universität wien

DISSERTATION

Titel der Dissertation

Determination of a general mode of microRNA action and functional characterization of specific microRNAs in *Drosophila melanogaster*

angestrebter akademischer Grad

Doktorin der Naturwissenschaften (Dr. rer.nat.)

Verfasserin / Verfasser:	Natascha Bushati
Matrikel-Nummer:	9725217
Dissertationsgebiet (lt. Studienblatt):	Molekulare Biologie
Betreuerin / Betreuer:	Univ.-Prof. Dr. Renée Schröder

Heidelberg, am 10. Januar 2008

Table of Contents

Table of Contents

Acknowledgements

First of all, I thank Stephen Cohen for having me in his lab, being highly supportive, for sharing his enthusiasm and almost boundless optimism, for his understanding, his flexibility, for working in shifts in different time zones, and for reassuring me when things seemed so scary.

I would like to thank the members of my thesis advisory committee, Pernille Rørth and Lars Steinmetz for discussions and helpful advice. I also would like to thank Milanka Stojkovic for her straightforward help with PhD program and contract issues. I am grateful to my supervisor at the University of Vienna, Renée Schröder, and to Barbara Hamilton for her flexibility and help with university issues.

I am deeply grateful to Alex Stark. It was great working with him and I thank him for our fruitful collaborations, for calculating last minute P-values while traveling and for being a **friend.

From the Cohen Lab, I would like to thank Barry Thompson for lots of advice, both scientific and personal, even if I rarely took it, and for being a friend. I especially thank Ville Hietakangas for many late-night discussions, sharing Finnish ‘news’ and pedometers with me, for fixing my car tyre with me in the rain (unforgettable!), for caring and understanding and being a friend. Thanks to Eva Löser for experimental help, even when far away, and for fun conversations. Thanks to Yawen Chen and Sandrine Pizette for the friendly atmosphere in our ‘chick’s bay’. I thank Julius Brennecke, Aurelio Teleman, Janina Karres, Yawen Chen, Thomas Sandmann, Sandrine Pizette, Dagmar Hipfner and Russ Collins for their help and advice, especially in the beginning. During the last months at EMBL, the lab changed a lot, and I am very happy to have been part of the ‘merged leftovers’. Thanks to Valérie Hilgers, Janina Karres, Sandra Mueller and Ambra Bianco for the girl power in the last months in Heidelberg, it was just what I needed. Thanks also to Valérie Hilgers and Clemens Grabher for proof-reading parts of the thesis.

The time in Heidelberg was not the happiest in my life. Nevertheless, or maybe because of this, I made wonderful friendships, some of which I believe will last a long time. Apart from those who were luckily part of the Cohen lab and therefore (almost) always easily accessible (before they all left me back), I am indebted to Uli Elling, Helena Jambor, Sandra Mueller, Lukas Neidhart, Adam Cliffe and Kat Brown. I especially thank Clemens Grabher for his support during the last year.

I thank my parents for their continuous support and for their refreshing visits in Heidelberg, thanks to my mother for long phone conversations, for understanding and never losing her faith in me in every respect.

Publications

This thesis describes work carried out from June 2004 to January 2008 in the laboratory of Dr. Stephen Cohen at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany.

The thesis is written as a cumulative dissertation. The following publications are presented. Publication 4 was substantially incorporated into the introduction.

Publication 1:

Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution

Alexander Stark, Julius Brennecke, **Natascha Bushati**, Robert B. Russell and Stephen M. Cohen (2005), Cell 123, 1133-46

Publication 2:

A single Hox locus in *Drosophila* produces functional microRNAs from opposite DNA strands

Alexander Stark[1], **Natascha Bushati[1]**, Calvin Jan, Pouya Kheradpour, Emily Hodges, Julius Brennecke, David P. Bartel, Stephen M. Cohen and Manolis Kellis (2008), Genes Dev 22, 8-13

Publication 3:

Temporal reciprocity of microRNAs and their targets during the maternal to zygotic transition in *Drosophila*

Natascha Bushati, Alexander Stark, Julius Brennecke and Stephen M. Cohen (submitted)

Publication 4:

MicroRNA Functions (Review)

Natascha Bushati and Stephen M. Cohen (2007), Annu Rev Cell Dev Biol 23: 175-205

[1] with equal contribution

Summary

The accurate regulation of gene expression is essential for animal development. The discovery of microRNAs (miRNAs) as post-transcriptional regulators added a new layer of complexity to gene expression programs. miRNAs potentially target a large fraction of the genome. However, identification of physiologically relevant targets requires meticulous functional characterization of individual miRNAs. In the work presented I describe efforts to elucidate the biological roles of miRNAs in the development of *Drosophila melanogaster*.

I was involved in a global analysis of the expression patterns of miRNAs and their predicted targets, which revealed that many miRNAs are not co-expressed with their targets. This suggested that miRNAs might act by preventing the expression of unwanted transcripts that are present only at very low levels in the miRNA expressing cell.

In collaboration with Alexander Stark, we found more evidence for this hypothesis when we investigated the *miR-iab-4* locus and its role in Hox gene regulation. Homeobox - containing (Hox) transcription factors are encoded by highly conserved genes fundamentally required during animal development. We found that the *miR-iab-4* locus gives rise to two distinct miRNAs, one derived from the sense, and the other from the antisense strand. These miRNAs are expressed in nonoverlapping domains and are required to support the transcriptional repression conferred by Hox genes among each other. Therefore, the transcriptional regulation is the primary, and miRNA-mediated regulation the secondary level to reinforce the Hox protein expression domains.

Furthermore, I investigated the role of the *miR-309* cluster during early embryogenesis by generating a deletion mutant of this cluster. The *mir-309* cluster contains six distinct miRNAs, which are induced during the onset of zygotic transcription. Using microarray profiling, I identified a set of in vivo targets of the *miR-309* cluster and could show that these miRNAs are required for the timely

downregulation of maternally deposited transcripts during the maternal to zygotic transition. Again, the miRNAs and their targets were expressed in a temporally nonoverlapping manner. In this case, the miRNAs substantially contribute to downregulation of the target's transcript levels, thereby helping to shape the nonoverlapping expression pattern. Intriguingly, a comparable role was described for the unrelated miRNA miR-430 in zebrafish, suggesting convergent evolution.

Taken together, I was able to identify some of the complex regulatory relationships between selected miRNAs and their targets in *Drosophila*. I found evidence that the miRNAs' roles in development can be subtle, but may serve to ensure robustness of the developmental program, which is of significant importance from an evolutionary point of view.

Zusammenfassung

Die präzise Regulierung der Genexpression ist grundlegend in der Entwicklung eines Organismus. Die Entdeckung von microRNAs (miRNAs) als post-transkriptionelle Regulatoren erweiterte die Komplexität der Genexpression um eine neue Dimension. miRNAs haben das Potential, einen großen Teil des Genoms zu regulieren. Aus diesem Grund bedarf es sorgfältiger funktioneller Charakterisierung von einzelnen miRNAs, um herauszufinden, welche dieser möglichen Zielgene ("Targets") tatsächlich physiologisch relevant sind. In der hier vorgelegten Arbeit beschreibe ich Bemühungen, die Bedeutung von miRNAs in der Entwicklung von *Drosophila Melanogaster* aufzuklären.

Ich war an einer globalen Analyse der Expressionsmuster von miRNAs und ihrer vorhergesagten Targets beteiligt, die aufzeigte, dass viele miRNAs nicht mit ihren Targets co-exprimiert werden. Dies deutete darauf hin, dass miRNAs die Expression von unerwünschten Genen, die nur in sehr niedriger Menge in der Zelle vorhanden sind, verhindern.

In Zusammenarbeit mit Alexander Stark fanden wir weitere Anhaltspunkte für diese Hypothese, als wir den *miR-iab-4* Locus und seine Rolle in der Hox-Gen Regulierung untersuchten. Die hochkonservierten Homöobox (Hox)–Gene kodieren für Transkriptionsfaktoren, die für die Entwicklung vielzelliger Organismen unentbehrlich sind. Unsere Analyse ergab, dass vom *miR-iab-4* Locus zwei verschiedene miRNAs exprimiert werden, jeweils vom Plus- und vom Minusstrang. Diese miRNAs werden unterschiedlich exprimiert, und unterstützen so die transkriptionelle Regulierung der Hoxgene untereinander. Folglich wird die korrekte Expression von Hoxproteinen primär transkriptionell und sekundär zusätzlich durch die miRNAs des *miR-iab-4* Locus reguliert.

Weiters charakterisierte ich die Rolle des *miR-309* Clusters während der frühen Embryonalentwicklung, indem ich eine Mutante erzeugte, der dieser Cluster fehlte. Das *miR-309* Cluster kodiert für sechs verschiedene miRNAs, deren Expression zu Beginn der zygotischen Genexpression aktiviert wird. Durch Analyse von

Zusammenfassung

Microarrayprofilen gelang es mir, eine Reihe von in vivo Targets des *miR-309* clusters zu identifizieren. Ich konnte damit zeigen, dass diese miRNAs für die rechtzeitige Entfernung maternaler Transkripte notwendig sind. Auch diese miRNAs sind nicht mit ihren Targets gemeinsam, sondern zeitlich versetzt exprimiert. In diesem Fall tragen die miRNAs stark zur Reduktion der Transkriptmenge bei und festigen somit das einander ausschließende Expressionsmuster. Interessanterweise wurde eine vergleichbare Rolle auch für eine andere miRNA, miR-430, in Zebrafish beschrieben, was auf konvergente Evolution hindeutet.

Zusammengenommen ist es mir in dieser Arbeit gelungen, einige der zahlreichen regulatorischen Funktionen von ausgewählten miRNAs im Zusammenspiel mit ihren physiologiosch relevanten Targets zu charakterisieren. Die Aufgaben von miRNAs in der Entwicklung können subtil sein, aber sie können wesentlich dazu beitragen, ein genetisch robustes Entwicklungsprogramm sicherzustellen. Unter einem evolutionären Gesichtspunkt betrachtet erweisen sich miRNAs daher als ein ausgesprochen nützliches Werkzeug der Natur.

Introduction

1 miRNA Discovery, Biogenesis and Silencing Mechanism

1.1 miRNA Discovery

Since the discovery of the founding members of the microRNA family, *lin-4* and *let-7*, both encoding ~22-nucleotide non-coding RNAs (Lee et al 1993, Reinhart et al 2000, Wightman et al 1993), hundreds of microRNAs (miRNAs) have been identified in plants, animals and viruses by molecular cloning and bioinformatic approaches (Berezikov et al 2006, Lagos-Quintana et al 2001, Lau et al 2001, Lee & Ambros 2001, Ruby et al 2006, Sandmann & Cohen 2007, Stark et al 2007). miRNAs were found to downregulate gene expression by base pairing with the 3'UTRs of target messenger RNAs (mRNAs) (Lee et al 1993, Reinhart et al 2000, Slack et al 2000, Wightman et al 1993). These discoveries indicated that this widespread class of non-coding RNA molecules may constitute a new layer of regulatory control over gene expression programs in many organisms.

1.2 miRNA Biogenesis

Most miRNA genes are transcribed by RNA polymerase II (Pol II) to generate a stem-loop-containing 'primary miRNA' (pri-miRNA), which can range in size from several hundred nucleotides (nt) to tens of kilobases (kb) (Cai et al 2004, Lee et al 2004a) (Figure 1). An exception to this rule are miRNAs lying within Alu repetitive elements which have been reported to be transcribed by RNA polymerase III (Borchert et al 2006). Like mRNAs, Pol II transcribed pri-miRNAs contain 5' cap structures, are polyadenylated and may be spliced (Bracht et al 2004, Cai et al 2004). The pri-miRNA is processed within the nucleus by a multi-protein complex called the Microprocessor, which is composed of the RNAse III enzyme Drosha and the double-stranded RNA binding domain (dsRBD) protein

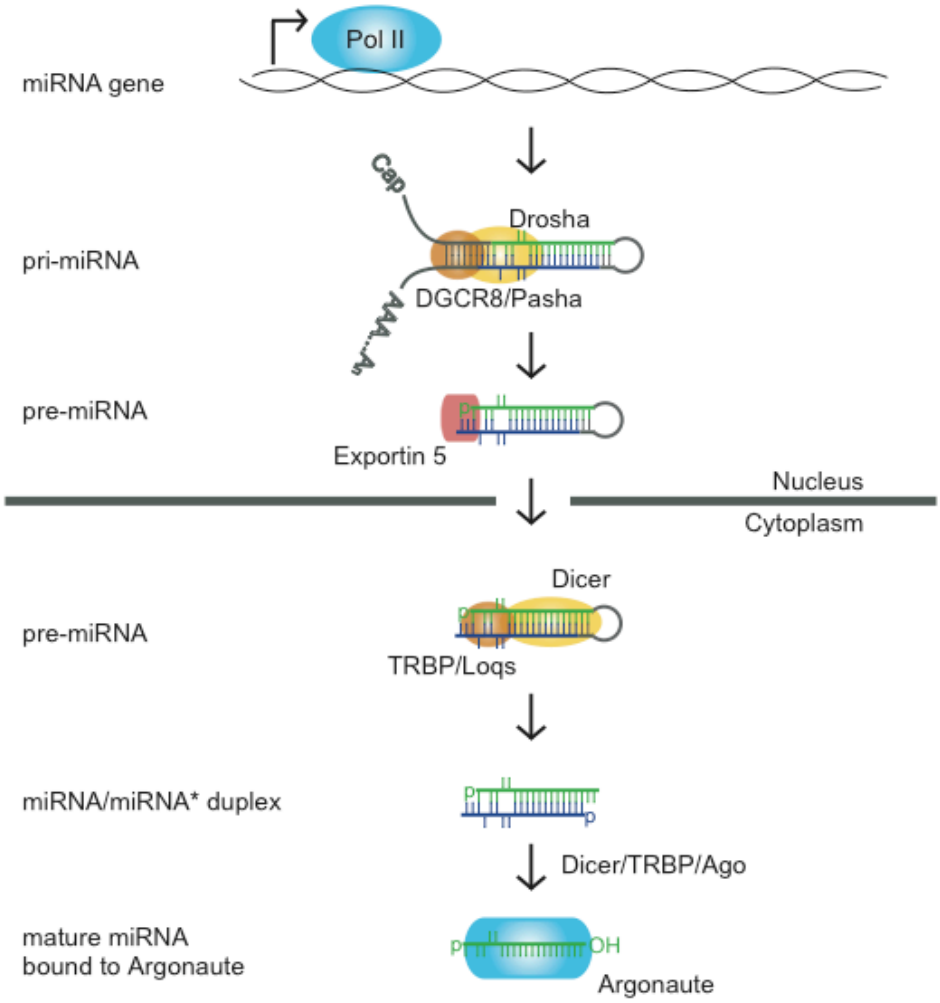


Figure 1. miRNA biogenesis. A miRNA gene is transcribed, generally by RNA polymerase II, generating the primary miRNA (pri-miRNA). In the nucleus, the RNAse III endonuclease Drosha and the double-stranded RNA-binding domain (dsRBD) protein DGCR8/Pasha cleave the pri-miRNA to produce a 2-nt 3' overhang containing the ~70-nt precursor miRNA (pre-miRNA). Exportin-5 transports the pre-miRNA into the cytoplasm, where it is cleaved by another RNAse III endonuclease, Dicer, and its partner, the dsRBD protein TRBP/Loquacious, releasing the 2-nt 3' overhang-containing ~21-nt miRNA:miRNA* duplex. The miRNA strand is loaded into an Argonaute-containing RNA-induced silencing complex (RISC), whereas the miRNA* strand is typically degraded.

DGCR8/Pasha (Denli et al 2004, Gregory et al 2004, Han et al 2004, Landthaler et al 2004, Lee et al 2003). This complex cleaves the pri-miRNA stem, producing the ~70nt hairpin precursor miRNA (pre-miRNA, Figure 1). Some pre-miRNAs, so-called mirtrons, derive directly from short hairpin introns and are produced by the splicing

machinery instead of Drosha (Berezikov et al 2007, Okamura et al 2007, Ruby et al 2007a). The 2-nt 3'overhang, characteristic of RNAse III mediated cleavage and also present in invertebrate mirtrons (Ruby et al 2007a), is recognized by Exportin-5, which binds the pre-miRNA and transports it into the cytoplasm via a Ran-GTP dependent mechanism (Bohnsack et al 2004, Lund et al 2004, Yi et al 2003). The final step in miRNA biogenesis is cleavage of the pre-miRNA into the mature ~22nt miRNA:miRNA* duplex (Figure 1). This event is mediated by another RNAse III enzyme, Dicer, which interacts with the dsRBD proteins TRBP/Loquacious and, in human cells, PACT (Chendrimada et al 2005, Forstemann et al 2005, Hutvagner et al 2001, Jiang et al 2005, Ketting et al 2001, Lee et al 2006, Saito et al 2005). Subsequently, TRBP/Loquacious recruits the Argonaute protein, and together with Dicer they form a trimeric complex that initiates the assembly of the RNA-induced silencing complex (RISC), a ribonucleoprotein (RNP) complex (Gregory et al 2005, Maniataki & Mourelatos 2005, Tomari et al 2007). In *Drosophila*, miRNA/miRNA* duplexes containing a central mismatch are loaded into an Ago1 containing RISC (Tomari et al 2007), whereas those whose central region is base paired are loaded into both, Ago1 and Ago2 containing RISCs (Forstemann et al 2007). Based on the relative stability of the two ends of the duplex, the miRNA strand, with relatively unstable base pairs at the 5' end, remains incorporated in the RISC, whereas the passenger strand, or miRNA* strand, is degraded (Leuschner & Martinez 2007, Matranga et al 2005, Schwarz et al 2003).

1.3 Silencing Mechanism

Once incorporated within the RISC complex, the miRNA directs RISC to downregulate expression of target mRNAs. Depending on the degree of complementarity between the miRNA and the target sequence, messages are either cleaved and degraded (perfect or near perfect complementarity) or their translation is repressed (imprecise complementarity) (Hutvagner & Zamore 2002, Martinez & Tuschl 2004). In the case of precise complementarity and target cleavage ('slicing'), the RISC must contain an Argonaute protein capable of endonucleolytic cleavage. Ago2 is the sole enzyme conferring this activity in mammals and is the major enzyme in flies (Liu et al 2004, Meister et al 2004, Okamura et al 2004). However, only a few

endogenous animal miRNAs act in this 'slicing' mode (Yekta et al 2004). In *Drosophila*, most miRNA/miRNA* duplexes have a central unpaired region, which is rejected by Ago2 and is preferentially bound by Ago1, the Argonaute protein capable of non-endonucleolytic repression of mRNAs containing partially complementary miRNA binding sites in their 3'UTR (Forstemann et al 2007, Tomari et al 2007). Most animal miRNAs base pair imprecisely with their targets and promote translational repression, rather than cleavage and degradation. In this case, target mRNAs are not actively degraded but can be moderately destabilized due to deadenylation and subsequent decapping (Giraldez et al 2006, Jackson & Standart 2007, Wu et al 2006).

The mechanism of translational repression by miRNAs remains unclear (Standart & Jackson 2007). Indeed, the precise step at which miRNAs block translation is controversial. Evidence has been presented that miRNAs inhibit translation initiation (Humphreys et al 2005, Pillai et al 2005), but other studies suggest that they block elongation (Maroney et al 2006, Nottrott et al 2006, Petersen et al 2006). Recent studies using different cell-free translation systems strongly point towards inhibition of translation initiation as the main miRNA-induced silencing mechanism (Mathonnet et al 2007, Thermann & Hentze 2007, Wakiyama et al 2007, Wang et al 2006). Moreover, Ago proteins contain a highly conserved motif which shows similarity to the m^7G-cap-binding domain of the translation initiation factor eIF4E, and this motif is required for translational repression (Kiriakidou et al 2007). It has been proposed that the Argonaute protein competes with eIF4E for cap binding, which is supported by the observation that addition of excess eIF4E can reverse miRNA-mediated translational repression in a mammalian cell-free system (Mathonnet et al 2007). In *Drosophila* embryo extracts, miR-2 induced silencing of a luciferase reporter led to the formation of so-called pseudo-polysomes, dense messenger ribonucleoprotein particles (mRNPs) that shift towards the sucrose gradient fractions, which contain polyribosomes (polysomes) (Thermann & Hentze 2007). This also occured when ribosome complex formation or translation elongation were blocked, indicating that these complexes were indeed of non-polysomal nature. The occurrence of these pseudo-polysomes might explain some of the discrepancies between the different results reported. The observation of co-sedimentation of miRNA-regulated mRNAs

with polyribosomes might have been misinterpreted as polysome association. However, it is also conceivable that miRNAs act distinctly on different targets or at different times. Inhibition of cap-dependent translation initiation seems to be the earliest event inflicted by a miRNA upon its target, but the possibility of additional miRNA-mediated inhibition during steps after initiation cannot be excluded.

The pseudo-polysomes described above are large protein-mRNA aggregates, which might resemble processing bodies (P-bodies). P-bodies are cytoplasmic foci that exclude ribosomal components and may therefore serve as sites in which mRNAs can be stored without translation or be degraded. Indeed, Argonaute proteins bound to miRNAs and their target mRNAs accumulate in processing bodies (Liu et al 2005a, Pillai et al 2005, Sen & Blau 2005). Several proteins found in P-bodies (including GW182, the Dcp1/Dcp2 decapping complex and the RCK/p54 helicase) can bind to mammalian Argonaute proteins, and this interaction mediates translational repression (Behm-Ansmant et al 2006, Chu & Rana 2006, Eulalio et al 2007b, Liu et al 2005a, Liu et al 2005b, Rehwinkel et al 2005). These components of the mRNA degradation machinery also induce accelerated degradation of miRNA targets (Bagga et al 2005, Behm-Ansmant et al 2006, Giraldez et al 2006, Wu et al 2006). However, doubts remain over the importance of cytoplasmic relocation to P-bodies in target repression. Disrupting P-bodies does not have an effect on the degree of translational repression, but blocking the miRNA pathway prevents P-body formation, indicating the P-body localization and even formation are a consequence rather than the cause of repression (Chu & Rana 2006, Eulalio et al 2007a, Jackson & Standart 2007). Moreover, miRNA-mediated repression and P-body localization have been demonstrated to be reversible (Bhattacharyya et al 2006), indicating that the P-bodies might merely serve as sites of temporary storage of translationally repressed mRNAs.

Many human mRNAs that encode proteins whose levels are under tight control have AU-rich elements (AREs) in their 3'UTRs. A specific miRNA, miR-16, is required for the rapid turnover of mRNAs containing such elements in their 3'UTRs, to which miR-16 binds (Jing et al 2005). However, this does not seem to involve site-specific endonucleolytic cleavage, since none of the decay intermediates expected

from slicer activity have been observed. Presumably, miR-16 collaborates via RISC binding with TPP, an ARE-binding protein required for ARE mRNA degradation. Recently, a surprising new role for miRNAs and AU-rich elements has been reported (Vasudevan & Steitz 2007, Vasudevan et al 2007). A miR-369-3-Ago2 complex was found to target the ARE of the *TNFα* 3'UTR. In proliferating cells, this complex confers translational repression, but upon cell cycle arrest, it recruits the RNA binding protein FXR1 to the mRNA to upregulate translation. Other miRNAs were also shown to possess upregulation activity upon cell cycle arrest, indicating that miRNAs can enhance or repress translation, depending on the physiological conditions in the cell, which lead to recruitment of different regulatory proteins to the targeted mRNA.

Although to date most attention has been paid to miRNA action in the cytoplasm, one report shows that mature miR-29b contains a 6-nt motif at its 3' terminus which directs import of the mature miRNA into the nucleus (Hwang et al 2007), raising intriguing possibilities for other modes of miRNA function.

2 Other small non-coding RNAs

In this introduction I will focus on animal miRNAs and their biological functions. However, I want to mention the other kinds of small non-coding RNAs with distinct properties.

2.1 Small interfering RNAs (siRNAs)

Small interfering RNAs (siRNAs) differ from miRNAs mainly in their origin: They derive from endogenous or exogenous long double-stranded RNAs, and are processed into siRNAs by Dicer. siRNAs usually act to induce cleavage of their targets via RNA interference (RNAi), loaded preferentially onto an Ago2-containing RISC (Tomari et al 2007). However, siRNAs can also act as miRNAs on targets with imperfect complementarity and induce translational repression. Moreover, depending on the number and position of the mismatches, RISC-mediated cleavage of mismatched targets can also occur, albeit at a lower rate (Martinez & Tuschl 2004, Meister & Tuschl 2004).

2.2 Piwi-interacting RNAs (piRNAs)

The germline specific Piwi-interacting RNAs (piRNAs) associate with members of the Piwi protein family, a subtype of Argonaute proteins, in mouse, zebrafish and flies. (Aravin et al 2006, Girard et al 2006, Grivna et al 2006, Gunawardane et al 2007, Houwing et al 2007, Lau et al 2006, Saito et al 2006, Vagin et al 2006, Watanabe et al 2006, Brennecke et al 2007). piRNAs are 26-31 nt in length, and they are produced by a Dicer-independent mechanism (Houwing et al 2007, Vagin et al 2006), which presumably involves the Piwi proteins themselves (Brennecke et al 2007, Gunawardane et al 2007). Since mouse piRNAs accumulate at the onset of male meiosis and sperm maturation arrests at different stages in *Mili and Miwi,* the mouse Piwis, knockout mice (Carmell et al 2007, Deng & Lin 2002, Kuramochi-Miyagawa et al 2004), mouse piRNAs are thought to play an essential role during gametogenesis. The previously identified class of repeat-associated siRNAs (Aravin et al 2003) is the major class of *Drosophila* piRNAs (Brennecke et al 2007, Gunawardane et al 2007, Saito et al 2006, Vagin et al 2006). These are involved in silencing transposons in the male and female germline. Discrete piRNA-generating loci are composed of defective transposon copies, and an amplification cycle is thought to boost piRNAs against actively transcribed transposons (Brennecke et al 2007, Gunawardane et al 2007).

2.3 Plant miRNAs

Plant miRNAs differ from animal miRNAs in their biogenesis and in their mode of target regulation: Most plant miRNAs display perfect or near perfect complementarity to their target mRNAs, in both coding regions and 3'UTRs, and therefore induce mRNA cleavage (Vaucheret 2006). However, the plant miRNA family *miR-854* was reported to act by translational repression via imperfect binding sites and to be conserved in animals (Arteaga-Vazquez et al 2006). This observation is surprising because miRNAs are thought to have evolved independently in the plant and animal kingdoms (Chen & Rajewsky 2007). However, the recent discovery of miRNA precursors and genes involved in miRNA biogenesis in the unicellular green algae

Chlamydomonas reinhardtii suggests an older origin for miRNAs in multicellular organisms, but none of the miRNAs identified display sequence similarity to known plant or animal miRNAs (Molnar et al 2007, Zhao et al 2007a).

3 Target Identification

The first animal miRNA targets were identified by genetics in *Caenorhabditis elegans* (*C. elegans*). The heterochronic miRNA *lin-4* displays a mutant phenotype that can be suppressed by a second mutation in its target mRNA *lin-14* (Lee et al 1993, Wightman et al 1993). These genetic interactions combined with sequence analysis led to the identification of complementary target sites in the 3'UTR regions of *lin-14* (Figure 2). However, this forward genetic approach will likely only help identify targets whose individual loss is sufficient to suppress phenotypes caused by loss of the miRNA. This may be useful in cases where the miRNA has one or a few biologically important targets. It is unclear at present whether this sort of relationship will be more the exception than the rule.

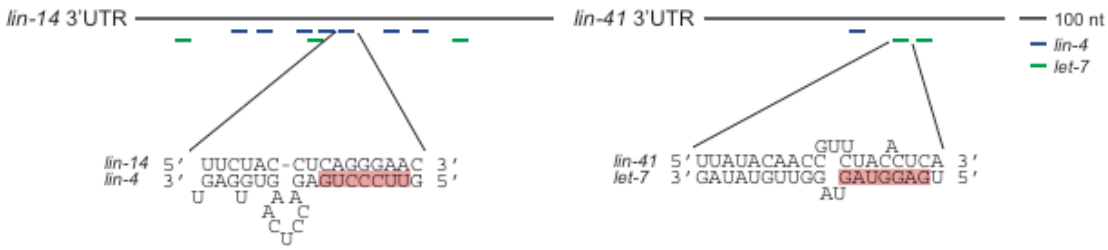


Figure 2. miRNA-target basepairing. The miRNAs lin-4 and let-7 confer regulation to their targets via imprecise base-pairing with the target's 3'UTRs. The miRNA seed region is highlighted in red.

Groundbreaking genetic studies of this kind allowed the criteria for functional miRNA-target interactions to be deciphered by mutation of known miRNA target sites and testing for function in miRNA misexpression-based assays (Brennecke et al 2005b, Doench et al 2003, Doench & Sharp 2004, Kiriakidou et al 2004, Kloosterman et al 2004, Lewis et al 2003). In principle, miRNA target sites were categorized into two classes: One class, called 5' dominant sites, base-pair precisely to the so-called "seed" (Figure 2) of the miRNA, with or without 3' pairing support. Target sites of the

other class, called 3' compensatory sites, have weak 5' pairing but strong pairing with the miRNAs' 3' region.

