miRNA-dependent gene regulation. The other bottleneck derived from this model is the rate with which target proteins are degraded. From the analysis of the experimental data in which both levels of mRNA and protein were measured after miRNA transfections, we observed that changes in the mRNA levels were more profound than changes in proteins. This indicates that on the time scale of miRNA activity (miRNAs being typically transfected in these experiments) and the slow rate of protein decay buffers the effect of the miRNA. Thus, the full extent of protein downregulation is not reached before the transfected miRNAs are degraded. The model that we derived will further help in designing experiments to identify miRNA targets and characterize miRNA function.

Studies based on the data generated from high-throughput approaches like CLIP indicate that CDS sites are as many as those of 3'UTRs [138, 140]. Yet the functionality of these sites is still questioned because at least at the mRNA level, their impact seems to be small. In our study "Analysis of CDS-located sites miRNA targets sites suggests that they can effectively inhibit translation", we have analyzed a variety of data sets that included mRNA [131, 132, 158-160] and protein expression profiling [131, 132] after miRNA transfections, and AGO-CLIPs [161] to characterize putative miRNA target sites in CDS regions. We have employed the ElMMo model [126], which was described before and is considered to be one of the most accurate methods for predicting miRNA target sites to identify putative target sites in both CDS and 3'UTR regions. A subsequent comparative analysis revealed that both 3'UTR and CDS sites are under evolutionary selection pressure, and the properties that were found to characterize functional 3'UTR sites also characterized the functional CDS sites. Furthermore, target sites in CDS and 3'UTR regions possess similar properties. In particular, they are both located preferentially in U-rich and structurally accessible environments characterize functional miRNA sites irrespective of their location in 3'UTR or CDS. We also proposed that several miRNA families, including those expressed in embryonic context, induce strong effects by acting both on 3'UTRs and

CDS sites. Among these are miRNA families that are expressed highly in embryonic cells. Our comparative analysis based on the number of genes containing high-scoring 3'UTR and CDS sites revealed a significant number of genes that were co-targeted in 3'UTR and CDS regions suggesting that miRNAs can functionally co-target 3'UTR and CDS sites. Upon analysis of mRNA and ribosome profiling data, for the first time our analysis revealed that CDS-located miRNA binding sites are more effective in inducing translational repression than the sites in 3'UTRs. We validated this observation with other data sets as well as using psiCHECK2 dual luciferase reporter system. Although more work is needed to understand the role that the position of miRNA-complementary sites has in their function, our study was the first to show that CDS sites may be predominantly involved in regulating translation rates.

Chapter 2

6. Tissue specific splicing regulators promote somatic cell reprogramming

6.1 Introduction

Mammalian development starts with fertilization that involves the fusion of two transcriptionally silent and morphologically distinct gametes to form a totipotent zygote. The term "totipotent" denotes the competence to give rise to an entire organism including the extra-embryonic tissues. The single cell zygote undergoes several divisions to form a 8-16-cell morula [162, 163]. It is at this stage when the first lineage restriction occurs: the morula compacts, polarizes and transforms into a blastocyst that consists of two distinct cell populations: the trophectoderm (TE), an outer layer of epithelium surrounding the blastocyst, and the inner cell mass (ICM). Just before the implantation of the embryo, the ICM undergoes another lineage segregation giving rise to a primitive endoderm (PE) cells and epiblast cells. The PE separates the fluid-filled cavity called blastocoel from epiblast, which forms as a cluster of cells at one side of the blastocyst. Cells of this ICM-derived cluster are competent to give rise to cells of all three germ layers (mesoderm, ectoderm and endoderm) and hence called pluripotent or embryonic stem cell (ESC)s [164].

The lineage segregation during the formation of blastocyst arises from robust transcriptional regulatory events, which are themselves due to specific transcription factors (TFs). Both *in vivo* and in vitro studies showed that the octamer binding transcription factor 4 (OCT4, encoded by *Pou5f1* gene) is essential in the maintenance of ESC pluripotency, whereas the caudal type homeobox protein 2 (CDX2) is essential for the maintenance of the TE cell lineages [165-167]. Just before the formation of the blastocyst, these two TFs act upon each other through a positive feedback mechanism whereby the expression of CDX2 terminates the expression of OCT4 and vice versa. From gene expression and lineage-tracing experiments it has become apparent that early ICM cells are in fact two distinct population, expressing the homeobox TF NANOG and GATA binding factor 6 (GATA6) [168], respectively. GATA6-expressing cells form PE [169] and NANOG acts to maintain the pluripotency of ICM [170]. Studies of the TEA-domain family member 4 (TEAD4) null mutant embryos have shown that Tead4 TF is associated with upstream mechanisms driving CDX2 expression [171, 172]. A schematic representation of the early

lineage segregations during the blastocyst formation and origin of ESCs is shown in Figure 12.

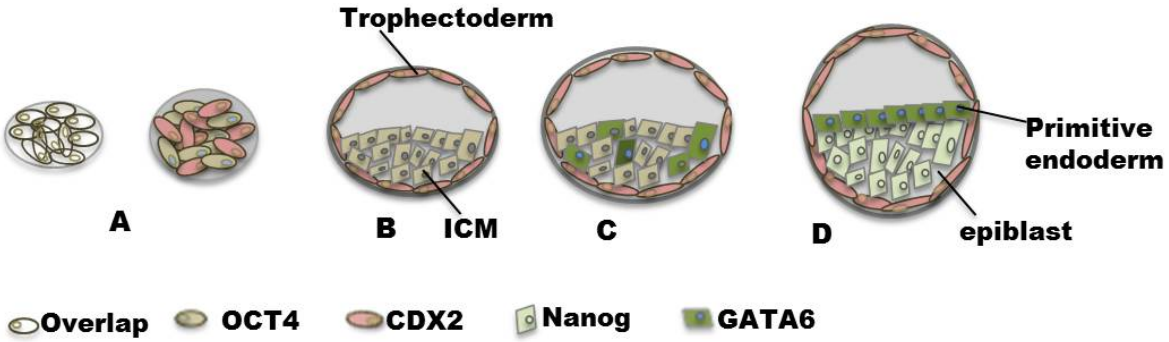


Figure 12: Early lineage differentiation during the embryogenesis under the control of transcription factors. (A) During the 8 to 16 cell morula stage of the embryogenesis, the overlap of the transcription factors CDX2 and OCT4 is seen, (B) which later together determine the first lineage restriction of the cells into either trophectoderm or inner cell mass respectively. This is followed by the formation of the blastocyst with trophectoderm forming the outer layer of cells and ICM residing at one side of the blastocyst. (C) During the early blastocyst, a second lineage restriction results in the formation of epiblast and primitive endoderm, with the activation of NANOG and GATA6 transcription factors. (D) During the late blastocyst stage, GATA6 positive cells form the surface layer of the ICM called primitive endoderm and NANOG cells facilitate in the maintenance of pluripotent epiblast cells. Concept of the image is adapted from Arnold et al. [164].

6.1.1 Pluripotent stem cells - Origins and properties

It was in the year 1970 that a striking observation was reported from two different groups. Namely, that early mouse embryos, when grafted in extrauterine or ectopic sites such as under kidney produced malignant teratocarcinomas or benign teratomas [173, 174]. Teratocarcinomas are multidifferentiated tumors with substantial populations of undifferentiated stem cells called embryonal carcinoma (EC) cells. Spontaneous occurrence of teratocarcinomas originating from germ cells, such as testicular teratocarcinoma was much more common in strain 129 mice.

When isolated clonally and propagated in cultures, EC cells retained their differentiation potential, and they were also involved in the formation of embryos when implanted in the ICM of the blastocyst, being capable of producing the derivatives of ectoderm, mesoderm and endoderm [60]. However, because EC cells showed some chromosomal aberrations, possessed poor differential potential or could only differentiate under certain circumstances and, most importantly, due to their incompetence to undergo meiosis to produce mature gametes, they were less amenable for experimental analyses [175].

Eventually, by culturing EC cells on feeders of mitotically inactivated mouse embryonic fibroblasts (MEF), researchers were successful in maintain these cells in culture for longer time periods [176]. It was in the year 1981 when Evans and Martin, with different approaches, succeeded in establishing ESC lines directly from the ICM of the early blastocyst without the need to produce teratocarcinomas [177, 178].

Evans's laboratory further demonstrated that ESCs are more consistent and competent in producing viable mice when injected in the blastocysts [179]. ESCs are also able to differentiate into various somatic cell types in vitro [180] and Nagy and colleagues showed in 1991 with a tetraploid complementation assay that ESCs alone are sufficient to generate entire fetus. In this assay ESCs are injected in a tetraploid embryo, in which tetraploid cells only contribute to the extra embryonic tissue, whereas ESCs contribute to the development of the embryo eventually giving rise to a viable mouse [181, 182]. The co-culture of ESCs on feeder layers was long considered essential for indefinite expansion and for the maintenance of the undifferentiated phenotype. However, it was later shown that feeders could be replaced by the addition of a single cytokine, leukemia inhibiting factor (LIF) in the media. LIF is a member of interleukin (IL)-6 family of cytokines that acts through a membrane-bound gp130 receptor to activate a downstream signaling process, in which tyrosine kinase janus (JAK) activates and recruits the latent transcription factor signal transducer and activator of transcription3 (STAT3) [183]. Other molecular signals triggered by LIF lead to the activation of two other signaling pathways: the phosphatidylinositol 3-kinase (PI3K) mediated pathway and the mitogen activated protein kinase (MAPK) pathway. These converge on the activation of the core pluripotency regulating TFs OCT4, SOX2 and NANOG through different mechanisms. STAT3 activates Kruppel like factor 4 (KLF4), a positive regulator of *Sox2* gene and the concurrent activation of PI3K pathway lead to the activation of a gene, which encodes a TF TBX3, which further activates the *Nanog* gene. The MAPK pathway activated by LIF-signaling is known to act on *Tbx3* and *Nanog* genes [184]. The serum that is required during ES cell derivation and clonal expansions can be replaced by a growth factor called bone morphogenetic protein (BMP)4, which is functional in the presence of LIF [185] . Cell surface markers such as stage specific embryonic antigen (SSEA)-1 in mouse or SSEA4 in human ESCs, membrane bound receptor gp130, as well as enzymatic activities such as alkaline phosphatase (AP) and telomerase, define the molecular basis of ESCs [186]. In vitro culture of mouse primordial germ cells in presence of fibroblast growth factor (FGF)2, lead to the conversion of these cells into cells that were largely indistinguishable from ESCs known as embryonic germ (EG)cells. One important characteristic of EG cells is that they retain the capacity of parental germ cells to erase imprints [187].

Induction of pluripotency in somatic cells constituted a big landmark in both understanding the principles of pluripotency and the use of this technique towards therapeutics. Somatic cell reprogramming was first achieved through a technique called somatic cell nuclear transfer (SCNT). The technique, introduced by John Gurdon in 1962, involved the transfer of an adult nucleus from an intestinal cell to an enucleated oocyte. This generated a swimming tadpole, although it was not clear that the organism indeed developed from the adult nucleus [188]. This was proven later, by the transfer of the nucleus from an antibody producing lymphocyte [189]. This technique was successfully used by Ian Wilmut to clone a sheep in 1997 [190]. Other landmark discoveries of reprogramming were generated by the work on Drosophila, wherein overexpression of the Antennapedia (ANTP) TF induced the formation of legs instead of antennae [191]. In the same year, another work showed the transformation of fibroblasts to myoblasts when a muscle specific TF MyoD was overexpressed in fibroblasts [5]. A more recent and practical method of inducing pluripotency was introduced by Shinya Yamanaka in the year 2006 and is considered to be a groundbreaking discovery of the recent times [4]. The method involves ectopic expression of four embryonic stem cell transcription factors (OCT4, SOX2, KLF4 and c-MYC) in somatic cells. This reprogramming method was first employed in mouse fibroblasts, thus reprogrammed cells resembled mouse ESCs. They were called induced pluripotent stem cell (iPSC)s. A year later Yamanaka and colleagues reported the same phenomenon in human somatic cells [192]. At the same time James Thomson came up with a different set of factors (OCT4, SOX2 NANOG and LIN28), which were also able to induce pluripotency in human dermal fibroblasts [193].

6.1.2 Induced pluripotent stem cells

Based on the previous studies based on SCNT, Yamanaka hypothesized that it is a combination of transcriptional regulators that is expressed in oocytes or ESCs, and is responsible for the reprogramming of somatic nuclei or somatic cells. Searching for such combinations, Yamanaka and colleagues ended with a set of four embryo-specific TFs: OCT4, SOX2, KLF4 and c-MYC. This combination is now known as OSKM or Yamanaka cocktail. Overexpression of these TFs reprogrammed mouse embryonic and adult fibroblasts to cells that resembled ESCs and were called induced pluripotent stem cell (iPSC)s [4, 192, 193]. Compared to other approaches for obtaining pluripotent cells, the iPSC approach is relatively simple and reproducible. However, its efficiency is relatively low. iPSCs are also not hampered by the same ethical barrier as embryonic tissues, enabling studies of developmental processes and holds great potential for translational research. At the beginning, iPSC induction involved retroviruses or lentiviruses. This is a problem for translational applications because there is the risk of insertional mutagenesis. Nonetheless, the approach can be used in studies that will lead to a thorough

understanding of the mechanisms underlying pluripotency induction, which in turn will enable the development of safer and more efficient methods. This part of the thesis describes my work towards understanding the role of splicing factors and miRNAs in increasing the efficiency of pluripotency induction in somatic cells. I will first review the current state of knowledge concerning the molecular mechanism of pluripotency induction.

6.1.3 Transcription factors OCT4, SOX2, KLF4, MYC, and NANOG

The TFs OCT4, SOX2 and NANOG play a central role in sustaining the pluripotency of both human ESCs (hESCs) and mouse ESCs (mESCs) cells. They regulate transcription of a broad spectrum of pluripotency-related target genes and by auto regulating their own expression. The genomic organization of *Oct4*, *Sox2* and *Nanog* and the structural domains of these proteins are conserved between mouse and human. OCT4, encoded by *Pou5f1* gene, is a homeodomain transcriptional regulator of the POU (Pit, Oct, Unc) family that regulates expression of its targets by binding to an octameric sequence motif "ATGCAAAT". The OCT4 protein consists of three domains: N-terminal domain, POU domain, and a C-terminal domain. The POU domain consists of two structurally independent subdomains, an amino terminal POU specific (POU_S) domain and a carboxyl terminal homeodomain (POU_{HD}), which both bind to DNA in a helix-turn-helix structure connected by a linker [194]. Inhibition of OCT4 expression by RNAi in both human and mouse ESCs resulted in the differentiation toward trophoblast and endoderm cells [195, 196]. Although expression of OCT4 is important in the lineage commitment of epiblast, its expression is not sufficient for the maintenance of pluripotency phenotype of ESCs. Rather, it has been shown that ESCs expressing constitutive *Oct4* are dependent on LIF and tend to differentiate as normal ESCs do [167], thus suggesting that there are other intrinsic and extrinsic factors required in the maintenance of pluripotency.

Along with OCT4, the Sex determining region Y-box2 (SOX2) TF is considered a defining factor for ESCs. SOX2 contains the DNA-binding high mobility group (HMG) domain. SOX2 mutant embryos die at the implantation stage, and a normal expression of *SOX2* gene is necessary for the self-renewal of both human and mice ESCs. Knockdown or overexpression of *SOX2* results in hESC differentiation [197]. NANOG is also one of the TFs to play an important role in maintaining ESC pluripotency. NANOG is a homeodomain protein, whose expression is one of the characteristic markers of ESCs. NANOG has several functional elements, which help in transactivation: the centrally located homeodomain, and the three C-terminal elements C-terminal domains (CD)1, CD2 and a tryptophan (W) repeat [198]. Overexpression of *Nanog* enables ESCs to maintain their ES cell phenotype even in the absence of LIF. It is the W-repeat localized in between CD1 and CD2 involved

in NANOG protein dimerization, which helps in the maintenance of ESCs pluripotency in the absence of LIF [199]. The CD2 transactivation element is necessary for the maintenance of self-renewal property of ESCs [200].

Kruppel like factor 4(KLF4), a member of the Kruppel-like family of zinc finger transcription factors, is expressed in a variety of tissues and plays an important role in processes like proliferation, terminal differentiation, apoptosis and development. KLF4 can act as an activator or repressor of gene expression, depending on the cellular context, and it acts as well as an oncogene or tumor suppressor [201]. The expression of Klf4 is high in undifferentiated ESCs, and decreases dramatically during differentiation. Because over expression of Klf4 is sufficient for ESC to maintain their undifferentiated phenotype in the absence of LIF, Klf4 is considered a direct target of LIF [202]. Similar to KLF4, the Myelocytomatosis oncogene (c-MYC) also has context-dependent, acting as an activator or repressor of the transcription and playing important role in proliferation, cell growth, differentiation and apoptosis [203]. Knockout of c-MYC leads to malformation of different organs and reduced embryo size [204]. Although it was shown that KLF4 and c-MYC were important in the self-renewal of ESCs, they do not seem to be uniformly required. For example KLF4 could be replaced with ESSRB, an orphan nuclear receptor, in the reprogramming of mouse embryonic fibroblast [205] and NANOG and LIN28 were used along with OCT4 and SOX2 to reprogram human fibroblast to iPS cells [193].