3.1 Computational Prediction of miRNA Targets

Aside from direct experimental tests for site function, further indication of the importance of the seed region in miRNA target recognition has been inferred from computational studies showing significant overrepresentation of conservation of matches to miRNA seeds or, in some cases, avoidance of miRNA seed matches (Brennecke et al 2005b, Farh et al 2005, Krek et al 2005, Lewis et al 2005, Lewis et al 2003, Stark et al 2005, Xie et al 2005). Furthermore, an analysis of single nucleotide polymorphism (SNP) genotype data showed that polymorphism density was significantly lower in conserved target site regions that match the 5' portion of the miRNA (Chen & Rajewsky 2006). Largely based on these rules and on conservation, miRNA targets have been predicted using different computational approaches (Brennecke et al 2005b, Enright et al 2003, Grun et al 2005, Kertesz et al 2007, Krek et al 2005, Lewis et al 2003, Rajewsky 2006, Ruby et al 2007b, Stark et al 2003, Xie et al 2005). Roughly 30% of all animal genes are predicted to be targeted by miRNAs. An algorithm which attempts to identify miRNA target sites without relying on cross-species conservation or miRNA sequences (Miranda et al 2006) predicts even higher numbers of miRNA-regulated genes.

It is likely that ~85% of predicted conserved miRNA sites are functionally important in an evolutionary sense, as inferred from a computational study on SNPs within conserved miRNA sites (Chen & Rajewsky 2006). Their regulation by miRNAs must therefore improve fitness of the organism in some way, unless the sequences are coincidentally conserved for reasons unrelated to miRNAs. Therefore, whether these presumptive targets are in fact physiologically relevant remains an open question, and one that can only be addressed in vivo by examining miRNA mutants. Whether the biological functions of these miRNA target sites have been selected for will prove to be amenable to experimental analysis remains to be seen. Small differences in fitness may be difficult to study in the laboratory.

3.2 Target Validation

One obvious approach to validate miRNA target predictions is to misexpress a given miRNA and assay for regulation of the putative target. Overexpression of miRNAs in tissue culture followed by mRNA profiling demonstrates that the seed region can indeed be sufficient for target down-regulation (Lim et al 2005). The converse high throughput approach, bulk depletion of miRNAs followed by mRNA profiling, demonstrates upregulation of many predicted target mRNAs (Rehwinkel et al 2006). Thus, a global picture of which target mRNAs react to miRNA expression can be obtained. However, it is unclear what fraction of mRNAs change levels upon miRNA regulation; targets whose mRNA levels are not affected appear as false negatives, whereas those affected indirectly through secondary effects are false positives. On the protein level, miRNA-target interactions are not amenable to high throughput approaches, and must be tested one by one in reporter assays. Yet, it must be stressed that misexpression is an artificial situation and therefore while useful to test whether regulation is possible, such experiments are not sufficient to draw conclusions about miRNA-target relationships in vivo.

3.3 Biochemical Target Identification

To identify miRNA-target relationships in vivo, several approaches have been taken to isolate microribonucleoprotein complexes (miRNPs) with their associated miRNAs and mRNAs and to analyze their contents. One approach used highly specific monoclonal antibodies against human Ago1 and Ago2 to co-immunoprecipitate the bound mRNAs from human cells (Beitzinger et al 2007). These transcripts were then identified by generation of a cDNA library and subsequent sequencing. About 60% of the Ago-bound mRNAs were predicted to be miRNA targets by different algorithms. Two other studies used tagged versions of Argonaute proteins (Easow et al 2007, Karginov et al 2007). In one study, the mRNA content of Ago2 containing miRNPs from cells transfected with miR-124a was compared to that from untransfected cells (Karginov et al 2007). It is unclear how accurately the associated RNAs reflect the in vivo target spectrum of miR-124a, since these cells do not normally express miR-124a. In *Drosophila*, Ago1 containing miRNPs from *miR-1* mutant embryos were

immunoprecipitated and their mRNA composition was compared to that obtained from wild-type Ago1-miRNPs, identifying authentic in vivo miR-1 targets (Easow et al 2007). In both studies, a substantial number of miRNA targets containing seed matches were recovered, which did not change significantly in mRNA levels, as assessed by microarray profiling from total RNA. Therefore, it seems that a quite large fraction of miRNA targets are not significantly destabilized at the mRNA level, or that the magnitude of repression conferred by the miRNA is small. This will be the case for miRNA-target relationships in which the miRNA is not absolutely required to reduce the target expression to an inconsequential level, but to buffer it to optimal levels, when required.

3.4 Target Types: Switching off vs Tuning Target Expression Levels

A few years ago Bartel and Chen (2004) proposed the terms “switch target” and “tuning target” to describe two classes of relationship between miRNA and targets (Figure 3). These terms were coined to describe how the miRNA affects its target, and should not be confused with how regulation of the target affects the cell or organism.

A switch target is one for which the miRNA reduces target expression below a level at which it has any meaningful activity in the cell, effectively switching it off. Targets which are normally expressed at very low levels in the miRNA expressing cells would likely qualify as switch targets because the function of the miRNA appears to be to reduce their potential leaky expression to inconsequential levels. In such a case, the miRNA does not appear to act as developmental switch. On the other hand, important miRNA targets, including those of the heterochronic miRNAs lin-4 and let-7, may act as developmental switches (Lee et al 1993, Wightman et al 1993). Although these are also likely to be switch targets, it remains to be demonstrated that the residual expression after miRNA-mediated repression is without function.

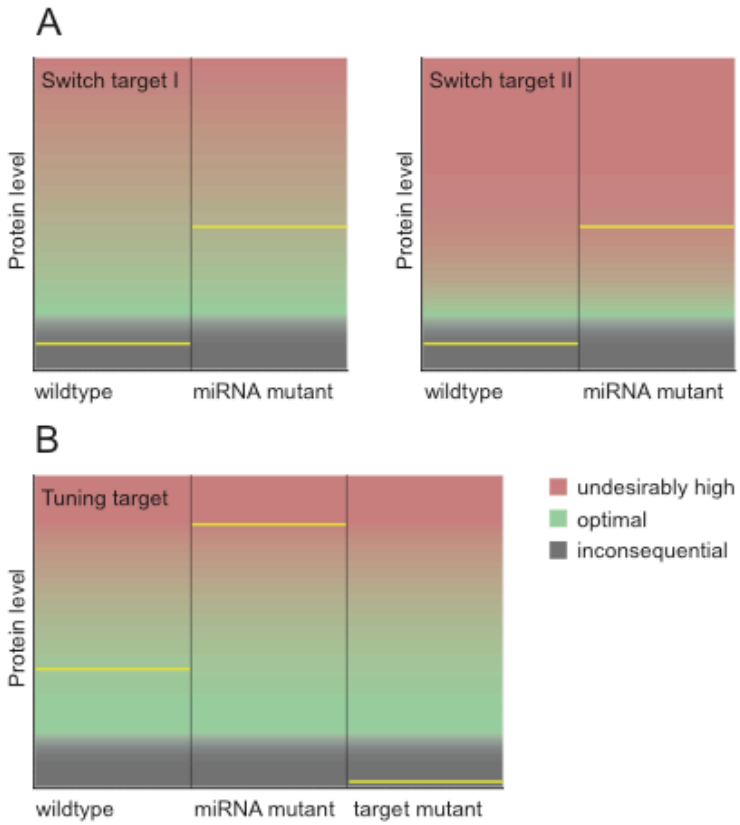


Figure 3. Switch and tuning targets. Red, green, and grey indicate areas in which protein levels are undesirably high, optimal, and inconsequential, respectively. (A) Switch targets are downregulated to inconsequential levels. In the absence of the miRNA, these targets are expressed at levels that may be harmless (switch target I) or at detrimental levels (switch target II). The consequences can be subtle or severe, depending on the type of target gene and its relative level of misexpression. (B) miRNAs adjust the expression of tuning targets to optimal levels. In the absence of the miRNA, the tuning target is expressed at a detrimental level. In contrast to switch targets, tuning targets have a function in the miRNA-expressing cell, so reducing them further has a negative effect.

The notion of a tuning target reflects the role of the miRNA in setting a defined level of target expression, while being co-expressed with the miRNA. Tuning targets can make use of miRNAs to smooth out fluctuations in their expression or, by altering miRNA level, ensure that target levels are suitable for the prevailing conditions. The critical distinction between the switch and tuning modes lies in whether the residual level of target expression is required, i.e. has a specific function, in that cell, or if the job of the miRNA is merely to reduce target expression to an inconsequential level.

3.5 Context Dependence

Evidence is accumulating to suggest that miRNA-target regulation can be context dependent. For example, miRNAs are capable of up- or downregulating their target genes, depending on the composition of the protein complex recruited to the mRNA, which seems to be different under distinct physiological conditions (Vasudevan et al 2007). Other, less dramatic, incidences of context dependence have been reported. For example, (a) zebrafish miR-430 directly contributes to the repression of residual maternal *nanos1* in somatic cells (Mishima et al 2006). Though both miR-430 and *nanos1* are also present in primordial germ cells, *nanos1* can overcome the regulation conferred by the miRNA and is robustly expressed. The RNA-binding protein Dead end (Dnd1) was very recently reported to be required for this resistance (Kedde et al 2007). Dnd1 binds to uridine-rich regions in the vicinity of miRNA binding sites in 3'UTRs, thereby preventing the association of the miRNA and its target. This mechanism does not seem to be an oddity of zebrafish primordial germ cells, since Dnd1 was also shown to protect *LATS2* from miR-372 targeting in a human teratoma-derived cell line (Kedde et al 2007). (b) In mammalian hepatocarcinoma cells, stress-induced derepression of the miR-122 target *CAT-1* depends on the presence of a binding site for the HuR protein in the 3'UTR, which is distinct from the miRNA sites (Bhattacharyya et al 2006). Derepression did not occur if the HuR sites were removed in a reporter construct, implying context dependence. (c) In *C. elegans*, a minimal lsy-6 target site has been reported to be functional when embedded in the 3'UTR of its endogenous target but not when embedded in a heterologous 3'UTR (Didiano & Hobert 2006). This was taken as evidence that target site function may depend on the specific 3'UTR context. However, this conclusion is weakened because the presence of a second site in the endogenous 3'UTR that may contribute to its regulation was not considered.

4 miRNA Expression Patterns

Obviously miRNA target prediction programs do not take into account whether a miRNA is ever expressed in the same cell as its predicted target, as would be required to make regulation feasible. While this is less likely to be a concern for conserved miRNA target sites, it is an issue for non-conserved sites since as little as 7-8 nt homology can in principle confer repression. Therefore, information about miRNA and target mRNA expression patterns can help to assess the likelihood that a predicted miRNA-target relationship is relevant in vivo.

Developmental miRNA profiles were generated by small RNA cloning and northern blotting (Aravin et al 2003, Berezikov et al 2006, Chen et al 2005, Lagos-Quintana et al 2001, Lau et al 2001, Lee & Ambros 2001, Ruby et al 2006, Ruby et al 2007b). The great advantage of cloning and high-throughput sequencing is that new miRNAs can be identified and quantitative information obtained about their expression at different stages of embryonic development. miRNA microarrays (Baskerville & Bartel 2005, Miska et al 2004, Nelson et al 2004, Thomson et al 2004), quantitative reverse transcriptase-PCR (qRT-PCR) methods and in situ data have provided further insight into tissue-specific expression of pri- and mature miRNAs during development (Aboobaker et al 2005, Ason et al 2006, Kloosterman et al 2006, Wienholds et al 2005). It should be noted that the expression pattern of the pri-miRNA does not necessarily reflect that of the mature miRNA, since some mammalian miRNAs have been shown to be regulated post-transcriptionally (Obernosterer et al 2006, Thomson et al 2006). The available expression data still clearly demonstrates that many miRNAs conserved in their sequences also display similar expression patterns across species, which may reflect conservation of their functions. Prominent examples are the muscle-specific miR-1 or the nervous system-specific miR-124, which are stably tissue specific. miRNAs can also be expressed in highly dynamic patterns, e.g. miR-279 or miR-9a in *Drosophila* (Stark et al 2005).

5 Functions in Animal Development

5.1 Inferred from Bioinformatic Approaches

Transfection of the highly tissue-specific miRNAs miR-1 and miR-124 into HeLa cells shifts the cells' mRNA expression profiles towards those of the respective miRNA expressing tissues, affecting ~100-200 transcripts (Lim et al 2005). The interesting insight that emerged from this study was that the transcripts targeted in a heterologous cell type were those that would normally be expressed at a low level, if at all, in the cell type in which the miRNA was expressed. For example, the mRNAs targeted by the muscle specific miR-1 were normally low or absent in muscle. A similar conclusion for several highly tissue-specific miRNAs was drawn in global analyses of the relations of miRNA expression patterns to those of their conserved targets (Farh et al 2005, Sood et al 2006, Stark et al 2005). One of these analyses is presented in this thesis (Stark et al 2005).

The outcomes of these studies are essentially that (1) the average miRNA has target sites in hundreds of genes, and only few predicted targets contain multiple conserved binding sites for a single miRNA, indicating that stringent, switch-like, regulation by a miRNA is likely to be exceptional; (2) most of the abundant mRNAs are under selective pressure to circumvent regulation by coexpressed miRNAs ("antitargets"); (3) genes involved in basic cellular processes have been selected for short 3'UTR length in order to prevent appearance of unwanted miRNA sites, whereas genes involved in developmental processes show statistical evidence of enrichment for miRNA sites.

An additional observation made in *Drosophila* is that many miRNAs and their targets are expressed in a mutually exclusive manner, either temporally or spatially (Stark et al 2005) (Figure 4). In the latter case, targets are typically present in domains adjacent to the miRNA expressing tissue. However, in human and mouse tissues, many conserved targets are present in the tissue expressing the miRNA, albeit at significantly lower levels than in most other tissues (Farh et al 2005, Sood et al 2006). In *Drosophila* this possibility is not excluded, although by in situ

hybridization the targets were not detectable in the miRNA expressing cell. This apparent difference probably reflects differences in sensitivity of the methodology, since selection for miRNA sites would not emerge if targets were not even present at low levels.

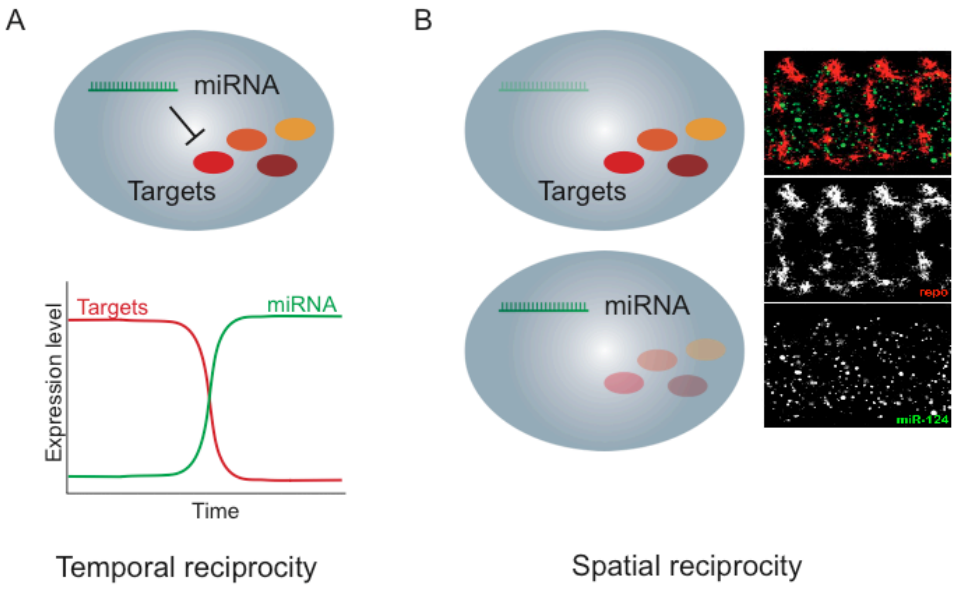


Figure 4. Mutually exclusive expression of miRNAs and their targets. (A) miRNAs and their targets can be expressed in a temporally reciprocal manner: Target transcription may be turned off, concurrent with the turning on of miRNA expression. miRNA expression may also actively contribute to target repression, without an underlying change in target transcription. (B) In the case of spatial reciprocity, the miRNA and its targets, e.g., miR-124 (green) and its target repo (red) in the *Drosophila* central nervous system, are often expressed in adjacent domains (from Stark et al. 2005).

These studies, together, indicated that the average animal miRNA has ~200 conserved targets in the genome and that some miRNAs and their predicted targets are expressed in a largely nonoverlapping manner. Therefore, it was suggested that such miRNAs merely help maintain and define cell types by dampening expression of unwelcome transcripts, which arise either from pre-existing mRNAs after developmental transitions or simply from leaky transcription. As a cell changes fate in time, transcription of a miRNA is turned on and the transcription of its targets is turned off, producing apparently mutually exclusive patterns in space (Figure 4). Thus, the transcriptional regulation is primary and miRNA regulation is a “second

level" reinforcing the transition as the targets are on their way out (or already "gone": as in leaky transcripts). In that way, miRNAs would provide a failsafe mechanism and thereby confer robustness and/or accuracy to underlying gene expression programs.

It is important to bear in mind that the trend of anti-correlation between miRNA and target expression has been documented only for a few highly tissue-specific miRNAs. Although this relationship is common enough to emerge from a global computational study of miRNA and target expression patterns it does not follow that all miRNA-target relationships will be of this type. Further in-depth analysis of miRNAs and their targets as well as of miRNA mutants will be needed to reveal how prevalent this mode of regulation is. Mutants lacking miRNAs for this class of regulation would be expected to suffer from many subtle changes in target levels. The accumulated consequences of many small misregulations might lead to relatively mild and difficult-to-characterize defects, rather than obvious phenotypes in such mutants.

5.2 Depletion of all miRNAs

An approach to examine the global role of miRNAs in development is to eliminate the production of mature, functional miRNAs. Dicer mutants have proven quite informative in the study of the miRNAs expressed earliest during development. Moreover, the conditional inactivation of Dicer in specific mouse tissues has shed some light on the functions of a few differentially expressed miRNAs.

Animals that are unable to produce mature miRNAs do not survive or reproduce (Bernstein et al 2003, Forstemann et al 2005, Ketting et al 2001, Wienholds et al 2003). However, the degree to which development is impaired differs between species. In *C. elegans*, *dicer-1 (dcr-1)* mutant animals display defects in germline development and a burst vulva phenotype reminiscent of that seen in *let-7* mutants (Ketting et al 2001, Knight & Bass 2001). RNAi of *dcr-1* in *dcr-1* mutant embryos, which also eliminates maternal contribution, induces an embryonic lethal phenotype, suggesting a requirement of miRNAs during *C. elegans* embryogenesis (Grishok et al 2001). In *Drosophila*, depletion of Loquacious, the partner of Dicer-1, leads to

female sterility (Forstemann et al 2005), and *dicer-1 (dcr-1)* mutant germline stem cells display cell division defects (Hatfield et al 2005). Unfortunately, the full phenotype of *dcr-1* mutant flies has not been described, but *dcr-1* mutant eyes are reported to be small, with disorganized ommatidial arrays and missing interommatidial bristles (Lee et al 2004b). However, *ago1, dcr-1* double mutants, unlike the single mutants, have been reported to exhibit strong segment polarity defects and to lack Wingless protein expression (Meyer et al 2006), indicating that in *Drosophila*, miRNAs might have a role in embryonic patterning.

A hint towards the requirement of miRNAs for the establishment of long-lasting memory in *Drosophila* comes from analysis of the putative RNA helicase Armitage, which is involved in RISC maturation (Tomari et al 2004). Armitage is colocalized at synapses with CaMKII, a kinase required for memory (Ashraf et al 2006). Upon neural stimulation by olfactory-avoidance learning, Armitage is degraded and CaMKII translation increases, leading to the establishment of a stable memory. The *CaMKII* 3'UTR contains some miRNA binding sites and is required for this regulation. Moreover, in *armitage* and *dicer-2* mutants, CaMKII expression is significantly higher. However, *Drosophila dicer-2* has been implicated in the biogenesis of siRNAs rather than hairpin-derived miRNAs (Lee et al 2004b, Tomari et al 2007), indicating that it might not be the canonical miRNA pathway which is involved in the establishment of long-term memory.

Dicer mutant mice are depleted of Oct4 positive pluripotent stem cells and die at embryonic day 7.5 (Bernstein et al 2003). Their development arrests during gastrulation, before axis formation. In contrast, zebrafish mutant for *dicer-1* display relatively mild effects only. *dicer-1* mutant embryos die by 2-3 weeks of age without obvious defects, except for a general growth arrest (Wienholds et al 2003). Since the *dicer-1* mutant germline is fully functional, it can be transferred into a wild-type embryo and give rise to maternal and zygotic *dicer-1* mutants (Giraldez et al 2005). These animals do not produce any mature miRNAs, and still axis formation and patterning of the embryo were not severely perturbed. Abnormalities were seen in morphogenetic processes during gastrulation, somitogenesis, and heart and brain development, ultimately leading to late-embryonic lethality.

In studies using conditional mouse *Dicer* alleles, embryonic stem cells were impaired in their ability to proliferate (Murchison et al 2005) and those selected for survival failed to differentiate (Kanellopoulou et al 2005). In contrast to the observations made in zebrafish, dicer-dependent miRNA biogenesis in the mouse oocyte is essential to support normal embryonic development (Murchison et al 2007, Tang et al 2007). Other *Dicer* depleted mouse tissues were able differentiate and pattern properly. Yet, the morphogenesis of hair follicles, the lung epithelium and limbs was perturbed (Andl et al 2006, Harfe et al 2005, Harris et al 2006, Yi et al 2006), and apoptosis was induced. Specific deletion of *Dicer* in the mouse heart led to cardiac failure due to a spectrum of developmental defects, but markers of initial cardiac differentiation and patterning were expressed normally (Zhao et al 2007b), consistent with the phenotype described in zebrafish (Giraldez et al 2005). Specific deletion of *Dicer* in mouse postmitotic midbrain dopaminergic neurons (DNs) led to a progressive loss of of midbrain DNs and subsequently to reduced locomotion (Kim et al 2007). Moreover, specific deletion of *Dicer* in the mouse thymus early in T cell development reduced the proliferation and survival of $\alpha\beta$ T cells, but the CD4/CD8 lineage choice was not affected (Cobb et al 2005). These studies suggest that some miRNAs play important roles in these tissues, but leave open the question of what aspect of their biology is being regulated. Identifying which miRNAs act, and on which targets, will be needed for a deeper understanding of their roles in these developmental processes.

5.3 Specific miRNA Functions

The biological roles of most individual miRNAs are unknown. Considering the abundance of this class of genes, one would anticipate the isolation of many miRNA mutants in the numerous genetic screens done in *Drosophila* or *C. elegans*. However, only a small number of miRNA mutants have been recovered by forward genetics in classical mutagenesis screens in *C. elegans* (Chalfie et al 1981, Johnston & Hobert 2003, Reinhart et al 2000) or gain-of-function screens by P-element driven overexpression in *Drosophila* (Hipfner et al 2002, Teleman et al 2006, Xu et al 2003). Why were all these miRNA genes missed?

For chemical mutagenesis, one obvious reason might be the small size of miRNAs. In most cases, the seven nucleotides that constitute the seed must be affected to lead to a complete loss of miRNA function, making miRNAs difficult-to-hit targets. However, this explanation does not hold true for screens based on transposon insertion for loss- or gain-of-function. In screens of this sort miRNA genes may have been overlooked in past since researchers usually searched for protein-coding regions. Moreover, redundancy between miRNAs that share the same seed sequence could mask loss of function phenotypes of single miRNAs (Abbott et al 2005). It might as well be that many miRNA mutants just show relatively subtle defects, which might be difficult to identify in high-throughput genetic screens, or which might be missed if the penetrance is low.

To address these issues, a large-scale mutagenesis project was undertaken to identify mutants for almost all *C. elegans* miRNAs (Miska et al 2007). Relatively few miRNA genes individually generated robust abnormal phenotypes. In some cases, this could be attributed to overlapping functions; multiple mutants that remove related miRNAs may reveal stronger defects than the single mutants. However, some miRNA multiple mutants still produce few or not discernible defects. This group may reflect those miRNAs that confer precision to developmental processes rather than making developmental decisions.

Therefore, to study a specific miRNA's role in development, reverse genetics is a valuable tool to generate animals lacking single miRNAs, but it is risky because one never knows what phenotype, if any, will result. So how should one decide which miRNA(s) to remove? Besides considering obvious factors like strong conservation, interesting expression patterns and avoidance of possible redundancy, it is tempting, though risky, to speculate about a miRNA's function based on the functions of predicted targets and/or phenotypes obtained by misexpression of the miRNA.

5.3.1 Misexpression and Overexpression Analyses

When a miRNA is misexpressed, it will have the potential to regulate many targets that it might never encounter in its endogenous expression domain. As mentioned above, many miRNA targets are highly expressed in domains spatially and/or temporally distinct of those of their respective miRNAs (Farh et al 2005, Sood et al 2006, Stark et al 2005). Shutting down such targets by introduction of high miRNA levels can be expected to lead to strong phenotypes. But since the miRNA and target may not normally be significantly coexpressed, it is difficult to extrapolate from the misexpression effect to the normal function of the miRNA in the organism. miRNA misexpression can produce intriguing defects that have limited relevance to what was learned from mutants lacking the miRNA (Karres et al 2007, Teleman et al 2006, Varghese & Cohen 2007)

In *Drosophila*, two families of Notch target genes are clustered in two genomic locations: the *Enhancer of split* complex and the *Bearded* complex. Most of these genes contain conserved 6-7nt motifs in their 3'UTRs, which are complementary to the seed sequences of a family of related miRNAs (Lai 2002, Stark et al 2003). Some of these miRNAs, when misexpressed, can induce phenotypes reminiscent of those seen in Notch pathway loss of function mutants (Lai et al 2005, Stark et al 2003). Similarly, misexpression of miR-iab-5p downregulates endogenous Ubx protein levels, potentially via 7 target sites, and therefore induces a homeotic phenotype (Ronshaugen et al 2005). But it remains to be determined if mutants lacking these miRNAs impact Notch signaling or Ubx function in vivo.

In cell culture, miRNA overexpression is a popular and sensible approach to find candidate targets, when antisense-mediated depletion of the respective miRNA is done in parallel and leads to the reciprocal effect, namely upregulation of these putative targets.

An example is the pancreatic islet-specific miR-375. When overexpressed, it inhibits glucose-induced insulin secretion. This can be mimicked by knock-down of its

target *myotrophin* (Poy et al 2004). Importantly, depletion of miR-375 increases *myotrophin* levels and enhances glucose-stimulated insulin secretion, indicating that miR-375 is an inhibitor of glucose-stimulated insulin secretion. Using a similar approach in cultured hippocampal neurons, miR-134 was found to regulate dendritic spine size by inhibiting translation of *Limk1* (Schratt et al 2006).

In hematopoietic stem cells, enforced expression of the B-cell specific miRNA miR-181 stimulates their differentiation to B-lineage cells (Chen et al 2004), but it remains to be determined if loss of *miR-181* impedes differentiation. However, opposing effects on antigen sensitivity in T cells have been reported upon overexpression and depletion of miR-181 (Li et al 2007).

miR-1 and miR-133 are strongly upregulated upon differentiation of myoblasts into myotubes, but not present in undifferentiated myoblasts (Boutz et al 2007). miR-1 misexpression can accelerate myoblast differentiation by targeting *histone deacetylase 4* (*HDAC4*), a repressor of muscle differentiation (Chen et al 2006). Depletion of miR-1 impedes differentiation, as indicated by a decrease in myogenic markers. Consistently, tissue-specific overexpression of miR-1 in the developing mouse heart induces premature differentiation of myocytes (Zhao et al 2005). *miR-1* and *miR-133* form one genomic cluster and are co-expressed in the heart and skeletal muscle. It has been reported that miR-133 can promote myoblast proliferation by targeting *serum response factor* (*SRF*) (Chen et al 2006). However, in a different experimental setup myoblast differentiation was not observed (Boutz et al 2007). miR-133 was rather required during differentiation to downregulate nPTB protein, a repressor of alternative splicing. This downregulation resulted in splicing of a group of silenced exons in mature myotubes. From this data, it seems that miR-1 and miR-133 are required to define and maintain the properties of differentiated muscle cells, in agreement with the conclusions drawn from misexpression of miR-1 in HeLa cells (Lim et al 2005).

In *C. elegans*, overexpression of miR-84, a member of the let-7 family, can rescue the multivulva phenotype caused by *let-60/RAS* gain of function alleles (Johnson et al 2005). This is consistent with the observation that *let-7* mutants display a burst

vulva phenotype (reinhart 2000, slack 2000), which can be suppressed by RNAi of *let-60/RAS*. *let-60/RAS* contains several putative miR-84/let-7 binding sites, and can be downregulated by these miRNAs. These results implicate let-7 as an important in vivo regulator of *let-60/RAS*, but final conclusions about the relevance of miR-84 await mutant analysis.

5.3.2 Modes of miRNA Function: Roles of Individual miRNAs

How essential are the contributions of specific miRNAs to developmental programs? Because many miRNAs are highly conserved, their functions will be advantageous either for an individual's life or to provide a selective advantage at the population level. Clearly, as is the case for every class of regulatory molecules, individual miRNAs will influence development to different degrees. Whether a general mode of action is attributable to most miRNAs is still subject of discussion (Bartel & Chen 2004, Hornstein & Shomron 2006, Stark et al 2005). Most likely, one size won't fit all. Comparison of the insights gained from experimental analysis of miRNAs and computational studies suggest that there may be several conceptually distinct modes in which miRNAs act.

5.3.2.1 MiRNAs Acting as Developmental Switches

miRNAs required for developmental decision-making can be expected to cause strong phenotypes when deleted. The first miRNAs identified, *C. elegans'* lin-4 and let-7, are of this kind. Indeed, it could hardly be otherwise, given that they were found in forward genetic screens for mutant phenotypes during development (Chalfie et al 1981, Lee et al 1993, Reinhart et al 2000, Wightman et al 1993). Without *lin-4*, the animal is unable to make the transition from the L1 to the L2 larval stage due to a differentiation defect. This developmentally retarded phenotype is the result of a failure to posttranscriptionally downregulate the heterochronic gene *lin-14*, which harbors 7 lin-4 target sites in its 3'UTR (Figure 2) (Lee et al 1993, Wightman et al 1993). In the adult worm, lin-4 mediated downregulation of *lin-14* has been implicated in life span regulation (Boehm & Slack 2005).

Depletion of *let-7*, a highly conserved miRNA, also leads to a heterochronic defect, precisely a failure of larval-to-adult transition (Reinhart et al 2000). Several genes are direct targets of let-7 during this transition, the heterochronic genes *lin-41*, *hbl-1* and *daf-12* and the forkhead transcription factor *pha-4* (Abrahante et al 2003, Grosshans et al 2005, Lin et al 2003, Slack et al 2000). let-7 therefore substantially regulates the L4 to adult transition in *C. elegans*.

Moreover, the let-7 family miRNAs miR-48, miR-84 and miR-241 function in a cooperative manner to control the L2-to-L3 transition, most probably by targeting *hbl-1* (Abbott et al 2005, Li et al 2005). Additionally, in adult worms miR-48 and miR-84 are coordinately required to bring the larval molting cycle to an end (Abbott et al 2005). These *C. elegans* miRNAs act as key regulators of developmental timing. This regulatory relationship might also exist in other organisms, since the temporal regulation of let-7 during development and the presence of a target site in *lin-41* (Figure 2) are conserved (Pasquinelli et al 2000).

Loss of *lsy-6* in *C. elegans* induces a cell fate switch: Instead of one ASE left (ASEL) and one ASE right (ASER) neuron, two ASER neurons are specified (Johnston 2003). The neuronal asymmetry is lost due to a failure to repress cog-1 expression specifically in ASEL via its lsy-6 miRNA target sites. die-1, present only in ASEL, is required for lsy-6 expression (Chang et al 2004) (Figure 5). die-1 in turn is downregulated in ASER by another miRNA, miR-273, which is expressed predominantly in ASER, since it is activated there by the lsy-6 target *cog-1* (Johnston et al 2005). Therefore, the cell fate decision and stabilization of the gustatory ASE neurons is controlled via a double-negative feedback loop in which two transcription factors use miRNAs to repress each other (Figure 5).

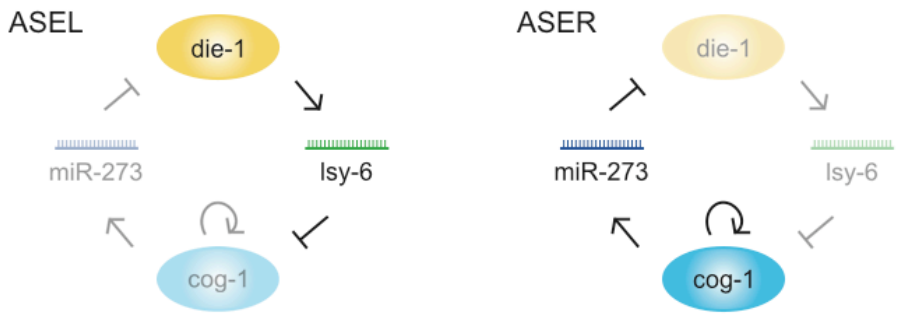


Figure 5. miRNas acting in a double negative feedback loop control neuronal asymmetry in *C. elegans*. In *C. elegans*, lsy-6, induced by the transcription factor die-1, determines adoption of the ASEL fate by repressing cog-1 expression. In the ASER neuron, cog-1 positively regulates its own expression and induces miR-273, which suppresses die-1 and, consequently, *lsy-6*.

MiR-181 expression is upregulated during terminal differentiation of myoblasts, and its depletion impedes their differentiation, partly due to upregulation of one of its targets, Hox-A11, an inhibitor of differentiation (Naguibneva et al 2006). However, miR-181 overexpression is not sufficient to induce differentiation. In adult muscle, miR-181 is expressed at very low levels, suggesting that it is involved in muscle establishment, not maintenance. Whether or not this miRNA acts as a developmental switch in vivo awaits analysis of a knockout mouse.

Mice lacking one of the two *miR-1* genes, *miR-1-2*, were reported to have defects in heart development and function (Zhao et al 2007b). Half of the mutants died during embryonic development due to heart morphogenesis defects. Some survivors died at 2-3 months of age from heart defects, and some survived to adulthood with apparently normal heart morphology, but physiological defects. A fraction of these adults suffered sudden death, indicating that miR-1-2 might have important functions not only in the developing, but also in the adult heart. 22 of the 45 mRNAs upregulated in *miR-1-2* mutant hearts contained miR-1 seed matches, suggesting that miR-1-2 regulates a spectrum of targets in this tissue. Removing the second *miR-1* gene, *miR-1-1*, might result in a stronger and more penetrant phenotype,

although the single mutant phenotype already provides robust evidence for the requirement of miR-1 in cardiogenesis.

Nodal signaling, which is required for endoderm and mesoderm formation in vertebrates, has been shown to be regulated by miRNAs in zebrafish and *Xenopus laevis* (Choi et al 2007, Martello et al 2007), but the effects of depleting this regulation differ significantly between the two organisms. In *Xenopus*, miR-15 and miR-16, which have the same seed, were shown to repress the Nodal type II receptor *Avcr2a* (Martello et al 2007). Antisense-mediated depletion of miR-15 and miR-16 in the *Xenopus* embryo led to strong expansion of the Spemann's organizer, which patterns the germ layers. Wnt/β-catenin signaling establishes the site of the organizer. pre-miR-15 and pre-miR-16 processing was shown to be inhibited by Wnt/β-catenin signaling, leading to asymmetric expression of the mature miRNAs, being enriched on the ventral side of the embryo, where the organizer forms. Subsequently, the miRNAs reinforce the complementary gradient of Nodal signaling by repressing *Avcr2a*, thereby making dorsal cells more responsive to Nodal signaling. This dorsal peak of Nodal signaling is required for the proper formation of the organizer. miR-15 and mir-16 therefore have an essential role in limiting the size of the Spemann's organizer during early embryonic development in *Xenopus*. Intriguingly, the miR-15 family / Nodal type II receptor interaction appears to be conserved from amphibians to humans, but not in teleosts as zebrafish, which might explain the discrepancies between the severity of global miRNA depletion in zebrafish and mice mentioned above.

5.3.2.2 Fine-Tuning of Developmental Programs: Conferring Robustness

Some miRNAs are expected to regulate a large set of targets expressed at low levels, thereby providing a second layer of regulation to reinforce and/or stabilize underlying gene expression programs. miRNA mutants of this class may be expected to have rather subtle phenotypes, and perhaps phenotypes that are difficult to study, due to the heterogeneous nature of their targets.

Surprisingly, most *Drosophila* larvae lacking the highly conserved, muscle specific miRNA miR-1 develop apparently normal, functional muscles (Sokol & Ambros 2005). 20% of the mutant embryos do not hatch but no obvious defects were observed (Brennecke et al 2005a), and the remaining 80% die with massively disrupted muscles when the mutants begin the rapid phase of larval growth upon feeding. However, if the mutant larvae are fed on sugar, conditions under which they live but do not grow, muscle function is perfectly normal and the animals survive. It is unclear which targets are misregulated in the *miR-1* mutant and how the muscle degeneration during growth is caused. This phenotype could reflect the requirement for miR-1 activity in the larval muscles as they grow rapidly. Alternatively, this defect might be the consequence of several subtle errors happening earlier during development, when miR-1 might be required to confer robustness to the identity and/or physiology of muscle cells.

The majority of flies depleted of the conserved miRNA *miR-8* are viable and fertile (Karres et al 2007). However, a small fraction of mutants died during pupal or early adult stages. Most of those and a substantial number of survivors had malformed, but properly patterned, legs and wings. The surviving morphologically normal flies displayed elevated apoptosis in the brain and a behavioral defect. Failure to downregulate *atrophin* to an optimal, but not inconsequential, level by miR-8 was shown to be responsible for the mutant phenotype, providing strong evidence for a tuning target relationship, which might be conserved in mammals.

Similarly, *miR-150* knockout mice were viable, fertile, and did not display any morphological defects. However, the number of B1 cells, a specific B cell subpopulation, was increased. Upon overexpression of miR-150, B1 cell formation was reduced, and so were the levels of the transcription factor c-Myb, which is required for B1 cell formation. The *c-Myb* 3'UTR could be regulated by miR-150 in a reporter assay. Heterozygous *c-Myb* knockout mice display comparable reduction of c-Myb protein levels as seen upon miR-150 overexpression, and B1 cell formation is comparably reduced, indicating the miR-150 may be required to downregulate c-Myb to a specific level at which it induces formation of a precise number of B1 cells. Again, this hints towards a tuning target relationship between miR-150 and c-Myb.

Zebrafish miR-214 is expressed during early somitogenesis and throughout embryogenesis. Antisense oligonucleotide-mediated depletion of miR-214 leads to an overall decrease in Hedgehog signaling in the presomitic mesodermal cells immediately adjacent to the notochord and therefore reduced the number of slow muscle cells (Flynt et al 2007). During normal development, these cells experience high Hedgehog levels, whereas their more lateral neighbors are exposed to lower amounts of Hedgehog. Su(fu), a negative regulator of Hedgehog signaling, is upregulated in miR-214 depleted embryos, probably due to the lack of regulation by miR-214. Su(fu) participates in Hedgehog signaling by retaining both the activator and repressor forms of the Gli transcription factors in the cytoplasm, and miRNA-mediated downregulation of Su(fu) is thought to maximize the response to different levels of Hedgehog signaling.

As mentioned above, Nodal signaling is regulated by a miRNA in zebrafish. The approach taken to tackle this topic was quite unconventional, in that the putative miR-430 target sites in the 3'UTRs of the Nodal ligand *squint* and the Nodal inhibitor *lefty2* were specifically blocked by antisense morpholino oligonucleotides, inhibiting the miRNA from binding and therefore releasing the mRNAs from repression (Choi et al 2007). The mRNA levels increased upon this treatment, and opposite phenotypes characteristic of increased, or decreased, Nodal signaling were observed for either the Nodal ligand, or the inhibitor. This analysis therefore elegantly demonstrated that miR-430 is required to keep Nodal signaling in balance by simultaneously dampening the Nodal ligand and inhibitors, classifying these miR-430 targets as tuning targets.

Antisense-mediated silencing of the abundant liver-specific miR-122 in the adult mouse liver leads to significant upregulation of more than 100 mRNAs containing miR-122 seed matches in their 3'UTRs. Phenotypically these mice are healthy and display no gross morphology changes, but their levels of total plasma cholesterol and triglycerides are substantially reduced. MiR-122 therefore has a function in hepatic lipid metabolism, and might be required to fine-tune this process (Esau et al 2006, Krutzfeldt et al 2005).

Mice lacking *miR-208*, which is expressed mainly in the heart, were viable and did not display discernible heart defects (van Rooij et al 2007). However, when the mutants were subjected to cardiac stress, they did not react with the stress-dependent cardiomyocyte growth control animals showed in this situation. miR-208 therefore seems to be required for cardiac growth in response to stress.

Members of the highly abundant miR-430 family are expressed at the onset of zygotic transcription in zebrafish embryos. Supplying maternal zygotic *dicer* mutant embryos (see above) with these miRNAs by injecting miR-430 duplexes rescues the brain morphogenesis defects seen in embryos lacking all miRNAs (Giraldez et al 2005). Moreover, the comparison of expression profiles of maternal zygotic *dicer* embryos to those of embryos resupplied with miR-430 duplexes revealed that several hundred transcripts, which are likely to be direct miR-430 targets, are misregulated in absence of miR-430 (Giraldez et al 2006). About 40% of those are maternally deposited mRNAs, suggesting that miR-430 is required to accelerate the clearance of pre-existing maternal mRNAs in the embryo when zygotic transcription starts (Figure 6). This study provides experimental evidence that a miRNA targets hundreds of transcripts in vivo and serves to ensure fidelity or robustness of the developmental program. In this thesis I present a similar function for the *miR-309* cluster in *Drosophila*, which contributes to the clearance of maternal mRNAs at the onset of zygotic transcription.

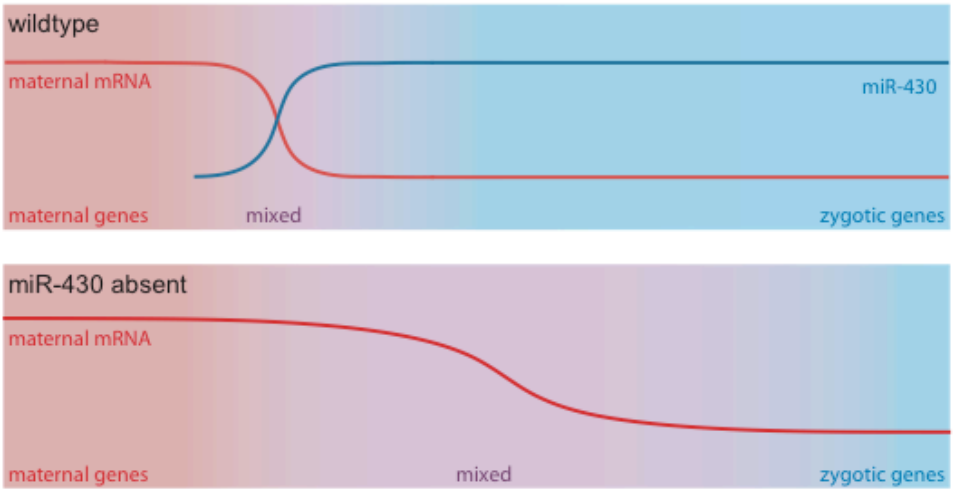


Figure 6. Fine-tuning embryonic development. In early zebrafish embryogenesis, miR-430 regulates the transition from maternal to zygotic mRNA transcription by targeting maternal mRNAs for degradation. In the absence of miR-430, maternal mRNA expression overlaps with zygotic gene expression.

miR-124 expression is highly CNS specific. The miRNA is not present in neural progenitors, but robustly induced upon differentiation, indicating that it might have a role in removing transcripts left behind from progenitor stages. In chick embryos, one study reported that neither antisense-mediated inhibition nor misexpression of miR-124 in the neural tube affected differentiation (Cao et al 2007). However, ectopic miR-124 led to disruptions of the basal lamina which surrounds the neural tube, and laminin γ1 and integrin β1 were shown to be the miR-124 targets responsible for this defect. Another chick study reported a mild stimulatory effect of miR-124 on neuronal differentiation, and consistently mild reduction of neuronal differentiation upon antisense-mediated mir-124 depletion (Visvanathan et al 2007). It is likely that repression of the anti-neural factor SCP1 by miR-124 is at least partly responsible for this effect. In neuroblastoma cell lines, miR-124 was also shown to induce neural differentiation and to target PTBP1, which represses alternative splicing in non-neuronal cells (Makeyev et al 2007). The targets identified in all of these studies are highly expressed in neural progenitors and repressed upon neuronal differentiation, consistent with the hypothesis that miRNAs and their targets are often expressed in a reciprocal manner, in either time or space (Stark et al 2005) (Figure 4).

In mouse embryos, *miR-196*, a miRNA gene located in all 4 mammalian HOX clusters in several copies, cleaves its target HOXB8 (Yekta et al 2004). Expression of *Hoxb8* and *Hoxa7*, a putative translational target of miR-196 (Hornstein et al 2005), is induced by retinoic acid in the vertebrate forelimb, but not in the hindlimb, where miR-196 is expressed. Thus, miR-196 functions to suppress the occurrence of unwanted transcripts in domains where these are already repressed transcriptionally, supporting the primary, transcriptional level of regulation to prevent *Hoxb8* and *Hoxa7* activation in the hindlimb. This provides another in vivo example of how apparently spatially reciprocal expression domains of miRNAs and targets can reflect an important regulatory relationship.

A similar relationship exists between miR-278 and its target, *expanded*, in adipose tissue in *Drosophila* (Teleman et al 2006). Although miR-278 overexpression results in overgrowth, the body size of mutant flies is normal. However, insulin levels are elevated and the mutants are lean due to insulin insensitivity of the adipose tissue, where miR-278 is predominantly expressed. A similar defect results in mice with adipose-tissue specific knockout of insulin responsiveness. *expanded* mRNA and protein levels are normally very low in adipose tissue but increase considerably in the *miR-278* mutant. Genetic analysis has shown that *expanded* overexpression causes the lean phenotype. This miRNA-target relationship is therefore an unusual example of spatial exclusion, since the miRNA contributes strongly to downregulation of the target's transcript levels, thereby helping to shape its expression domain.

5.3.2.3 Immune Response

miR-155 has been shown to be required for a functional immune response (Rodriguez et al 2007, Thai et al 2007). Mice which lack the *miR-155* gene were viable and fertile, but with age, these animals developed structural changes of the lung, similar to those seen in lung fibrosis, which often occurs in autoimmune processes with lung involvement. Most remarkably, *mir-155* mutants were immunodeficient: When they were vaccinated against a pathogen, they failed to develop protective immunity and hence died upon exposure to the pathogen. The

mutant's T cells, B cells and dendritic cells did not exert their proper functions, causing the immunodeficiency. It is unclear which targets are responsible for these phenotypes, but a wide spectrum of possible mir-155 targets were upregulated in the mutant.

5.3.2.4 Proliferation and Apoptosis

The *Drosophila* miRNA *bantam* was identified in a gain-of-function screen for regulators of tissue growth (Hipfner et al 2002), where it induces overgrowth of wings and eyes. *bantam* mutants are small and die as early pupae (Brennecke et al 2003). To allow tissue growth in development, apoptosis needs to be overcome. The 3'UTR of the proapoptotic gene *hid* contains 5 bantam target sites, and therefore, one part of the ability of bantam to promote growth can be explained by this regulatory relationship. But this is not sufficient to explain the mutant phenotype, since the rate of proliferation is reduced without inducing apoptosis. The targets involved in the generation of a positive growth regulatory output by bantam, eg. by inhibition of a negative growth regulator, have not yet been identified. Recently, the Hippo signaling pathway, which coordinately controls cell proliferation and apoptosis, has been shown to regulate bantam expression (Nolo et al 2006, Thompson & Cohen 2006).

5.3.2.5 Feedback Loops

Several miRNAs are part of regulatory, positive or negative, feedback loops. in those, their effects can be profound, when miRNA-mediated regulation is required for a cell fate decision, as is the case for *C. elegans* lsy-6 and miR-273, which are thought to act in a double-negative feedback loop to specify left-right asymmetry of gustatory neurons (Chang et al 2004, Johnston & Hobert 2003) (Figure 5). In other cases, the roles they play can be subtle, when the miRNA merely reinforces and stabilizes decisions made by other factors.

In the *Drosophila* eye, miR-7 reinforces a developmental decision via a reciprocal negative feedback loop (Li & Carthew 2005) (Figure 7). During photoreceptor differentiation, a transient EGF signal leads to degradation of the protein Yan and, by

releasing Yan-mediated repression of *miR-7* transcription, to elevated expression of *miR-7*. Subsequently, miR-7 represses Yan. Thereby a stable change in cell fate from progenitor to photoreceptor is achieved, and the differentiated fate is maintained by the presence of the miRNA. Consistently, the overall eye development of *miR-7* mutant flies appears normal. Only when the system was sensitized by introduction of a *Yan* allele, which does not respond normally to EGF receptor, developmental defects were observed, demonstrating that the miRNA does not control the feedback loop but reinforces an upstream decision. Furthermore, this is another example of mutually exclusive expression of a miRNA and its target, since *Yan* is downregulated when miR-7 expression is induced.

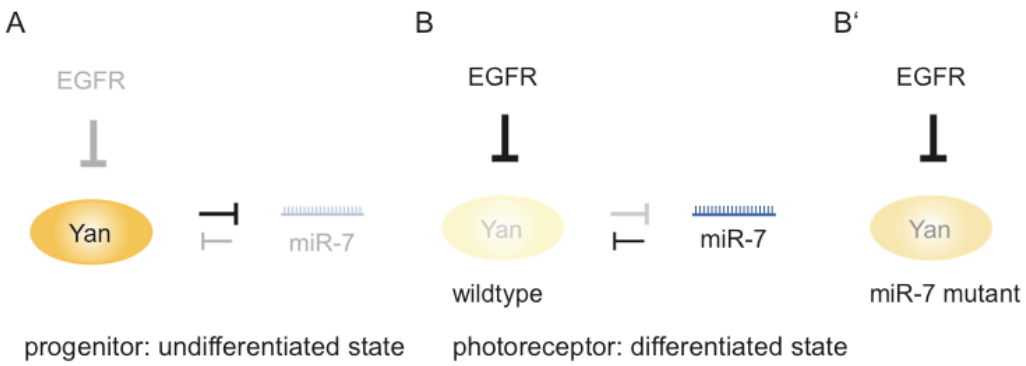


Figure 7. miR-7 reinforces photoreceptor differentiation. (A) Yan represses *miR-7* transcription in the progenitor. (B, B') During photoreceptor differentiation, epidermal growth factor receptor (EGFR) signaling induces Yan degradation, which leads to the release of *miR-7* repression. (B) miR-7 then reinforces Yan downregulation. (B') In the absence of miR-7, Yan is still repressed by EGFR, and therefore photoreceptor differentiation occurs normally.

A similar negative feedback loop has been found to operate during granulocytic differentiation (Fazi et al 2005): In undifferentiated myeloid precursors, the transcription factor NFI-A keeps levels of *miR-223* expression low. When ectopically expressed, miR-223 enhances the differentiation of myeloid precursors into granulocytes. Upon stimulation with retinoid acid, which induces differentiation, C/EBPα replaces NFI-A and induces high levels of *miR-223* transcription. miR-223 in turn represses NFI-A translation via a binding site in its 3'UTR, thus ensuring its own expression. Depletion of miR-223 followed by stimulation with retinoic acid leads to

reduced expression of a granulocytic differentiation marker (Fazi et al 2005). However, in a different study, cloning of the *mir-223* primary transcript revealed a highly conserved promoter with binding sites for the myeloid transcription factors PU.1 and C/EBPα and not NFI-A (Fukao et al 2007). PU.1 and C/EBPα were observed to be required for the myeloid expression of miR-223. In this study, the element for C/EBPα and NFI-A competition was assigned to an intronic region, implying that it might act as an intronic regulatory element. Whether one of the two proposed mechanisms of regulation predominates awaits further investigation.

In mouse postmitotic midbrain dopaminergic neurons (DNs), *miR-133b* transcription is regulated by the transcription factor Pitx3 (Kim et al 2007). At the same time, Pitx3 is targeted by miR-133b, forming a negative feedback loop. It was proposed that this feedback circuit increases response time and stability in the context of the dynamic changes in midbrain DN function.

In *C. elegans*, *miR-61* is directly transcriptionally activated in secondary vulval precursor cells by LIN-12/Notch (Yoo & Greenwald 2005). In turn, miR-61 represses Vav-1, which then releases its repression of LIN-12/Notch, thereby inducing a positive feedback loop, reinforcing LIN-12 activation and specification of secondary vulval cell fate. In contrast to miR-7 in *Drosophila*, where the miRNA only reinforces a decision made by other players, miR-61 may play a major role in cell fate specification.

5.3.2.6 Thresholding

Flies lacking *mir-9a*, a conserved nervous system-specific miRNA gene, produce extra sense organs (Li et al 2006). This defect is due to failure to repress the mir-9a target *senseless*. Indeed, the *miR-9a* mutant phenotype is suppressed by the removal of one copy of the *senseless* gene, which limits the extent to which *senseless* can be overexpressed. An interesting aspect of this mutant is that the absence of this regulation leads to a sporadic rather than a fully penetrant phenotype, with up to 40% of animals being affected. Why is that so? During selection of the sensory organ precursor (SOP) cell from cells of a proneural cluster,

senseless expression is turned on by proneural proteins and feedbacks positively to reinforce their expression. The resulting high level of proneural activity in the presumptive SOP cell leads to repression of proneural genes in surrounding cells, where mir-9a is expressed. There, the miRNA presumably makes sure that proneural gene expression is not induced by keeping *senseless* expression turned off, consistent with the model of mutual exclusion (Stark et al 2005). Interestingly, mir-9a is broadly expressed in the neurectoderm during early proliferation. During this phase, mir-9a might set a threshold which *senseless* expression has to overcome in order to induce SOP fate, thereby selecting only the cell with the highest transcriptional peak. In animals lacking miR-9a, *senseless* levels sporadically reach the threshold, which leads to the adoption of SOP fate in an additional cell. Here, the miRNA is not part of a feedback loop but limits the level at which a transcriptional feedback system can be activated.

The situation is similar for miR-14 (Varghese & Cohen 2007). Most *Drosophila* pupae lacking the ubiquitously expressed miR-14 die, but flies which eclose show activation of the apoptotic effector caspase Drice and defects in fat metabolism, are stress sensitive and live shorter (Xu et al 2003). Some of these phenotypes can be attributed to elevated levels of the miR-14 target Ecdysone receptor (EcR), since removal of on functional copy of the *EcR* gene in the *miR-14* mutant background was shown to restore life span and survival to almost normal levels (Varghese & Cohen 2007). Ecdysone is a steroid hormone which induces the major developmental transitions in insects. During the *Drosophila* larval-to-pupal transition it acts via its receptor to induce *EcR* transcription, thereby producing a transcriptional autoregulatory loop. In the same way, it mediates down-regulation of *miR-14*, thereby relieving miR-14-mediated repression of EcR, inducing maximal EcR activity. Again, the miRNA is not part of the regulatory loop, but involved in the extent of activation.

An attempt to investigate mir-9a's and other miRNAs' functions early in development has been made by injection of complementary 2' O-methyl oligoribonucleotides into the fly embryo (Leaman et al 2005). However, the phenotypes observed by this method differ significantly in penetrance and in nature from those observed in the miRNA null mutants available. More specifically,

antisense-mediated depletion of endogenous miR-1, miR-9a, miR-6 and several miRNAs of the miR-310 family was reported to cause a variety of drastic embryonic defects, but in the corresponding null mutants, the vast majority of embryos survive and for some of these it is possible to make homozygous mutant strains that are viable and fertile (Li et al 2006, this thesis, Sokol & Ambros 2005). It is unclear how this discrepancy can be explained.

6 miRNAs in Human Disease

One of the human diseases in which miRNAs have been implicated is the neuropsychiatric disorder Tourette's syndrome (TS) (Abelson et al 2005). The 3'UTR of *SLITRK1* contains a miR-189 binding site, which is mutated in some TS patients. This polymorphism replaces a GU base-pair with AU pairing, leading to stronger regulation by the miRNA. In situ hybridization of *SLITRK1* mRNA and miR-189 revealed co-expression in neuroanatomical circuits most commonly implicated in TS. Obviously this mutation is only one out of many rare mutations leading to this complex disease, but it demonstrates how a miRNA can be involved in the establishment of a disease phenotype.

90% of patients suffering from DiGeorge syndrome lack one copy of the chromosomal region 22q11 (Lindsay 2001). Interestingly, the Drosha partner DGCR8 maps to this region (Denli et al 2004, Gregory et al 2004, Han et al 2004, Landthaler et al 2004, Shiohama et al 2003), but it remains to be determined if reduced miRNA levels are an underlying cause of DiGeorge syndrome.

Expansion of the polyglutamine repeats in Ataxin-3 has been linked to neurodegeneration. Depletion of *dicer* in human cells led to a significant enhancement of Ataxin-3 induced toxicity (Bilen et al 2006). Which miRNAs are involved in this human neurodegenerative disorder remains to be determined. In Parkinson's patient samples, expression of miR-133b was deficient in the midbrain, and depletion of *Dicer* in midbrain dopaminergic neurons (DNs) resulted in reduced locomotion, a phenotype reminiscent of Parkinson's disease. Whether the lack of miR-133b is indeed responsible for this phenotype remains to be determined.

6.1 miRNAs in Cancer

miRNA expression levels are altered in primary human tumors (Calin et al 2004, Lu et al 2005). Global miRNA expression was reported to be lower in cancer tissues than in normal tissues in one study (Lu et al 2005), but another did not find this trend (Volinia et al 2006). Since many miRNAs are expressed in differentiated cell types, global miRNA levels might reflect the differentiation state of the tissue, and perhaps might be attributable to a failure of Drosha-processing (Thomson et al 2006). Intriguingly, significantly differing miRNA profiles could be assigned to various types of tumors, suggesting that miRNA profiling has diagnostic and perhaps prognostic potential (Calin & Croce 2006, Lu et al 2005). However, for the majority of miRNAs it is unknown whether they actually play an active role in tumor formation.

6.1.1 miRNAs as tumor suppressors

Loss of miRNAs in cancer tissue could suggest a role as tumor suppressors. The *miR-15a* and *miR-16-1* genes are located in chromosome region 13q14, which is deleted in most cases of chronic lymphocytic leukemia (Calin et al 2002). These miRNAs target *B cell lymphoma 2* (*Bcl2*), an antiapoptotic gene, suggesting that loss of miR-15a and miR-16-1 in B cells might lead to inhibition of apoptosis, giving rise to malignancies (Cimmino et al 2005). However, causality has not been established, and the correlation is not universal. The same miRNAs have been found to be overexpressed in a type of pancreatic tumor (Volinia et al 2006).

Recently, it has been shown that activation of the c-Myc oncogenic transcription factor led to repression of most miRNAs, including miR-15a, miR-16, miR-34a and let-7 family members. Enforced expression of some repressed miRNAs even reduced the tumorigenic potential of lymphoma cells (Chang et al 2007). Loss of let-7 mediated repression of the oncogene *Hmga2*, which is caused by a chromosomal translocation of the *Hmga2* locus, enhances oncogenic transformation (Lee & Dutta 2007, Mayr et al 2007), implying a role for let-7 as a tumor suppressor.