The transcription factor trio OCT4, SOX2, and NANOG is the core pluripotency network, which maintains the expression of a broad spectrum of targets including TFs, genes involved in signaling pathways and miRNAs responsible for self-renewal and pluripotency. To identify these targets, a recent study used biotinylated chromatin immunopreciptation (ChIP) to identify target genes whose promoters were occupied by more than one TF. This study also concluded that the genes with binding sites for 5 different TFs have a tendency to be active and the genes bound by only few TFs are repressed in ESCs [206]. In addition, OCT4 and SOX2 have been shown to act in conjunction on several targets that are involved in ES cell pluripotency. OCT4 binds to SOX2 to form a heterodimer that activates *Sox2, Pou5f1* and *Nanog* transcription, while NANOG regulates the transcription of *Pou5f1* and *Sox2* genes [207-210]. In another key study using ChiP-sequencing, mapping of 13 different TF's binding sites involved in ESC pluripotency maintenance and somatic cell reprogramming revealed two major clusters: OCT4-SOX2-NANOG cluster and MYC-specific cluster, wherein each TF from one cluster occupied their own and each other's genomic targets forming a complex regulatory network. For instance OCT4-SOX2-NANOG co-occupy regulatory elements in the *Stat3* and *Smad1* genes, which are the downstream targets of LIF and BMP4 signaling. The latter two factors regulate in turn *Oct4*, *Sox2* and *Nanog*. Several groups used both expression and TF footprinting to define the interactions

that are important for pluripotency. For example, by knocking down the expression of *POU5F1*, generating microarray data, and comparing them with already available knockdown of *POU5F1* data, a study revealed 137 orthologous genes which were downregulated in both human and mouse ESCs, with about half of these targets being conserved and possessing binding sites for either OCT4 or NANOG [211]. Through ChIP coupled with DNA microarrays a study revealed that OCT4-SOX2-NANOG form feed-forward loops which include at least 353-protein coding and 2 miRNA genes. Feed-forward loops are gene expression regulator networks with a regulator controlling the expression of another regulator, as well as of targets that are jointly targeted by the tow regulators. When the two regulators have the same effect on their common targets, the loops are called consistent, while when the regulators have opposite effect, the loops are called incoherent [210]. OCT4-SOX2-NANOG TFs form pluripotency regulatory networks with miRNAs by occupying the promoters of 55 miRNA transcription units, including three polycistronic clusters [212].

6.1.4 Methods to generate iPSCs

In their initial iPSC study [4], Takahashi and Yamanaka found an efficiency of 0.02% of iPSC formation, 14-21 days post transduction with retroviruses individually expressing OSKM. In the following year, Yamanaka's lab reported a similar efficiency of 0.02% after 30 days post transduction in human dermal fibroblasts [192]. James Thomson and colleagues reported similar efficiencies of 0.03-0.05 in fetal fibroblasts and 0.01% in the newborn foreskin fibroblasts at 20 days post-lentiviral transduction with a different set of genes, *OCT4 SOX2*, *NANOG* and *LIN28* [193].

One of the major concerns from the initial experimental designs was the use of four individual viral vectors that may have lead to suboptimal results. Thus, a single reprogramming cassette containing the four reprogramming factors separated by self-cleaving peptides was introduced [213]. To overcome the problem of stable viral integration in the host genome, a Cre-Lox system was used. In this system the reprogramming factors are flanked by loxP sites, which lead, upon Cre expression, to the excision of the transgene cassette. This method was first successfully used in cells from patients with the Parkinson's disease, at an efficiency of 0.5% [214]. Other approaches included the use of adenoviruses, which are non-integrating, but this method further dropped the efficiency of the technique to 0.001-0.0001% in mouse and to 0.0002% in human cells [215]. RNA viruses such as the Sendai virus, which does not enter the nucleus, were also used at an efficiency of 1%. However, but this approach has the disadvantage of lengthy passaging (10 passages) needed to dilute the viruses in cells, Furthermore cells have to be grown at high temperatures [216]. Use of reprogramming factors as purified

proteins was also reported but with an efficiency of 0.006% in mouse and 0.001% in human. This was technically challenging, because it required purification of large amounts of bioactive proteins, able to cross the plasma membrane [217]. Other reprogramming methods that do not involve viruses used instead episomal plasmids [218] and mini circles [219] but they were both reported to have relatively low efficiency, ranging from 0.001 to 0.005%.

With mature mRNAs of the reprogramming factors in human fibroblasts, an efficiency of 1.4% was achieved. The further addition of LIN28 in presence of valproic acid (VPA) gave an efficiency of 4.4% but the method was considered to be labor intensive [220]. PiggyBac vectors, which can be integrated and excised with the expression of transposase enzyme and doxycycline inducible promoters, were also proposed [221]. Small molecules, which act on different signaling pathways involved in the reprogramming process or epigenetic modifications, were also used to enhance the efficiency of the method. For example, VPA and sodium butyrate inhibit the activity of histone deacetylase, while vitamin C enhances the activity of epigenetic modifiers such as histone demethylases Jhdm1a/1b [222]. Addition of several embryo-specific miRNAs to OSKM factors enhanced reprogramming 10 to 15 fold. Mimics of miR-302b and miR-372 were transfected along with lentiviruses expressing OSKM to enhance reprogramming efficiency in human fibroblast cells [223]. Recently, it has also been reported that the miR302/367 cluster is sufficient to reprogram the fibroblasts without the addition of OSKM factors. This study further claimed an increase in efficiency by two orders of magnitude [224].

One of the major concerns in these experiments is that large variations are expected, because the expression of the reprogramming factors is difficult to control. Thus, to facilitate studies in the field, secondary systems were developed. The secondary reprogramming systems evolved from individual OSKM vectors [225] to polycistronic OSKM vectors [213]. Rudolf Jaenisch's group pioneered this type of systems and proposed various approaches to generate them. In this type of systems, a polycistronic-reprogramming cassette with a doxycycline-inducible promoter is introduced by gene targeting in ESCs that constitutively express other components required for the induction such as rtTA (reverse tetracycline transactivators) or iPSCs are obtained with inducible viral vector systems. The transgenic ESCs or iPSCs thus generated, with integrated inducible OSKM cassettes, are further used to obtain chimeric animals from which secondary somatic cells are isolated. Upon the addition of doxycycline, reprogramming factors are expressed in the transgenic cells to eventually reprogram these cells [213]. A recent study published a new transgenic mouse line wherein a doxycycline inducible Cre-recombinase and inducible polycistronic OSKM transgene is inserted in Rosa26 locus [226]. Use of such systems strongly reduced heterogeneity and lead to 25- 50-fold

increase in the efficiency of pluripotency induction compared to primary reprogramming systems [227].

6.1.5 Phases of somatic cell reprogramming

Reprogramming is a dedifferentiation process, in which the ectopic expression of OSKM in the right stoichiometry leads to changes in the transcriptional and epigenetic status of the somatic cells and eventual acquisition of the undifferentiated state and self-renewing ability. Studies of gene expression along the time course of reprogramming revealed three phases of reprogramming and two major transcriptional waves [228, 229]. The first transcriptional wave is driven by c-MYC/KLF4 and the second transcriptional wave is driven by OCT4/SOX2/KLF4. Cells that do not reprogram fail to initiate second transcriptional wave, but an elevated expression of OSKM can direct these cells towards pluripotency. In the initial transcriptional wave the bivalent domains of gene expression are also gradually established [229]. Studies of the initial transcriptional phase revealed that cells assume stochastically one of the possible cell fates: apoptosis, senescence, transdifferentiation or reprograming [230]. The cells that initiated reprogramming undergo increased proliferation, changes in histone marks, activation of DNA repair, and initiate mesenchymal to epithelial transition (MET). OCT4 downregulates the epithelial to mesenchymal transition (EMT) regulator Snai1/2 via TGFβ2 and TGFβR3, whereas OCT4 and SOX2 activate the miR-200 cluster of miRNAs, which in turn represses another EMT regulator Zeb2 subsequently facilitating MET. The MET process is assessed based on the expression of cadherin 1 (*Cdh1*) and epithelial cell adhesion molecule (*Epcam*) [231]. Besides MET, ESC-like proliferation is also acquired during the early phase of reprogramming [232]. In general, c-MYC enhances the transcription of proliferation-related genes, and a recent study showed that c-MYC acts as a general amplifier of the gene expression contributing to the first transcriptional wave [233]. On the other hand KLF4 is known to assist c-MYC in enhancing transcription by jointly binding to promoter regions [228]. OCT4 and SOX2 are mostly known to bind to distal elements of both active and repressed genes. OSK factors also known to open chromatin regions to activate the genes responsible for the maintenance of pluripotent state whereas c-MYC only facilitates this process [234]. To look for further changes occurring in the early phase of reprograming, several groups also proposed to follow the loss of cell surface markers such Thy1 and CD44 and gain of SSEA1 and alkaline phosphatase [229, 235].

The second phase of reprogramming is less understood, but it is believed that during this phase, pluripotency related genes are activated to facilitate the activation of core pluripotency network. KLF4 assists OCT4 and SOX2 in the activation of glycolysis and transient activation of developmental genes [229]. This phase has less transcriptional

changes, but some of the pluripotency genes detected during this phase are *Fbxo15, Sall4, endogenous Oct4,* embryonic cell transcription factor (*Utf1),* followed by *Nanog and* oestrogen related receptor beta (*Essrb*) [228]. In general single cell transcriptomics and other techniques like clonal analysis indicate that the pluripotency associated genes are activated in a sequential way, some during the second phase as mentioned above, and others like Sox2 and surface marker ICAM1 in the late, maturation phase [230, 235]. At the end of this phase, expression of *Sox2* and developmental pluripotency associated 2 (*Dppa2*) genes are stabilized and changes in the DNA methylation occurs [229, 236]. After the activation of Sox2, the events involved are more deterministic. Other major changes at the end of this phase include the silencing of the transgenes. The cells then transit into the final phase, with little dependence on the transgene expression [236].

The final stabilization phase denotes changes in iPSCs after the acquisition of pluripotency. The iPSCs are then further expanded and tested by blastocyst injection or other in vitro differentiation assays. Stabilization also involves epigenetic changes such as telomere elongation and resetting of the epigenetic memory, but some bias towards the somatic cell type of origin remains [237, 238]. In support of these different phases of reprogramming, a quantitative proteomics analysis of fibroblast reprogramming revealed that proteome changes occur mainly at early 3 days and last 3 days of reprogramming. In the early phase, proteins related to cell cycle, metabolism, RNA processing and chromatin organization were strongly induced, whereas the EMT-associated protein were oppositely regulated between early and late phases of reprogramming [239]. Figure 13 shows the representation of the different phases observed during somatic cell reprogramming.

6.1.6 Epigenetic changes during reprogramming

During reprogramming, the epigenome of the cells has to change as well to reach a state similar to that characteristic of ES cells. Epigenetic changes occur at the level of chromatin remodeling, histone modifications and DNA methylation. Chromatin structure plays an important role in the establishment of specific expression patterns responsible for a specific cellular identity. Essential structural changes in chromatin state make biding sites accessible for OSK factors, which act in a cooperative manner, binding of one factor destabilizing nucleosomes, thereby allowing other transcription factor to access sites which were hindered earlier [240]. Although OSK factors individually can target sites in the closed chromatin as well, the ability of c-MYC to access sites in closed chromatin is dependent on OSK occupancy. In turn, MYC facilitates the binding of OSK to the target sites present in closed chromatin [234]. Depletion of Mbd3, a core member of the Mbd3/NuRD (nucleosome remodeling and deacetylation) complex in concert with OSKM

transduction results in synchronized and deterministic reprogramming of both mouse and human cells [241].

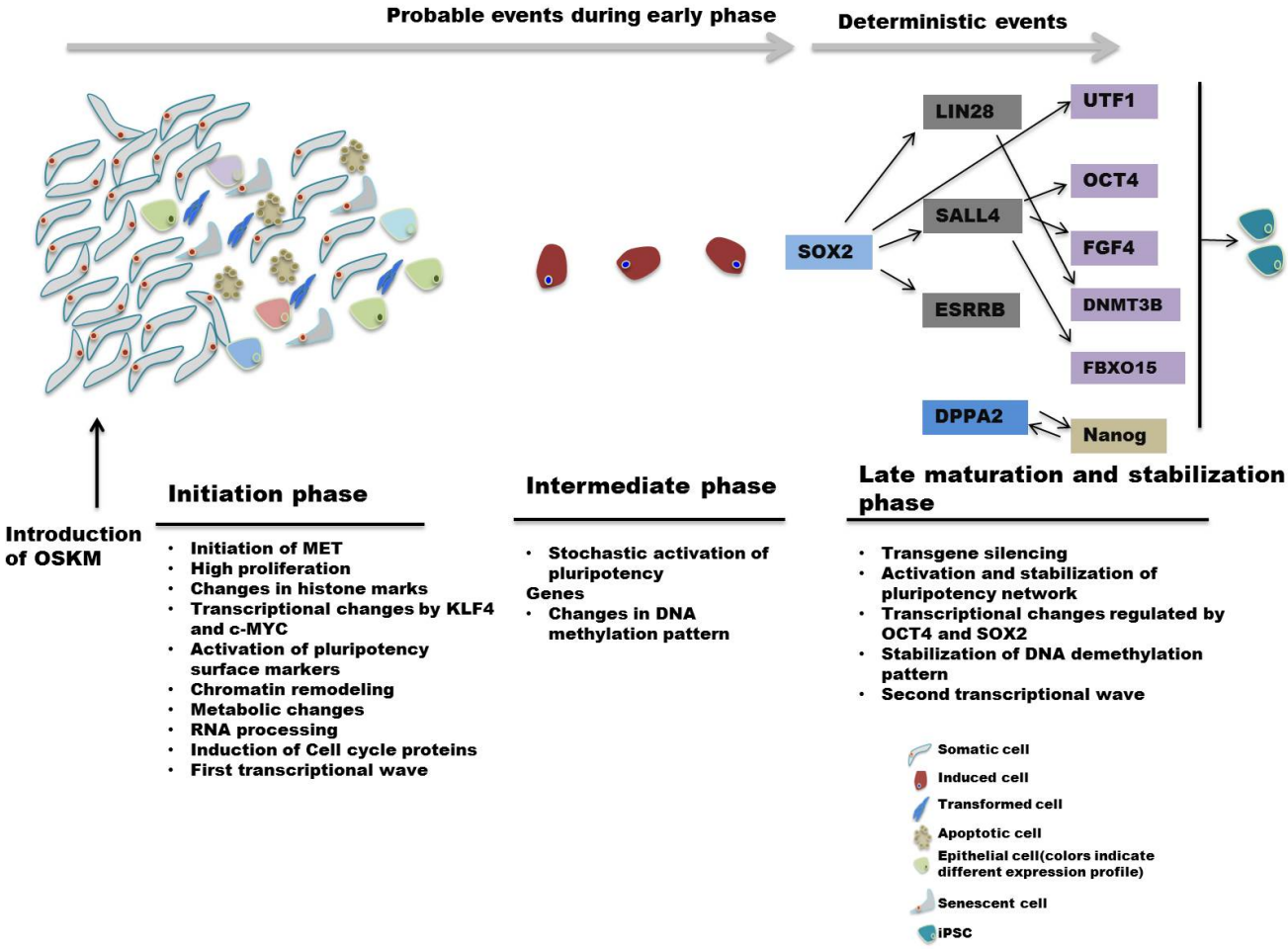


Figure 13: Phases of somatic cell reprogramming. In general, major changes during reprogramming are observed during the initiation phase where a first transcriptional wave occurs and at the late maturation phase, with a second transcriptional wave. In the initiation phase, high proliferation of cells, accompanied by the activation of mesenchymal to epithelial transition occurs. Changes in the histone marks of the differentiation genes begins, which prepare them to become silent, and *de novo* changes in the chromatin marks of the pluripotency genes, prepares them to get activated during the later phases of reprogramming. Upon the overexpression of reprogramming genes, stochastic activation of gene expression leads a cell to adapt any of the various phenotypes such as apoptotic, senescent, transformed, and induced towards epithelial phenotype. Several pluripotency specific surface markers such as SSEA1 and CD44 are also detected in the early phase. The cells that engage on the reprogramming path pass through a long latent intermediate phase, wherein the stochastic activation of pluripotency and developmental genes occurs. In the late maturation phase, more deterministic changes occur in the pluripotency network after the activation of Sox2 and developmental gene DPPA2. These lead to the activation of the endogenous core pluripotency network. Changes in the chromatin marks of pluripotency genes facilitate the transcriptional activation of these genes. In the last phase, expression of pluripotency genes is stabilized, the epigenetic status is reset to the pluripotent state which can be maintained independent of transgene expression. Concept of the image is adapted from Buganim et al. [242].