The important tumor suppressor p53, which is mutated in nearly all cancers, has been shown to directly regulate the transcription of miR-34 family miRNAs (Bommer et al 2007, He et al 2007). Overexpression of miR-34 family members in four different tumor cell lines led to downregulation of mRNAs significantly enriched in miR-34 seed matches. Genes involved in the control of the cell cycle were strongly overrepresented in this set, consistent with the anti-proliferative potential of p53, indicating that this potential is partly mediated by miR-34 family miRNAs.

6.1.2 miRNAs as Oncogenes

miRNAs with oncogenic potential are expressed from the *miR17-92* locus 13q31, which is amplified in some tumors, eg. 10-fold in B-cell lymphoma samples (He et al 2005). Overexpression of this cluster in a mouse model of human B-cell lymphoma accelerated c-Myc induced tumorigenesis, and the apoptosis normally seen in c-Myc induced tumors was suppressed (He et al 2005). c-Myc has been reported to directly transcriptionally activate the *miR17-92* cluster and the pro-apoptotic E2F1, which is in turn targeted by 2 miRNAs of the cluster (O'Donnell et al 2005). These miRNAs would therefore support a shift from apoptosis toward proliferation, by downregulating E2F1. Moreover, c-Myc induced activation of the *miR17-92* cluster leads to enhanced tumor angiogenesis in mouse colonocytes, probably via direct, miRNA mediated downregulation of anti-angiogenic proteins (Dews et al 2006).

Primary human fibroblasts expressing the related miRNAs miR-372 and miR-373 were able to overcome oncogenic Ras mediated arrest, and therefore induced tumorigenesis (Voorhoeve et al 2006). In part, this effect is mediated by targeting the tumor suppressor LATS2. miR-372 and miR-373 were found to be expressed specifically in testicular germ cell tumors.

Interestingly, both miRNA clusters are highly expressed in embryonic stem cells (Suh et al 2004, Thomson et al 2004), which suggests that they contribute to tumorigenesis by exerting their normal function at the wrong time and place in the organism. Another miRNA with oncogenic potential is miR-155, which is overexpressed in several kinds of B-cell lymphomas and can induce preleukaemic

pre-B-cell proliferation in mice, when overexpressed specifically in B cells (Costinean et al 2006).

miR-10b is highly and specifically expressed in metastatic breast cancer cells, but not in primary breast tumors (Ma et al 2007). It leads to invasion and metastasis when overexpressed in otherwise non-metastatic breast cancer cells *in vitro* and in vivo. mir-10b represses HOXD10, which normally represses RHOC, a key player in metastasis. Releasing this repression by strongly reducing HOXD10 levels seems to be the main downstream event caused by high miR-10b levels in metastasis.

6.2 miRNAs and viruses

Viruses use miRNAs in their effort to control their host cell; reciprocally, host cells use miRNAs to target essential viral functions. miRNAs have been found in nuclear DNA viruses like the herpesvirus, but to date none have been found in RNA viruses (Cullen 2006, Pfeffer et al 2005, Pfeffer et al 2004). Most RNA viruses are restricted to the cytoplasm and are therefore not expected to encode miRNAs, since miRNA transcripts need to be processed in the nucleus by Drosha.

The SV40 encoded miRNA miR-S1 helps keep the infected cell hidden from the immune system. It is expressed late in the viral replication cycle, when it acts to degrade early viral mRNAs encoding T antigen, limiting exposure of the infected cell to cytotoxic T lymphocytes (Sullivan et al 2005). A herpes virus miRNA, miR-UL112, inhibits translation of cellular MICB, a ligand for a receptor on natural killer cells of the immune system, again hiding the cell from the immune system (Stern-Ginossar et al 2007). *miR-K12-11*, which is encoded by Kaposi's–sarcoma-associatetd herpes virus, shares significant sequence homology with cellular mir-155 and therefore has the potential to exploit pathways normally regulated by miR-155 in the cell (Gottwein et al 2007). miR-LAT of herpes simplex virus-1 inhibits apoptosis of latently infected neurons by targeting the pro-apoptotic TGF-β and one of its mediators, SMAD3 (Gupta et al 2006). The miRNA thereby keeps its host cell alive.

Endogenous, cellular miRNAs that target viral RNAs have been reported as well. In one scenario, the cell uses the miRNA to impede viral replication. miR-32 restricts the replication of the retrovirus PFC-1 in cell culture (Lecellier et al 2005). In another scenario, the virus takes advantage of an endogenous cellular miRNA. Replication of Hepatitis C viral RNA is facilitated by binding of the liver-specific miRNA miR-122 to the 5' noncoding region of the viral genome (Jopling et al 2005). Intriguingly, interferon β was shown to downregulate *miR-122* and simultaneously induce the expression of many cellular miRNAs (Pedersen et al 2007). Eight of these have target sites in the hepatitis C virus genomic RNA and can inhibit Hepatitis C virus replication and infection. Therefore, modulation of the levels of cellular miRNAs which can inhibit or activate viral replication contributes elegantly to the antiviral effects of interferon β. A cluster of miRNAs enriched in resting $CD4^+$ T cells has been found to potentially repress HIV-1 mRNAs and thereby contribute to HIV-1 latency in these cells (Huang et al 2007). These miRNAs are therefore potential therapeutic targets, their inhibition could activate latent HIV-1 reservoirs in order to expose them to antiretroviral drugs. HIV-1 was also found to actively suppress the expression of the miRNA cluster *miR-17/92* in order to replicate efficiently, however, this did not involve targeting of the viral genome itself, but of a cellular cofactor for HIV-1 replication (Triboulet et al 2007).

Aims of the Thesis

When this work was initiated, very little was known about the biological roles of miRNAs. Few mutants were available, and few in vivo miRNA targets were known. In the mutants identified, misregulation of a small number of targets seemed to be attributable to the mutant phenotypes (Lee et al 1993, Reinhart et al 2000, Wightman et al 1993). This contrasted with the high number of computationally predicted targets for individual miRNAs (Rajewsky 2006). However, whether these predicted targets were in fact physiologically relevant was an open question. The aim of this thesis was to contribute to the understanding of the biological roles of different miRNAs in *Drosophila*.

I was involved in a global analysis of the expression patterns of miRNAs and their predicted targets (Publication 1), which revealed that many miRNAs are not co-expressed with their targets, indicating that miRNAs might act by preventing the expression of unwanted transcripts that are present only at very low levels in the miRNA expressing cell. This observation could explain the high number of conserved miRNA targets in the genome.

In a collaboration with Alexander Stark, I analyzed the expression pattern and possible biological function in Hox gene regulation of miR-iab-4AS, a miRNA derived from the antisense strand of the known miRNA gene *miR-iab-4*. This study is presented in Publication 2.

An unambiguous way to determine a miRNAs' role in development is to analyze the phenotypes of animals lacking the miRNA. Using a reverse genetic approach, I investigated the role of the *miR-309* cluster during early embryogenesis. I generated a mutant deleting the *miR-309* cluster using homologous recombination, which was a new method at that time. Using microarray profiling of mutant versus control embryos, I identified in vivo targets of *miR-309* cluster miRNAs. I could show that the miRNAs in this cluster are required for the timely downregulation of maternally deposited transcripts during the maternal zygotic transition (Publication 3).

Summaries of Publications

Publication 1:

Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution

Alexander Stark, Julius Brennecke, **Natascha Bushati**, Robert B. Russell and Stephen M. Cohen (2005), Cell 123, 1133-46

To assess the biological roles of a miRNA, it is of significant importance to know which gene(s) it regulates. In our study, miRNA target predictions were improved based on a previous systematic experimental analysis of functional target site architecture in *Drosophila* (Brennecke et al 2005b). Predictions were restricted to such sites conserved between *Drosophila melanogaster* and *Drosophila pseudoobscura* 3'UTRs in order to filter for functionally relevant sites. These predictions yielded an average of 179 target sites per miRNA. I validated the functionality of eight out of nine predicted target sites using a luciferase-based 3'UTR reporter assay in S2 cells. miRNA-target pairs which had been experimentally tested in other ways were also included in the analysis of the functionality of predicted target sites, resulting in a total of 57 tested miRNA-target pairs. 88% (50/57) of these were functional, demonstrating the high accuracy of the improved target predictions.

95% of the targeted 3'UTRs contain only one conserved site for individual miRNAs, indicating that strong regulation by a single miRNA via multiple binding sites is rare. But almost 50% of the targeted 3'UTRs contain target sites for two or more miRNAs with differing seed sequences, suggesting potential cooperative regulation by coexpressed miRNAs or complementary regulation by different miRNAs in different cells or conditions.

We found that genes which were coexpressed with an individual miRNA were significantly depleted of both conserved and non-conserved miRNA binding sites. This implies that these genes were under evolutionary pressure to avoid miRNA-mediated regulation. Indeed, they were annotated to be mainly involved in basic

cellular processes. We confirmed this correlation using a collection of annotated in situ gene expression patterns for *Drosophila* embryogenesis (Tomancak et al 2007): The set of genes annotated as ubiquitously expressed significantly avoid miRNA target sites.

Conversely, we found that many miRNAs and their targets were expressed in a largely nonoverlapping manner, either temporally or spatially. In the latter case, the genes containing the conserved target sites were typically expressed in domains adjacent to the miRNA expressing cells.

What do these observations mean regarding the possible functions of miRNAs? The target sites should be of advantage to the animal, otherwise they would not have been conserved. We proposed that miRNAs can act by preventing the expression of unwanted transcripts, either from preexisting mRNAs during developmental transitions or simply from leaky transcription. Thereby, the miRNAs would reinforce developmental decisions, acting as a second level of regulation after transcriptional regulation. miRNAs acting in this way would ensure accuracy and confer robustness to developmental programs.

Publication 2:

A single Hox locus in *Drosophila* produces functional microRNAs from opposite DNA strands

Alexander Stark[1], **Natascha Bushati[1]**, Calvin Jan, Pouya Kheradpour, Emily Hodges, Julius Brennecke, David P. Bartel, Stephen M. Cohen and Manolis Kellis (2008), Genes Dev. 22, 8-13

Homeobox - containing (Hox) transcription factors are highly conserved proteins fundamentally required during animal development. The spatial colinearity between Hox gene expression along the anteroposterior axis of the embryo and the order of genes along the chromosome is conserved (Duboule 1998, Pearson et al 2005). Hox clusters also encode several noncoding RNAs, including the miRNA genes *miR-10* and *miR-iab4/miR-196*. Strikingly, these miRNAs are located in analogous positions in flies and vertebrates (Yekta et al 2004). Ectopic miR-iab-4 expression was shown to directly repress the Hox transcription factor Ubx, proposing that miR-iab-4 is partly required for the correct spatial expression of Ubx (Ronshaugen et al 2005).

We found evidence that the reverse complement of the miR-iab-4 hairpin also folds into a precursor miRNA (pre-miRNA) hairpin. The mature product of this antisense transcript, miR-iab-4 anti-sense (miR-iab-4AS), was detected by high-throughput sequencing of small RNA libraries of *Drosophila* testes and ovaries. Using strand-specific RT-PCR, I detected the primary transcripts of both *miR-iab-4* and *miR-iab-4AS* during all stages of *Drosophila* development. In situ hybridization revealed that the two miRNA primary transcripts are expressed in nonoverlapping domains in the *Drosophila* embryo. As reported previously (Bae et al 2002, Ronshaugen et al 2005), *miR-iab-4* was expressed highly in abdominal segments A5-A7. Within the segments, cells containing the Hox gene *abd-A* exhibited higher, those expressing *Ubx* very low *miR-iab-4* levels. In contrast, *miR-iab-4AS* transcription was detected in abdominal segments A8 and A9, where neither *abd-A* nor *Ubx* are expressed, but only *Abd-B*. Abd-B transcriptionally represses the Hox genes *abd-A*, *Ubx* and *Antp* in these segments (Pearson et al 2005).

[1] with equal contribution

We found several highly conserved seed matches to miR-iab-4AS in the 3'UTRs of *Ubx*, *abd-A* and *Antp*. Overexpression of miR-iab-4AS conferred downregulation of Luciferase reporters carrying the *Ubx* or *abd-A* 3'UTRs, but not of reporters carrying 3'UTRs in which the seed matches were mutated, indicating that conserved sites were indeed functional. Moreover, misexpression of miR-iab-4AS in the fly haltere resulted in a strong haltere-to-wing transformation, which is the phenotype observed upon depletion of *Ubx* in the haltere. Therefore, endogenous Ubx is highly susceptible to regulation by miR-iab-4AS in vivo. An accompanying study reported derepression of Ubx protein in embryos depleted of the *mir-iab-4* locus (Bender 2008), demonstrating the in vivo significance of this regulatory relationship.

miR-iab-4AS therefore seems to support the transcriptional repression conferred by Abd-B upon *abd-A*, *Ubx* and *Antp*. miR-iab-4, in turn, is coexpressed with *abd-A*, and has target sites in *Ubx* and *Antp*, which are transcriptionally repressed by both Abd-B and abd-A. Therefore, these miRNAs work together to reinforce the regulatory relationships between Hox genes. Using high-throughput sequencing in *Drosophila* and mouse, more putative antisense miRNAs were identified. Hence, the regulatory and spatial relationships between sense/antisense miRNAs and their target spectra may constitute a more common mechanism to establish and maintain expression domains, in which miRNAs expressed from the same locus but on different strands are expressed in a nonoverlapping manner.

Publication 3:

Temporal reciprocity of microRNAs and their targets during the maternal to zygotic transition in *Drosophila*

Natascha Bushati, Alexander Stark, Julius Brennecke and Stephen M. Cohen (submitted)

During oogenesis, eggs are loaded with maternally provided transcripts that will be translated to produce new proteins in the developing embryo. A certain fraction of these transcripts needs to be degraded during the transition from purely maternal to mixed maternal-zygotic gene expression (maternal to zygotic transition). In *Drosophila*, two independent pathways act to ensure timely turnover of maternal transcripts (Bashirullah et al 1999). One is driven by maternally encoded factors, including SMAUG (Tadros et al 2007), whereas the other is activated at the onset of zygotic transcription, about 2 hours after fertilization.

Using quantitative real-time PCR, I found that the expression of the miRNAs in the *Drosophila miR-309* cluster was strongly induced at the onset of zygotic transcription. In situ hybridization of the *miR-309* cluster primary transcript showed that it is expressed ubiquitously at this stage, with the exception of the pole cells, which give rise to the germline. The *miR-309* cluster contains 8 miRNA genes, which encode 6 different miRNAs. Only two of these miRNAs have the same seed sequence, therefore the cluster has the potential to regulate a broad spectrum of target genes.

Using homologous recombination, I generated a mutant in which the *miR-309* cluster was deleted. Homozygous mutant embryos completed embryogenesis, but ~20% died at larval stages without discernible defects. The remaining mutants survived and were viable and fertile. However, they showed a developmental delay during larval stages. I was able to rescue these phenotypes by expression of a transgene containing a genomic fragment spanning the miRNA cluster. I did not observe the severe embryonic defects that were reported using anti-sense 2'-O-methyl antisense oligonucleotides injections to deplete the *miR-309* cluster miRNAs miR-286 or mir-6 (Leaman et al 2005).

When we compared expression of *miR-309* cluster miRNAs to a high-resolution temporal gene expression profile of early embryonic development (Pilot et al 2006), we found that mRNAs with a temporal expression profile most similar to that of miR-309 cluster miRNAs contained significantly fewer 7mers complementary to *miR-309* cluster miRNAs in their 3'UTRs. Conversely, maternal transcripts which were strongly downregulated when miRNA expression is induced contained significantly more 7mer sites in their 3'UTRs, suggesting that the miRNAs might contribute to this downregulation.

To address this possibility, I compared mRNA expression profiles of control and mutant embryos at 0-1h and 2-3h of embryonic development. During the first time point, *miR-309* cluster miRNAs are barely detectable, whereas they are strongly induced during the 2-3h interval. For our analysis we took advantage of two different sets of mRNAs classified as maternal transcripts. In one of the sets, maternal mRNAs were classified according to their degree of down-regulation during the maternal to zygotic transition in fertilized embryos (Arbeitman et al 2002). The other set classified mRNAs as stable or unstable maternal transcripts based on expression profiling RNA from unfertilized eggs (Tadros et al 2007).

Both, maternal mRNAs which are normally strongly downregulated in fertilized embryos as well as transcripts classified as unstable in unfertilized eggs, were significantly enriched among the genes upregulated >1.5 fold in mutant embryos at 2-3 hours, indicating that their downregulation is affected in the *miR-309* cluster mutants. Moreover, mRNAs containing 7mers complementary to the seed of one or more cluster miRNAs were enriched among the >1.5 fold upregulated genes at this time point. Transcripts classified as unstable in fertilized or unfertilized embryos which contained 7mer seed sites were even more enriched. Importantly, we did not observe any significant enrichment or depletion in 0-1h embryos. To confirm that the sites we had identified in the upregulated genes were indeed functional, I tested 32 3'UTRs of the affected genes in a luciferase based assay. 29 of the 32 reporters were significantly downregulated upon *miR-309* cluster expression. These results

indicate that *miR-309* cluster miRNAs contribute directly to the downregulation of maternal transcripts during the maternal to zygotic transition.

We also examined 7mer seed matches complementary to individual *miR-309* cluster miRNAs. Sites complementary to 4 of the 5 unique seeds were significantly enriched among the upregulated transcripts at 2-3 hours, indicating that the cluster miRNAs, with the possible exception of one, indeed act in concert to regulate a broad spectrum of targets.

Since SMAUG is a key component of the maternal system for maternal mRNA degradation (Tadros et al 2007), and the *miR-309* cluster acts zygotically on its targets, we were interested in the degree of overlap between the targets of these two systems. We found that SMAUG targets were present in the set of miRNA regulated genes to the same extent as unstable genes which were not targeted by SMAUG. Therefore, the maternal and zygotic systems seem to regulate distinct, but overlapping sets of maternal mRNAs.

Our results suggest that the zygotically expressed *miR-309* cluster acts to promote the turnover of many maternally deposited transcripts. Intriguingly, a comparable role has been reported for the miRNA miR-430 in zebrafish (Giraldez et al 2006), suggesting that the use of miRNAs to promote degradation of maternal mRNAs during the maternal to zygotic is a conserved phenomenon. Since the miRNAs in *Drosophila* and zebrafish are unrelated, this might be an example of convergent evolution.

Publication 4:

MicroRNA Functions (Review)

Natascha Bushati and Stephen M. Cohen (2007), Annu Rev Cell Dev Biol 23: 175-205

In this review article, we discussed recent work, principally from animal models, that reveals how miRNAs are generated and act to silence gene expression, how targets of miRNAs can be identified, and how the biological functions of miRNAs can be illuminated by knowledge of gene expression patterns, by mutant phenotypes of miRNAs, and by overexpression of their targets. We focused especially on the analysis of miRNA functions in vivo, which has begun to shed light on the types of biological processes that miRNAs regulate. Moreover, we outlined the variety of ways miRNAs can act in the context of other regulatory mechanisms. In summary, it appears that their biological roles in development and disease as well as their modes of action are diverse.

Contribution

This part was written by Dr. Stephen Cohen.

Publication 1:

Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution

Alexander Stark, Julius Brennecke, **Natascha Bushati**, Robert B. Russell and Stephen M. Cohen (2005), Cell 123, 1133-46

Natascha contributed important experiments to this paper.

Publication 2:

A single Hox locus in *Drosophila* produces functional microRNAs from opposite DNA strands

Alexander Stark[1], **Natascha Bushati[1]**, Calvin Jan, Pouya Kheradpour, Emily Hodges, Julius Brennecke, David P. Bartel, Stephen M. Cohen and Manolis Kellis (2008), Genes Dev 22, 8-13

This publication was based on a computational observation by Alex Stark. Natascha carried out all of the experimental work involving *Drosophila*. She contributed to analysis of the data, writing the paper and prepared figures. Other middle authors contributed microRNA sequence data. We discussed experiments and their design. Natascha and Alex share co-first author, reflecting that theirs are the major contributions.

[1] with equal contribution

Contribution

Publication 3:

Temporal reciprocity of microRNAs and their targets during the maternal to zygotic transition in *Drosophila*

Natascha Bushati, Alexander Stark, Julius Brennecke and Stephen M. Cohen (submitted)

Natascha carried out all of the experimental work. Natascha and I designed experiments and interpreted results. She wrote the first draft of the manuscript. Alex Stark carried out the computational analysis needed for interpretation of some of the results. Julius Brennecke made a small early contribution.

Publication 4:

MicroRNA Functions (Review)

Natascha Bushati and Stephen M. Cohen (2007), Annu Rev Cell Dev Biol 23: 175-205

Natascha wrote the first draft of the review entirely independently and we discussed revisions in depth.

Stephen M. Cohen

References

Abbott AL, Alvarez-Saavedra E, Miska EA, Lau NC, Bartel DP, et al. 2005. The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in *Caenorhabditis* elegans. *Dev Cell* 9: 403-14

Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, et al. 2005. Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* 310: 317-20

Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC. 2005. *Drosophila* microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proc Natl Acad Sci U S A* 102: 18017-22

Abrahante JE, Daul AL, Li M, Volk ML, Tennessen JM, et al. 2003. The *Caenorhabditis* elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. *Dev Cell* 4: 625-37

Andl T, Murchison EP, Liu F, Zhang Y, Yunta-Gonzalez M, et al. 2006. The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles. *Curr Biol* 16: 1041-9

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442: 203-7

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, et al. 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 5: 337-50

Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, et al. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science* 297: 2270-5

Arteaga-Vazquez M, Caballero-Perez J, Vielle-Calzada JP. 2006. A Family of MicroRNAs Present in Plants and Animals. *Plant Cell*

Ashraf SI, McLoon AL, Sclarsic SM, Kunes S. 2006. Synaptic protein synthesis associated with memory is regulated by the RISC pathway in *Drosophila*. *Cell* 124: 191-205

Ason B, Darnell DK, Wittbrodt B, Berezikov E, Kloosterman WP, et al. 2006. Differences in vertebrate microRNA expression. *Proc Natl Acad Sci U S A* 103: 14385-9

Bae E, Calhoun VC, Levine M, Lewis EB, Drewell RA. 2002. Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the *Drosophila* bithorax complex. *Proc Natl Acad Sci U S A* 99: 16847-52

Bagga S, Bracht J, Hunter S, Massirer K, Holtz J, et al. 2005. Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* 122: 553-63

Bartel DP, Chen CZ. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* 5: 396-400

Bashirullah A, Halsell SR, Cooperstock RL, Kloc M, Karaiskakis A, et al. 1999. Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in *Drosophila melanogaster*. *Embo J* 18: 2610-20

Baskerville S, Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna* 11: 241-7

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* 20: 1885-98

Beitzinger M, Peters L, Zhu JY, Kremmer E, Meister G. 2007. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol* 4: 76-84

Bender W. 2008. MicroRNAs in the *Drosophila* bithorax complex. *Genes Dev* 22: 14-9

Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. 2007. Mammalian mirtron genes. *Mol Cell* 28: 328-36

Berezikov E, Cuppen E, Plasterk RH. 2006. Approaches to microRNA discovery. *Nat Genet* 38 Suppl: S2-7

References

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, et al. 2003. Dicer is essential for mouse development. *Nat Genet* 35: 215-7

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: 1111-24

Bilen J, Liu N, Burnett BG, Pittman RN, Bonini NM. 2006. MicroRNA pathways modulate polyglutamine-induced neurodegeneration. *Mol Cell* 24: 157-63

Boehm M, Slack F. 2005. A developmental timing microRNA and its target regulate life span in *C. elegans*. *Science* 310: 1954-7

Bohnsack MT, Czaplinski K, Gorlich D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* 10: 185-91

Bommer GT, Gerin I, Feng Y, Kaczorowski AJ, Kuick R, et al. 2007. p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr Biol* 17: 1298-307

Borchert GM, Lanier W, Davidson BL. 2006. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* 13: 1097-101

Boutz PL, Chawla G, Stoilov P, Black DL. 2007. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev* 21: 71-84

Bracht J, Hunter S, Eachus R, Weeks P, Pasquinelli AE. 2004. Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. *Rna* 10: 1586-94

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, et al. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* 128: 1089-103

Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell* 113: 25-36

Brennecke J, Stark A, Cohen SM. 2005a. Not miR-ly muscular: microRNAs and muscle development. *Genes Dev* 19: 2261-4

Brennecke J, Stark A, Russell RB, Cohen SM. 2005b. Principles of microRNA-target recognition. *PLoS Biol* 3: e85

Cai X, Hagedorn CH, Cullen BR. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* 10: 1957-66

Calin GA, Croce CM. 2006. MicroRNA signatures in human cancers. *Nat Rev Cancer* 6: 857-66

Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, et al. 2002. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A* 99: 15524-9

Calin GA, Liu CG, Sevignani C, Ferracin M, Felli N, et al. 2004. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc Natl Acad Sci U S A* 101: 11755-60

Cao X, Pfaff SL, Gage FH. 2007. A functional study of miR-124 in the developing neural tube. *Genes Dev* 21: 531-6

Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, et al. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* 12: 503-14

Chalfie M, Horvitz HR, Sulston JE. 1981. Mutations that lead to reiterations in the cell lineages of *C. elegans*. *Cell* 24: 59-69

Chang S, Johnston RJ, Jr., Frokjaer-Jensen C, Lockery S, Hobert O. 2004. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* 430: 785-9

Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, et al. 2007. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell* 26: 745-52

Chen CZ, Li L, Lodish HF, Bartel DP. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303: 83-6

Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, et al. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* 38: 228-33

Chen K, Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 38: 1452-6

Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8: 93-103

Chen PY, Manninga H, Slanchev K, Chien M, Russo JJ, et al. 2005. The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev* 19: 1288-93

Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, et al. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436: 740-4

Choi WY, Giraldez AJ, Schier AF. 2007. Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. *Science* 318: 271-4

Chu CY, Rana TM. 2006. Translation Repression in Human Cells by MicroRNA-Induced Gene Silencing Requires RCK/p54. *PLoS Biol* 4: e210

Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, et al. 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* 102: 13944-9

Cobb BS, Nesterova TB, Thompson E, Hertweck A, O'Connor E, et al. 2005. T cell lineage choice and differentiation in the absence of the RNase III enzyme Dicer. *J Exp Med* 201: 1367-73

Costinean S, Zanesi N, Pekarsky Y, Tili E, Volinia S, et al. 2006. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A* 103: 7024-9

Cullen BR. 2006. Viruses and microRNAs. *Nat Genet* 38 Suppl: S25-30

Deng W, Lin H. 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* 2: 819-30

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* 432: 231-5

Dews M, Homayouni A, Yu D, Murphy D, Sevignani C, et al. 2006. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nat Genet*

Didiano D, Hobert O. 2006. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat Struct Mol Biol* 13: 849-51

Doench JG, Petersen CP, Sharp PA. 2003. siRNAs can function as miRNAs. *Genes Dev* 17: 438-42

Doench JG, Sharp PA. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev* 18: 504-11

Duboule D. 1998. Vertebrate hox gene regulation: clustering and/or colinearity? *Curr Opin Genet Dev* 8: 514-8

Easow G, Teleman AA, Cohen SM. 2007. Isolation of microRNA targets by miRNP immunopurification. *Rna* 13: 1198-204

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. 2003. MicroRNA targets in *Drosophila*. *Genome Biol* 5: R1

Esau C, Davis S, Murray SF, Yu XX, Pandey SK, et al. 2006. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab* 3: 87-98

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E. 2007a. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol* 27: 3970-81

Eulalio A, Rehwinkel J, Stricker M, Huntzinger E, Yang SF, et al. 2007b. Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev* 21: 2558-70

Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, et al. 2005. The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* 310: 1817-21

Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, et al. 2005. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. *Cell* 123: 819-31

Flynt AS, Li N, Thatcher EJ, Solnica-Krezel L, Patton JG. 2007. Zebrafish miR-214 modulates Hedgehog signaling to specify muscle cell fate. *Nat Genet*

Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. 2007. *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* 130: 287-97

Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, et al. 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* 3: e236

Fukao T, Fukuda Y, Kiga K, Sharif J, Hino K, et al. 2007. An evolutionarily conserved mechanism for microRNA-223 expression revealed by microRNA gene profiling. *Cell* 129: 617-31

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, et al. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* 308: 833-8

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, et al. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312: 75-9

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442: 199-202

Gottwein E, Mukherjee N, Sachse C, Frenzel C, Majoros WH, et al. 2007. A viral microRNA functions as an orthologue of cellular miR-155. *Nature* 450: 1096-9

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 123: 631-40

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, et al. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432: 235-40

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, et al. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106: 23-34

Grivna ST, Beyret E, Wang Z, Lin H. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* 20: 1709-14

Grosshans H, Johnson T, Reinert KL, Gerstein M, Slack FJ. 2005. The temporal patterning microRNA let-7 regulates several transcription factors at the larval to adult transition in *C. elegans*. *Dev Cell* 8: 321-30

Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N. 2005. microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput Biol* 1: e13

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, et al. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315: 1587-90

Gupta A, Gartner JJ, Sethupathy P, Hatzigeorgiou AG, Fraser NW. 2006. Anti-apoptotic function of a microRNA encoded by the HSV-1 latency-associated transcript. *Nature* 442: 82-5

Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* 18: 3016-27

Harfe BD, McManus MT, Mansfield JH, Hornstein E, Tabin CJ. 2005. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proc Natl Acad Sci U S A* 102: 10898-903

Harris KS, Zhang Z, McManus MT, Harfe BD, Sun X. 2006. Dicer function is essential for lung epithelium morphogenesis. *Proc Natl Acad Sci U S A* 103: 2208-13

Hatfield SD, Shcherbata HR, Fischer KA, Nakahara K, Carthew RW, Ruohola-Baker H. 2005. Stem cell division is regulated by the microRNA pathway. *Nature* 435: 974-8