Covalent modifications of histone proteins also alter the chromatin structure. Various histone-modifying enzymes such as histone methyltransferases, histone acetyltransferases, histone demethylases and histone deacetylases regulate different histone marks. The expression changes in the early phase of reprogramming are confined to genes with active chromatin marks such as H3K4me3, and H3K4me2. These chromatin marks appears *de novo* during early phase in many promoter regions of the pluripotency genes such as *Sall4* and fibroblast growth factor 4 (*Fgf4),* which are enriched for OCT4 and SOX2 binding sites [243]. The gain of H3K4me2 mark occurs on the nucleosome that covers the transcriptional start site, blocking the assembly of the transcriptional machinery. The depletion of the nucleosome after the acquisition of the active mark allows the transactivation of these genes later in reprogramming [234]. Chromatin changes at enhancers are considered to be more prominent in the early stages of the reprogramming and the active marks H3K4me1/2 and H3K27Ac surround both the promoters and enhancers of the pluripotency related genes [243]. H3K4 methylation is catalyzed by a complex involving the histone methyltransferase Set/M11, which is a homologue of Trithorax group (TrxG) proteins [244]. Removal H3K4 methylation perturbs reprogramming, whereas knockdown of a demethylase Kdm5b, which is responsible for the demethylation of active H3K4me1/2/3 enhances reprogramming efficiency [245]. Other active histone marks include H3K36me1/2, and H3K79. Global reduction of these marks also enhances reprogramming and a study showed that overexpression of the H3K36me2 demethylase Jhdm1b along with *OCT4* is sufficient to reprogram somatic cells [222]. H3K27me3 is associated with gene silencing, but when co-localized with H3K4me3, it marks bivalent domains. Bivalent domains are observed at elevated levels in the pluripotent stem cells and the genes with these domains can be quickly turned on when H3K27me3 is erased, whereas can be silenced upon the erasure of H3K4me3. Downregulation of somatic genes during early stages of reprogramming is mediated by the Polycomb repressive complex (PRC)2 that trimethylates H3K27. Loss of PRC2 components diminishes reprogramming efficiency [230]. H3K9 is another repressive chromatin mark that is associated with reprogramming via PRCs. Knockdown of H3K9 methyltransferases such Ehmt2, Setdb1, Suv39h1/2 enhances reprogramming efficiency [234].

DNA methylation patterns also restrict early reprogramming. DNA demethylation starts in the second phase and continues until the stabilization phase of reprogramming. DNA demethylation occurs by two mechanisms: passive demethylation takes place during DNA replication and active demethylation is mediated by the oxidative activity of ten eleven translocation methylcytosine dioxygenase (TET) enzymes. Tet1/2/3 enzymes catalyze the hydroxylation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), which is

converted to unmethylated cytosine by base-excision repair [246]. DNA demethylation leads to open chromatin regions, which are further accessed by TFs binding to the promoters of the genes involved in pluripotency. In the early phase of reprogramming Tet1 induces DNA hydroxylation of pluripotency related genes, leading to their transcriptional activation at the later phases of reprogramming [229, 243]. Tet1/2, when overexpressed with *Nanog,* promotes iPSC formation, while knockdown of DNA methyltransferase (DNMT)3a promotes human iPSC formation [247], suggesting that erasure of DNA methylation promotes the expression of pluripotency genes, which were silenced through hypermethylation by DNA methyltransferases in somatic cells.

6.1.7 miroRNAs and reprogramming

The importance of miRNAs during development was underscored by observations that knock out of genes encoding proteins that are necessary for miRNA biogenesis leads to the embryonic lethality in mice [248]. Knock out of these factors in ESCs makes the cells less able to proliferate and upon induction, they cannot differentiate [119]. Small RNA sequencing and microarrays studies revealed that several microRNA families are differentially expressed in human ESCs compared to differentiated cells. The microRNA families that are highly expressed in human ESCs include miR-302, miR-106, miR-372, miR17, miR-200, and miR-195 families [249, 250]. These miRNA families have homologues in mice with the same nomenclature except for the miR-372 family, which corresponds to the miR-290 family in mice. In human ESCs the let-7 family is expressed at low level and is upregulated upon differentiation [250]. miRNAs can either silence or stabilize the pluripotency genes of the ESCs. For example, miR-134, miR-296, and miR-470 can repress the core pluripotency genes *Oct4*, *Sox2*, and *Nanog* by binding to the coding regions of these genes [251] and miR-145 was shown to repress the expression of *Oct4*, *Sox2* and *Klf4* by binding to the 3'UTRs of these genes [252]. The EMT associated miR-200 family, which contains 5 members, is transcriptionally regulated by *c-Myc*. Overexpression of this family of miRNAs opposes the downregulation of pluripotency factors in ESCs grown without LIF [253].

The most abundantly expressed miRNA family in mouse ESCs is the miR-290 family. MiRNAs of this family, miR-291-3p, miR-294, and miR-295 and miR-302d, are associated with stem cell cycling and called embryonic stem cell cycle (ESCC) miRNAs. The Blelloch group has shown that this subset of miRNAs can enhance the generation of iPSCs along with OCT4, SOX2 and KLF4, but not in the presence of c-MYC. c-MYC binds to the promoters of these miRNAs which suggests that these miRNAs are the downstream targets of c-MYC [254]. The same microRNAs were also shown to promote the generation of human iPSCs [223]. The reprogramming process has to cross various roadblocks and

this subset of miR-290/302 family targets various pathways such as EMT, cell cycle, cell death and mitochondrial function at both early and later stages of reprogramming. [255]. One such pathway is the p53 pathway, which triggers cell cycle arrest, senescence or cell death. During reprogramming, p53 expression is upregulated by Klf4 and c-Myc, presumably as an attempt of the cells to maintain their genomic stability under stress. It was shown that miR-138 can fine-tune the expression of p53 and promote OSKM mediated iPSC generation [256]. Another early pathway observed during the reprogramming process is MET, which is controlled by TGF beta signaling and TFs like Zeb1/2, Snai 1/2, and Twist 1/2 [257]. MiRNAs of the miR-200 family downregulate the mesenchymal specific Zeb 1/2 to promote iPSCs formation [228]. It was further demonstrated that the miR-290/302 family and the miR-17~92 cluster significantly increase the reprogramming efficiency by negatively regulating the TGF beta pathway [223, 258]. A study from Blelloch's group screened 570 mouse miRNAs for their ability to promote iPSC generation and identified several other miRNAs such as miR-181, miR-19*, miR-34, miR-324, miR-451 and miR-677 that are able to promote reprogramming [255].

In 2011, three independent groups reported that microRNAs alone could reprogram the somatic cells to pluripotency. One group transiently transfected a combination of mature, double-stranded miR-200c, miR-302s and miR-369 miRNAs in mouse adipose stromal cells and after 15 days observed iPSC colonies with similar efficiency reported in the original article on OSKM mediated reprogramming. The same set of microRNAs also induced pluripotency in human adipose stromal cells and dermal fibroblasts cells 20 days after the first transfection [259]. The second study used only miR-302s family of miRNAs to induce pluripotency in human hair follicle cells. They stated that miR-302's mediated reprogramming in a dose dependent system wherein an inducible system was used to achieve 1.3-fold higher expression than it is normally expressed in human ES cell lines. It was also proposed that miR-302's suppressed four epigenetic regulators, among which repression of histone demethylase AOF2 lead to decreased levels DNMT1 which further enhanced global demethylation [260]. The same group previously reported that miR-302's can induce ES cell properties in human skin cancer cells and that 86% of the gene expression patterns were similar to the ES cell lines [261].

The third study reported that the miR-302/367 cluster could alone reprogram both human and mouse dermal fibroblasts with an efficiency of two orders of magnitude higher compared to the OSKM mediated reprogramming [224]. The miR-302/367 cluster is located in the intron 8 of the Larp7 gene on chromosome 3 in mouse and on chromosome 4 in human. The polycistronic cluster is transcribes to lead to expression of 5 miRNA: miR-302a/b/c/d and miR-367. MiR-302a/b/c/d have similar 5' end ('seed sequence') and only differ in the hexanucleotides located at the 3'end of the miRNAs [249]. This cluster was

initially cloned from mouse ESCs and was not detectable in adult cell lines. Quantitative PCR methods later supported the high expression of miR-302/367 in mouse and human ESCs [262]. In mouse ESCs it was shown that Oct4, Sox2 and Nanog bind to the promoter regions of miR-302/367, miR-290 and miR-160a/363, whereas OCT4 is bound to the same conserved promoter region of miR-302/367 in human ESCs [212]. It is also shown that OCT4 and SOX2 transactivate the miR-302/367 cluster in human ESCs and that the levels of miR-302/367 and OCT4 are downregulated upon differentiation [263]. In the reprogramming context, miR-302/367 activation by OCT4 and SOX2 occurs in different pathways involved in reprogramming. For example, it is shown that hsa-miR-302/367 targets BMP inhibitors like transducer of ERBB2 (TOB2), DAZ associated proteins 2(DAZAP2) and SLAIN motif family member 1 (SLAIN1) to promote BMP signaling which supports the growth of human ESCs [264]. miR-302 also promotes MET by acting on TGFβR2 and the Ras homologue family member RHOC, which are known to reverse the process of EMT [223]. In the reprogramming study of Anokye-Danso et al., it was noted that miR-367 and VPA are necessary for the higher efficiency of miR-302/367-mediated reprograming of mouse fibroblasts, whereas VPA, which target histone deacetylase Hdac2 is not necessary and did not affect the reprogramming efficiency of human fibroblasts cells, which have low levels of Hdac2. However, a study failed to reprogram mouse fibroblasts with miR-302/367 [265] alone and also human adipose stem cells with miR-302s [266]. Since no other reports recapitulated miRNA-only mediated reprogramming, one cannot yet consider miRNAs as a robust tool of reprogramming.

6.1.8 Small molecules and reprogramming

Several small molecules are also shown to enhance reprogramming of mouse and human cells. For example vitamin C (L-ascorbic acid), is an essential nutrient with antioxidant properties in humans, its deficiency leading to scurvy [267]. Vitamin C promotes cell proliferation and fibroblasts show an increase in their lifespan when vitamin C is present in the cell culture medium [268]. Vitamin C prevents the activation of INK4/ARF locus, which is an important tumor suppressor locus. Induction of this locus leads to the activation of the p53 pathway, which is known to be a roadblock in the somatic cell reprogramming [269]. This study also showed that vitamin C induces H3K36 me2/me3 chromatin marks by regulating two histone demethylases, Jhdm1a and Jhd1mb and that Jhd1mb and OCT4 are sufficient to induce pluripotency in fibroblasts. This reprogramming process also strongly induces expression of the miR-302/367 cluster. Jhd1mb decreases methylation levels of histone encompassing OCT4 binding sites in the miR-302/367 gene and facilitates gene expression [222]. In another study it was shown that vitamin C also regulates histone-demethylating dioxygenase Tet1 hydroxylase that is implicated in DNA demethylation. In the presence of vitamin C, Tet1 negatively regulates somatic cell

reprogramming by modulating MET [270]. Another molecule, TGB receptorI/activin like kinase 5 (Alk5) inhibitor, enhances reprogramming efficiency by acting as a cooperative factor to inhibit TGF beta signaling pathway and also acts as a replacement factor, which bypasses the use of c-MYC and SOX2 to reprogram mouse fibroblasts [271].

6.1.9 Splicing and reprogramming

Although alternative splicing was not extensively studied in the context of reprogramming, differences in isoform content between ESCs and differentiated cells have been recently reported. For instance OCT4, an important pluripotency factor has two isoforms, OCT4A, the canonical transcription factor known for its major role in the maintenance of pluripotency and OCT4B, which is expressed in the cytoplasm of differentiated cells and has unknown functions [272]. Other pluripotency factors such as Sall4 with its two isoforms collaborate with other factors to maintain pluripotency [273] and NANOG has three different isoforms that also assist in the maintenance of pluripotency with different efficiencies [274]. A very interesting example has been reported in the study of FOXP1 transcripts [275] FOXP1 is part of one of the four subfamilies of forkhead box (FOX) transcription factors and is known to regulate the genes involved in proliferation, development and differentiation. In this study a highly conserved alternative splicing event in FOXP1 transcript was identified. The transcript, named it as FOXP1-ES, contains an exon variant specifically expressed in ESCs and silenced during differentiation. This particular exon selection modifies the sequence of amino acids in the forkhead domain and changes the DNA binding specificity of the transcription factor. In ESCs this spliced isoform of canonical FOXP1 switches the transcriptional output that stimulates several pluripotency genes like *OCT4, NANONG, GDF3, and NR5A2*, concurrently silencing the genes involved in the differentiation. Overexpression of this isoform promoted self-renewal and its silencing inhibited iPSC formation. Thus alternative splicing may critically modify the specificity of pluripotency or differentiation-related transcription factors with very broad consequences for the expression profile of the cells.

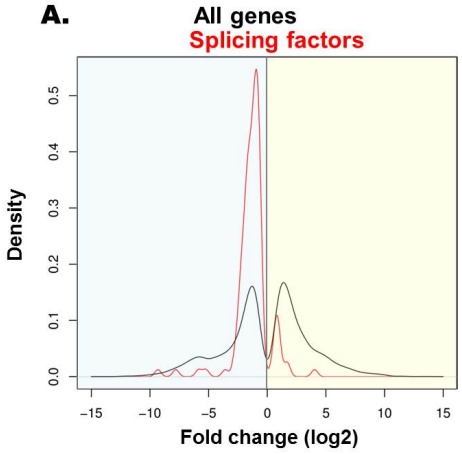
In a follow up study by the Blencowe group [276], involving an analysis of RNA-seq data from ESCs, iPSCs, and from differentiated tissues, revealed that the muscle blind like (MBNL) 1 and 2 RNA binding proteins are expressed at low level in ES and iPS cells compared to other cells and tissues. These proteins were hypothesized to repress ES cell-specific exons in non-ESCs. Knockdown of these proteins in differentiated cells switched ~half of alternative splicing pattern to ES-cell like and promoted the OSKM-mediated reprogramming. Conversely, overexpression of these proteins in ESCs promoted a differentiated cell-like alternative splicing pattern. These proteins act in part by repressing ESC specific FOXP1 isoform, which enhances the expression of pluripotency genes and

represses the genes related to differentiation. Using high-throughput RT-PCR technology, splicing changes that occurred upon induction of fibroblasts to pluripotency and their redifferentiation revealed that MBNL1 and RBFOX2 act in concert to bring about alternative splicing changes associated with differentiation into late mesoderm [277]. In a very recent report [278] it was shown that SFSR2 is the most enriched splicing factor in ESCs and that MBD2 is most differentially spliced with three isoforms (MBD2a and MBD2c). MBD2c is enriched in pluripotent cells and MBD2a in fibroblasts cells. The study also found that overexpression of MBD2c enhanced the reprogramming of fibroblasts. Furthermore, miR-302 specifically targets MBD2a in promoting reprogramming of the fibroblasts, which suggests that the effect of MBD2 on pluripotency is isoform specific.

6.2. Hypothesis

Recently, our group analyzed an extensive data set of mRNA-seq reads that were obtained from a somatic cell reprogramming system [279]. Strikingly, we found that splicing factors form one of the most significantly differentially expressed category of genes (Figure 14A). The majority of splicing factors are expressed at higher levels in iPSCs compared to fibroblasts. In particular, the expression of epithelial splicing regulatory proteins (Esrp)1, Esrp2 and musashi(Msi)1 was strongly upregulated during reprogramming (Figure 14B).

A.



B.

Gene	Symbol	Log2 fold change
Epithelial splicing regulatory protein 1	***Esrp1***	**12.9**
Epithelial splicing regulatory protein 2	***Esrp2***	**6.55**
Musashi 1	***Msi1***	**11.29**

Figure 14: (A) Differential gene expression analysis of mRNA-seq data sets of iPSCs and parental fibroblasts revealed that splicing factors are preferentially upregulated upon pluripotency induction. Note that only differentially expressed genes (false discovery rate ≤ 0.05) are plotted. (B) Expression fold changes of Esrp1, Esrp2 and Msi1 in iPSCs compared to fibroblasts. Analysis performed by Alexander Kanitz (Zavolan lab, Biozentrum, University of Basel)

6.2.1 Esrp1/2 and Msi1

Esrp1/2 are epithelial cell-specific splicing regulators, which were first identified from a high-throughput cDNA expression screen [67]. Espr1/2 are highly conserved paralogs that contain three RNA recognition motif (RRM) domains. Upon of induction of EMT in a mammary epithelial cell line *ESRP1/2* expression was downregulated, which subsequently switched the splicing pattern in the *FGFR2*, *CD44*, *CTNND1*, and *ENAH* from epithelial to mesenchymal. The same switch of splicing pattern in these four transcripts was observed when *ESRP1/2* were downregulated using siRNAs. When *Esrp1* was overexpressed in mouse mesenchymal cells it induced the opposing changes in these four transcripts, leading to an epithelial phenotype [67]. Since these four transcripts have well documented roles in EMT [67, 280], the regulation of their isoforms by Esrp1/2 suggests that Esrp1/2 have a major role in the EMT process. Moreover, knockdown of Esrp1 in mouse ESCs increases the expression levels of core pluripotency TFs, eventually enhancing the self-renewal property of ESCs, while the Esrp1 knockout ESCs showed impaired ability in early differentiation. Interestingly, ESRP1 was found to bind to the *Oct4* and *Sox2* mRNA [281].

MSI1 belongs to the family of musashi proteins, which are highly conserved proteins that are highly expressed in stem cells and cancer cells and are considered to be neural stem cell markers [282]. Msi1 is known to regulate a large number of targets involved in apoptosis, proliferation, and cell cycle. It is also a stem cell marker, which regulated the balance between self-renewal and differentiation status of the stem cells [283]. In a recent study it was shown that overexpression of Msi1 repressed the EMT process by inhibiting translation of an EMT factor, Jagged1. Knockdown of Msi1 in epithelial cancer cells supported the loss of epithelial phenotype, suggesting that Msi1 is contributes to the epithelial phenotype [284]. In breast cancer cells *Msi1* maintains the expression of ES cell specific markers c-*MYC*, *NANOG*, *SOX2*, *BMi1*, and *OCT4* [285]. Finally, overexpression of c-*MYC* in human breast epithelial cells lead to an increased expressed of *MSI1* [283].

This body of evidence supported our hypothesis that splicing factors, particularly those with epithelial pattern of expression (Esrp1/2 and Msi1) are major players in the EMT process, which is part of the early phase of reprogramming. Thus, we sought to determine whether Esrp1/2 and Msi1 could modulate the reprogramming efficiency of somatic cells.

Moreover, in another study we would like to decipher the mechanism underlying the enhancement of somatic cell reprogramming by miR-302/367 along with OSKM genes. For this we established the miR-302/367 mediated reprogramming and we will generate mRNA-seq libraries form the time course of reprogramming.

6.3. Methods

6.3.1 Cell culture

Hek 293TN (System biosciences) and Platinum-E cells were cultured in GMEM (Sigma Aldrich), supplemented with 10% ESC specific FCS, 1X non-essential amino acids (Sigma Aldrich), 1X Pen/Strep antibiotics, 1mM sodium pyruvate (Sigma Aldrich) and 0.1mM 2-mercaptoethanol (Gibco), 2mM L-glutamine (Sigma Aldrich) and 20ng/ml of LIF (from Kaji lab, SCRM, University of Edinburgh). This medium is hereafter denoted as "normal medium". TNG-MEFs and WT MEFs were cultured with normal medium supplemented with FGF-2 (5ng/ml) and heparin (1ng/ml). Reprogramming medium was composed of normal medium with 1µg/ml doxycycline (Sigma Aldrich), 10µg/ml vitamin C (Sigma Aldrich) and 500nm Alk-5i (Tocris Bioscience).

6.3.2 Viral vectors and transductions

A doxycycline inducible lentiviral expression vector pTetO-Fuw-OSKM (addgene ID: 20321) was modified as per requirement. The EcoRI flanked OSKM cassette was replaced with the respective mouse cDNAs of Esrp1, Esrp2, Msi1, primary miR-302/367 and Renilla luciferase gene using the blunt end cloning method. All four modified lentivectors are hereby denoted as pTetO-Fuw-cDNA. Esrp1 and Esrp2 cDNAs were PCR amplified from a mouse mammary gland "NMuMGg" epithelial cell line [286], primary miR-302/367 was PCR amplified from pCuo-302/367 vector (Biocat), and Msi1 cDNA was commercially synthesized by Eurofins Scientific AG. Renilla luciferase gene was PCR amplified from psiCHECK-2 vector (Promega AG). pTetO-Fuw-OSKM vector with OSKM insert was used as an empty vector infection control for few optimization experiments. The blunt end cloning approach is briefly described separately in this methods section. Second-generation lentiviral packaging plasmids were used to generate pseudo-lentiviral particles in a producer HEK293 cell line (System Biosciences). The packaging plasmids used were pCMVR8.74 (addgene ID: 22036) and pMD2.G (addgene ID: 12259). To generate pseudo-lentiviral particles, 2*10^6 HEK293 cells were seeded in 10cm culture dishes and the following day, pTetO-Fuw-cDNA, pCMVR8.74 and pMD2.G were co-transfected in a ratio of 5:3:2µg of plasmid DNA using 25µl of Lipofecatmine 2000 (Invitrogen). 8 hrs. after transfection, normal medium was replaced with fresh normal medium and 48 hrs. after transfection, culture medium containing pseudo-lentiviral particles was harvested, filtered using 0.45 micron filters. Polybrene (8µg/ml, Merck Millipore AG) was added to the filtered lentiviral medium to facilitate transduction. 2ml of such filtered viruses were transduced in TNG-MEFs/WT MEFs seeded on a 0.1% gelatinized one well of a six well

plate. After 4 hrs. of transduction, medium containing pseudo-viral particles was replaced with fresh medium of interest.

Retroviral vectors were constructed using attR1 R2 gateway cloning cassette (Invitrogen). Mouse cDNAs of Esrp1, Esrp2, Msi1 and Renilla luciferase gene were cloned into an entry pENTR2B vector (Life technologies) using EcoRI sites. After confirming the inserts free of mutations by sequencing (Microsynth), these inserts were gateway cloned into a pMXs retro-vector using LR clonase (Life technologies). To produce pseudo-retroviral particles, 2*10^6 platinum-E cells were seeded in 10 cm culture dish and the next day 10µg of pMXs retro-vector with transgenes were transfected with 25µl of lipofectamine 2000. 8 hrs. after transfection, the medium was replaced with fresh medium. After 48 hrs. of transfection, culture medium containing retroviral particles was harvested, filtered through 0.45 micron filters. Thus filtered medium with viruses was mixed with polybrene (final concentration 8µg/ml) and transduced in TNG-MEFs. 4hrs after transduction, medium was replaced with fresh medium of interest. The primers used for cloning splicing factors, Renilla luciferase and miR-302/367 are listed in table7.

6.3.3 Generation of secondary mouse embryonic fibroblasts (TNG-MEFs) and reprogramming method

Most of the TNG-MEFs and WT MEFs used in this study were kindly provided by Dr. Keisuke Kaji. Some batches of TNG-MEFs used to optimize lentiviral transductions and some preliminary experiments were done in the Zavolan lab at the Biozentrum, University of Basel. To obtain TNG-MEFs, a transgenic Nanog-GFP embryonic stem cell line (made in the Kaji lab, Figure 15A) was gene targeted in the Sp3 locus with a targeting plasmid containing a polycistronic cassette of reprogramming genes (Vector backbone represented in the Figure 15B), followed by IRES-mOrange, inducibly expressing from a single tetO2 promoter. The construct was used by the transgenic mouse core facility, Biozentrum, University of Basel to obtain chimaeric embryos by blastocyst injections. Embryos were collected at 12.5 days post coitum (d.p.c), decapitated, eviscerated and individually dissociated in 0.25% trypsin-EDTA, then cultured in normal medium. One tenth of the cells were treated with 1µg/ml doxycycline for two days. To estimate the percentage of transgenic MEFs, a FACS analysis was performed based on mOrange expression. A schema of the method is represented in the Figure 15C

6.3.4 Quantitative real-time PCR

MEFs that were transduced with retroviruses or lentiviruses were harvested 4 days after induction of reprogramming. Total RNA was extracted using TRI reagent (Sigma Aldrich) as per the manufacturer's protocol, 10µg of total RNA was subjected to DNase digestion with

RQ1DNase (Promega), followed by phenol-chloroform purification. 1µg of total RNA was used for cDNA synthesis using SuperScriptIII (invitrogen) reverse transcriptase according to the manufacturer's protocol. 10ng of cDNA per sample was used to in the qRT-PCR reaction using Power SYBR Green PCR Master Mix (Applied Biosystems) in triplicates. GAPDH levels were used as internal control and the samples from non-infected controls were used as external controls. mRNA levels of the transgenes were measured by Step One Plus real time PCR system (Applied Biosystems). Relative fold change values were calculated as $2^{-\Delta\Delta CT}$ [287]. Small RNA Northern was performed as described in the first part of this thesis.

6.3.5 Blunt end cloning

Since the pTetO Fuw lentivector and the entry vector pENTR used for gateway cloning had only EcoRI restriction site for cloning the inserts and because this restriction site was encoded in all of the splicing factors, we adopted the blunt end cloning approach. 2µg of vector was digested for two hours with EcoRI restriction enzyme, gel purified using QIAEX II beads (from Qiagen) and made blunt by T4 DNA polymerase and Klenow DNA polymerase for 30 minutes at 20^0C. The vector was column purified by QIAquick PCR purification kit (from Qiagen) followed by the dephosphorylation of the blunted end of the vectors by FastAP (Thermo Fisher Scientific) for 30 minutes at 37^0C, then at 70^0C for 10 mins to deactivate the enzyme. This was followed by the column purification of the reaction to elute blunted-dephosphorylated vector free of reaction components. On the other hand, PCR products were gel purified using QIAEX beads and phosphorylated by PNK (New England Biolabs) in presence of 10mM ATP for 1 hour at 37^0C, followed by column purification. The column-purified blunt vector and PCR products were ligated in 1:1 ratio, followed by the bacterial transformation. Positive clones were screened by sequencing.

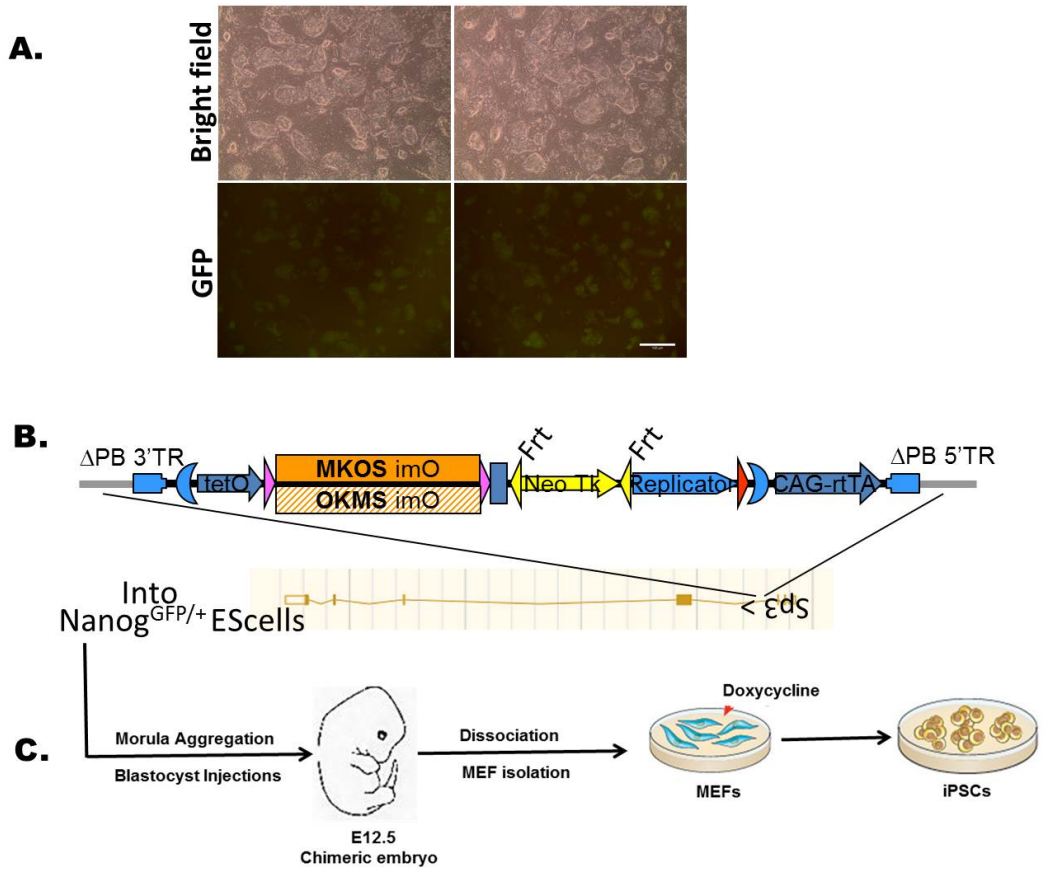


Figure 15**:** Experimental model used to study the impact of tissue specific splicing factors on OSKM mediated somatic cell reprogramming. (A) A Nanog-GFP embryonic stem cell line was gene targeted in Sp3 locus with a (B) targeting vector containing an inducible polycistronic cassette expressing MKOS genes tagged with mOrange and constitutively expressing reverse tetracycline transactivator protein. (C) These ES cells were further injected in blastocysts or formed aggregates with morula. These transformed blastocysts or morula aggregated were implanted in the uterus of foster mice to obtain chimeric embryos. 12.5 dpc embryos were harvested and enzymatically digested to derive embryonic fibroblasts. Thus obtained transgenic mouse embryonic fibroblasts cells can be induced with doxycycline to express MKOS factors, which eventually reprogram MEFs to iPSCs. Image B was kindly provided by Dr. Keisuke Kaji.

6.4. Results

6.4.1 Reprogramming of TNG-MEFs

To maintain optimal conditions of reprogramming, 3% of TNG-MEFs were seeded with 97% 129-WT MEFs constituting up to 1*10^5 cells for each 0.1% gelatinized well of a six well plate. TNG-MEFs were cultured in presence of reprogramming medium for 15 days. A general outline of the reprogramming experiment performed in this study is shown in Figure 16. A day after the induction with doxycycline, mOrange expression is observed, which indicates the expression of MKOS genes. By day 4 TNG-MEFs are seen with high proliferation, transiting through MET stage. By day 8, nice colonies are seen and some of

them start to express GFP, which indicates *de novo* expression of Nanog, a pluripotency marker. While the GFP expression becomes apparent, simultaneous silencing of mOrange expression is also observed. By day 12 most of the colonies are seen expressing GFP. Images of the TNG-MEFs undergoing transition into iPSCs are shown in Figure 17. By day 15, based on the number of Nanog-GFP colonies from the experimental condition are compared with the number of Nanog-GFP colonies observed in the controls to determine the reprogramming efficiency.

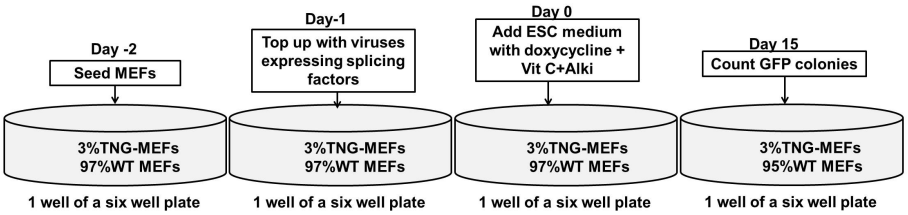


Figure 16: A schematic representation of the reprogramming experiments performed in this study. TNG-MEFs are seeded with WT-MEFs and are infected with viruses containing transgenes. Expression of MKOS is then induced with doxycycline. After induction, TNG-MEFs are grown in reprogramming medium, which is changed every two days. After 15 days, the reprogramming efficiency is estimated by GFP-colony counting.