He L, He X, Lim LP, de Stanchina E, Xuan Z, et al. 2007. A microRNA component of the p53 tumour suppressor network. *Nature* 447: 1130-4

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, et al. 2005. A microRNA polycistron as a potential human oncogene. *Nature* 435: 828-33
Hipfner DR, Weigmann K, Cohen SM. 2002. The bantam gene regulates *Drosophila* growth. *Genetics* 161: 1527-37
Hornstein E, Mansfield JH, Yekta S, Hu JK, Harfe BD, et al. 2005. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature* 438: 671-4
Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat Genet* 38 Suppl: S20-4
Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129: 69-82
Huang J, Wang F, Argyris E, Chen K, Liang Z, et al. 2007. Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes. *Nat Med* 13: 1241-7
Humphreys DT, Westman BJ, Martin DI, Preiss T. 2005. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc Natl Acad Sci U S A* 102: 16961-6
Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293: 834-8
Hutvagner G, Zamore PD. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297: 2056-60
Hwang HW, Wentzel EA, Mendell JT. 2007. A hexanucleotide element directs microRNA nuclear import. *Science* 315: 97-100
Jackson RJ, Standart N. 2007. How do microRNAs regulate gene expression? *Sci STKE* 2007: re1
Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. 2005. Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev* 19: 1674-9
Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, et al. 2005. Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* 120: 623-34
Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, et al. 2005. RAS is regulated by the let-7 microRNA family. *Cell* 120: 635-47
Johnston RJ, Hobert O. 2003. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis* elegans. *Nature* 426: 845-9
Johnston RJ, Jr., Chang S, Etchberger JF, Ortiz CO, Hobert O. 2005. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc Natl Acad Sci U S A* 102: 12449-54
Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* 309: 1577-81
Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, et al. 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, et al. 2007. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci U S A* 104: 19291-6
Karres JS, Hilgers V, Carrera I, Treisman J, Cohen SM. 2007. The conserved microRNA miR-8 tunes atrophin levels to prevent neurodegeneration in *Drosophila*. *Cell* 131: 136-45
Kedde M, Strasser MJ, Boldajipour B, Vrielink JA, Slanchev K, et al. 2007. RNA-Binding Protein Dnd1 Inhibits MicroRNA Access to Target mRNA. *Cell* 131: 1273-86
Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. *Nat Genet* 39: 1278-84
Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev* 15: 2654-9
Kim J, Inoue K, Ishii J, Vanti WB, Voronov SV, et al. 2007. A MicroRNA feedback circuit in midbrain dopamine neurons. *Science* 317: 1220-4

Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, et al. 2004. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev* 18: 1165-78

Kiriakidou M, Tan GS, Lamprinaki S, De Planell-Saguer M, Nelson PT, Mourelatos Z. 2007. An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell* 129: 1141-51

Kloosterman WP, Steiner FA, Berezikov E, de Bruijn E, van de Belt J, et al. 2006. Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res* 34: 2558-69

Kloosterman WP, Wienholds E, Ketting RF, Plasterk RH. 2004. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res* 32: 6284-91

Knight SW, Bass BL. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis* elegans. *Science* 293: 2269-71

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, et al. 2005. Combinatorial microRNA target predictions. *Nat Genet* 37: 495-500

Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, et al. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 438: 685-9

Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N, et al. 2004. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131: 839-49

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294: 853-8

Lai EC. 2002. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* 30: 363-4

Lai EC, Tam B, Rubin GM. 2005. Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev* 19: 1067-80

Landthaler M, Yalcin A, Tuschl T. 2004. The human DiGeorge syndrome critical region gene 8 and Its D. *melanogaster* homolog are required for miRNA biogenesis. *Curr Biol* 14: 2162-7

Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis* elegans. *Science* 294: 858-62

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, et al. 2006. Characterization of the piRNA complex from rat testes. *Science* 313: 363-7

Leaman D, Chen PY, Fak J, Yalcin A, Pearce M, et al. 2005. Antisense-mediated depletion reveals essential and specific functions of microRNAs in *Drosophila* development. *Cell* 121: 1097-108

Lecellier CH, Dunoyer P, Arar K, Lehmann-Che J, Eyquem S, et al. 2005. A cellular microRNA mediates antiviral defense in human cells. *Science* 308: 557-60

Lee RC, Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis* elegans. *Science* 294: 862-4

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: 843-54

Lee Y, Ahn C, Han J, Choi H, Kim J, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425: 415-9

Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN. 2006. The role of PACT in the RNA silencing pathway. *Embo J* 25: 522-32

Lee Y, Kim M, Han J, Yeom KH, Lee S, et al. 2004a. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* 23: 4051-60

Lee YS, Dutta A. 2007. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev* 21: 1025-30

Lee YS, Nakahara K, Pham JW, Kim K, He Z, et al. 2004b. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117: 69-81

Leuschner PJ, Martinez J. 2007. In vitro analysis of microRNA processing using recombinant Dicer and cytoplasmic extracts of HeLa cells. *Methods* 43: 105-9

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

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. *Cell* 115: 787-98

Li M, Jones-Rhoades MW, Lau NC, Bartel DP, Rougvie AE. 2005. Regulatory mutations of mir-48, a *C. elegans* let-7 family MicroRNA, cause developmental timing defects. *Dev Cell* 9: 415-22

Li QJ, Chau J, Ebert PJ, Sylvester G, Min H, et al. 2007. miR-181a is an intrinsic modulator of T cell sensitivity and selection. *Cell* 129: 147-61

Li X, Carthew RW. 2005. A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the *Drosophila* eye. *Cell* 123: 1267-77

Li Y, Wang F, Lee JA, Gao FB. 2006. MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes Dev* 20: 2793-805

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, et al. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433: 769-73

Lin SY, Johnson SM, Abraham M, Vella MC, Pasquinelli A, et al. 2003. The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Dev Cell* 4: 639-50

Lindsay EA. 2001. Chromosomal microdeletions: dissecting del22q11 syndrome. *Nat Rev Genet* 2: 858-68

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, et al. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305: 1437-41

Liu J, Rivas FV, Wohlschlegel J, Yates JR, 3rd, Parker R, Hannon GJ. 2005a. A role for the P-body component GW182 in microRNA function. *Nat Cell Biol* 7: 1261-6

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R. 2005b. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* 7: 719-23

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* 435: 834-8

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. 2004. Nuclear export of microRNA precursors. *Science* 303: 95-8

Ma L, Teruya-Feldstein J, Weinberg RA. 2007. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 449: 682-8

Makeyev EV, Zhang J, Carrasco MA, Maniatis T. 2007. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol Cell* 27: 435-48

Maniataki E, Mourelatos Z. 2005. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev* 19: 2979-90

Maroney PA, Yu Y, Fisher J, Nilsen TW. 2006. Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat Struct Mol Biol* 13: 1102-7

Martello G, Zacchigna L, Inui M, Montagner M, Adorno M, et al. 2007. MicroRNA control of Nodal signalling. *Nature* 449: 183-8

Martinez J, Tuschl T. 2004. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev*

Mathonnet G, Fabian MR, Svitkin YV, Parsyan A, Huck L, et al. 2007. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* 317: 1764-7

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. 2005. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* 123: 607-20

Mayr C, Hemann MT, Bartel DP. 2007. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* 315: 1576-9

References

Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* 15: 185-97

Meister G, Tuschl T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature* 431: 343-9

Meyer WJ, Schreiber S, Guo Y, Volkmann T, Welte MA, Muller HA. 2006. Overlapping functions of argonaute proteins in patterning and morphogenesis of *Drosophila* embryos. *PLoS Genet* 2: e134

Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, et al. 2006. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126: 1203-17

Mishima Y, Giraldez AJ, Takeda Y, Fujiwara T, Sakamoto H, et al. 2006. Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Curr Biol* 16: 2135-42

Miska EA, Alvarez-Saavedra E, Abbott AL, Lau NC, Hellman AB, et al. 2007. Most *Caenorhabditis* elegans microRNAs Are Individually Not Essential for Development or Viability. *PLoS Genet* 3: e215

Miska EA, Alvarez-Saavedra E, Townsend M, Yoshii A, Sestan N, et al. 2004. Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol* 5: R68

Molnar A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC. 2007. miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. *Nature* 447: 1126-9

Murchison EP, Partridge JF, Tam OH, Cheloufi S, Hannon GJ. 2005. Characterization of Dicer-deficient murine embryonic stem cells. *Proc Natl Acad Sci U S A* 102: 12135-40

Murchison EP, Stein P, Xuan Z, Pan H, Zhang MQ, et al. 2007. Critical roles for Dicer in the female germline. *Genes Dev* 21: 682-93

Naguibneva I, Ameyar-Zazoua M, Polesskaya A, Ait-Si-Ali S, Groisman R, et al. 2006. The microRNA miR-181 targets the homeobox protein Hox-A11 during mammalian myoblast differentiation. *Nat Cell Biol* 8: 278-84

Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. 2004. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat Methods* 1: 155-61

Nolo R, Morrison CM, Tao C, Zhang X, Halder G. 2006. The bantam microRNA is a target of the hippo tumor-suppressor pathway. *Curr Biol* 16: 1895-904

Nottrott S, Simard MJ, Richter JD. 2006. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol* 13: 1108-14

O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435: 839-43

Obernosterer G, Leuschner PJ, Alenius M, Martinez J. 2006. Post-transcriptional regulation of microRNA expression. *Rna* 12: 1161-7

Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. 2007. The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 130: 89-100

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* 18: 1655-66

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, et al. 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408: 86-9

Pearson JC, Lemons D, McGinnis W. 2005. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* 6: 893-904

Pedersen IM, Cheng G, Wieland S, Volinia S, Croce CM, et al. 2007. Interferon modulation of cellular microRNAs as an antiviral mechanism. *Nature* 449: 919-22

Petersen CP, Bordeleau ME, Pelletier J, Sharp PA. 2006. Short RNAs repress translation after initiation in mammalian cells. *Mol Cell* 21: 533-42

Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, et al. 2005. Identification of microRNAs of the herpesvirus family. *Nat Methods* 2: 269-76

Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, et al. 2004. Identification of virus-encoded microRNAs. *Science* 304: 734-6

Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, et al. 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* 309: 1573-6

Pilot F, Philippe JM, Lemmers C, Chauvin JP, Lecuit T. 2006. Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of *Drosophila* cellularisation. *Development* 133: 711-23

Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, et al. 2004. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432: 226-30

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

Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *Rna* 11: 1640-7

Rehwinkel J, Natalin P, Stark A, Brennecke J, Cohen SM, Izaurralde E. 2006. Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in *Drosophila melanogaster*. *Mol Cell Biol* 26: 2965-75

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis* elegans. *Nature* 403: 901-6

Rodriguez A, Vigorito E, Clare S, Warren MV, Couttet P, et al. 2007. Requirement of bic/microRNA-155 for normal immune function. *Science* 316: 608-11

Ronshaugen M, Biemar F, Piel J, Levine M, Lai EC. 2005. The *Drosophila* microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes Dev* 19: 2947-52

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, et al. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127: 1193-207

Ruby JG, Jan CH, Bartel DP. 2007a. Intronic microRNA precursors that bypass Drosha processing. *Nature* 448: 83-6

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. 2007b. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* 17: 1850-64

Saito K, Ishizuka A, Siomi H, Siomi MC. 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol* 3: e235

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, et al. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 20: 2214-22

Sandmann T, Cohen SM. 2007. Identification of Novel *Drosophila melanogaster* MicroRNAs. *PLoS ONE* 2: e1265

Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, et al. 2006. A brain-specific microRNA regulates dendritic spine development. *Nature* 439: 283-9

Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115: 199-208

Sen GL, Blau HM. 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat Cell Biol* 7: 633-6

Shiohama A, Sasaki T, Noda S, Minoshima S, Shimizu N. 2003. Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. *Biochem Biophys Res Commun* 304: 184-90

Slack FJ, Basson M, Liu Z, Ambros V, Horvitz HR, Ruvkun G. 2000. The lin-41 RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell* 5: 659-69

References

Sokol NS, Ambros V. 2005. Mesodermally expressed *Drosophila* microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes Dev* 19: 2343-54

Sood P, Krek A, Zavolan M, Macino G, Rajewsky N. 2006. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc Natl Acad Sci U S A* 103: 2746-51

Standart N, Jackson RJ. 2007. MicroRNAs repress translation of m7Gppp-capped target mRNAs in vitro by inhibiting initiation and promoting deadenylation. *Genes Dev* 21: 1975-82

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell* 123: 1133-46

Stark A, Brennecke J, Russell RB, Cohen SM. 2003. Identification of *Drosophila* MicroRNA Targets. *PLoS Biol* 1: E60

Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, et al. 2007. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res* 17: 1865-79

Stern-Ginossar N, Elefant N, Zimmermann A, Wolf DG, Saleh N, et al. 2007. Host immune system gene targeting by a viral miRNA. *Science* 317: 376-81

Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, et al. 2004. Human embryonic stem cells express a unique set of microRNAs. *Dev Biol* 270: 488-98

Sullivan CS, Grundhoff AT, Tevethia S, Pipas JM, Ganem D. 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* 435: 682-6

Tadros W, Goldman AL, Babak T, Menzies F, Vardy L, et al. 2007. SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Dev Cell* 12: 143-55

Tang F, Kaneda M, O'Carroll D, Hajkova P, Barton SC, et al. 2007. Maternal microRNAs are essential for mouse zygotic development. *Genes Dev* 21: 644-8

Teleman AA, Maitra S, Cohen SM. 2006. *Drosophila* lacking microRNA miR-278 are defective in energy homeostasis. *Genes Dev* 20: 417-22

Thai TH, Calado DP, Casola S, Ansel KM, Xiao C, et al. 2007. Regulation of the germinal center response by microRNA-155. *Science* 316: 604-8

Thermann R, Hentze MW. 2007. *Drosophila* miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature* 447: 875-8

Thompson BJ, Cohen SM. 2006. The Hippo pathway regulates the bantam microRNA to control cell proliferation and apoptosis in *Drosophila*. *Cell* 126: 767-74

Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* 20: 2202-7

Thomson JM, Parker J, Perou CM, Hammond SM. 2004. A custom microarray platform for analysis of microRNA gene expression. *Nat Methods* 1: 47-53

Tomancak P, Berman BP, Beaton A, Weiszmann R, Kwan E, et al. 2007. Global analysis of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol* 8: R145

Tomari Y, Du T, Haley B, Schwarz DS, Bennett R, et al. 2004. RISC assembly defects in the *Drosophila* RNAi mutant armitage. *Cell* 116: 831-41

Tomari Y, Du T, Zamore PD. 2007. Sorting of *Drosophila* small silencing RNAs. *Cell* 130: 299-308

Triboulet R, Mari B, Lin YL, Chable-Bessia C, Bennasser Y, et al. 2007. Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science* 315: 1579-82

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313: 320-4

van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN. 2007. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* 316: 575-9

Varghese J, Cohen SM. 2007. microRNA miR-14 acts to modulate a positive autoregulatory loop controlling steroid hormone signaling in *Drosophila*. *Genes Dev* 21: 2277-82

Vasudevan S, Steitz JA. 2007. AU-Rich-Element-Mediated Upregulation of Translation by FXR1 and Argonaute 2. *Cell* 128: 1105-18
Vasudevan S, Tong Y, Steitz JA. 2007. Switching from repression to activation: microRNAs can upregulate translation. *Science* 318: 1931-4
Vaucheret H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* 20: 759-71
Visvanathan J, Lee S, Lee B, Lee JW, Lee SK. 2007. The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. *Genes Dev* 21: 744-9
Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* 103: 2257-61
Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, et al. 2006. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* 124: 1169-81
Wakiyama M, Takimoto K, Ohara O, Yokoyama S. 2007. Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev* 21: 1857-62
Wang B, Love TM, Call ME, Doench JG, Novina CD. 2006. Recapitulation of short RNA-directed translational gene silencing in vitro. *Mol Cell* 22: 553-60
Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, et al. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* 20: 1732-43
Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, et al. 2005. MicroRNA expression in zebrafish embryonic development. *Science* 309: 310-1
Wienholds E, Koudijs MJ, van Eeden FJ, Cuppen E, Plasterk RH. 2003. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat Genet* 35: 217-8
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: 855-62
Wu L, Fan J, Belasco JG. 2006. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A* 103: 4034-9
Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434: 338-45
Xu P, Vernooy SY, Guo M, Hay BA. 2003. The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr Biol* 13: 790-5
Yekta S, Shih IH, Bartel DP. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304: 594-6
Yi R, O'Carroll D, Pasolli HA, Zhang Z, Dietrich FS, et al. 2006. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat Genet* 38: 356-62
Yi R, Qin Y, Macara IG, Cullen BR. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* 17: 3011-6
Yoo AS, Greenwald I. 2005. LIN-12/Notch activation leads to microRNA-mediated down-regulation of Vav in *C. elegans*. *Science* 310: 1330-3
Zhao T, Li G, Mi S, Li S, Hannon GJ, et al. 2007a. A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. *Genes Dev* 21: 1190-203
Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, et al. 2007b. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129: 303-17
Zhao Y, Samal E, Srivastava D. 2005. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436: 214-20

Curriculum Vitae

Name	Natascha Bushati
Date of Birth	July 24th, 1978
Birthplace	Langenhagen, Germany
Nationality	Austrian

Academic Education

since 06/2004	Pre-doctoral fellow at the European Molecular Biology Laboratory (EMBL), Heidelberg, Germany Supervisor: Dr. Stephen Cohen
02/2004	University degree “Magistra der Naturwissenschaften”
10/2002 - 10/2003	Diploma thesis “Regulation of Cytokinesis in *Caenorhabditis elegans*” in the laboratory of Dr. Michael Glotzer Research Institute of Molecular Pathology (IMP) Vienna, Austria
10/2001 – 02/2002	Erasmus semester in Barcelona, Spain Practical work in the laboratory of Dr. Pere Suau Universitat Autònoma de Barcelona, Spain
1999 – 02/2004	Advanced studies in chemistry, main subject biochemistry University of Vienna, Austria,
10/1997 – 1999	Basic studies in technical chemistry Vienna University of Technology, Austria
06/1997	High school graduation (Matura) with honors Bundesgymnasium Wien IX, Vienna, Austria

January 2008

Lebenslauf

Name	Natascha Bushati
Geburtsdatum	24. Juli 1978
Geburtsort	Langenhagen, Deutschland
Nationalität	Österreich

Ausbildung

Juni 2004 - Heute	Promotionsstipendium des Europäischen Laboratoriums für Molekularbiologie (EMBL) in Heidelberg, Deutschland Betreuer: Dr. Stephen Cohen
Feb. 2004	Abschluss des Chemiestudiums “Magistra der Naturwissenschaften” (mag.rer.nat.)
Okt. 2002 – Okt. 2003	Diplomarbeit “Regulierung der Zytokinese in *Caenorhabditis elegans*” in der Arbeitsgruppe von Dr. Michael Glotzer Research Institute of Molecular Pathology (IMP) Wien, Österreich
Okt. 2001 – Okt. 2002	Erasmus Semester in Barcelona, Spanien Praktikum in der Arbeitsgruppe von Dr. Pere Suau Universitat Autònoma de Barcelona
1999 – Feb. 2004	Zweiter Studienabschnitt des Studiums der Chemie, Schwerpunkt Biochemie Universität Wien, Österreich
Okt.1997 – 1999	Erster Studienabschnitt des Studiums der Technischen Chemie Technische Universität Wien, Österreich
Juni 1997	Matura mit Auszeichnung Bundesgymnasium Wien IX, Wien, Österreich

Januar 2008

Appendix

Cell

Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3′UTR Evolution

Alexander Stark,[1,2,3] Julius Brennecke,[1,2] Natascha Bushati,[1] Robert B. Russell,[1] and Stephen M. Cohen[1,*]
[1] European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany
[2] These authors contributed equally to this work.
[3] Present address: Broad Institute of MIT and Harvard, Cambridge, MA 02141, USA and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Contact: cohen@embl.de
DOI 10.1016/j.cell.2005.11.023

SUMMARY

MicroRNAs are small noncoding RNAs that serve as posttranscriptional regulators of gene expression in higher eukaryotes. Their widespread and important role in animals is highlighted by recent estimates that 20%–30% of all genes are microRNA targets. Here, we report that a large set of genes involved in basic cellular processes avoid microRNA regulation due to short 3′UTRs that are specifically depleted of microRNA binding sites. For individual microRNAs, we find that coexpressed genes avoid microRNA sites, whereas target genes and microRNAs are preferentially expressed in neighboring tissues. This mutually exclusive expression argues that microRNAs confer accuracy to developmental gene-expression programs, thus ensuring tissue identity and supporting cell-lineage decisions.

INTRODUCTION

Regulation of gene expression at the transcriptional level plays a central role in defining cell fates and controlling organ formation. But the importance of posttranscriptional gene regulation is increasingly recognized. microRNAs (miRNAs) confer a novel layer of posttranscriptional regulation, widely used in plants and animals. miRNAs are small noncoding RNAs that repress gene expression by recruiting effector complexes (miRNPs) to miRNA complementary sites on mRNAs (Bartel, 2004; Zamore and Haley, 2005). miRNP recruitment in plants requires extensive sequence complementarity and typically leads to target mRNA cleavage (e.g., Schwab et al., 2005). Animal miRNAs are only partially complementary to their targets and repress their expression, likely by blocking translation initiation and by recruiting miRNPs to processing bodies where degradation might occur (Bagga et al., 2005; Lim et al., 2005; Liu et al., 2005; Pillai et al., 2005; Rehwinkel et al., 2005; Sen and Blau, 2005).

miRNAs are estimated to comprise 1%–5% of animal genes (Bartel, 2004; Bentwich et al., 2005; Berezikov et al., 2005), making them one of the most abundant classes of regulators. Their importance is evidenced by evolutionary conservation and by the many biological processes in which they are implicated, including developmental timing, cell proliferation, apoptosis, metabolism, cell differentiation, and morphogenesis (Alvarez-Garcia and Miska, 2005; Ambros, 2004). Current ideas about animal miRNA functions have been influenced by the handful of genetically identified miRNAs and their targets. These miRNAs have been described as developmental switches, repressing a few target genes. Indeed, some miRNA mutant phenotypes can largely be explained by increased expression of a single target (Ambros, 2004).

New insights are challenging the view of animal miRNAs as switches for a few targets and suggest a more complex picture. (1) Recent estimates indicate that an average miRNA may regulate hundreds of genes (Brennecke et al., 2005; Grün et al., 2005; Krek et al., 2005; Lewis et al., 2005; Xie et al., 2005). (2) Most targets contain only single sites that might not be sufficient to confer strong repression, making a switch-like relationship unlikely. (3) Despite striking tissue-specific expression patterns of miRNAs in zebrafish (Wienholds et al., 2005), a general role as developmental switches in patterning or organogenesis has been excluded by analysis of embryos lacking all miRNAs (Giraldez et al., 2005).

This suggests that miRNAs might not primarily be involved in developmental decision-making. This view has gained initial support from the finding that overexpressing *miR-1* and *miR-124* in HeLa cells downregulated many mRNAs, which are of low abundance in the tissues expressing these miRNAs (Lim et al., 2005). miRNAs might thus help to maintain and define cell types by suppressing expression of unwanted transcripts. However, it remained unclear if the downregulated transcripts are representative of physiological

targets and whether the insights gained can be generalized to a new role for miRNAs in animal development.

Here, we address the role of miRNAs in developmental gene expression programs, particularly their relationship to the large number of conserved targets. We combined improved miRNA target prediction with information on gene function and expression in *Drosophila*. We present evidence that many genes are under selective pressure to avoid miRNA regulation. The existence of such ''antitargets'' had been proposed on theoretical grounds (Bartel and Chen, 2004). We find that antitargets are involved in basic processes common to all cells, whereas targets are mainly involved in developmental processes. For individual miRNAs, coexpressed genes avoid regulation, whereas predicted targets and miRNAs are preferentially expressed in adjacent domains. When considered in the temporal and spatial context of development, this relationship of miRNA, target, and antitarget expression suggests that miRNAs confer accuracy to gene-expression programs. Our findings indicate that miRNAs have had a profound impact on 3′UTR evolution, reflected in the observed patterns of site avoidance and enrichment.

RESULTS

Target-Site Prediction with High Specificity and Coverage

We predict miRNA targets based on a systematic experimental analysis of the structural requirements for target site function in vivo (Brennecke et al., 2005). Briefly, we identified sites via complementarity to miRNA 5′ ends and evaluated 5′ and 3′ pairing. We restricted the search to sites conserved in an alignment of the orthologous *D. melanogaster* and *D. pseudoobscura* 3′UTRs as these are more likely biologically relevant. This yielded 179 conserved target sites per average miRNA. We estimated the overall significance of these predictions by analyzing whether target sites for real miRNAs are better conserved than those for shuffled control miRNAs (Lewis et al., 2003). While 34% of sites for real miRNAs identified in *D. melanogaster* were conserved, only 14% were conserved for shuffled miRNAs, yielding a highly significant p value and a signal-to-noise ratio of 2.4:1. This signal was abolished when two nucleotides at the 5′ end of the miRNAs were changed, indicating the validity of the approach and control (Figure 1A). An average *Drosophila* miRNA thus targets over 100 sites above noise. Note that this does not imply that the other sites are false, only that they cannot be distinguished from noise. The false-positive rate can only be assessed experimentally. We consider it likely that most identified sites are functional because all comply with our rules and are conserved.

We tested 9 of the top 25 predictions using a luciferase-reporter assay in S2 cells and found eight to be significantly regulated ($p < 0.01$; Figures 1B and 1C). Including the previously validated *bantam* target *hid* (Brennecke et al., 2003), this suggests a 90% success rate for top predictions. Interestingly, over half encode transcription factors, whose misregulation could have severe consequences. To assess the performance at different ranks and the improvement over our previous study (Stark et al., 2003), we evaluated the predictions with a large number of experimentally tested miRNA-target pairs (Figure 1D). 88% (50/57) of the new predictions were functional, which is a substantial improvement in specificity as the number of false positives was reduced by 65% (7 versus 20). This did not come at the cost of lower sensitivity because we now predict 50% more functional pairs (50/62 versus 34/62; see Figure S1 and Table S1 in the Supplemental Data available with this article online for comparison to other *Drosophila* target predictions). In summary, the experimental and statistical results show that our method and the one by Grün et al. (2005) predict functional, biologically relevant sites with high accuracy, which is critical for the analysis below. Our predictions are available at www.miRNA.embl.de.

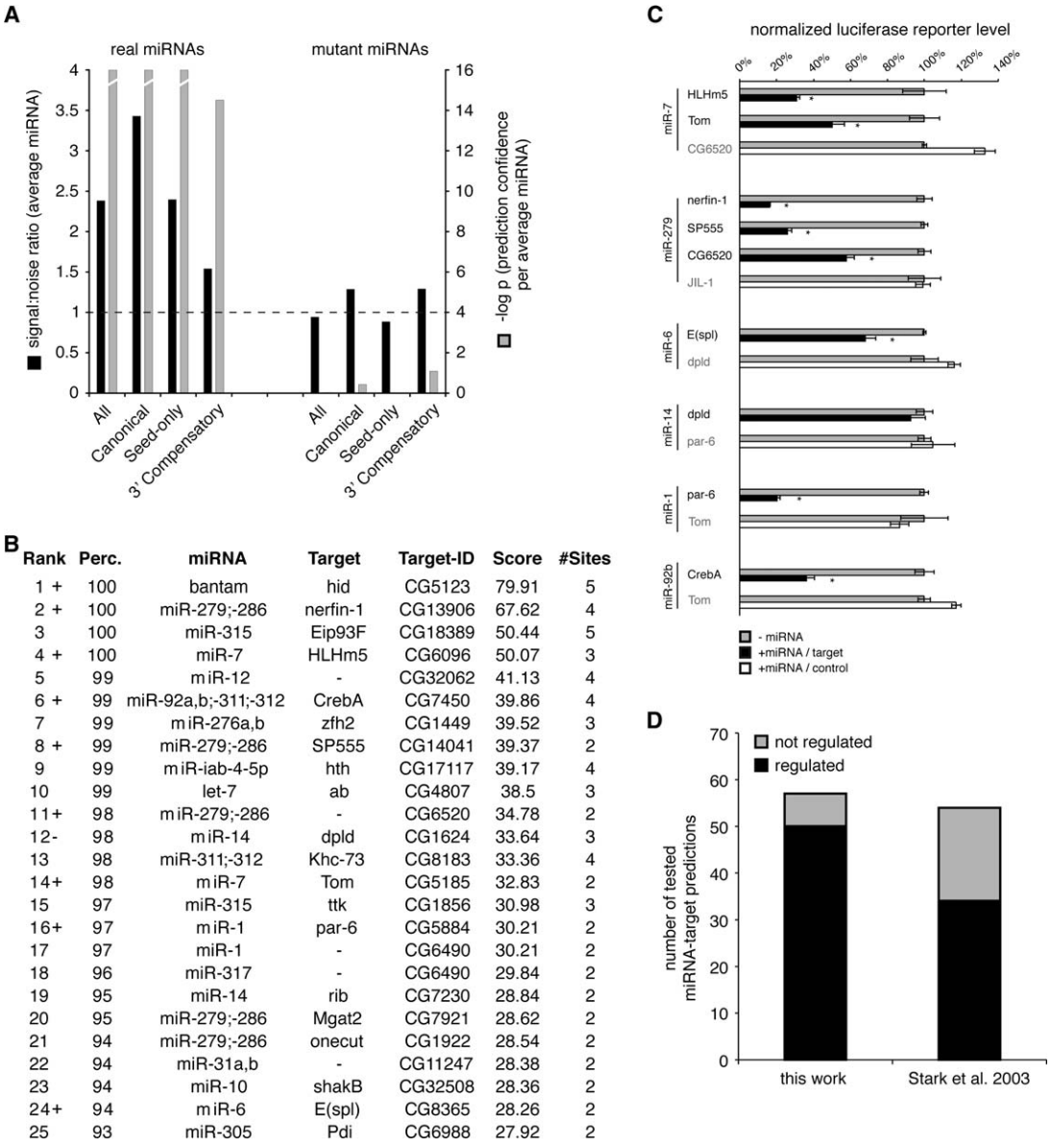
Extensive Cooccurrence of Sites for Different miRNAs

Only 5% of all predicted targets contain more than one conserved site for any single miRNA, indicating that stringent regulation by one miRNA is rare. In contrast, we observed extensive cooccurrence of sites for different miRNAs in target 3′UTRs (Figure S2; Enright et al., 2003; Grün et al., 2005; Krek et al., 2005). The 9487 binary interactions correspond to 3125 different 3′UTRs. Almost 50% of target 3′UTRs have sites for two or more 5′ unique miRNAs and some have sites for up to 12. In contrast, 5129 3′UTRs had no conserved site, indicating that target sites are distributed highly asymmetrically across different genes. A detailed analysis revealed that genes with more miRNA sites have on average longer 3′UTRs but also significantly more sites/kb of 3′UTR sequence (Figure S2). Reciprocally, genes with few sites have short 3′UTRs and lower site densities. These two trends are not seen together in random controls and indicate that 3′UTRs have been under selection to acquire or eliminate miRNA target sites.