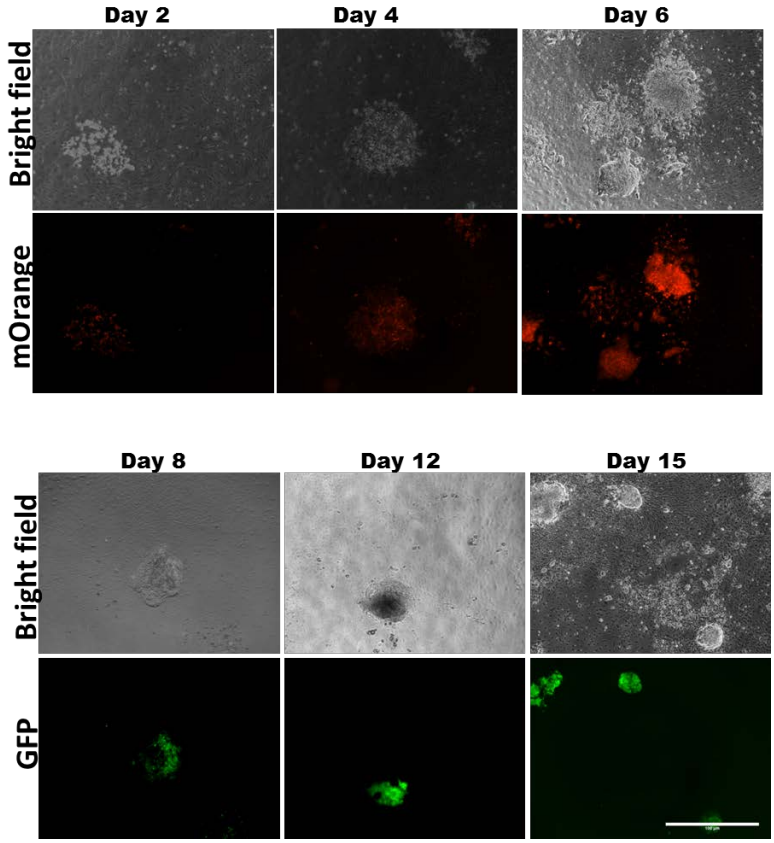


Figure 17: Microscopic images taken during the time course of reprogramming. Top panel shows the images for the first six days where mOrange expression was tracked. Bottom panel shows the images taken until

day 15, wherein expression of GFP was monitored. After the induction with doxycycline, mOrange expression is observed and formation of colonies begins with high proliferation. By day 8 few colonies start to express GFP, indicating the expression of Nanog. By day 15 stable GFP colonies are observed which can be counted to estimate the reprogramming efficiency between different conditions. Scale bar represents 100µm

6.4.2 Tissue specific splicing factors promote reprogramming

6.4.2.1 With retroviral top ups

To address the hypothesis that the splicing factors, whose expression levels were drastically changed during reprogramming, may promote the reprogramming process, we have overexpressed the *Esrp1*, *Esrp2* and *Msi1* in TNG-MEFs. These splicing factors were cloned in the retroviral vectors. TNG-MEFs and 129 WT MEFs were seeded as described in the above section in a well of a six well plate. The next day, TNG-MEFs were infected with 2ml of retroviruses encoding *Esrp1*, *Esrp2, and Msi1* genes for 4 hours, followed by the medium change with reprogramming medium. For all the experiments, infection of retroviruses with Renilla luciferase gene was used as a negative control along with non-infection control. For each splicing factor and controls, three such wells were considered in the experiment, wherein cells from one well were harvested after 4days of infection to check the transgene overexpression by qRT-PCR, and the rest of the two wells were considered in the actual experiment. Transgene overexpression was observed more than 30 fold for all the splicing factors and luciferase (Figure 18A), when compared to the endogenous levels of the non-infection control and GAPDH gene was used an internal control. Expression of mOrange and GFP were monitored throughout the reprogramming experiment, and at day 15, total number of Nanog-GFP colonies was counted. The average taken over two independent experiments conducted in two different labs showed an increase in the reprogramming efficiency by splicing factors. *Esrp1* and *Msi1* showed ~1.6 fold, whereas *Esrp2* showed ~1.9 fold increase in the number of Nanog-GFP colonies compared to the infected Renilla luciferase and non-infected controls (Figure 18B). No increase was seen with Renilla luciferase gene compared to non-infected control. This indicates that the tissue specific factors promote reprogramming along with cognate reprogramming genes MKOS. Colonies were counted manually either from the images taken by Celigo cell imaging cytometer or under the microscope with GFP filter. Representative images of Nanog-GFP colonies of a single well of a six well plate taken by Celigo cell imaging cytometer are shown in Figure 18C**.**

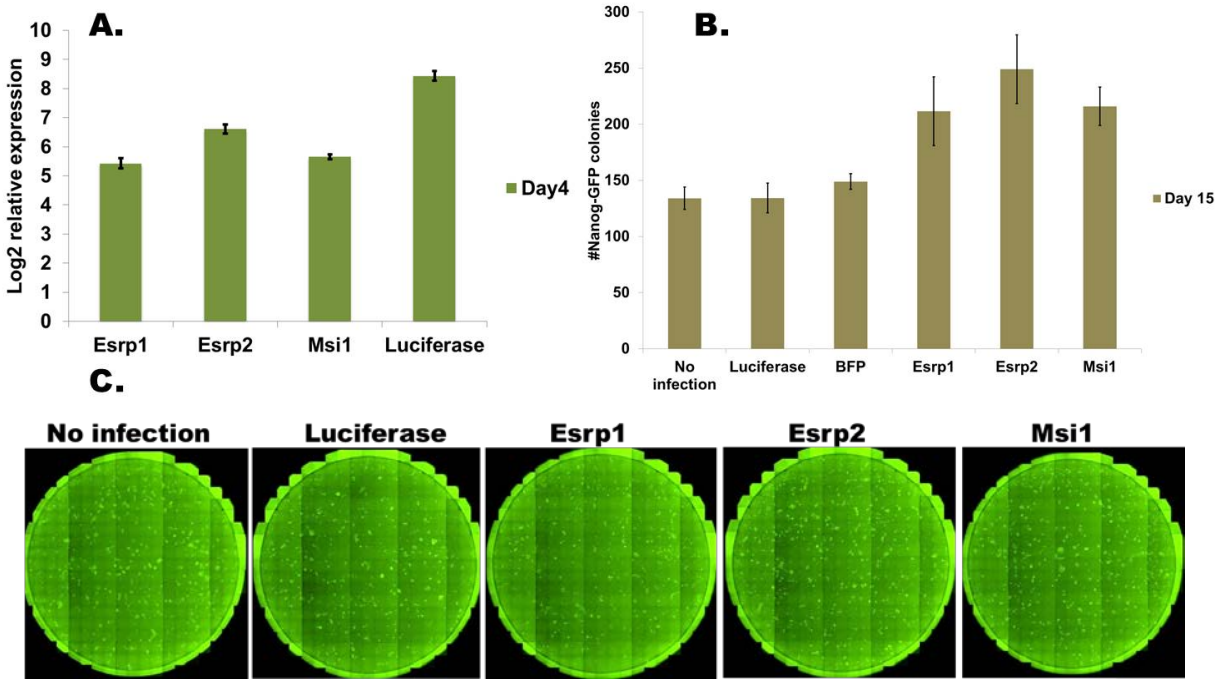


Figure 18: Tissue specific splicing factors enhance reprogramming efficiency. TNG-MEFs infected with retroviruses containing *Esrp1, Esrp2, Msi1* showed increased in the Nanog-GFP colony number compared to infected Renilla luciferase and non-infected controls. (A) Transgene overexpression was measured by qRT-PCR from the cells with 4 days of infection with retroviruses encoding different transgenes. An overexpression of over 30 fold of the splicing factors and luciferase gene compared to the endogenous levels from non-infected controls was observed. GAPDH gene was used as an internal control and error bars represent ± standard deviation of the three technical replicates. (B) Total Nanog-GFP colonies counted from two independent experiments of the splicing factors showed an increase of more than ~1.6 fold compared infected Renilla luciferase gene and non-infected controls. Error bars represent ± standard deviation of two biological replicates. (C) Images of single wells infected with transgenes along with non-infected controls taken by Celigo imaging cell cytometer showing the total number of Nanog-GFP colonies per well.

6.4.2.2 miR-302/367 and tissue specific splicing factors promote reprogramming with lentiviral top ups

An efficient way to deliver the transgenes in primary cells is through lentiviruses. Assuming that increasing the transgene overexpression could increase the efficiency of splicing factors in promoting somatic cell reprogramming, an inducible lentiviral system was used. To validate the lentivectors containing Esrp1, Esrp2, Msi1, and miR-302/367, TNG-MEFs seeded as described above were infected with these lentiviruses. TNG-MEFs were harvested after 4 days of induction with 1µg/ml of doxycycline for further analysis of transgene expression. miR-302/367 expression was validated by small RNA Northern (Figure 19A), whereas the relative expression levels of Esrp1, Esrp2 and Msi1 were measured with qRT-PCR (Figure 19B). Transgene overexpression was higher than what was observed with retroviruses. Lentiviruses containing splicing factors and miR-302/367

were transduced in the same experimental set up as described above for the reprogramming experiment. Lentiviruses expressing Renilla luciferase gene and non-infected cells were used as negative controls and viruses that expressed *Nanog* gene were considered as a positive control. With miR-302/367 lentiviruses we observed ~3 fold increase and with *Nanog* about ~2.5 fold increase in the number of Nanog-GFP colonies. With *Esrp1*, there was ~1.4 fold increase in the number of Nanog-GFP colonies while with Msi1 ~1.6 fold increase in the number of Nanog-GFP colonies. These numbers were similar to what was observed with retroviruses. *Esrp2* showed up to 2-fold increase in the increase in the number of Nanog-GFP colonies, which is slightly greater than what was observed with retroviruses. Graphical representation of the number of Nanog-GFP colonies is shown in the Figure 19C and the images taken by Celigo imaging cell cytometer are shown in Figure 19D. To confirm and improve the results a better setup of lentiviral packaging and further repetitions of the experiment are needed.

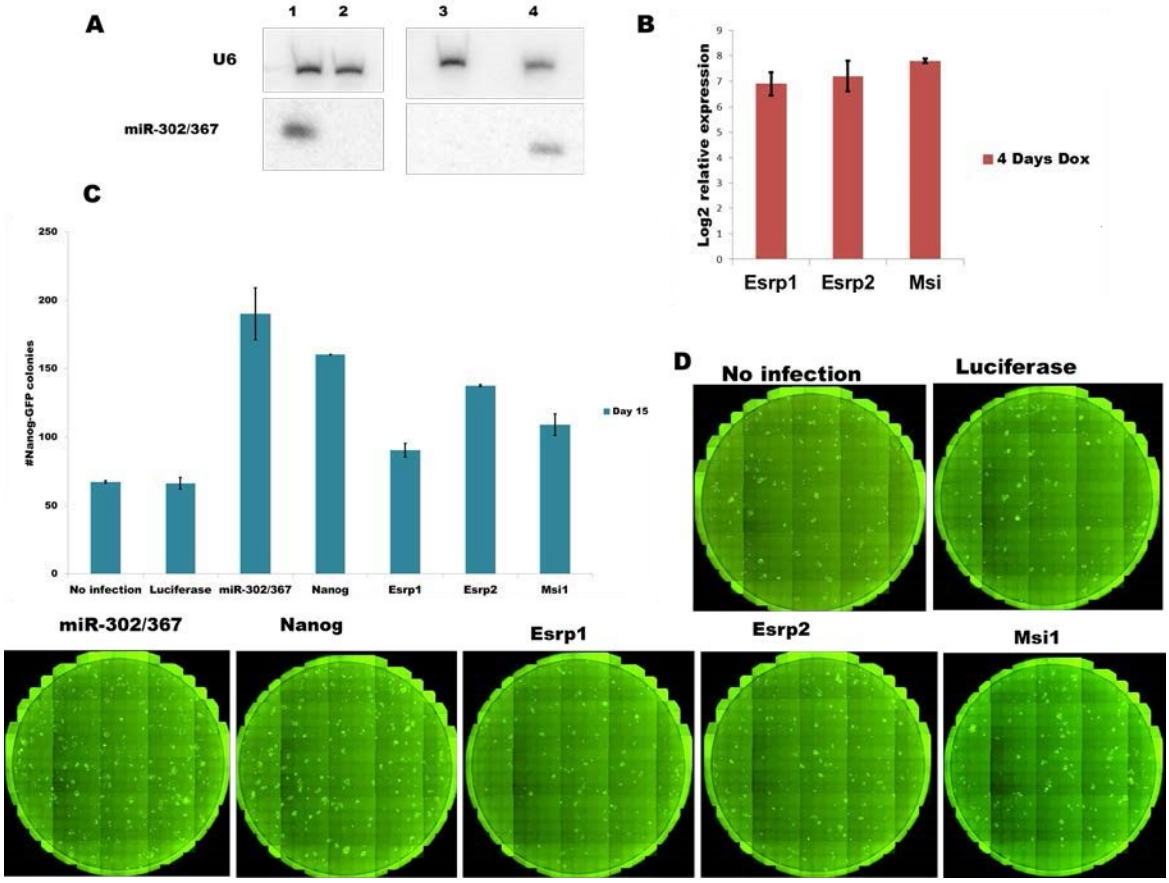


Figure 19: Reprogramming experiment with lentiviruses expressing splicing factors and miR-302/367. Transgene expression from the lentiviruses was validated from TNG-MEFs-WT MEFs infected with

lentiviruses. 4 days after doxycycline induction in the TNG-MEFs infected with lentiviruses expressing miR-302/367 were harvested and its expression was confirmed with small RNA Northern. (A) Northern blot showing miR-302/367 cluster expression. Lane 1 and 4 were probed with the probes specific to miR-302b and miR-367, lanes 2 and 3 were loaded with total RNA from the cells without induction with doxycycline. Top panel blot shows the expression of U6 snRNA used as loading control and bottom panel shows miR-302b and miR-367 expression. (B) Relative expression levels of splicing factors from the TNG-MEFs infected with lentiviruses containing splicing factors as transgenes were measured with qRT-PCR. Relative expression values are represented in log2 values, GAPDH gene was used an internal control. Error bars represent the ±standard deviation of three technical replicates (C) Total Nanog-GFP colony number counted at day 15 of the reprogramming experiment observed from the splicing factors and miR-302/367 along infected and non-infected controls. Error bars represent the ±standard deviation from two technical replicates. (D) Images of the reprogramming experiment with splicing factors and miR-302/367 taken with Celigo imaging and cell cytometer.

6.5. Discussion

The differential gene expression analysis of 21 human iPSCs and 7 parental fibroblasts mRNA-seq libraries [279] revealed that 24% of all genes and 54% of splicing factors were differentially expressed between iPSCs and their parental fibroblasts. Among the splicing factors, tissue specific splicing factors showed a striking change in their expression levels. Esrp1, and Msi1 were the two splicing factors which were upregulated in iPSCs with high expression fold changes .In contrast, muscle blind like proteins MBNL1 and MBNL2 were downregulated with high expression fold changes. Just after our analysis, Blencowe and colleagues reported that MBNL proteins are negative regulators of alternative splicing events that are differentially expressed between ES cells and other differentiated cells types, and further showed that knockdown of these proteins in secondary MEFs enhanced the reprogramming efficiency [276]. Given the role of Esrp1/2 and Msi1 in the EMT process, their association with several pluripotency genes [67, 281, 283-285], the evidence showing the role of splicing factors in regulating pluripotency and reprogramming [275, 276], and our own analysis of mRNA-seq data sets, we sought to investigate the role of Esrp1/2 and Msi1 in reprogramming.

To address the role of Espr1/2 and Msi1 in the reprogramming of somatic cells we first asked whether these factors have an impact on somatic cell reprogramming. To look into this aspect, we chose a secondary reprogramming model to estimate the reprogramming efficiency after the overexpression these splicing factors in TNG-MEFs with cognate reprogramming genes MKOS. Splicing factors were overexpressed using retroviruses or lentiviruses in TNG-MEFs and the induction of MKOS was induced with doxycycline. The number of Nanog-GFP colonies with Esrp1 and Msi1 was more than 1.6 fold higher compared to controls and with Esrp2 it was up to 1.9 fold, indicating the role of tissue

specific splicing factors in somatic cell reprogramming. This was confirmed from two independent experiments

Since the observed difference in the expression fold changes with Esrp1 and Msi1 from our differential gene expression analysis was more than 1000 fold, we assumed that increasing overexpression efficiency of the transgenes in TNG-MEFs, would probably be a better experimental design. For this we tried a lentiviral system, lentiviruses being known to be efficient in infecting both dividing and non-dividing cells. Our preliminary experiments with lentiviruses were problematic due to the toxicity of the constructs for cells [288], which thereby showed inconsistency in the number of Nanog-GFP colonies in infected controls of dsRED and empty vector controls. Further Optimization of the lentiviral system was carried out and reducing the infection time lead to consistent numbers of Nanog-GFP colonies with infected and non-infected controls. With no difference in the Nanog-GFP colony number between infected with Renilla luciferase and non-infected controls, Msi1 showed similar difference as observed with retroviruses. Esrp2 increased the number of Nanog-GFP colonies up to 2 fold, which is slightly more than 1.9 fold as observed with retroviruses. Esrp1 also showed almost similar difference compared to what has been seen with retroviruses. Transgene expression was not verified with qRT-PCR in this particular experiment and thus additional experiments with lentiviruses are needed.

With the knockdown of MBNL proteins, it was shown that the reprogramming efficiency could be increased up to 2 fold in the secondary reprogramming system [276], supporting our computational analysis in which MBNL proteins were the most downregulated splicing factors. Thus, we believe that our data indicates that splicing factors with epithelial pattern of expression can improve the reprogramming efficiency, probably because they regulate splicing events specific to pluripotency.

Overexpression of miR-302/367 cluster alone in MEFs was reported to reprogram MEFs to iPSCs with an efficiency of two orders of magnitude compared to the reprogramming with OSKM alone [224]. In an attempt to decipher the underlying mechanisms of miRNA mediated reprogramming using a time series mRNA-seq data, we tried to re-establish this reprogramming method in the lab, but as reported by other labs [265, 266], we also failed to reprogram MEFs with miR-302/367 alone. Instead, overexpressing miR-302/367 along with OSKM genes enhanced reprogramming efficiency as reported by others [223]. Using this as experimental model we aim to uncover the role of miR-302/367 in somatic cell reprogramming by generating and analyzing a time series of mRNA expression.

One of the important observations during the preliminary optimization experiments with lentiviruses was the bigger size of the iPSC colony with upon overexpression of Msi1 as shown in Figure 20. Msi1, which is also known as a self-renewing gene in neural stem cells [289], has the ability to increase the self-renewal ability during reprogramming of TNG-MEFs that eventually showed the colonies with bigger size. Quantification of the number of cells based on pluripotency marker expression at the end of the reprogramming experiment would most likely indicate an even higher reprogramming efficiency with Msi1. This is another aspect we would like to follow up in further studies.

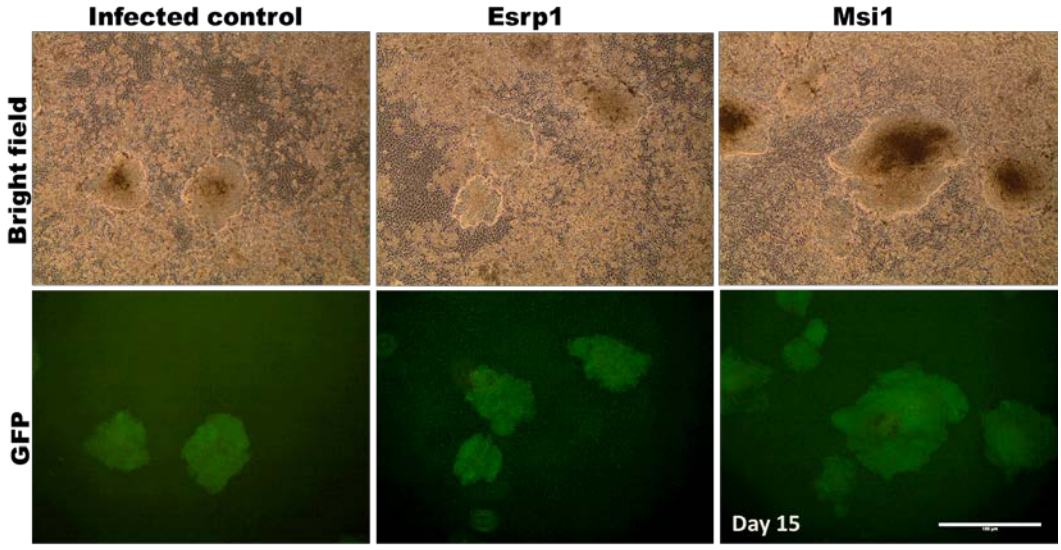


Figure 20: Overexpression with Msi1 leads to a phenotype of bigger Nanog-GFP colonies. Microscopic images of iPSC colonies with Nanog-GFP of infected control, Esrp1 and Msi1. Lentiviruses encoding Esrp1, empty vector control and Msi1 were transduced in TNG-MEFs and images were taken on day 15 of the reprogramming experiment. Top panel shows the images taken with bright field and bottom panel shows the images taken with GFP filter. Scale bar represents 100µm.

6.6. Conclusion and future prospects

In this study we have experimentally investigated our hypothesis that we formulated based on a bioinformatics analysis of mRNA-seq data obtained from iPSCs and their parental fibroblasts. The hypothesis that tissue specific splicing factors promote somatic cell reprogramming has been validated. We have used inducible secondary MEFs as our basic reprogramming model, and the tissue specific splicing factors were delivered via lentiviruses or retroviruses for their overexpression. With retroviral vectors, tissue specific splicing factors Esrp1, Esrp2 and Msi1 were able to promote the efficiency of TNG-MEFs reprogramming by over 1.6 fold and the highest improvement in efficiency, of 1.9 fold, was observed with Esrp2. With lentiviruses as well, Esrp2 showed the highest 2-fold increase, whereas Msi1 and Esrp1 showed changes similar to retroviruses of 1.6 and 1.4

fold. The experiments with lentiviruses will be repeated to validate the increase in reprogramming efficiency and to further characterize the mechanism.

As a follow up of these preliminary results, I would like to decipher the mechanisms involved in the facilitation of reprogramming by tissue specific splicing factors by preparing the mRNA-seq libraries of the early and late time series of the reprogramming and analyzing this data computationally. Following the kinetics of the reprogramming mediated by tissue specific splicing factors with the reprogramming associated cell surface markers [235] would also open new avenues to splicing and reprogramming field. I also intend to functionally characterize the iPSCs obtained with splicing factors and OSKM genes by in vitro and *in vivo* assays. In vitro assays include alkaline phosphatase assay of iPSC colonies and differentiation of these iPSCs into cells of three germ layers and *in vivo* assays include teratomas assay and checking the ability of these iPSCs to form chimeric mice by blastocyst injections.

Table 7: List of primers used to clone the splicing factors, Renilla luciferase and miR-302/367 cluster in pTetO FUW and pMXs vectors. qRT-PCR primers used for to check the transgene overexpression. DNA probes used for miR-302b and 367 Northern.

Primers used for cloning into pTetO Fuw and PMXs vectors		
Gene Symbol	Forward primer (5'-3')	Reverse primer (5'-3')
Esrp1	AAAAGAATTC ACCATGACGGCGTCTCCGGATTACTTGGT G	CCCC GAATTC TTAAATACAAACCCATTCTTTGGGTA
Esrp2	CCCC GAATTC ACCATGACTCCGCCGCCGCCGCCGCCCCC A	GGAA GAATTCC TACAAACACACCCACTCCTTAGGGGCT
Msi1	AAGG GAATTC ACCATGGAGACTGACGCGCCCCAGCCCGG C	CCCC GAATTC TCAGTGGTACCCATTGGTGAAGGCTGTG G
Renilla luciferase	TGCACGCAAA GAATTC ACC ATGGCTTCCAAGGTGTACGACCCCGAGC	TGCACGCAAAA GAATTC TTACTGCTCGTTCTTCAGCACGCGCTCC
miR-302/367 cluster	AAAA GAATTC GAAGTCCCTGCCTTTTACCC	AAAA GCGGCCGC TGAAGAGGAAAAGGATACTGGA
Primers used for qRT-PCR		
Esrp1	CCCCAGGTTCGCCTAATAGT	AGTCCATCCTCGGTTGCATA
Esrp2	TCCTACTGCCTTGGCCTCTA	AGGGGCTTGTAACACAGTGC
Msi1	GGCTTCCTAGGGACCACAA	TACCCATTGGTGAAGGCTGT
Renilla luciferase	TCGTCCATGCTGAGAGTGTC	CTAACCTCGCCCTTCTCCTT
DNA probes used miR-302/367 Northerns		
miR-302b	CTACTAAAACATGGAAGCACTTA	
miR-367	TCACCATTGCTAAAGTGCAATT	

7. References

1. **Kaiser, D., *Coupling cell movement to multicellular development in myxobacteria.* Nat Rev Microbiol, 2003. 1(1): p. 45-54.**
2. **Britten, R.J. and E.H. Davidson, *Gene regulation for higher cells: a theory.* Science, 1969. 165(3891): p. 349-57.**
3. **Whitfield, M.L., et al., *Common markers of proliferation.* Nat Rev Cancer, 2006. 6(2): p. 99-106.**
4. **Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. 126(4): p. 663-76.**
5. **Davis, R.L., H. Weintraub, and A.B. Lassar, *Expression of a single transfected cDNA converts fibroblasts to myoblasts.* Cell, 1987. 51(6): p. 987-1000.**
6. **Eguchi, G. and T.S. Okada, *Differentiation of lens tissue from the progeny of chick retinal pigment cells cultured in vitro: a demonstration of a switch of cell types in clonal cell culture.* Proc Natl Acad Sci U S A, 1973. 70(5): p. 1495-9.**
7. **Lage, K., et al., *A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes.* Proc Natl Acad Sci U S A, 2008. 105(52): p. 20870-5.**
8. **Monani, U.R., et al., *A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2.* Human Molecular Genetics, 1999. 8(7): p. 1177-1183.**
9. **Bishop, J.M., *The molecular genetics of cancer.* Science, 1987. 235(4786): p. 305-11.**
10. **Lal, A., et al., *A public database for gene expression in human cancers.* Cancer Res, 1999. 59(21): p. 5403-7.**
11. **Crick, F., *Central dogma of molecular biology.* Nature, 1970. 227(5258): p. 561-3.**
12. **Lazarides, E., *Modern cell biology: molecular cell biology.* Science, 1986. 234(4782): p. 1448.**
13. **Sawadogo, M. and A. Sentenac, *Rna Polymerase-B (Ii) and General Transcription Factors.* Annual Review of Biochemistry, 1990. 59: p. 711-754.**
14. **Alberts, B., *Molecular biology of the cell 5th edition*. 5th ed. Reference Collection2008, New York: Garland Science.**
15. **Levine, M. and R. Tjian, *Transcription regulation and animal diversity.* Nature, 2003. 424(6945): p. 147-151.**
16. **Maston, G.A., S.K. Evans, and M.R. Green, *Transcriptional regulatory elements in the human genome.* Annu Rev Genomics Hum Genet, 2006. 7: p. 29-59.**
17. **Chapman, R.D., M. Conrad, and D. Eick, *Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation.* Mol Cell Biol, 2005. 25(17): p. 7665-74.**
18. **Saunders, A., L.J. Core, and J.T. Lis, *Breaking barriers to transcription elongation.* Nat Rev Mol Cell Biol, 2006. 7(8): p. 557-67.**
19. **Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD.* Genes Dev, 2006. 20(21): p. 2922-36.**
20. **Halbeisen, R.E., et al., *Post-transcriptional gene regulation: from genome-wide studies to principles.* Cellular and molecular life sciences : CMLS, 2008. 65(5): p. 798-813.**
21. **Thomas, M.C. and C.M. Chiang, *The general transcription machinery and general cofactors.* Crit Rev Biochem Mol Biol, 2006. 41(3): p. 105-78.**

22. Lewis, J.D. and E. Izaurralde, *The role of the cap structure in RNA processing and nuclear export.* Eur J Biochem, 1997. 247(2): p. 461-9.
23. Rasmussen, E.B. and J.T. Lis, *In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes.* Proc Natl Acad Sci U S A, 1993. 90(17): p. 7923-7.
24. Herrick, D. and A. Jacobson, *A Coding Region Segment Is Necessary, but Not Sufficient for Rapid Decay of the His3 Messenger-Rna in Yeast.* Gene, 1992. 114(1): p. 35-41.
25. Christofori, G. and W. Keller, *3' Cleavage and Polyadenylation of Messenger-Rna Precursors Invitro Requires a Poly(a) Polymerase, a Cleavage Factor, and a Snrnp.* Cell, 1988. 54(6): p. 875-889.
26. Davis, R. and Y. Shi, *The polyadenylation code: a unified model for the regulation of mRNA alternative polyadenylation.* J Zhejiang Univ Sci B, 2014. 15(5): p. 429-37.
27. Gruber, A.R., et al., *Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors.* Wiley Interdiscip Rev RNA, 2014. 5(2): p. 183-96.
28. Elkon, R., A.P. Ugalde, and R. Agami, *Alternative cleavage and polyadenylation: extent, regulation and function.* Nat Rev Genet, 2013. 14(7): p. 496-506.
29. Rosenfeld, M.G., et al., *Altered expression of the calcitonin gene associated with RNA polymorphism.* Nature, 1981. 290(5801): p. 63-5.
30. Cascino, I., et al., *Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing.* J Immunol, 1995. 154(6): p. 2706-13.
31. Park, J.W. and B.R. Graveley, *Complex alternative splicing.* Adv Exp Med Biol, 2007. 623: p. 50-63.
32. Valdivia, H.H., *One gene, many proteins: alternative splicing of the ryanodine receptor gene adds novel functions to an already complex channel protein.* Circ Res, 2007. 100(6): p. 761-3.
33. Schmucker, D., et al., *Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity.* Cell, 2000. 101(6): p. 671-684.
34. Keller, W., *The RNA world - 2nd edition.* Science, 1999. 285(5428): p. 668-669.
35. Smith, C.W., et al., *Scanning from an independently specified branch point defines the 3' splice site of mammalian introns.* Nature, 1989. 342(6247): p. 243-7.
36. Colgan, D.F. and J.L. Manley, *Mechanism and regulation of mRNA polyadenylation.* Genes Dev, 1997. 11(21): p. 2755-66.
37. Padgett, R.A., et al., *Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors.* Science, 1984. 225(4665): p. 898-903.
38. Grabowski, P.J., R.A. Padgett, and P.A. Sharp, *Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate.* Cell, 1984. 37(2): p. 415-27.
39. Ruskin, B., et al., *Excision of an Intact Intron as a Novel Lariat Structure during Pre-Messenger Rna Splicing Invitro.* Cell, 1984. 38(1): p. 317-331.
40. Sharp, P.A., *On the origin of RNA splicing and introns.* Cell, 1985. 42(2): p. 397-400.
41. Cech, T.R., *The generality of self-splicing RNA: relationship to nuclear mRNA splicing.* Cell, 1986. 44(2): p. 207-10.
42. Wahl, M.C., C.L. Will, and R. Luhrmann, *The spliceosome: design principles of a dynamic RNP machine.* Cell, 2009. 136(4): p. 701-18.
43. Jackson, I.J., *A reappraisal of non-consensus mRNA splice sites.* Nucleic Acids Res, 1991. 19(14): p. 3795-8.
44. Lacadie, S.A. and M. Rosbash, *Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast.* Mol Cell, 2005. 19(1): p. 65-75.
45. Zamore, P.D. and M.R. Green, *Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor.* Proc Natl Acad Sci U S A, 1989. 86(23): p. 9243-7.
46. Crispino, J.D., et al., *Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing.* RNA, 1996. 2(7): p. 664-73.

47. Tacke, R. and J.L. Manley, *Determinants of SR protein specificity.* Curr Opin Cell Biol, 1999. 11(3): p. 358-62.
48. Zhu, J., A. Mayeda, and A.R. Krainer, *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins.* Mol Cell, 2001. 8(6): p. 1351-61.
49. Bourgeois, C.F., et al., *Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation.* Mol Cell Biol, 1999. 19(11): p. 7347-56.
50. Zuo, P. and T. Maniatis, *The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing.* Genes Dev, 1996. 10(11): p. 1356-68.
51. Longman, D., et al., *Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of C. elegans.* Curr Biol, 2001. 11(24): p. 1923-33.
52. Tacke, R. and J.L. Manley, *Functions of SR and Tra2 proteins in pre-mRNA splicing regulation.* Proc Soc Exp Biol Med, 1999. 220(2): p. 59-63.
53. Feng, Y., M. Chen, and J.L. Manley, *Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator.* Nat Struct Mol Biol, 2008. 15(10): p. 1040-8.
54. Shin, C., Y. Feng, and J.L. Manley, *Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock.* Nature, 2004. 427(6974): p. 553-8.
55. Dreyfuss, G., et al., *hnRNP proteins and the biogenesis of mRNA.* Annu Rev Biochem, 1993. 62: p. 289-321.
56. House, A.E. and K.W. Lynch, *An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition.* Nat Struct Mol Biol, 2006. 13(10): p. 937-44.
57. Forch, P., et al., *The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing.* Mol Cell, 2000. 6(5): p. 1089-98.
58. Expert-Bezancon, A., et al., *hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B.* J Biol Chem, 2004. 279(37): p. 38249-59.
59. Crawford, J.B. and J.G. Patton, *Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F.* Mol Cell Biol, 2006. 26(23): p. 8791-802.
60. Min, H., R.C. Chan, and D.L. Black, *The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event.* Genes Dev, 1995. 9(21): p. 2659-71.
61. Chou, M.Y., et al., *hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells.* Mol Cell Biol, 1999. 19(1): p. 69-77.
62. Castle, J.C., et al., *Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines.* Nature Genetics, 2008. 40(12): p. 1416-1425.
63. Li, Q., J.A. Lee, and D.L. Black, *Neuronal regulation of alternative pre-mRNA splicing.* Nat Rev Neurosci, 2007. 8(11): p. 819-31.
64. Boutz, P.L., et al., *A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons.* Genes Dev, 2007. 21(13): p. 1636-52.
65. Dredge, B.K., et al., *Nova autoregulation reveals dual functions in neuronal splicing.* EMBO J, 2005. 24(8): p. 1608-20.
66. Dredge, B.K. and R.B. Darnell, *Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer.* Mol Cell Biol, 2003. 23(13): p. 4687-700.
67. Warzecha, C.C., et al., *ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing.* Mol Cell, 2009. 33(5): p. 591-601.
68. Mallinjoud, P., et al., *Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin.* Genome Research, 2014. 24(3): p. 511-521.
69. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C-Elegans Heterochronic Gene Lin-4 Encodes Small Rnas with Antisense Complementarity to Lin-14.* Cell, 1993. 75(5): p. 843-854.