A striking example of site cooccurrence is the transcript for the nervous system-specific transcription factor Nerfin-1 (Stivers et al., 2000), whose 3′UTR contains 15 target sites for 10 different miRNAs (Figure S2D). Consistent with the presence of multiple miRNA sites, a ubiquitously transcribed *nerfin-1* 3′UTR reporter is repressed by miRNAs, as lack of *Dicer-1* in cells of various tissues showed strong reporter upregulation (Figure S2E). Intriguingly, expression of this reporter is normally detectable only in the nervous system (data not shown). Thus, by virtue of the miRNA target sites it contains, the *nerfin-1* 3′UTR appears to carry information about the tissue in which *nerfin-1* is expressed and required.

Presence and Absence of Target Sites Correlate with Gene Function

To ask if the presence or absence of miRNA target sites correlates with gene function we compared the 3125 predicted targets with all genes lacking conserved sites. We determined if these sets contain more or fewer genes from any given gene ontology (GO) or KEGG category than expected, given the category's frequency in the 3′UTR database (''gene enrichment,'' see Supplemental Data). Table 1 shows categories that are most significant in terms of containing more, or fewer, target genes than expected (see Figure S3 for top



B

Rank	Perc.	miRNA	Target	Target-ID	Score	#Sites
1 +	100	bantam	hid	CG5123	79.91	5
2 +	100	miR-279;-286	nerfin-1	CG13906	67.62	4
3	100	miR-315	Eip93F	CG18389	50.44	5
4 +	100	miR-7	HLHm5	CG6096	50.07	3
5	99	miR-12	-	CG32062	41.13	4
6 +	99	miR-92a,b;-311;-312	CrebA	CG7450	39.86	4
7	99	miR-276a,b	zfh2	CG1449	39.52	3
8 +	99	miR-279;-286	SP555	CG14041	39.37	2
9	99	miR-iab-4-5p	hth	CG17117	39.17	4
10	99	let-7	ab	CG4807	38.5	3
11+	98	miR-279;-286	-	CG6520	34.78	2
12-	98	miR-14	dpld	CG1624	33.64	3
13	98	miR-311;-312	Khc-73	CG8183	33.36	4
14+	98	miR-7	Tom	CG5185	32.83	2
15	97	miR-315	ttk	CG1856	30.98	3
16+	97	miR-1	par-6	CG5884	30.21	2
17	97	miR-1	-	CG6490	30.21	2
18	96	miR-317	-	CG6490	29.84	2
19	95	miR-14	rib	CG7230	28.84	2
20	95	miR-279;-286	Mgat2	CG7921	28.62	2
21	94	miR-279;-286	onecut	CG1922	28.54	2
22	94	miR-31a,b	-	CG11247	28.38	2
23	94	miR-10	shakB	CG32508	28.36	2
24+	94	miR-6	E(spl)	CG8365	28.26	2
25	93	miR-305	Pdi	CG6988	27.92	2

Figure 1. Summary of Target Validation

(A) Signal to noise ratio (black bars) and statistical significance (−log p value; gray bars) for miRNA target predictions compared to shuffled controls. Signals obtained for 5′ nonredundant miRNAs (real) are lost for variants with 2 nt changes in the 5′ end (mutant). Both measures are displayed for all predicted sites; sites with good seed matches and strong (canonical) or weak (seed-only) 3′ pairing energy (>90% of sites are seed only); 3′ compensatory sites with weak seed matches and strong 3′ pairing.

(B) Top 25 predictions ranked by UTR score (± indicates functionality, see Supplemental Data). Percentile (Perc.) indicates the rank relative to the best UTR scores for shuffled miRNAs.

(C) Experimental validation of top predictions using luciferase 3′UTR reporters. Normalized luciferase activity for nine predicted target 3′UTRs ± coexpressed miRNA. Error bars: standard deviation; asterisks: $p < 0.01$, $n = 3$; double-sided t test. Control 3′UTR reporters lacking predicted sites for the miRNA were not repressed (white bars).

(D) Performance of our predictions based on all *Drosophila* miRNA target interactions tested (see Supplemental Data for details). Functional pairs are in black, false positives in gray.

50). We refer to these as ''target'' and ''antitarget'' categories (after Bartel and Chen [2004]). Out of several thousand GO and KEGG categories, the top target categories were dominated by developmental processes (consistent with previous studies, e.g., Enright et al., 2003; Grün et al., 2005), whereas the top antitarget categories were exclusively basic processes common to all cells.

To assess the basis for this highly asymmetric distribution of target sites, we analyzed 3′UTR characteristics that might influence target-site occurrence (considering all genes in

Table 1. miRNA Target and Antitarget Categories

Category	Description	# Genes	p(over) in Targets	p(under) in Antitargets
GO:0009887	Organogenesis	646	2.1E-34	7.3E-26
GO:0007399	Neurogenesis	364	2.2E-23	5.4E-19
GO:0007165	Signal transduction	791	2.7E-19	2.5E-14
GO:0030154	Cell differentiation	213	2.0E-11	5.4E-09
GO:0009790	Embryonic development	228	5.4E-11	1.4E-08
GO:0045165	Cell fate commitment	146	1.2E-10	3.8E-09
GO:0045449	Regulation of transcription	448	1.4E-09	2.8E-06
GO:0002009	Morphogenesis of an epithelium	104	1.0E-08	3.0E-08
GO:0007422	Peripheral nervous system development	95	4.5E-08	3.9E-07
GO:0009795	Embryonic morphogenesis	101	1.1E-07	5.2E-07
GO:0007498	Mesoderm development	135	3.5E-07	2.0E-04
Category	**Description**	**# Genes**	**p(over) in Antitargets**	**p(under) in Targets**
GO:0030529	Ribonucleoprotein complex	200	3.7E-06	1.3E-11
GO:0005840	Ribosome	128	2.4E-05	1.1E-11
GO:0006412	Protein biosynthesis	289	4.1E-03	3.8E-04
GO:0016070	RNA metabolism	190	7.4E-03	7.7E-04
GO:0016591	DNA-directed RNA polymerase II, holoenzyme	62	7.7E-03	5.6E-05
GO:0006119	Oxidative phosphorylation	61	1.8E-02	2.3E-04
GO:0006281	DNA repair	70	2.2E-02	4.7E-04
GO:0000502	Proteasome complex (sensu Eukarya)	37	2.6E-02	4.1E-04
GO:0006259	DNA metabolism	203	2.8E-02	3.9E-03
GO:0008380	RNA splicing	78	3.9E-02	1.4E-02

Best target and antitarget GO categories, whose genes are significantly over- or underrepresented among all 3125 predicted targets (obviously redundant categories were removed). Similar results are obtained when asking for under- or overrepresentation among the 5129 antitargets lacking predicted target sites. p values indicate the probability that the over- or underrepresentation occurred randomly. (Top) GO categories overrepresented among miRNA targets (target categories). (Bottom) GO categories underrepresented among miRNA targets (antitarget categories). Multiple testing correction factors were determined by using shuffled gene–GO term assignments. For the different statistics (left to right, top to bottom), the factors are 719957, 3879, 8.6, and 30.1, respectively.

each category, not only predicted targets and antitargets). Figure 2 shows this analysis for two representative categories: the target category *neurogenesis* and the antitarget category *ribosome* (comparable results were obtained for most other target and antitarget categories, Figure S3). Given that sites with as little as 7–8 nucleotides complementarity are functional (Brennecke et al., 2005), longer 3′UTRs likely contain more sites. Indeed, average 3′UTR length differs considerably: genes-encoding ribosomal proteins have ~6-fold shorter 3′UTRs than *neurogenesis* genes (Figure 2A). Selection against long 3′UTRs could be an effective means to limit miRNA regulation. As 3′UTR lengths might differ for reasons unrelated to miRNAs, we tested whether site densities differed in both categories. We observed a marked difference in that *ribosomal* genes have 4.3× fewer sites per kb of 3′UTR than *neurogenesis* genes (Figure 2B), whereas both categories showed comparable site densities for shuffled miRNAs. This difference could reflect *ribosomal* genes having fewer sites than expected or *neurogenesis* genes having more given their 3′UTR lengths, so we tested whether *ribosomal* genes specifically avoid miRNA target sites compared to random sequences (''site enrichment,'' see Supplemental Data). We found significantly fewer miRNA complementary sites than sites for shuffled miRNAs; Figure 2C; $p = 7 \times 10^{-7}$), indicating specific avoidance of miRNA regulation. In contrast, 3′UTRs of *neurogenesis* genes are specifically enriched for sites ($p = 2 \times 10^{-5}$).

The degree of 3′UTR conservation also influences the gene-enrichment analysis, as genes with more conserved 3′UTR sequence are more likely to be predicted as targets. Although the overall degree of 3′UTR conservation is comparable (Figure 2D), we observed a striking difference in the conservation of miRNA complementary sites only (''selective conservation,'' see Supplemental Data). Sites in 3′UTRs of *neurogenesis* genes are much better conserved than expected given the overall 3′UTR conservation ($p < 10^{-150}$; Figure 2E), whereas those in *ribosomal* genes are not ($p = 0.6$).

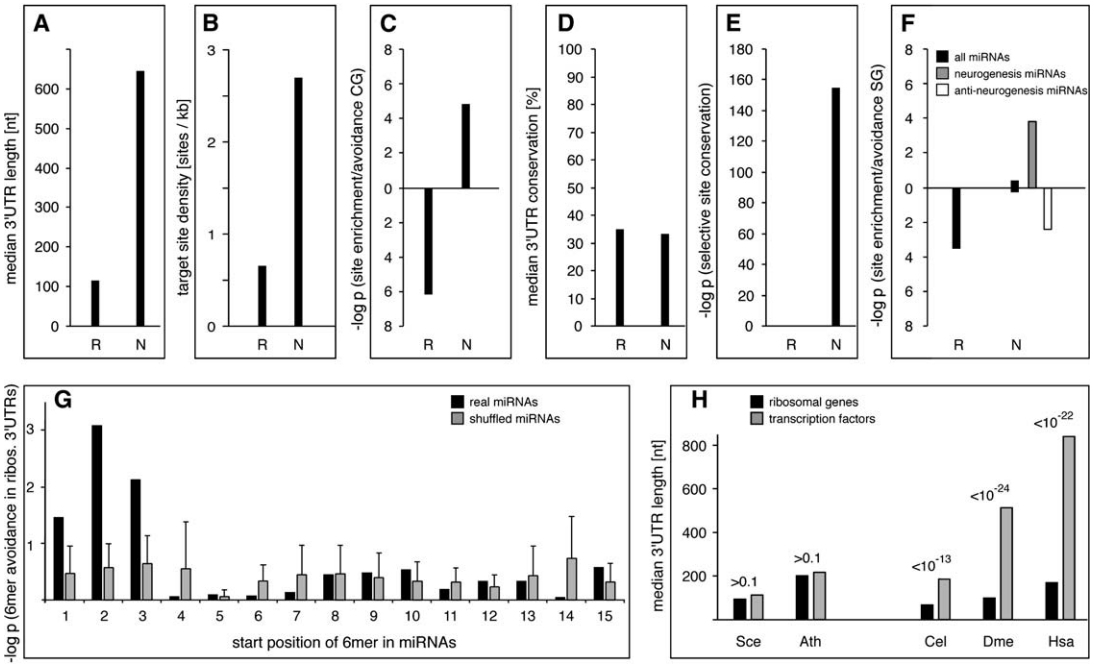


Figure 2. Properties of Target and Antitarget 3′UTRs

Values are based on all genes in the GO categories *ribosome* (GO:0005840; R) and *neurogenesis* (GO:0007399; N).

(A) median 3′UTR length; (B) target-site density (predicted sites/kb 3′UTR sequence); (C) specific enrichment (up) or avoidance (down) of predicted sites (CG: conserved genome analysis; −log p values; statistic corrects for 3′UTR length); (D) overall 3′UTR conservation counting nucleotides in conserved blocks of ≥6; (E) selective site conservation (−log p values); (F) specific enrichment (up) or avoidance (down) of sites in *D. melanogaster* 3′UTRs (SG: single genome; −log p values). Values for neurogenesis are split: all miRNAs (black), top 10 neurogenesis-miRNAs (gray), top 10 antineurogenesis miRNAs (white).

(G) Underrepresentation (−log p values) of miRNA complementary 6mers in *ribosomal* 3′UTRs versus position in the miRNAs: all miRNAs (black), average and standard deviation of 10 shuffled miRNAs per real miRNA (gray).

(H) median 3′UTR lengths of ribosomal (black) and transcription factor genes (gray) for *S. cerevisiae* (Sce), *A. thaliana* (Ath), *C. elegans* (Cel), *D. melanogaster* (Dme), and *H. sapiens* (Hsa). p values indicate the significance of length differences for each species (double-sided t test).

Selection for and against miRNA Target Sites

A key finding is that genes in antitarget categories specifically avoid miRNA sites. If miRNAs had no influence on antitarget 3′UTRs, we would expect random site occurrence rather than avoidance. Target site avoidance indicates that miRNA-mediated regulation of genes in antitarget categories would be detrimental and that it has been subject to selection during evolution. On this basis, we expect avoidance of miRNA sites in antitargets, whether conserved or not. This was confirmed by examining 3′UTRs without requiring site conservation. We found significantly fewer sites in 3′UTRs of *ribosomal* genes than expected given their lengths ($p = 3 \times 10^{-4}$; Figure 2F). To test for avoidance independent of our predictions, we performed a 6mer "seed walk," where we assessed the avoidance of 6mers along the sequence of all miRNAs. This further illustrated the specificity of target site avoidance: only 6mers complementary to the 5′ region of real miRNAs—the crucial element for target recognition—are avoided in *ribosomal* 3′UTRs (Figure 2G, "seed avoidance"; see Supplemental Data).

Although *neurogenesis* genes are enriched in conserved target sites, the single-genome analysis failed to detect a similar trend. This could reflect opposing influences of different miRNAs: while it is easy to imagine that *ribosomal* genes avoid sites for all miRNAs, *neurogenesis* genes might enrich for sites for some miRNAs but avoid sites for others. We asked which individual miRNAs predominantly target *neurogenesis* genes and which do not (gene-enrichment statistics) and then repeated the single-genome site-enrichment analysis for these sets of miRNAs separately. This confirmed that *neurogenesis* genes enrich sites for neurogenesis miRNAs (e.g., miR-9; Figure 2F) but reciprocally avoid sites for anti-neurogenesis miRNAs (e.g., miR-124).

Our results indicate that antitargets circumvent miRNA-mediated regulation by limiting 3′UTR length and by selective avoidance of target sites. In contrast, target genes have longer 3′UTRs that are enriched in evolutionarily conserved sites. The single-genome analysis reveals a more complex picture where 3′UTRs of target categories are enriched in sites for some miRNAs but depleted for others, consistent with individual miRNAs regulating specific sets of functionally related genes. Note that the preceding analyses are based on all genes in each category, not only on the predicted targets and antitargets, for which the trends would be even stronger.

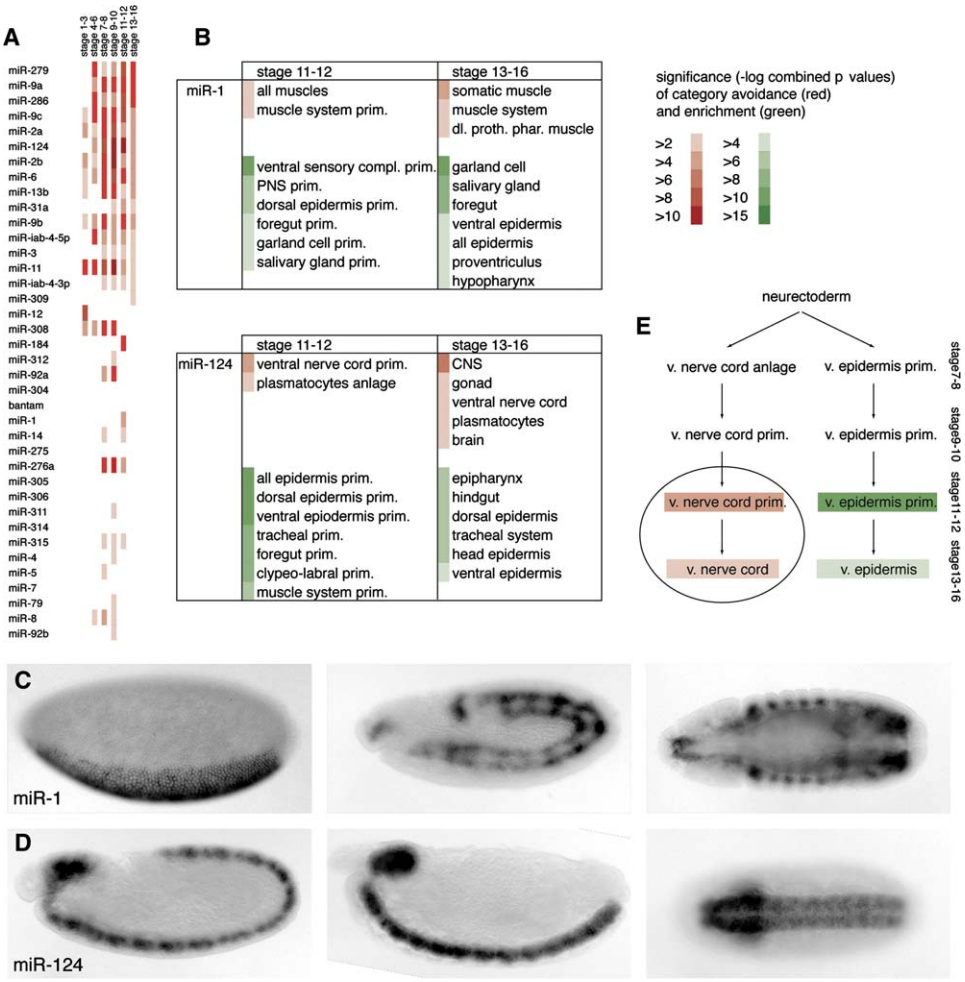


Figure 3. *miR-1* and *miR-124* Target/Antitarget Expression

(A) Ubiquitously expressed genes significantly avoid sites for many embryonic miRNAs; no miRNA shows a preference for these genes at any stage of embryonic development.

(B) Avoidance and enrichment patterns for *miR-1* (top) and *miR-124* (bottom) for embryonic stages 11/12 and 13–16. The most significant tissues are color-coded according to p values that combine gene- and site-enrichment statistics. Abbreviations: prim., primordium; dl. proth. phar., dorsal prothoracic pharyngeal; compl., complexes.

(C and D) miRNA primary transcript in situ hybridization. Lateral views, anterior left, unless otherwise indicated.

(C) *miR-1* expressed in the presumptive mesoderm at blastoderm stage (left); in visceral and somatic mesoderm at stage 10/11 (middle); in somatic, visceral, and pharyngeal muscles at stage 17 (right, dorsal view).

(D) *miR-124* expression is detected only in brain and ventral nerve cord. From left: stages 12, 13, 16 (ventral view).

(E) Common developmental origin of nervous system and epidermis and their *miR-124* avoidance/enrichment patterns. Circle: *miR-124* expression.

Mutually Exclusive Expression of miRNAs and Their Targets

The simplest explanation for significant site avoidance is that antitargets are required in the miRNA-expressing cells and miRNA-mediated repression would be detrimental. Consistently, genes involved in basic cellular processes required in all cells avoid sites for all miRNAs. As many miRNAs show pronounced spatial and temporal expression patterns, we investigated whether a similar avoidance pattern could be found among genes that are developmentally coexpressed with specific miRNAs. We used an extensive collection of annotated in situ gene expression patterns for *Drosophila* embryogenesis (Tomancak et al., 2002). We tested whether sets of genes expressed in specific tissues or organs avoid regulation by individual miRNAs, by combining the gene- and site-enrichment statistics introduced above.

As expected, 3′UTRs of genes classified as ubiquitously expressed significantly avoid sites for many embryonically expressed miRNAs (Figure 3A). This gene set overlaps considerably with the antitarget categories above. In contrast, predicted targets for most miRNAs are preferentially expressed at later stages when organogenesis takes place. Hence, we investigated the tissue distribution of predicted targets at embryonic stages 11/12 and 13–16 and asked if target site avoidance correlates with miRNA expression.

The only tissue-specific gene sets that significantly avoid *miR-1* regulation are those for muscle (Figure 3B). Strikingly, *miR-1* is expressed exclusively in the presumptive mesoderm in the early embryo and subsequently in developing muscle (Figure 3C; Sokol and Ambros, 2005). Similarly, *miR-124* is expressed exclusively in the central nervous system (CNS) and CNS genes most significantly avoid *miR-124* sites (Figures 3B and 3D). Both findings are consistent with the analysis of Lim et al. (2005), who showed that overexpression of human *miR-1* or *miR-124* in HeLa cells led to preferential downregulation of nonmuscle or nonbrain transcripts, respectively. This indicates that these two miRNAs are conserved not only in sequence and spatial expression from flies to vertebrates but also in their tendency to avoid coexpressed genes (Lim et al., 2005; Wienholds et al., 2005; Zhao et al., 2005; Sokol and Ambros, 2005).

We also found tissue-specific gene sets that are significantly enriched for *miR-1* and *miR-124* targets. Genes expressed in ectodermal derivatives, especially epidermal tissues, enrich for *miR-124* sites (Figure 3B). This is intriguing in view of the common developmental origin of epidermal and neural cells (Figure 3E). Neural progenitor cells are selected from the neurectoderm in a stochastic process and change their identity to neuronal. *miR-124* is expressed exclusively in neuronal cells as they begin to differentiate (Figures 5B–5E), and genes expressed in these cells avoid *miR-124* sites. By repressing epithelial genes in neurons, *miR-124* may help ensure that the cell-type transition occurs with high fidelity and that neuronal identity is guaranteed. For miR-1, the strongest enrichment signal was for genes expressed in garland cells, which develop from the mesoderm (Figure 3B). *miR-1* may limit expression of garland-cell genes prior to separation of these cells from their mesodermal progenitors.

Figure 4A shows significant patterns of tissue avoidance and enrichment for *miR-9a* and *miR-279*. Genes expressed in ectodermal tissues avoid *miR-9a* sites, whereas genes expressed in the CNS and the peripheral nervous system (PNS) are highly enriched for them. Expression analysis showed that *miR-9a* is ectoderm specific. Early in development, it is expressed in the presumptive ectoderm and neurectoderm, but not in the presumptive mesoderm (Figure 4B). At later stages, it is expressed in a dynamic pattern in the ectoderm, but not in neural progenitors, sensory system progenitors, or the definitive nervous system, consistent with the target site avoidance and enrichment patterns.

Likewise, epidermal genes selectively avoid *miR-279* sites, while genes expressed in PNS and CNS enrich for them (Figure 4A). *miR-279* is expressed in a complex dynamic pattern during embryogenesis (Figure 4C). Strongest expression was seen in the head epidermis in regions adjacent to where the sensory organ progenitors form. In addition, we observed a complex expression pattern in trunk segments, reminiscent of the PNS (Figure 4D). Double labeling with the sensory cell marker *couch potato* (Bellen et al., 1992) showed that *miR-279*-expressing cells are closely associated with PNS cells, with limited overlap at the edges (Figures 4E and 4F). Consistent with the enrichment signal, neuronal cells in the PNS lack *miR-279* expression, suggesting that *miR-279* limits the neuronal character of cells in PNS and CNS to adjacent cell populations. *miR-279* is also expressed in the gonad (Figure 4C), again in agreement with site avoidance.

How Exclusive Is *Mutual Exclusion*?—Some Examples

Despite the overall mutual exclusion in the expression of miRNAs and targets, a number of genes with predicted target sites are annotated as being coexpressed with the miRNA. To challenge our model, we analyzed several of these cases in detail. For example, 16 genes with *miR-124* binding sites are annotated as being expressed in the CNS. In most cases, a close examination of the in situ data and/or relevant literature showed that their expression is high in nonneuronal tissues and low or not detectable in the CNS, resolving the apparent conflict (e.g., *thickveins*, *Amalgam*, and *RhoBTB*). From the remaining cases where absence of expression in neurons was less clear, we analyzed *lethal of scute* (*l(1)sc*), *reversed polarity* (*repo*), and *Gliotactin* (*Gli*) (Figure 5). All three contain *miR-124* sites that are conserved in eight *Drosophila* genomes, and regulation of their respective 3′UTRs has been verified in cell culture experiments (Robins et al., 2005; data not shown). *l(1)sc* encodes a transcription factor that is highly expressed in delaminating neuroblasts (Martin-Bermudo et al., 1991) but whose expression is lost upon onset of neuronal differentiation, when *miR-124* expression is first detected. Double in situs showed that *l(1)sc* and *miR-124* are expressed in the same cell lineage but predominantly at different developmental stages (temporal mutual exclusion; Figures 5B and 5C). Simultaneous expression is only detected during germ-band retraction, yet never in the same cells as determined by confocal analysis (Figure 5E).

repo encodes a transcription factor, whose expression in the CNS is restricted to lateral glia (Xiong et al., 1994). Double labeling showed that the neuron-specific *miR-124* is absent in neighboring *repo*-positive glia (spatial mutual exclusion; Figures 5H–5J).

Gliotactin encodes a transmembrane protein that is broadly expressed in most epidermal cells and becomes more refined at later stages (Figures 5L–5N). *Gli* is never seen in neurons but is expressed in exit glia closely associated with the ventral nerve cord (Auld et al., 1995), indicating spatial mutual exclusion with the neuron-specific *miR-124*.

We performed a similar analysis for the muscle-specific *miR-1*. The presence of two *miR-1* sites in the muscle gene *Tropomyosin 1* (*Tm1*) seems in clear conflict with our model. However, there are several isoforms of *Tm1* (Hanke and Storti, 1988; Figure 6A). The three isoforms expressed in muscle lack *miR-1* sites. The two highly conserved *miR-1* sites are found in the 3′UTR of the ''cytoplasmic'' isoform (*cTm1*) and confer regulation by *miR-1* in a luciferase-reporter assay (data not shown). cTm1 is involved in motility of nonmuscle cells, and expression is detected in gut, epidermis, and brain, but not in muscle (Figures 6B and 6C). cTm1 differs considerably from the muscle isoforms, and its misexpression might interfere with assembly of functional muscle fibers. Strikingly, the existence of Tropomyosin isoforms with distinct functions is conserved in vertebrates

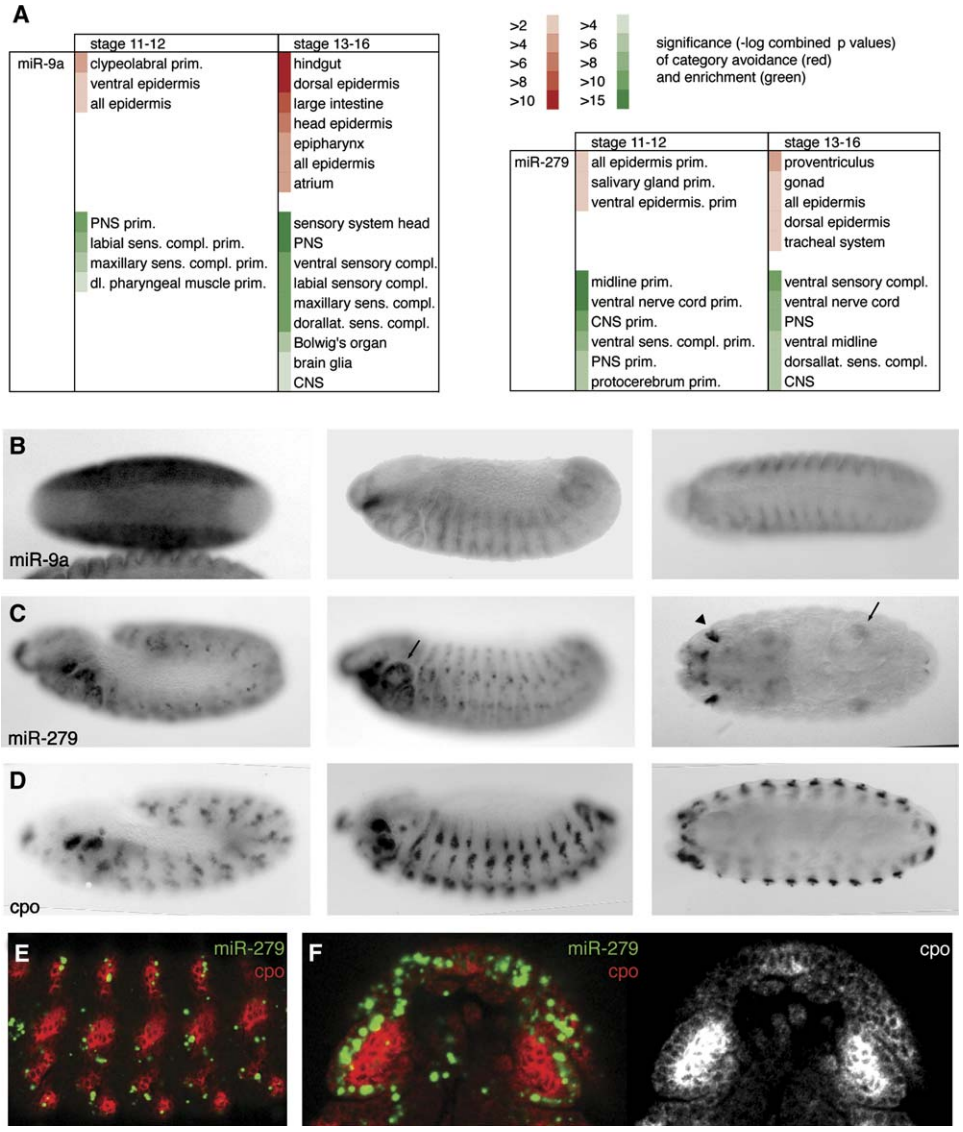


Figure 4. *miR-9a* and *miR-279* Target/Antitarget Expression

(A) Avoidance and enrichment patterns for *miR-9a* (left) and *miR-279* (right) for stages 11/12 and 13–16.