70. **Reinhart, B.J., et al.,** ***The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.*** **Nature, 2000. 403(6772): p. 901-6.**
71. **Sempere, L.F., et al.,** ***Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation.*** **Genome Biol, 2004. 5(3): p. R13.**
72. **Brennecke, J., et al.,** ***bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila.*** **Cell, 2003. 113(1): p. 25-36.**
73. **Wienholds, E., et al.,** ***The microRNA-producing enzyme Dicer1 is essential for zebrafish development.*** **Nat Genet, 2003. 35(3): p. 217-8.**
74. **Kozomara, A. and S. Griffiths-Jones,** ***miRBase: annotating high confidence microRNAs using deep sequencing data.*** **Nucleic Acids Res, 2014. 42(Database issue): p. D68-73.**
75. **Monteys, A.M., et al.,** ***Structure and activity of putative intronic miRNA promoters.*** **Rna-a Publication of the Rna Society, 2010. 16(3): p. 495-505.**
76. **Ozsolak, F., et al.,** ***Chromatin structure analyses identify miRNA promoters.*** **Genes & Development, 2008. 22(22): p. 3172-3183.**
77. **Lee, Y., et al.,** ***MicroRNA genes are transcribed by RNA polymerase II.*** **EMBO J, 2004. 23(20): p. 4051-60.**
78. **Bartel, D.P.,** ***MicroRNAs: Genomics, biogenesis, mechanism, and function.*** **Cell, 2004. 116(2): p. 281-297.**
79. **Chen, K. and N. Rajewsky,** ***Deep conservation of microRNA-target relationships and 3'UTR motifs in vertebrates, flies, and nematodes.*** **Cold Spring Harb Symp Quant Biol, 2006. 71: p. 149-56.**
80. **Marco, A., M. Ninova, and S. Griffiths-Jones,** ***Multiple products from microRNA transcripts.*** **Biochemical Society Transactions, 2013. 41: p. 850-854.**
81. **Kim, V.N.,** ***MicroRNA biogenesis: coordinated cropping and dicing.*** **Nat Rev Mol Cell Biol, 2005. 6(5): p. 376-85.**
82. **Pfeffer, S., et al.,** ***Identification of microRNAs of the herpesvirus family.*** **Nat Methods, 2005. 2(4): p. 269-76.**
83. **Lee, Y., et al.,** ***The nuclear RNase III Drosha initiates microRNA processing.*** **Nature, 2003. 425(6956): p. 415-9.**
84. **Gregory, R.I., et al.,** ***The Microprocessor complex mediates the genesis of microRNAs.*** **Nature, 2004. 432(7014): p. 235-40.**
85. **Han, J., et al.,** ***The Drosha-DGCR8 complex in primary microRNA processing.*** **Genes Dev, 2004. 18(24): p. 3016-27.**
86. **Ha, M. and V.N. Kim,** ***Regulation of microRNA biogenesis.*** **Nat Rev Mol Cell Biol, 2014. 15(8): p. 509-24.**
87. **Bohnsack, M.T., K. Czaplinski, and D. Gorlich,** ***Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs.*** **Rna-a Publication of the Rna Society, 2004. 10(2): p. 185-191.**
88. **Krol, J., I. Loedige, and W. Filipowicz,** ***The widespread regulation of microRNA biogenesis, function and decay.*** **Nat Rev Genet, 2010. 11(9): p. 597-610.**
89. **He, L. and G.J. Hannon,** ***MicroRNAs: small RNAs with a big role in gene regulation.*** **Nat Rev Genet, 2004. 5(7): p. 522-31.**
90. **Hutvagner, G., et al.,** ***A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.*** **Science, 2001. 293(5531): p. 834-8.**
91. **Chendrimada, T.P., et al.,** ***TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing.*** **Nature, 2005. 436(7051): p. 740-4.**

92. Okamura, K., N. Liu, and E.C. Lai, *Distinct mechanisms for microRNA strand selection by Drosophila Argonautes.* Mol Cell, 2009. 36(3): p. 431-44.
93. Czech, B., et al., *Hierarchical rules for Argonaute loading in Drosophila.* Mol Cell, 2009. 36(3): p. 445-56.
94. Khvorova, A., A. Reynolds, and S.D. Jayasena, *Functional siRNAs and miRNAs exhibit strand bias.* Cell, 2003. 115(2): p. 209-16.
95. Chiang, H.R., et al., *Mammalian microRNAs: experimental evaluation of novel and previously annotated genes.* Genes Dev, 2010. 24(10): p. 992-1009.
96. Wu, H., et al., *Alternative processing of primary microRNA transcripts by Drosha generates 5' end variation of mature microRNA.* PLoS One, 2009. 4(10): p. e7566.
97. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.* Nature, 1998. 391(6669): p. 806-811.
98. Hammond, S.M., et al., *Argonaute2, a link between genetic and biochemical analyses of RNAi.* Science, 2001. 293(5532): p. 1146-50.
99. Huntzinger, E. and E. Izaurralde, *Gene silencing by microRNAs: contributions of translational repression and mRNA decay.* Nat Rev Genet, 2011. 12(2): p. 99-110.
100. Meister, G., *Argonaute proteins: functional insights and emerging roles.* Nat Rev Genet, 2013. 14(7): p. 447-59.
101. Elkayam, E., et al., *The structure of human argonaute-2 in complex with miR-20a.* Cell, 2012. 150(1): p. 100-10.
102. Kiriakidou, M., et al., *An mRNA m7G cap binding-like motif within human Ago2 represses translation.* Cell, 2007. 129(6): p. 1141-51.
103. Hutvagner, G. and M.J. Simard, *Argonaute proteins: key players in RNA silencing.* Nat Rev Mol Cell Biol, 2008. 9(1): p. 22-32.
104. Lewis, B.P., et al., *Prediction of mammalian microRNA targets.* Cell, 2003. 115(7): p. 787-98.
105. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions.* Cell, 2009. 136(2): p. 215-33.
106. Ameres, S.L. and P.D. Zamore, *Diversifying microRNA sequence and function.* Nat Rev Mol Cell Biol, 2013. 14(8): p. 475-88.
107. Obernosterer, G., H. Tafer, and J. Martinez, *Target site effects in the RNA interference and microRNA pathways.* Biochem Soc Trans, 2008. 36(Pt 6): p. 1216-9.
108. Kedde, M., et al., *RNA-binding protein Dnd1 inhibits microRNA access to target mRNA.* Cell, 2007. 131(7): p. 1273-86.
109. Hamilton, A.J. and D.C. Baulcombe, *A species of small antisense RNA in posttranscriptional gene silencing in plants.* Science, 1999. 286(5441): p. 950-952.
110. Merrick, W.C., *Cap-dependent and cap-independent translation in eukaryotic systems.* Gene, 2004. 332: p. 1-11.
111. Thermann, R. and M.W. Hentze, *Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation.* Nature, 2007. 447(7146): p. 875-8.
112. Kim, J., et al., *Identification of many microRNAs that copurify with polyribosomes in mammalian neurons.* Proceedings of the National Academy of Sciences of the United States of America, 2004. 101(1): p. 360-365.
113. Nottrott, S., M.J. Simard, and J.D. Richter, *Human let-7a miRNA blocks protein production on actively translating polyribosomes.* Nat Struct Mol Biol, 2006. 13(12): p. 1108-14.
114. Rehwinkel, J., et al., *Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster.* Mol Cell Biol, 2006. 26(8): p. 2965-75.

115. Behm-Ansmant, I., et al., *MRNA degradation by miRNAs and GW182 requires both CCR4 : NOT deadenylase and DCP1 : DCP2 decapping complexes.* Genes & Development, 2006. 20(14): p. 1885-1898.
116. Wu, L.G., J.H. Fan, and J.G. Belasco, *MicroRNAs direct rapid deadenylation of mRNA.* Proceedings of the National Academy of Sciences of the United States of America, 2006. 103(11): p. 4034-4039.
117. Parker, R. and H. Song, *The enzymes and control of eukaryotic mRNA turnover.* Nat Struct Mol Biol, 2004. 11(2): p. 121-7.
118. Eulalio, A., et al., *P-body formation is a consequence, not the cause, of RNA-mediated gene silencing.* Molecular and Cellular Biology, 2007. 27(11): p. 3970-3981.
119. Kanellopoulou, C., et al., *Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing.* Genes Dev, 2005. 19(4): p. 489-501.
120. Chen, J.F., et al., *Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure.* Proc Natl Acad Sci U S A, 2008. 105(6): p. 2111-6.
121. Tay, Y., et al., *MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation (vol 445, pg 1124, 2008).* Nature, 2009. 458(7237): p. 538-538.
122. Lim, L.P., et al., *Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.* Nature, 2005. 433(7027): p. 769-773.
123. Elmen, J., et al., *Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver.* Nucleic Acids Research, 2008. 36(4): p. 1153-1162.
124. Frankel, L.B., et al., *Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells.* Journal of Biological Chemistry, 2008. 283(2): p. 1026-1033.
125. Krutzfeldt, J., et al., *Silencing of microRNAs in vivo with 'antagomirs'.* Nature, 2005. 438(7068): p. 685-689.
126. Gaidatzis, D., et al., *Inference of miRNA targets using evolutionary conservation and pathway analysis (vol 8, pg 248, 2007).* Bmc Bioinformatics, 2007. 8.
127. Grun, D., et al., *microRNA target predictions across seven Drosophila species and comparison to mammalian targets.* PLoS Comput Biol, 2005. 1(1): p. e13.
128. John, B., et al., *Human MicroRNA targets.* Plos Biology, 2004. 2(11): p. 1862-1879.
129. Stark, A., et al., *Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures.* Nature, 2007. 450(7167): p. 219-232.
130. Vinther, J., et al., *Identification of miRNA targets with stable isotope labeling by amino acids in cell culture.* Nucleic Acids Research, 2006. 34(16): p. e107.
131. Selbach, M., et al., *Widespread changes in protein synthesis induced by microRNAs.* Nature, 2008. 455(7209): p. 58-63.
132. Baek, D., et al., *The impact of microRNAs on protein output.* Nature, 2008. 455(7209): p. 64-71.
133. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. 74(12): p. 5463-7.
134. Mortazavi, A., et al., *Mapping and quantifying mammalian transcriptomes by RNA-Seq.* Nat Methods, 2008. 5(7): p. 621-8.
135. Pan, Q., et al., *Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing.* Nat Genet, 2008. 40(12): p. 1413-5.
136. Martin, G., et al., *Genome-wide Analysis of Pre-mRNA 3 ' End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3 ' UTR Length.* Cell Reports, 2012. 1(6): p. 753-763.
137. Ule, J., et al., *CLIP identifies Nova-regulated RNA networks in the brain.* Science, 2003. 302(5648): p. 1212-5.

138. **Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.* Nature, 2009. 460(7254): p. 479-86.**
139. **Kishore, S., et al., *Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing.* Genome Biology, 2013. 14(5).**
140. **Hafner, M., et al., *PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins.* J Vis Exp, 2010(41).**
141. **Khorshid, M., et al., *A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets.* Nat Methods, 2013. 10(3): p. 253-5.**
142. **Bornkamm, G.W., et al., *Stringent doxycycline-dependent control of gene activities using an episomal one-vector system.* Nucleic Acids Res, 2005. 33(16): p. e137.**
143. **Lal, A., et al., *miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements.* Mol Cell, 2009. 35(5): p. 610-25.**
144. **Abdelmohsen, K., et al., *miR-519 suppresses tumor growth by reducing HuR levels.* Cell Cycle, 2010. 9(7): p. 1354-9.**
145. **Lanet, E., et al., *Biochemical evidence for translational repression by Arabidopsis microRNAs.* Plant Cell, 2009. 21(6): p. 1762-8.**
146. **Brodersen, P., et al., *Widespread translational inhibition by plant miRNAs and siRNAs.* Science, 2008. 320(5880): p. 1185-90.**
147. **Forman, J.J., A. Legesse-Miller, and H.A. Coller, *A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence.* Proceedings of the National Academy of Sciences of the United States of America, 2008. 105(39): p. 14879-14884.**
148. **Huang, F.W., et al., *Target prediction and a statistical sampling algorithm for RNA-RNA interaction.* Bioinformatics, 2010. 26(2): p. 175-81.**
149. **Reczko, M., et al., *Functional microRNA targets in protein coding sequences.* Bioinformatics, 2012. 28(6): p. 771-6.**
150. **Fang, Z. and N. Rajewsky, *The impact of miRNA target sites in coding sequences and in 3'UTRs.* PLoS One, 2011. 6(3): p. e18067.**
151. **Guo, H., et al., *Mammalian microRNAs predominantly act to decrease target mRNA levels.* Nature, 2010. 466(7308): p. 835-40.**
152. **Epanchintsev, A., et al., *Inducible microRNA expression by an all-in-one episomal vector system.* Nucleic Acids Res, 2006. 34(18): p. e119.**
153. **Sambrook, J. and D.W. Russell, *The condensed protocols from Molecular cloning : a laboratory manual*2006, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. v, 800 p.**
154. **Pall, G.S. and A.J. Hamilton, *Improved northern blot method for enhanced detection of small RNA.* Nat Protoc, 2008. 3(6): p. 1077-84.**
155. **Wee, L.M., et al., *Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties.* Cell, 2012. 151(5): p. 1055-67.**
156. **Shimoni, Y., et al., *Regulation of gene expression by small non-coding RNAs: a quantitative view.* Mol Syst Biol, 2007. 3: p. 138.**
157. **Khan, A.A., et al., *Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs.* Nat Biotechnol, 2009. 27(6): p. 549-55.**
158. **Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing.* Mol Cell, 2007. 27(1): p. 91-105.**
159. **Hausser, J., et al., *Relative contribution of sequence and structure features to the mRNA binding of Argonaute/EIF2C-miRNA complexes and the degradation of miRNA targets.* Genome Res, 2009. 19(11): p. 2009-20.**

160. Karginov, F.V., et al., *A biochemical approach to identifying microRNA targets.* Proc Natl Acad Sci U S A, 2007. 104(49): p. 19291-6.
161. Kishore, S., et al., *A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins.* Nat Methods, 2011. 8(7): p. 559-64.
162. Kelly, S.J., *Studies of Developmental Potential of 4-Cell and 8-Cell Stage Mouse Blastomeres.* Journal of Experimental Zoology, 1977. 200(3): p. 365-376.
163. Tarkowski, A.K., *Experiments on the Development of Isolated Blastomeres of Mouse Eggs.* Nature, 1959. 184(4695): p. 1286-1287.
164. Arnold, S.J. and E.J. Robertson, *Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo.* Nat Rev Mol Cell Biol, 2009. 10(2): p. 91-103.
165. Niwa, H., et al., *Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation.* Cell, 2005. 123(5): p. 917-29.
166. Strumpf, D., et al., *Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst.* Development, 2005. 132(9): p. 2093-102.
167. Niwa, H., J. Miyazaki, and A.G. Smith, *Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells.* Nat Genet, 2000. 24(4): p. 372-6.
168. Chazaud, C., et al., *Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway.* Developmental Cell, 2006. 10(5): p. 615-624.
169. Fujikura, J., et al., *Differentiation of embryonic stem cells is induced by GATA factors.* Genes Dev, 2002. 16(7): p. 784-9.
170. Chambers, I., et al., *Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells.* Cell, 2003. 113(5): p. 643-55.
171. Yagi, R., et al., *Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development.* Development, 2007. 134(21): p. 3827-36.
172. Nishioka, N., et al., *Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos.* Mech Dev, 2008. 125(3-4): p. 270-83.
173. Solter, D., N. Skreb, and I. Damjanov, *Extrauterine Growth of Mouse Egg-Cylinders Results in Malignant Teratoma.* Nature, 1970. 227(5257): p. 503-&.
174. Stevens, L.C., *The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos.* Dev Biol, 1970. 21(3): p. 364-82.
175. Papaioannou, V.E., et al., *Fate of Teratocarcinoma Cells Injected into Early Mouse Embryos.* Nature, 1975. 258(5530): p. 70-73.
176. Martin, G.R., L.M. Wiley, and I. Damjanov, *The development of cystic embryoid bodies in vitro from clonal teratocarcinoma stem cells.* Dev Biol, 1977. 61(2): p. 230-44.
177. Evans, M.J. and M.H. Kaufman, *Establishment in culture of pluripotential cells from mouse embryos.* Nature, 1981. 292(5819): p. 154-6.
178. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. 78(12): p. 7634-8.
179. Bradley, A., et al., *Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines.* Nature, 1984. 309(5965): p. 255-6.
180. Doetschman, T.C., et al., *The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium.* J Embryol Exp Morphol, 1985. 87: p. 27-45.
181. Nagy, A., et al., *Embryonic stem cells alone are able to support fetal development in the mouse.* Development, 1990. 110(3): p. 815-21.
182. Nagy, A., et al., *Derivation of completely cell culture-derived mice from early-passage embryonic stem cells.* Proc Natl Acad Sci U S A, 1993. 90(18): p. 8424-8.

183. **Smith, A.G., et al., *Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides.* Nature, 1988. 336(6200): p. 688-90.**
184. **Niwa, H., et al., *A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells.* Nature, 2009. 460(7251): p. 118-22.**
185. **Ying, Q.L., et al., *BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3.* Cell, 2003. 115(3): p. 281-92.**
186. **Wobus, A.M. and K.R. Boheler, *Embryonic stem cells: prospects for developmental biology and cell therapy.* Physiol Rev, 2005. 85(2): p. 635-78.**
187. **Matsui, Y., K. Zsebo, and B.L. Hogan, *Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture.* Cell, 1992. 70(5): p. 841-7.**
188. **Gurdon, J.B., *The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles.* J Embryol Exp Morphol, 1962. 10: p. 622-40.**
189. **Wabl, M.R., R.B. Brun, and L. Du Pasquier, *Lymphocytes of the toad Xenopus laevis have the gene set for promoting tadpole development.* Science, 1975. 190(4221): p. 1310-2.**
190. **Wilmut, I., et al., *Viable offspring derived from fetal and adult mammalian cells.* Nature, 1997. 385(6619): p. 810-3.**
191. **Schneuwly, S., R. Klemenz, and W.J. Gehring, *Redesigning the body plan of Drosophila by ectopic expression of the homoeotic gene Antennapedia.* Nature, 1987. 325(6107): p. 816-8.**
192. **Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. 131(5): p. 861-72.**
193. **Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. 318(5858): p. 1917-20.**
194. **van den Berg, D.L., et al., *An Oct4-centered protein interaction network in embryonic stem cells.* Cell Stem Cell, 2010. 6(4): p. 369-81.**
195. **Velkey, J.M. and K.S. O'Shea, *Oct4 RNA interference induces trophectoderm differentiation in mouse embryonic stem cells.* Genesis, 2003. 37(1): p. 18-24.**
196. **Hay, D.C., et al., *Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells.* Stem Cells, 2004. 22(2): p. 225-35.**
197. **Avilion, A.A., et al., *Multipotent cell lineages in early mouse development depend on SOX2 function.* Genes Dev, 2003. 17(1): p. 126-40.**
198. **Wang, S.H., et al., *A novel NK-type homeobox gene, ENK (early embryo specific NK), preferentially expressed in embryonic stem cells.* Gene Expr Patterns, 2003. 3(1): p. 99-103.**
199. **Mullin, N.P., et al., *The pluripotency rheostat Nanog functions as a dimer.* Biochemical Journal, 2008. 411: p. 227-231.**
200. **Wang, Z., et al., *Aromatic residues in the C-terminal domain 2 are required for Nanog to mediate LIF-independent self-renewal of mouse embryonic stem cells.* J Biol Chem, 2008. 283(8): p. 4480-9.**
201. **Rowland, B.D. and D.S. Peeper, *KLF4, p21 and context-dependent opposing forces in cancer.* Nat Rev Cancer, 2006. 6(1): p. 11-23.**
202. **Zhang, P., et al., *Kruppel-like factor 4 (Klf4) prevents embryonic stem (ES) cell differentiation by regulating Nanog gene expression.* J Biol Chem, 2010. 285(12): p. 9180-9.**
203. **Blackwell, T.K., et al., *Binding of myc proteins to canonical and noncanonical DNA sequences.* Mol Cell Biol, 1993. 13(9): p. 5216-24.**
204. **Trumpp, A., et al., *c-Myc regulates mammalian body size by controlling cell number but not cell size.* Nature, 2001. 414(6865): p. 768-773.**
205. **Feng, B., et al., *Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb.* Nat Cell Biol, 2009. 11(2): p. 197-203.**

206. Kim, J., et al., *An extended transcriptional network for pluripotency of embryonic stem cells.* Cell, 2008. 132(6): p. 1049-61.
207. Kuroda, T., et al., *Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression.* Mol Cell Biol, 2005. 25(6): p. 2475-85.
208. Okumura-Nakanishi, S., et al., *Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells.* J Biol Chem, 2005. 280(7): p. 5307-17.
209. Rodda, D.J., et al., *Transcriptional regulation of nanog by OCT4 and SOX2.* J Biol Chem, 2005. 280(26): p. 24731-7.
210. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. 122(6): p. 947-56.
211. Kunarso, G., et al., *Transposable elements have rewired the core regulatory network of human embryonic stem cells.* Nat Genet, 2010. 42(7): p. 631-4.
212. Marson, A., et al., *Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.* Cell, 2008. 134(3): p. 521-33.
213. Carey, B.W., et al., *Reprogramming of murine and human somatic cells using a single polycistronic vector.* Proc Natl Acad Sci U S A, 2009. 106(1): p. 157-62.
214. Soldner, F., et al., *Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors.* Cell, 2009. 136(5): p. 964-77.
215. Zhou, W. and C.R. Freed, *Adenoviral gene delivery can reprogram human fibroblasts to induced pluripotent stem cells.* Stem Cells, 2009. 27(11): p. 2667-74.
216. Fusaki, N., et al., *Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome.* Proc Jpn Acad Ser B Phys Biol Sci, 2009. 85(8): p. 348-62.
217. Kim, D., et al., *Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins.* Cell Stem Cell, 2009. 4(6): p. 472-6.
218. Okita, K., et al., *Generation of mouse induced pluripotent stem cells without viral vectors.* Science, 2008. 322(5903): p. 949-53.
219. Narsinh, K.H., et al., *Generation of adult human induced pluripotent stem cells using nonviral minicircle DNA vectors.* Nat Protoc, 2011. 6(1): p. 78-88.
220. Warren, L., et al., *Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA.* Cell Stem Cell, 2010. 7(5): p. 618-30.
221. Woltjen, K., et al., *piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells.* Nature, 2009. 458(7239): p. 766-70.
222. Wang, T., et al., *The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner.* Cell Stem Cell, 2011. 9(6): p. 575-87.
223. Subramanyam, D., et al., *Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells.* Nat Biotechnol, 2011. 29(5): p. 443-8.
224. Anokye-Danso, F., et al., *Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency.* Cell Stem Cell, 2011. 8(4): p. 376-88.
225. Wernig, M., et al., *A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types.* Nat Biotechnol, 2008. 26(8): p. 916-24.
226. Friedrich, G. and P. Soriano, *Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice.* Genes Dev, 1991. 5(9): p. 1513-23.
227. Haenebalcke, L., et al., *The ROSA26-iPSC mouse: a conditional, inducible, and exchangeable resource for studying cellular (De)differentiation.* Cell Rep, 2013. 3(2): p. 335-41.
228. Samavarchi-Tehrani, P., et al., *Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming.* Cell Stem Cell, 2010. 7(1): p. 64-77.

229. **Polo, J.M., et al., *A molecular roadmap of reprogramming somatic cells into iPS cells.* Cell, 2012. 151(7): p. 1617-32.**
230. **Buganim, Y., et al., *Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase.* Cell, 2012. 150(6): p. 1209-22.**
231. **Wang, G., et al., *Critical regulation of miR-200/ZEB2 pathway in Oct4/Sox2-induced mesenchymal-to-epithelial transition and induced pluripotent stem cell generation.* Proc Natl Acad Sci U S A, 2013. 110(8): p. 2858-63.**
232. **Utikal, J., et al., *Immortalization eliminates a roadblock during cellular reprogramming into iPS cells.* Nature, 2009. 460(7259): p. 1145-8.**
233. **Lin, C.Y., et al., *Transcriptional amplification in tumor cells with elevated c-Myc.* Cell, 2012. 151(1): p. 56-67.**
234. **Soufi, A., G. Donahue, and K.S. Zaret, *Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome.* Cell, 2012. 151(5): p. 994-1004.**
235. **O'Malley, J., et al., *High-resolution analysis with novel cell-surface markers identifies routes to iPS cells.* Nature, 2013. 499(7456): p. 88-91.**
236. **Golipour, A., et al., *A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network.* Cell Stem Cell, 2012. 11(6): p. 769-82.**
237. **Polo, J.M., et al., *Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells.* Nat Biotechnol, 2010. 28(8): p. 848-55.**
238. **Kim, K., et al., *Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells.* Nat Biotechnol, 2011. 29(12): p. 1117-9.**
239. **Hansson, J., et al., *Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency.* Cell Rep, 2012. 2(6): p. 1579-92.**
240. **Zhu, J., et al., *Genome-wide chromatin state transitions associated with developmental and environmental cues.* Cell, 2013. 152(3): p. 642-54.**
241. **Rais, Y., et al., *Deterministic direct reprogramming of somatic cells to pluripotency.* Nature, 2013. 502(7469): p. 65-70.**
242. **Buganim, Y., D.A. Faddah, and R. Jaenisch, *Mechanisms and models of somatic cell reprogramming.* Nat Rev Genet, 2013. 14(6): p. 427-39.**
243. **Koche, R.P., et al., *Reprogramming factor expression initiates widespread targeted chromatin remodeling.* Cell Stem Cell, 2011. 8(1): p. 96-105.**
244. **Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nat Struct Mol Biol, 2006. 13(8): p. 713-9.**
245. **Kidder, B.L., et al., *Extended self-renewal and accelerated reprogramming in the absence of Kdm5b.* Mol Cell Biol, 2013. 33(24): p. 4793-810.**
246. **Kohli, R.M. and Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation.* Nature, 2013. 502(7472): p. 472-9.**
247. **Onder, T.T., et al., *Chromatin-modifying enzymes as modulators of reprogramming.* Nature, 2012. 483(7391): p. 598-602.**
248. **Bernstein, E., et al., *Dicer is essential for mouse development.* Nature Genetics, 2003. 35(3): p. 215-217.**
249. **Suh, M.R., et al., *Human embryonic stem cells express a unique set of microRNAs.* Dev Biol, 2004. 270(2): p. 488-98.**
250. **Bar, M., et al., *MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries.* Stem Cells, 2008. 26(10): p. 2496-505.**
251. **Tay, Y., et al., *MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation.* Nature, 2008. 455(7216): p. 1124-8.**

252. Xu, N., et al., *MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells.* Cell, 2009. 137(4): p. 647-58.
253. Gill, J.G., et al., *Snail and the microRNA-200 family act in opposition to regulate epithelial-to-mesenchymal transition and germ layer fate restriction in differentiating ESCs.* Stem Cells, 2011. 29(5): p. 764-76.
254. Judson, R.L., et al., *Embryonic stem cell-specific microRNAs promote induced pluripotency.* Nat Biotechnol, 2009. 27(5): p. 459-61.
255. Judson, R.L., et al., *MicroRNA-based discovery of barriers to dedifferentiation of fibroblasts to pluripotent stem cells.* Nat Struct Mol Biol, 2013. 20(10): p. 1227-35.
256. Zhao, Y., et al., *Two supporting factors greatly improve the efficiency of human iPSC generation.* Cell Stem Cell, 2008. 3(5): p. 475-9.
257. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* J Clin Invest, 2009. 119(6): p. 1420-8.
258. Li, Z., et al., *Small RNA-mediated regulation of iPS cell generation.* EMBO J, 2011. 30(5): p. 823-34.
259. Miyoshi, N., et al., *Reprogramming of Mouse and Human Cells to Pluripotency Using Mature MicroRNAs.* Cell Stem Cell, 2011. 8(6): p. 633-638.
260. Lin, S.L., et al., *Regulation of somatic cell reprogramming through inducible mir-302 expression.* Nucleic Acids Research, 2011. 39(3): p. 1054-1065.
261. Lin, S.L., et al., *Mir-302 reprograms human skin cancer cells into a pluripotent ES-cell-like state.* Rna-a Publication of the Rna Society, 2008. 14(10): p. 2115-2124.
262. Houbaviy, H.B., M.F. Murray, and P.A. Sharp, *Embryonic stem cell-specific MicroRNAs.* Dev Cell, 2003. 5(2): p. 351-8.
263. Card, D.A., et al., *Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells.* Mol Cell Biol, 2008. 28(20): p. 6426-38.
264. Lipchina, I., et al., *Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response.* Genes Dev, 2011. 25(20): p. 2173-86.
265. Lu, D., et al., *MiR-25 regulates Wwp2 and Fbxw7 and promotes reprogramming of mouse fibroblast cells to iPSCs.* PLoS One, 2012. 7(8): p. e40938.
266. Hu, S., et al., *MicroRNA-302 increases reprogramming efficiency via repression of NR2F2.* Stem Cells, 2013. 31(2): p. 259-68.
267. Peterkofsky, B., *Ascorbate requirement for hydroxylation and secretion of procollagen: relationship to inhibition of collagen synthesis in scurvy.* Am J Clin Nutr, 1991. 54(6 Suppl): p. 1135S-1140S.
268. Hwang, W.S., et al., *Ascorbic acid extends replicative life span of human embryonic fibroblast by reducing DNA and mitochondrial damages.* Nutr Res Pract, 2007. 1(2): p. 105-12.
269. Esteban, M.A., et al., *Vitamin C enhances the generation of mouse and human induced pluripotent stem cells.* Cell Stem Cell, 2010. 6(1): p. 71-9.
270. Chen, J., et al., *Vitamin C modulates TET1 function during somatic cell reprogramming.* Nature Genetics, 2013. 45(12): p. 1504-9.
271. Maherali, N. and K. Hochedlinger, *Tgfbeta signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc.* Curr Biol, 2009. 19(20): p. 1718-23.
272. Atlasi, Y., et al., *OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells.* Stem Cells, 2008. 26(12): p. 3068-74.
273. Cheong, C.Y. and T. Lufkin, *Alternative splicing in self-renewal of embryonic stem cells.* Stem Cells Int, 2011. 2011: p. 560261.
274. Das, S., S. Jena, and D.N. Levasseur, *Alternative splicing produces Nanog protein variants with different capacities for self-renewal and pluripotency in embryonic stem cells.* J Biol Chem, 2011. 286(49): p. 42690-703.

275. **Gabut, M., et al., *An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming.* Cell, 2011. 147(1): p. 132-46.**
276. **Han, H., et al., *MBNL proteins repress ES-cell-specific alternative splicing and reprogramming.* Nature, 2013. 498(7453): p. 241-5.**
277. **Venables, J.P., et al., *MBNL1 and RBFOX2 cooperate to establish a splicing programme involved in pluripotent stem cell differentiation.* Nat Commun, 2013. 4: p. 2480.**
278. **Lu, Y., et al., *Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells.* Cell Stem Cell, 2014. 15(1): p. 92-101.**
279. **Abyzov, A., et al., *Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells.* Nature, 2012. 492(7429): p. 438-42.**
280. **Yanagisawa, M., et al., *A p120 catenin isoform switch affects Rho activity, induces tumor cell invasion, and predicts metastatic disease.* J Biol Chem, 2008. 283(26): p. 18344-54.**
281. **Fagoonee, S., et al., *The RNA binding protein ESRP1 fine-tunes the expression of pluripotency-related factors in mouse embryonic stem cells.* PLoS One, 2013. 8(8): p. e72300.**
282. **Sakakibara, S., et al., *Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell.* Dev Biol, 1996. 176(2): p. 230-42.**
283. **Glazer, R.I., D.T. Vo, and L.O. Penalva, *Musashi1: an RBP with versatile functions in normal and cancer stem cells.* Front Biosci (Landmark Ed), 2012. 17: p. 54-64.**
284. **Katz, Y., et al., *Musashi proteins are post-transcriptional regulators of the epithelial-luminal cell state.* Elife, 2014. 3: p. e03915.**
285. **Wang, X.Y., et al., *Musashi1 regulates breast tumor cell proliferation and is a prognostic indicator of poor survival.* Mol Cancer, 2010. 9: p. 221.**
286. **Maeda, M., K.R. Johnson, and M.J. Wheelock, *Cadherin switching: essential for behavioral but not morphological changes during an epithelium-to-mesenchyme transition.* J Cell Sci, 2005. 118(Pt 5): p. 873-87.**
287. **Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. 25(4): p. 402-8.**
288. **Nethercott, H.E., D.J. Brick, and P.H. Schwartz, *Derivation of induced pluripotent stem cells by lentiviral transduction.* Methods Mol Biol, 2011. 767: p. 67-85.**
289. **Nikpour, P., et al., *The stem cell self-renewal gene, Musashi 1, is highly expressed in tumor and non-tumor samples of human bladder.* Indian J Cancer, 2013. 50(3): p. 214-8.**