(B and C) miRNA primary transcript in situ hybridization. Lateral views, anterior left, unless otherwise indicated.

(B) *miR-9a* expressed in ectoderm but not presumptive mesoderm at blastoderm (left, ventral view). At stage 12 expression is in ectodermal cells at segment edges and in stomodeum, but not in CNS (middle; right, ventral view).

(C) *miR-279* expression is first detected during germband elongation. Stage 11: expression in head epidermis and in segmentally repeated groups of cells in the trunk (left). Stage 13: more complex pattern in the trunk (middle). Note strong expression outlining the edges of the head segments and the absence of label centrally, where sense organ primordia form (arrow: maxillary segment). Stage 17: strong labeling in anterior spiracles (arrowhead), gonads (arrow), and a complex signal in head segments (right; dorsal view).

(D) In situ hybridization for the PNS marker *couch potato*. From left: stage 11, 13 with strong labeling of head and trunk sense organs, and 16 (dorsal view).

(E and F) Confocal images of fluorescent in situ labeling for *miR-279* (nuclear, green) and *cpo* (cytoplasmic, red). *miR-279* expression generally flanks *cpo* expressing PNS cells. Embryonic trunk region ([E], lateral view); head region ([F], ventral view).

and again, only the cytoplasmic isoform of Tropomyosin 3, contains a predicted *miR-1* site (Krek et al., 2005; Lewis et al., 2005). This example indicates that miRNAs might not only reduce noise arising from erroneous transcription but also from imprecise splicing.

Another interesting example is the V-ATPase complex. Some of its subunits are annotated as being expressed in muscle but contain *miR-1* sites conserved in flies, worms, and vertebrates (Figure 6D; Krek et al., 2005; Lewis et al., 2005). In contrast to the F-ATPase, which is important for ATP synthesis in all cells (not targeted by *miR-1*), V-ATPase generates a proton gradient across membranes and regulates the pH of certain organelles. Reexamination of the expression patterns (e.g., Figure 6E) and the literature (Allan

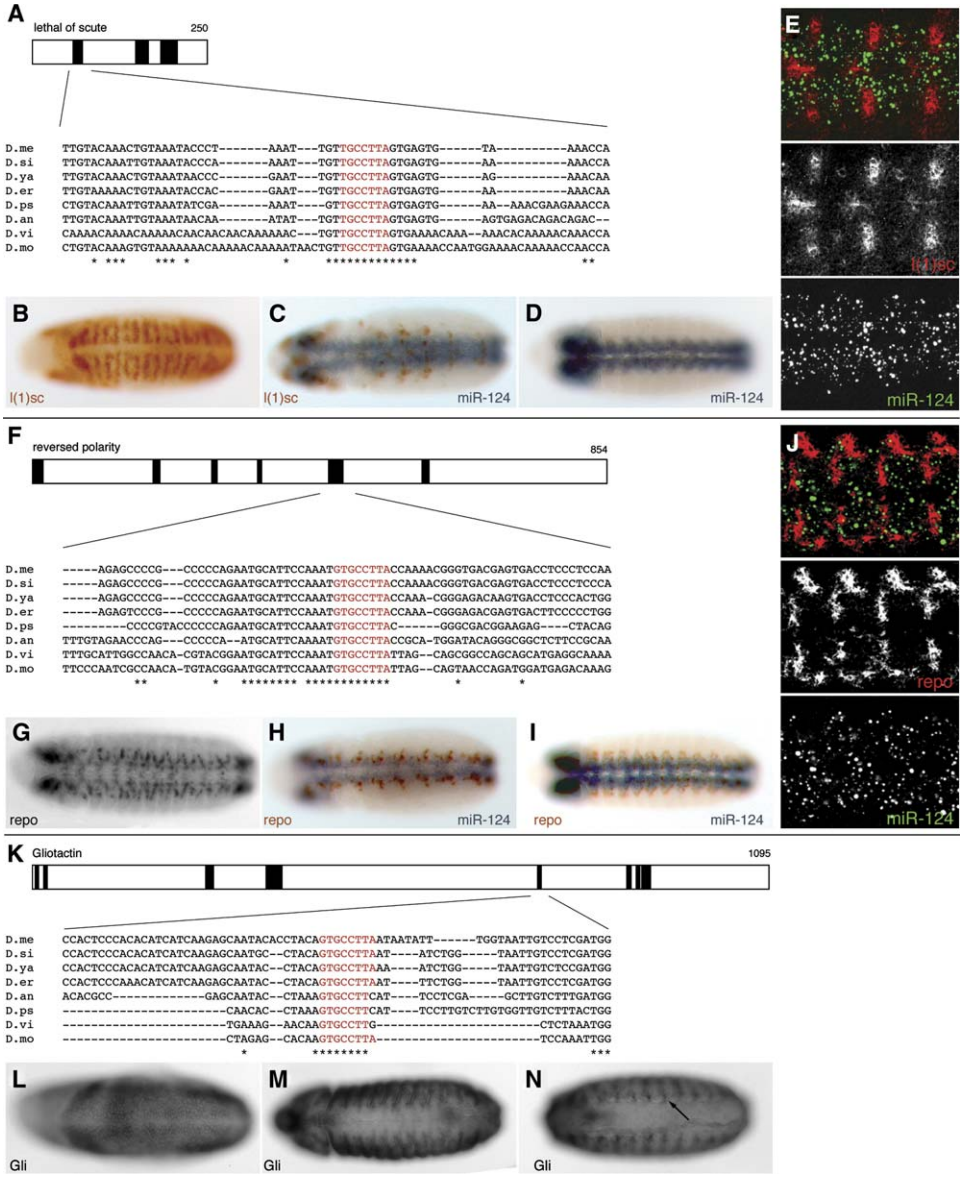


Figure 5. Proximity of *miR-124*/Target Expression

(A, F, and K) Evolutionarily conserved sequence blocks in 3′UTRs (black). Local sequence alignments highlight *miR-124* complementary seed sequences (red).

(B, C, and D) Double in situ hybridization for *miR-124* (blue) and l(1)sc (brown); development progresses from (B) (stage 9; dorsal view) to (C) to (D) (stages 11, 14; ventral views).

(E) Confocal analysis of fluorescent double in situ (ventral nerve cord, stage 11) showing *l(1)sc*-positive neuroblasts (red) and *miR-124*-positive neurons (green).

(G) In situ hybridization for *repo* (stage 13) labels all lateral glia cells.

(H and I) Double labeling for *miR-124* (blue) and *repo* (brown); development progresses from (H) to (I) (stages 11, 13; ventral views).

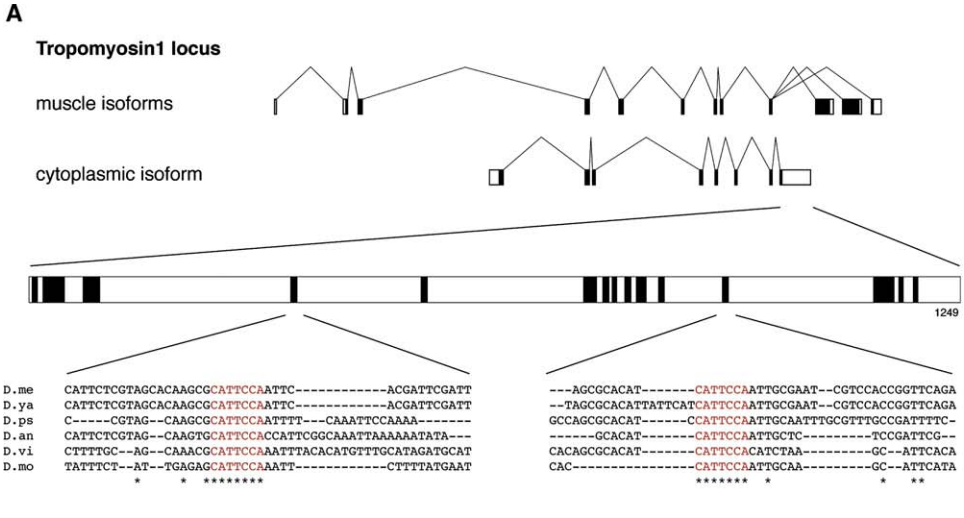
(J) Confocal analysis of fluorescent double in situ (ventral nerve cord, stage 13) showing *repo*-positive glia (red) and *miR-124*-positive neurons (green).

(L, M, and N) In situ hybridization verifies absence of *Gli* expression in neurons.

(L) Stage 11; dorsal view; (M) stage 13; (N) stage 16, arrow: exit glia.

et al., 2005) indicates that V-ATPase is not expressed in muscle but in tissues with high rates of membrane traffic (e.g., gut or malphigian tubules), so that *miR-1* might prevent its potentially deleterious ectopic expression in muscle.

Thus, the examples that were analyzed because they apparently conflicted with our model actually support it. They illustrate mutually exclusive expression of miRNAs with their targets, either temporally or spatially.



D

subunit	D. melanog.	H. sapiens	C. elegans
A	vha68-1	ATP6V1A	vha13
	vha68-2		
	vha68-3		
B	vha55	ATP6V1B1	vha12
		ATP6V1B2	
SFD	vhaSFD	ATP6V1H	vha15
C	vha44	ATP6V1C1	vha11
		ATP6V1C2	
D	vha36-1	ATP6V1D	vha14
	vha36-2		
	vha36-3		
E	vha26	ATP6V1E1	vha8
		ATP6V1E2	
F	vha14-1	ATP6V1F	vha9
	vha14-2		
G	vha13	ATP6V1G1	vha10
		ATP6V1G2	
		ATP6V1G3	
a	vha100-1	ATP6V0A1	vha5
	vha100-2	ATP6V0A2	vha6
	vha100-3	TCIRG1	vha7
	vha100-4	ATP6V0A4	unc-32
	vha100-5		
c	vha16-1	ATP6V0C	vha1
	vha16-2		vha2
	vha16-3		vha3
	vha16-4		
	vha16-5		
PPA1	vhaPPA1-1	ATP6V0B	vha4
	vhaPPA1-2		
M9.7	vhaM9.7-1	ATP6V0E	vha17
	vhaM9.7-2		
	vhaM9.7-3		
	vhaM9.7-4		
AC39	vhaAC39-1	ATP6V0D1	vha16
	vhaAC39-2	ATP6V0D2	
AC45	vhaAC45	ATP6AP1	-
M8.9	vhaM8.9	ATP6AP2	Y55H10A.1

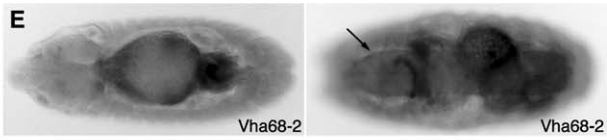


Figure 6. *miR-1* and Muscle Physiology

(A) *Tropomyosin 1* transcript isoforms (CDS, black; UTRs, white); 3′UTR of the cytoplasmic isoform (*cTm1*) is enlarged with conserved sequence blocks in black; miR-1 complementary seed sequences in red.

(B) In situ hybridization showing *cTm1* expression in epidermis (left; stage 8), nervous system and gut (right; stage 14).

(C) In situ hybridization for muscle isoforms (*mTm1*) shows expression in muscles; stage 14 ventral view; (D) V-ATPase subunit genes in *D. melanogaster*, *C. elegans*, and *H. sapiens*. Orange: those with predicted *miR-1* target sites (this work; Lewis et al., 2005, Krek et al., 2005)

(E) In situ hybridization for *Vha68-2* showing strong expression in gut, malphigian tubules, and CNS (left: dorsal view stage 13; right: ventral view stage 16; arrow: ventral nerve cord).

DISCUSSION

In this report, we provide evidence that animal miRNAs have wide-ranging effects on diverse sets of genes: (1) many genes have been subject to selection during evolution to enrich for or avoid miRNA binding sites by changes in 3′UTR length and in site density. (2) Genes that avoid miRNA regulation tend to be expressed ubiquitously and are involved in basic cellular processes. In contrast, target genes show tissue-specific expression with roles in developmental processes. (3) Target site avoidance for individual miRNAs correlates with miRNA expression, indicating that miRNA and target expression are largely nonoverlapping during development. (4) Targets of individual miRNAs tend to be expressed in tissues spatially or temporally flanking the miRNA-expression domain. This mutually exclusive expression allows us to propose a model in which miRNAs confer robustness to gene-expression programs.

Site Number and Cooperativity

Ninety-five percent of genes with conserved target sites have just one site for one miRNA. ''Switch'' targets (Bartel and Chen, 2004) that are stringently regulated via multiple sites for one miRNA are rare. Interestingly, most known targets of genetically identified miRNAs contain multiple sites (Abrahante et al., 2003; Brennecke et al., 2003; Lee et al., 1993; Lin et al., 2003; Reinhart et al., 2000; Wightman et al., 1993). We speculate that their genetic identification was in fact possible because of strong target derepression in the absence of the miRNA. Stringent target regulation might reflect the potential damage from misregulation (e.g., the proapoptotic gene *hid* or the many transcription factors among our top predictions). For genes with single sites, meaningful regulation is likely restricted to situations when transcript levels are low relative to miRNA levels. Our analysis of miRNA and target expression suggests that this type of regulatory relationship is common in vivo. However, we do not exclude the possibility that expression of some genes is more subtly modulated (tuning targets; Bartel and Chen, 2004). Weak sites are also a prerequisite for combinatorial regulation by several miRNAs. We observed extensive cooccurrence of sites for different miRNAs, which suggests cooperative regulation by coexpressed miRNAs or complementary regulation by different miRNAs in different cells.

miRNAs and 3′UTR Evolution

The widespread impact of animal miRNAs on many target genes results from the flexibility of target-site recognition, where as little as a 7mer seed can confer regulation (Brennecke et al., 2005; Doench and Sharp, 2004). Indeed, levels of many RNAs change when siRNAs or miRNAs are introduced into animal cells (Jackson et al., 2003; Lim et al., 2005). Functional sites will thus appear frequently during evolution, and our data suggest that genes confronted with miRNAs have been under selection to specifically avoid sites or take advantage of the regulation. Avoidance is expected for genes for which miRNA-mediated repression would be detrimental and for genes expressed at high levels, which could interfere by titrating miRNAs off their genuine targets (Bartel and Chen, 2004).

We find that selection has acted both to limit 3′UTR length and to specifically eliminate miRNA complementary sites. Reciprocally, 3′UTR length and site density increases with the number of miRNA binding sites. This suggests that miRNAs have had a profound impact on 3′UTR evolution. Remarkably, ~50% of conserved 8mer blocks in vertebrate 3′UTRs are complementary to known miRNAs (Xie et al., 2005). This predicts that the differences in 3′UTR length between target and antitarget categories observed in *Drosophila* should also be present in other animals but absent in species lacking miRNAs (yeast) or in plants where the requirements for miRNA pairing are higher, precluding prevalent off-target effects (Schwab et al., 2005). Indeed, genes coding for transcription factors have significantly longer 3′UTRs than ribosomal genes in nematodes, flies, and humans, whereas they are of similar lengths in yeast and *Arabidopsis* (Figure 2H).

Mutual Exclusion

A key outcome of this work is the perspective on miRNA function that emerges from the relationships between miRNAs and both their targets and antitargets. miRNAs and their targets are expressed in a largely nonoverlapping manner, whereas miRNAs and antitargets tend to be coexpressed. Most intriguing is the finding that miRNAs preferentially target genes expressed in neighboring tissues (spatially or temporally). We call this *mutual exclusion* to emphasize that miRNAs prevent unwanted expression of target transcripts, which should be absent in the miRNA-expressing cell. The evolutionary conservation of target sites and the observation that targets often have similar function or expression profiles argues that certain genes are predisposed of being misexpressed. Although not detectable by in situ hybridization, these transcripts might thus be present at low levels, allowing the miRNA to repress their expression to inconsequential levels.

Our findings suggest that only the combined analysis of targets and antitargets can reveal miRNA function, and that inferring function solely from targets might be misleading. For example, although we predict *miR-9* to target many neurogenesis genes, it is likely not involved in neurogenesis. In contrast, the patterns of target avoidance and the miRNA expression suggest that *miR-9* confers epidermal identity by suppressing erroneously transcribed neural genes. We were not able to visualize expression of *miR-9b* and *c*. It is possible that they are expressed in proliferating neuronal precursor cells as in vertebrates (Wienholds et al., 2005), where they might suppress premature differentiation, consistent with the *miR-9* target spectrum.

Our analysis is based on in situ expression data and indicates that the transcription of miRNA and targets is generally mutually exclusive. However, recent data show that animal miRNAs can also destabilize target mRNAs (Bagga et al., 2005; Lim et al., 2005), suggesting that miRNAs could shape transcript patterns. If the degree of miRNA-mediated downregulation were strong, this could explain mutual exclusion of miRNA and targets. However, the weight of available

evidence does not support this view: (1) the pattern of gene expression defined by in situ hybridization generally reflects the expression of enhancer traps that place *lacZ* under the control of an endogenous promoter. For predicted miRNA targets such as *repo* or *Gli*, the two patterns have been reported to be indistinguishable (Auld et al., 1995; Xiong et al., 1994; Figure 5). (2) We visualized the nascent transcripts for the predicted targets *repo*, *Gli*, and *l(1)sc* using intron probes or by confocal analysis and found them to be comparable to the mature mRNA. (3) If target-expression patterns were strongly influenced by miRNAs, ectopic expression should occur in the absence of miRNAs. However, Sokol and Ambros (2005) did not detect ectopic expression of predicted *miR-1* targets in *miR-1* mutant flies. Nor did Giraldez et al. (2005) find evidence for altered expression of important developmental genes in zebrafish embryos lacking all miRNAs. This is consistent with reports that the effects of miRNA expression on target mRNA levels are generally <50% (Lim et al., 2005; Pillai et al., 2005). Bagga et al. (2005) recently showed that the mRNA levels of the *lin-4* and *let-7* targets, *lin-14*, *lin-28*, and *lin-41*, were strongly downregulated upon miRNA expression. However, target mRNA levels were also reduced in *lin-4* and *let-7* mutants, albeit to a lesser extent, indicating independent transcriptional downregulation. It may be that even these ''switch-like'' miRNAs support, rather than dictate, target-gene repression.

The mode of mutually exclusive expression is likely important in developmental decisions where cells need to make transitions from one state to another. Progenitor cells must maintain their identity while being able to efficiently initiate a new developmental program. This might come at the cost of leaky transcription, and a miRNA expressed in the progenitor population could help to prevent premature expression of genes needed during differentiation (e.g., *miR-9a* in neurectoderm). Reciprocally, miRNAs expressed in the daughter lineage provide a rapid and effective means to repress residual mRNAs while the transcriptional program of the cell is changing (e.g., *miR-124* in neurons). We observed complementary patterns between miRNAs in spatial expression and identity of targets and antitargets, indicating reciprocal roles for different miRNAs (e.g., *miR-124* and *miR-9*).

Although we think that the model of mutual exclusion applies to many or all miRNAs, this might be obscured in some cases. Some miRNAs have highly dynamic expression patterns that do not coincide with tissues or organs. miRNAs likely target different genes at different times or in different tissues, but this temporal and spatial resolution is not reflected in the lists of predicted targets. Some miRNAs come in families with identical or near-identical sequences and consequently very similar target lists. As some are expressed from different genetic loci in different tissues (A. Boutla, personal communication), comparison of target prediction and spatial expression cannot be resolved for individual family members. In addition, complex organs often contain various cell types that express different miRNAs (e.g., nerve cells versus glia) so that reciprocal avoidance and enrichment signals for the whole organ might cancel each other. Our model might even apply to miRNAs that seem to be expressed ubiquitously, as those are likely not ubiquitous over time but could support developmental transitions as systemic timers.

Finally, we derived this model for miRNAs that are conserved and abundant during normal fly development. It is possible that recently evolved species-specific miRNAs are more involved in fine-tuning gene expression to adapt organisms to different environments rather than supporting more ancient developmental programs.

Perspective

We suggest that miRNAs confer precision and robustness to developmental processes. This view is based on several findings: (1) miRNAs regulate a large number of targets with diverse molecular and physiological functions rather than few key factors; (2) most targets contain only single sites for individual miRNAs insufficient for stringent regulation; (3) miRNAs and their targets are generally expressed in a mutually exclusive manner; (4) although miRNAs have recently been reported to show striking tissue- and organ-specific expression in zebrafish embryos (Wienholds et al., 2005), a general role for miRNAs as developmental switches in patterning or organogenesis was excluded by analyzing *Dicer* mutants (Giraldez et al., 2005). This is consistent with our proposal that miRNAs confer fidelity to developmental processes and leads to the expectation that a considerable proportion of mutants lacking single miRNAs might show only relatively mild defects, e.g., increased developmental variability. During evolution, developmental robustness is, however, crucial, and indeed numerous miRNAs are deeply conserved in insects, nematodes, and vertebrates. The ease with which novel miRNAs and miRNA target sites can be acquired or lost, with the ensuing consequences in developmental variation makes miRNAs powerful tools during evolution.

EXPERIMENTAL PROCEDURES

Target Prediction

Orthologous pairs of unique *D. melanogaster* and *D. pseudoobscura* 3′UTRs were aligned as described (Brennecke et al., 2005). For each cloned *Drosophila* miRNA (Aravin et al., 2003) we found all 8 to 4mers complementary to the 5′ end of the miRNA that were 100% conserved allowing for positional alignment errors of ±2 nt. For 8mers, we allowed one nt loop in the miRNA or target and one mismatch, for 7mers one G:U mismatch. For each match, we extracted the 3′ adjacent sequence for both genomes, paired it to the miRNA 3′ end starting at nt 10 with RNAhybrid (Rehmsmeier et al., 2004) and used the worse score. For 8mers with a G:U mismatch or loop on the target side, we required 3′ pairing to be ≥50% of the maximally possible pairing energy; 60% was required for 8mers with a mismatch or loop on the miRNA side, as well as for 7mers with a G:U mismatch and for 6mers, 70% for 5mers, and 80% for 4mers; none was required for 8 and 7mers. We normalized the 5′ and 3′ pairing energy calculated by RNAhybrid separately using Z scores (Stark et al., 2003). Based on the statistical signal obtained from pure seed matching for the individual seed types (Brennecke et al., 2005), the 5′ scores for 8mers were weighted by 2.8×, 7mers by 2×, and 6mers and target-side loops by 1.2 × .5′ and 3′ scores were added to give the individual site score. The UTR score is the sum of all sites with nonoverlapping seeds.

Shuffled miRNA Controls

We used 10 random (shuffled) miRNAs for each of the 39 cloned 5′ nonredundant miRNAs. We shuffled the entire miRNA sequence and required

the random sequences to have an equal number of matches (±15%) in the *D. melanogaster* 3′UTRs. Targets were predicted for shuffled miRNAs as above.

Site Cooccurrence

We counted the number of predicted sites for all cloned miRNAs per gene. For the random controls, we distributed an identical number of sites as obtained in our analysis for each miRNA randomly across all genes by a "drawing experiment with replacement" and counted number of times each gene was chosen. Note that comparison to predictions for shuffled miRNAs are uninformative (see Supplemental Data). For the drawing experiment, we first assumed an identical a priori probability (i.e., database frequency) of being a target for all genes and randomly drew genes (with replacement) from our database. We then corrected (multiplied) the a priori probability for each gene with its 3′UTR length, as the number of short matches in a long sequence depends linearly on the sequence length. For each distribution, we binned all genes according to the number of sites and calculated the median 3′UTR length and site density (number of sites per kb 3′UTR sequence) within each bin.

Comparison of 3′UTR Lengths

3′UTRs and GO annotations were obtained from ENSEMBL (*H. sapiens*), TAIR (www.Arabidopsis.org, *A. thaliana*), and Wormbase (*C. elegans*). For *S. cerevisiae*, GO annotations were obtained from SGD (www.yeastgenome.org), and 3′UTR length information for 2214 genes was kindly provided by Lars Steinmetz. Median lengths for all genes in the respective GO categories were calculated from the 3′UTR length average of all splice forms per gene, and the significance of the group differences was assessed with a two-tailed t test.

Functional Clustering

We obtained annotations from the Gene Ontology consortium, the KEGG database, and the BDGP in situ expression database (second release kindly provided by P. Tomancak and V. Hartenstein, personal communication). We added all parent categories to each gene's annotation to allow functional comparison at every level of the hierarchy. We tested for enrichment and avoidance of genes corresponding to all categories within our predictions by three different, complementary measures (see Supplemental Data for details). Briefly, we tested (1) whether a category is over/underrepresented among predicted targets, (2) whether 3′UTRs in a category contain more sites than expected given their length and conservation, and (3) whether target sites are better conserved than average 3′UTR sequences in a category. All measures are based on binomial p values that assess the deviation from random where small p values close to zero are significant.

UTR Assays

3′UTRs of predicted targets were cloned downstream of firefly luciferase (reporter plasmids). miRNAs were expressed from plasmids containing 500 bp genomic DNA including the hairpin. Reporter and miRNA plasmids contained the tubulin promoter. S2 cells were transfected in six-well plates with 0.1μg of the firefly luciferase reporter plasmid, 0.1 μg of Renilla luciferase transfection control, and 1 μg of miRNA expression plasmid or empty vector. Transfections were performed in triplicate. Dual luciferase assays were performed 2.5 days after transfection according to the manufacturer's protocol (Promega).

miRNA In Situs

pri-miRNA transcript in situs were as described (Kosman et al., 2004) except for the following: embryos were not treated with Xylene; probes were labeled with DIG-11 UTP but not hydrolyzed; probes were detected with AP-coupled anti-DIG Fab fragments (Roche #1093274; 1:2000; 2 hr RT) and visualized with NBT/BCIP (Roche #1682326; 30–120 min). For double in situs, the probes were hybridized together (labels—pri-miRNA, DIG; mRNA, Fluorescein-12-UTP), pri-miRNA was detected as above and the antibody removed with Glycine/HCl (0.1M; pH2.2). mRNA was then detected with AP-coupled anti-Fluorescein Fab fragments (1:2000; Roche #1426346) and visualized with INT/BCIP (Roche #1681460). Fluorescent double in situs were as follows: primary antibodies (anti-DIG POD [1:200; Roche #1207733] and AP-coupled anti-Fluorescein) were incubated together (2 hr RT). miRNA was detected first with the tyramide signal amplification method (Molecular Probes #T-20939; 2 hr RT). If needed, a second amplification round was performed using HRP coupled anti-oregon green (1:400; Molecular Probes #A21253). Subsequent mRNA detection was performed as above but using FastRed (Roche #3019560).

Supplemental Data

Supplemental Data include three figures, one table, and supplemental text and can be found with this article online at http://www.cell.com/cgi/content/full/123/6/1133/DC1/.

ACKNOWLEDGMENTS

We thank A. Teleman, J. Varghese, J. Karres, and U. Weihe for target tests; P. Tomancak and V. Hartenstein for providing the ImaGO database before publication; N. Sokol for sharing unpublished data on miR-1 and advice on miRNA in situs; H. Robins and N. Rajewsky for sharing miRNA target predictions; S. Eckert, E. Loeser, and A.-M. Voie for technical assistance; L.J. Jensen for helpful discussions on the statistics; and E. Izaurralde for comments on the manuscript.

Received: July 15, 2005
Revised: September 27, 2005
Accepted: November 14, 2005
Published online: December 1, 2005

REFERENCES

Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A., and Rougvie, A.E. (2003). The *Caenorhabditis elegans* hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. Dev. Cell *4*, 625–637.

Allan, A.K., Du, J., Davies, S.A., and Dow, J.A. (2005). Genome-wide survey of V-ATPase genes in Drosophila reveals a conserved renal phenotype for lethal alleles. Physiol. Genomics *22*, 128–138.

Alvarez-Garcia, I., and Miska, E.A. (2005). MicroRNA functions in animal development and human disease. Development *132*, 4653–4662.

Ambros, V. (2004). The functions of animal microRNAs. Nature *431*, 350–355.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during *Drosophila melanogaster* development. Dev. Cell *5*, 337–350.

Auld, V.J., Fetter, R.D., Broadie, K., and Goodman, C.S. (1995). Gliotactin, a novel transmembrane protein on peripheral glia, is required to form the blood-nerve barrier in *Drosophila*. Cell *81*, 757–767.

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

Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell *116*, 281–297.

Bartel, D.P., and Chen, C.Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat. Rev. Genet. *5*, 396–400.

Bellen, H.J., Kooyer, S., D'Evelyn, D., and Pearlman, J. (1992). The *Drosophila* couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins. Genes Dev. *6*, 2125–2136.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. *37*, 766–770.

Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H., and Cuppen, E. (2005). Phylogenetic shadowing and computational identification of human microRNA genes. Cell *120*, 21–24.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. Cell *113*, 25–36.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol. *3*, e85. 10.1371/journal.pbio.0030085.

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

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

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., and Schier, A.F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. Science *308*, 833–838.

Grün, D., Wang, Y.L., Langenberger, D., Gunsalus, K.C., and Rajewsky, N. (2005). microRNA target predictions across seven Drosophila species and comparison to mammalian targets. PLoS Comput. Biol. *1*, e13. 10.1371/journal.pcbi.0010013.

Hanke, P.D., and Storti, R.V. (1988). The Drosophila melanogaster tropomyosin II gene produces multiple proteins by use of alternative tissue-specific promoters and alternative splicing. Mol. Cell. Biol. *8*, 3591–3602.

Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat. Biotechnol. *21*, 635–637.

Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W., and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. Science *305*, 846.

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

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

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

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

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769–773.

Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F.J. (2003). The *C. elegans* hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev. Cell *4*, 639–650.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat. Cell Biol. *7*, 719–723.

Martin-Bermudo, M.D., Martinez, C., Rodriguez, A., and Jimenez, F. (1991). Distribution and function of the lethal of scute gene product during early neurogenesis in Drosophila. Development *113*, 445–454.

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

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

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. RNA *11*, 1640–1647.

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

Robins, H., Li, Y., and Padgett, R.W. (2005). Incorporating structure to predict microRNA targets. Proc. Natl. Acad. Sci. USA *102*, 4006–4009.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. Dev. Cell *8*, 517–527.

Sen, G.L., and Blau, H.M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. Nat. Cell Biol. *7*, 633–636.

Sokol, N.S., and Ambros, V. (2005). Mesodermally expressed Drosophila microRNA-1 is regulated by Twist and is required in muscles during larval growth. Genes Dev. *19*, 2343–2354.

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila microRNA targets. PLoS Biol. *1*, e60. 10.1371/journal.pbio.0000060.

Stivers, C., Brody, T., Kuzin, A., and Odenwald, W.F. (2000). Nerfin-1 and -2, novel Drosophila Zn-finger transcription factor genes expressed in the developing nervous system. Mech. Dev. *97*, 205–210.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *3*, RESEARCH0088.

Wienholds, E., Kloosterman, W.P., Miska, E., Alvarez-Saavedra, E., Berezikov, E., de Bruijn, E., Horvitz, R.H., Kauppinen, S., and Plasterk, R.H. (2005). MicroRNA expression in zebrafish embryonic development. Science *309*, 310–311.

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

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature *434*, 338–345.

Xiong, W.C., Okano, H., Patel, N.H., Blendy, J.A., and Montell, C. (1994). repo encodes a glial-specific homeo domain protein required in the Drosophila nervous system. Genes Dev. *8*, 981–994.

Zamore, P.D., and Haley, B. (2005). Ribo-gnome: the big world of small RNAs. Science *309*, 1519–1524.

Zhao, Y., Samal, E., and Srivastava, D. (2005). Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature *436*, 214–220.

Supplemental Data

Animal MicroRNAs Confer Robustness to Gene Expression and Have a Significant Impact on 3′UTR Evolution

Alexander Stark, Julius Brennecke, Natascha Bushati, Robert B. Russell, and Stephen M. Cohen

Table of Contents

Supplemental Results

Comparison of All Available *Drosophila* miRNA Target-Prediction Methods

Table S1 shows 133 *Drosophila* miRNA target pairs that have been experimentally tested by the Cohen (84), Lai (20), Padgett (16), and Macdonald (7) labs, or a combination of these labs (6; Brennecke et al., 2003; Brennecke et al., 2005; Burgler and Macdonald, 2005; Lai et al., 2005; Robins et al., 2005; Stark et al., 2003). Of all tested pairs, 71 were functional, i.e. showed miRNA dependent regulation of the reporter protein levels and 62 were not. The table further indicates which miRNA targets have been predicted by all of the publicly available miRNA target predictions. To summarize the performance of the methods, we calculated the number and percentage of functional and non-functional miRNA-target pairs that were predicted (Fig. S1, Table S1).

From all genome-wide predictions, MovingTargets (Burgler and Macdonald, 2005) has the highest **specificity** as it predicts the smallest number of non-functional targets (2%). It requires conservation in two fly genomes and highly stringent criteria such as a minimum of 3 sites per gene. Note that the high specificity comes at the cost of low sensitivity, as only 11% of all functional pairs are detected (see below). This has been specifically anticipated by the authors that sought for high-likelyhood predictions. Interestingly, the specific settings of PicTar (Grun et al., 2005) that require extensive site conservation but have otherwise relaxed criteria, achieve about the same level of specificity (3%) with a much higher sensitivity (48%).

To assess the **accuracy** or trustworthiness of the predictions, we calculated the percentage of predictions that were functional for each method. Methods by three groups were in the 90% range: The specific (94%), medium (91%) and sensitive (88%) settings of PicTar, MovingTargets (88%), and our new predictions (88%, this work). The remainder of more recent methods (miRanda 2005 update based on Enright et al. (2003); RNAhybrid (Rehmsmeier et al., 2004) and the method by Robins et al. (2005)) have accuracy values around 70%. This indicates that our understanding of miRNA target sites has improved compared to the first predictions that were based on empirical rules derived from very few examples (Stark et al. (2003), 55%; Enright et al. (2003), 39%; Rajewsky and Socci (2004), not assessed because only 1 miRNA/target pair overlapped with the experimental set).

We assessed the **sensitivity** or coverage of the methods by calculating the percent of functional sites that were predicted. Our new predictions correctly identified 81% of all functional sites, followed by the sensitive settings of PicTar (69%) and miRanda 2005 (58%). Targets predicted by miRanda, RNAhybrid and the method by Robins et al. (2005) have limited overlap with targets predicted by us or tested experimentally (see below). Thus many of their predictions could not be assessed and the sensitivity of these predictions might be an underestimate.

We also compared the **global overlap of predictions** for all different methods to our newest method considering all miRNA-target gene pairs for all cloned 5'non-redundant *Drosophila* miRNAs (Aravin et al., 2003). This shows that PicTar predictions are very similar to ours and have by far the highest overlap. For other methods, a substantial fraction of the predictions does not overlap with ours.

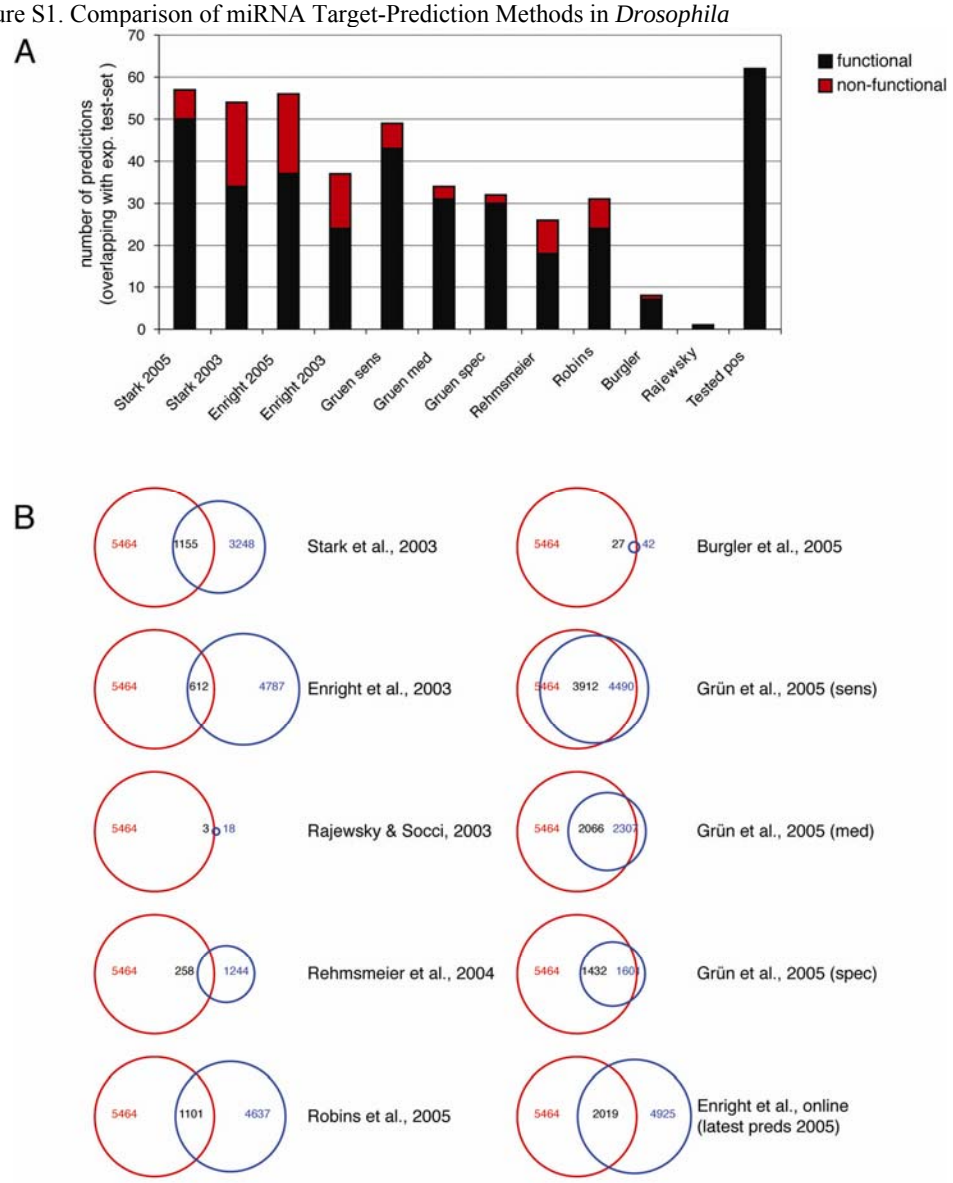
Although we use all currently available miRNA target pairs that have been experimentally tested, we are aware of the shortcomings of this comparison due to the incompleteness and bias in this set: The specificity of a method can in principle only be estimated if predictions at different ranks are systematically tested. As experimental tests are often biased towards the top-ranking predictions, it is difficult to assess the performance at different ranks (note that we tested many of our predictions at different ranks and even some miRNA/gene pairs that were not predicted). In addition, as most tested predictions are from our lab (63%), we can assess the other method's specificity only on those pairs that overlap with our predictions and tests, i.e. comply with our site rules. One could argue that the overlap of two or more prediction methods leads to better predictions, and we might thus overestimate the accuracy of other methods by assessing their performance on this overlap. In contrast, we risk underestimating the other methods' sensitivity as predictions that are not overlapping with the tests might be functional but cannot be considered. We also cannot assess sensitivity on an absolute level as nearly all miRNA target pairs were tested because they were predicted (i.e. predictions and tests are not independent). Finally, the poor overlap of predictions by methods other than PicTar and MovingTargets with our own predictions could reflect that these three methods miss a substantial part of valid targets with pairing characteristics not complying with our rules. However, it could also mean that the other methods substantially overpredict non-functional sites. We think it is crucial that the authors of the respective methods clarify on a large, representative set whether the non-overlapping

predictions are functional. If they were, we would be able to extend our understanding of site requirements. If not, this knowledge would help to improve these methods immensely and would reconcile the different attempts to target prediction, whose discrepancies have rightly been criticized (Lai, 2004).

We generally found a tradeoff between sensitivity and specificity, where methods with high specificity often have lower sensitivity, and feel that future methods have to work on improving this balance. Our comparison using the tested miRNA target pairs and the global overlap indicates that our new predictions and those by Grun et al. (2005) are very similar and seem to perform best. Notably, the rules about target site structure underlying our method and the current methods by the Rajewsky lab (Grun et al., 2005; Krek et al., 2005) are also very similar to those that were independently developed for target prediction in vertebrates by the Burge and Bartel labs (Lewis et al., 2005; Lewis et al., 2003).

If we were to provide a final statement about which predictions to consider, we would like to recommend our predictions and those by Grun et al. (2005). For the specific purpose of having a high confidence set that potentially misses valid targets, we recommend the specific PicTar predictions (94% accuracy and 48% coverage). Note however, that these are restricted to evolutionary ancient targets conserved across many fly species. In addition, the authors showed that requiring conservation across many species is currently hampered by methodological shortcomings, e.g. errors in genomic alignments or in the mapping of orthologs, which lead to lower sensitivity. If high sensitivity is wanted while guaranteeing good specificity, we recommend our predictions (88% accuracy and 81% coverage). This is especially important when working on a specific miRNA and/or phenotype and all putative target genes need to be considered. High sensitivity is also crucial for the purpose of extracting information about the complete target complement, as done in this work. Especially our finding that the signal for target site avoidance and enrichment is apparent in a single genome analysis without requiring any conservation argues that for this type of analysis the coverage is more important than a modest improvement in specificity.

Figure S1. Comparison of miRNA Target-Prediction Methods in *Drosophila*



(A) Performance of the different published methods on 133 experimentally tested miRNA target pairs (see Table S1). As most methods did not consider 3'UTRs that were not confirmed by ESTs, we excluded these pairs (gray shade in Table S1). Shown are the absolute numbers of tested miRNA target pairs that are predicted, where black bars indicate the prediction of functional pairs (*true positives*; '*Tested pos'* shows the total number of functional pairs) and red bars the prediction of non-functional pairs (*false positives*). The *miR-278* predictions of miRanda 2005 were not available. As most non-functional tests were for this miRNA and excluding it would have artificially penalized miRanda 2005, we decided to use the miRanda 2003 predictions for *miR-278*.

(B) Total overlap of predicted genes for all 5'non-redundant cloned Drosophila miRNAs (Aravin et al., 2003). Considered were the miRNAs *bantam, miR-1, -11, -12, -124, -14, -184, -210, -263b, -275, -276a, -277, -279, -281, -285, -2a, -3, -304, -305, -306, -307, -308, -311, -314, -315, -316, -317, -31a, -34, -4, -5, -7, -8, -92a, -9a, -iab-4-3p, -iab-4-5p* (*miR-278* and *-303* were not considered as predictions were not available from miRanda 2005 or PicTar, respectively). Note that the absolute number of predictions might differ from those in the respective papers because of the restriction to these miRNAs. Rajewsky and Socci (2004) for example predict 39 pairs, of which 21 are not considered as the respective miRNAs are not cloned (e.g. *miR-P323/vas*) or were not chosen as family representative (e.g. *miR-6/tll*, which is in fact recovered by us and PicTar).

Table S1. Comparison of miRNA Target-Prediction Methods in *Drosophila*

	Accession	Gene	Functional	Stark et al. 2005	Stark et al. 2003	Enright 2005	Enright et al. 2003	Gruen et al. 2005 sens	Gruen et al. 2005 med	Gruen et al. 2005 spec	Rehmsmeier et al. 2004	Robins et al. 2005	Burgler et al. 2005	Rajewsky & Socci 2004	Assay	Tested by
bantam	CG12187	-	**0**	0	1	0	1	0	0	0	0	0	0	n	W	Cohen
bantam	CG4385	S	**1**	1	1	1	1	0	0	0	0	0	0	n	W	Cohen
bantam	CG5123	W	**1**	1	1	1	1	1	1	1	1	1	1	n	W	Cohen
bantam	CG5919	-	**0**	0	0	1	0	0	0	0	0	0	0	n	W	Cohen
miR-1	CG5185	Tom	**0**	0	0	0	0	0	0	0	0	0	0	n	C	Cohen
miR-1	CG5884	par-6	**1**	1	1	1	1	1	0	0	0	1	0	n	C	Cohen
miR-11	CG13701	skl	**0**	0	0	1	0	0	0	0	0	0	0	n	W	Cohen
miR-11	CG4345	grim	**0**	0	0	1	0	0	0	0	0	0	0	n	W	Cohen
miR-11	CG7902	bap	**1**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-14	CG10473	-	**0**	0	0	0	0	0	0	0	1	0	0	n	W	Cohen
miR-14	CG1624	dpld	**0**	1	0	1	0	1	1	1	0	0	0	n	C	Cohen
miR-14	CG1765	EcR	**1**	1	1	1	1	0	0	0	1	0	0	n	W	Cohen
miR-14	CG4125	rst	**0**	1	1	1	0	1	1	0	0	0	0	n	W	Cohen
miR-14	CG4889	wg	**1**	1	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-14	CG5884	par-6	**0**	0	0	0	0	0	0	0	0	0	0	n	C	Cohen
miR-14	CG8127	Eip75B	**0**	1	0	1	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG10043	rtGEF	**1**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG10443	Lar	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG11539	-	**0**	0	0	n	n	n	n	0	0	n	0	n	W	Cohen
miR-278	CG1154	-	**0**	0	1	1	1	0	0	0	0	0	0	n	W	Cohen
miR-278	CG12187	-	**0**	0	1	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG14960	-	**1**	0	0	0	0	0	0	0	1	0	0	n	W	Cohen
miR-278	CG15097	-	**1**	0	1	1	1	0	0	0	0	0	0	n	W	Cohen
miR-278	CG15715	-	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG15861	-	**0**	0	0	0	0	0	0	0	1	0	0	n	W	Cohen
miR-278	CG16975	-	**0**	0	1	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG1709	Vha100-1	**0**	0	0	1	1	0	0	0	0	0	0	n	W	Cohen
miR-278	CG18042	Img	**0**	0	1	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG18815	-	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG1919	-	**0**	0	0	1	1	0	0	0	1	0	0	n	W	Cohen
miR-278	CG30089	-	**0**	0	0	1	1	0	0	0	0	0	0	n	W	Cohen
miR-278	CG3026	-	**0**	0	0	1	1	0	0	0	0	0	0	n	W	Cohen
miR-278	CG30483	Prosap	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG31651	-	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG31841	-	**0**	0	0	n	n	n	n	0	0	n	0	n	W	Cohen
miR-278	CG32057	-	**0**	0	0	1	1	0	0	0	0	0	0	n	W	Cohen
miR-278	CG32688	Hk	**0**	0	1	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG33006	-	**0**	0	1	0	0	0	0	0	0	0	1	n	W	Cohen
miR-278	CG3694	Ggamma30A	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG4444	px	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG6930	-	**0**	0	0	0	0	0	0	0	1	0	0	n	W	Cohen
miR-278	CG7368	-	**0**	0	0	n	n	n	n	0	0	n	0	n	W	Cohen
miR-278	CG7576	Rab3	**0**	1	1	0	0	1	0	0	0	0	0	n	W	Cohen
miR-278	CG7656	-	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-278	CG9339	-	**0**	0	1	1	1	0	0	0	0	0	0	n	W	Cohen
miR-278	CG9652	DopR	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-279	CG13906	nerfin-1	**1**	1	1	1	1	1	1	1	1	0	1	n	C	Cohen
miR-279	CG6297	JIL-1	**0**	0	0	0	0	0	0	0	0	0	0	n	C	Cohen
miR-279	CG6520	-	**1**	1	1	1	0	1	1	1	0	0	0	n	C	Cohen
miR-2b	CG11293	-	**0**	0	1	0	0	0	0	0	0	0	0	n	W	Cohen
miR-2b	CG13701	skl	**1**	1	1	1	1	1	1	1	1	1	0	n	W	Cohen
miR-2b	CG1969	-	**0**	0	1	0	0	0	0	0	0	0	0	n	W	Cohen

miR-2b	CG4269	-	**1**	1	1	1	0	1	0	1	1	0	0	n	W	Cohen
miR-2b	CG4319	rpr	**1**	1	1	1	1	1	0	0	1	1	0	n	W	Cohen
miR-2b	CG4345	grim	**1**	1	1	1	1	0	0	0	1	0	0	n	W	Cohen
miR-2b	CG4604	GLaz	**1**	1	1	1	1	1	1	1	0	1	0	n	W	Cohen
miR-2b	CG7902	bap	**1**	1	0	0	0	1	0	1	0	0	0	n	W	Cohen
miR-2b	CG7956	-	**0**	0	0	0	0	0	0	0	0	0	0	n	W	Cohen
miR-4	CG7902	bap	**1**	1	0	0	0	1	1	1	0	0	0	n	W	Cohen
miR-6	CG13701	skl	**1**	1	1	1	1	1	1	1	0	1	0	n	W	Cohen
miR-6	CG1624	dpld	**0**	0	0	0	0	0	0	0	0	0	0	n	C	Cohen
miR-6	CG4345	grim	**0**	0	0	1	1	0	0	0	0	0	0	n	W	Cohen
miR-6	CG8365	E(spl)	**1**	1	0	n	n	n	n	0	1	n	0	n	C	Cohen
miR-7	CG6494	h	**1**	1	1	1	1	1	0	0	1	1	0	1	W	Cohen
miR-7	CG6520	-	**0**	0	0	0	0	0	0	0	0	0	0	n	C	Cohen
miR-79	CG7902	bap	**1**	1	0	0	0	1	1	1	0	0	0	n	W	Cohen
miR-8	CG10917	fj	**1**	1	0	0	0	1	1	0	1	1	0	n	W	Cohen
miR-8	CG11518	pygo	**0**	0	1	1	0	0	0	0	0	0	0	n	W	Cohen
miR-8	CG11924	Cf2	**0**	0	1	1	1	0	0	0	0	0	0	n	W	Cohen
miR-8	CG18622	-	**0**	0	1	0	0	1	0	0	1	0	0	n	W	Cohen
miR-8	CG1882	-	**1**	1	0	1	1	1	0	1	0	0	0	n	W	Cohen
miR-8	CG2019	disp	**1**	0	1	1	1	0	0	0	0	0	0	n	W	Cohen
miR-8	CG32466	rn	**1**	1	0	0	0	1	1	1	1	0	0	n	W	Cohen
miR-8	CG4125	rst	**0**	0	1	1	0	0	0	0	1	0	0	n	W	Cohen
miR-8	CG4484	-	**0**	1	1	1	1	0	0	0	0	0	0	n	W	Cohen
miR-8	CG4889	wg	**1**	1	1	0	0	1	1	1	0	0	0	n	W	Cohen
miR-8	CG5735	-	**1**	1	1	1	0	1	0	1	0	1	0	n	W	Cohen
miR-8	CG6210	-	**1**	1	1	0	0	0	0	0	0	0	0	n	W	Cohen
miR-8	CG6456	MIP	**1**	1	1	0	0	1	0	0	0	0	0	n	W	Cohen
miR-8	CG6634	-	**1**	1	1	1	1	1	1	1	0	0	0	n	W	Cohen
miR-8	CG7525	Tie	**1**	1	1	1	1	1	1	0	1	0	0	n	W	Cohen
miR-8	CG7555	Nedd4	**0**	0	1	0	1	0	0	0	0	0	0	n	W	Cohen
miR-8	CG8544	sd	**0**	0	1	0	0	0	0	0	0	0	0	n	W	Cohen
miR-92b	CG5185	Tom	**0**	0	0	0	0	0	0	0	0	0	0	n	C	Cohen
miR-7	CG5185	Tom	**1**	1	1	1	1	1	1	1	1	1	0	n	C, W	Cohen, Lai
miR-7	CG6099	m4	**1**	1	0	n	n	*	*	0	0	n	0	n	W	Cohen, Lai
miR-7	CG8346	HLHm3	**1**	1	1	1	1	1	1	1	1	0	0	n	W	Cohen, Lai
miR-92b	CG7450	CrebA	**1**	1	0	1	0	1	1	1	0	0	1	n	C	Cohen, Macdon.
miR-279	CG14041	SP555	**1**	1	1	1	1	1	1	1	1	1	0	n	C	Cohen, Padgett
miR-7	CG6096	HLHm5	**1**	1	0	0	0	1	1	1	0	1	1	n	C, W	Coh., Padg., Lai
miR-11	CG6099	m4	**1**	0	0	n	n	n	n	0	0	n	0	n	W	Lai
miR-11	CG8328	HLHmdelta	**1**	1	0	0	0	1	1	0	0	0	0	n	W	Lai
miR-11	CG8337	malpha	**1**	0	0	n	n	n	n	0	0	n	0	n	W	Lai
miR-2a	CG8328	HLHmdelta	**1**	1	0	1	0	1	1	0	0	0	0	n	W	Lai
miR-2a	CG8337	malpha	**1**	1	0	n	n	n	n	0	0	n	0	n	W	Lai
miR-4	CG3096	Brd	**1**	1	0	0	0	1	0	0	0	0	0	n	W	Lai
miR-4	CG5185	Tom	**1**	1	0	0	0	1	1	1	0	0	0	n	W	Lai
miR-4	CG6096	HLHm5	**1**	1	0	0	0	1	1	1	0	1	0	n	W	Lai
miR-4	CG6099	m4	**1**	1	0	n	n	n	n	0	0	n	0	n	W	Lai
miR-4	CG8328	HLHmdelta	**1**	1	0	0	0	1	1	0	0	0	0	n	W	Lai
miR-4	CG8333	HLHmgamma	**1**	0	0	0	0	0	0	0	0	0	0	n	W	Lai
miR-4	CG8337	malpha	**1**	1	0	n	n	n	n	0	0	n	0	n	W	Lai
miR-7	CG3096	Brd	**1**	1	0	1	1	1	0	0	0	0	0	n	W	Lai
miR-7	CG3396	Ocho	**0**	0	0	n	n	n	n	0	0	n	0	n	W	Lai
miR-7	CG8328	HLHmdelta	**1**	0	1	1	1	0	0	0	0	0	0	n	W	Lai
miR-7	CG8333	HLHmgamma	**1**	1	1	1	0	1	1	0	1	0	0	n	W	Lai
miR-79	CG6096	HLHm5	**1**	1	0	0	0	1	1	1	0	1	0	n	W	Lai
miR-79	CG6099	m4	**1**	1	0	n	n	n	n	0	0	n	0	n	W	Lai
miR-79	CG8333	HLHmgamma	**1**	0	0	0	0	0	0	0	0	0	0	n	W	Lai
miR-79	CG8337	malpha	**1**	1	0	n	n	n	n	0	0	n	0	n	W	Lai
let-7	CG4807	ab	**1**	1	1	1	1	1	1	1	1	0	1	n	C	McDonald
miR-276b	CG7210		**0**	0	0	0	0	0	0	0	0	0	0	n	C	McDonald
miR-312	CG1856	ttk	**1**	1	0	1	0	1	1	1	0	0	0	n	C	McDonald
miR-312	CG7450	CrebA	**1**	1	1	1	0	1	1	1	0	1	1	n	C	McDonald
miR-34	CG32180	Eip74EF	**1**	1	1	1	1	1	1	1	0	1	1	n	C	McDonald
miR-92b	CG1856	ttk	**1**	1	1	1	0	1	1	1	0	1	0	n	C	McDonald

miR-92b	CG4807	ab	**1**	1	0	1	0	1	0	0	0	0	0	n	C	McDonald
bantam	CG12399	Mad	**1**	0	1	1	0	0	0	0	0	1	0	n	C	Padgett
miR-1	CG15427	tutl	**1**	1	1	1	0	0	0	0	0	1	0	n	C	Padgett
miR-12	CG6097	rt	**1**	0	0	0	0	0	0	0	0	1	0	n	C	Padgett
miR-124	CG3903	Gli	**1**	1	1	1	0	1	1	1	0	1	0	n	C	Padgett
miR-278	CG10619	tup	**0**	1	0	0	0	1	0	0	0	1	0	n	C	Padgett
miR-286	CG8285	boss	**0**	1	1	0	0	1	1	1	1	1	0	n	C	Padgett
miR-287	CG1411	CRMP	**1**	0	1	1	0	0	0	0	0	1	0	n	C	Padgett
miR-287	CG15367	Dip1	**0**	0	0	0	0	0	0	0	0	0	0	n	C	Padgett
miR-288	CG32057	-	**0**	0	0	1	1	0	0	0	0	1	0	n	C	Padgett
miR-303	CG14991	-	**0**	0	0	0	0	0	0	0	0	1	0	n	C	Padgett
miR-310	CG5576	imd	**1**	1	1	0	0	1	0	1	1	1	0	n	C	Padgett
miR-316	CG31795	ia2	**0**	0	0	0	0	0	0	0	0	1	0	n	C	Padgett
miR-317	CG4182	yellow-c	**0**	0	0	0	0	0	0	0	1	1	0	n	C	Padgett
miR-318	CG13380	-	**0**	0	0	0	0	0	0	0	0	1	0	n	C	Padgett
miR-34	CG8013		**1**	0	0	0	0	0	0	0	0	0	0	n	C	Padgett
miR-7	CG10580	fng	**1**	1	1	0	1	0	0	0	0	1	0	n	C	Padgett

Stark et al. 2005	Stark et al. 2003	Enright 2005	Enright et al. 2003	Gruen et al. sens 2005	Gruen et al. med 2005	Gruen et al. spec 2005	Rehmsmeier et al. 2004	Robins et al. 2005	Burgler et al. 2005	Rajewsky & Socci 2004	Tested	
7	**20**	**19**	**13**	**6**	**3**	**2**	**8**	**7**	**1**	**0**	**58**	**Non-Functional**
50	**34**	**37**	**24**	**43**	**31**	**30**	**18**	**24**	**7**	**1**	**62**	**Functional**
12	**34**	**33**	**22**	**10**	**5**	**3**	**14**	**12**	**2**	**0**	**100**	**Non-Functional %**
81	**55**	**60**	**39**	**69**	**50**	**48**	**29**	**39**	**11**	**2**	**100**	**Functional %**
88	**63**	**66**	**65**	**88**	**91**	**94**	**69**	**77**	**88**	**nd**	**52**	**Success %**

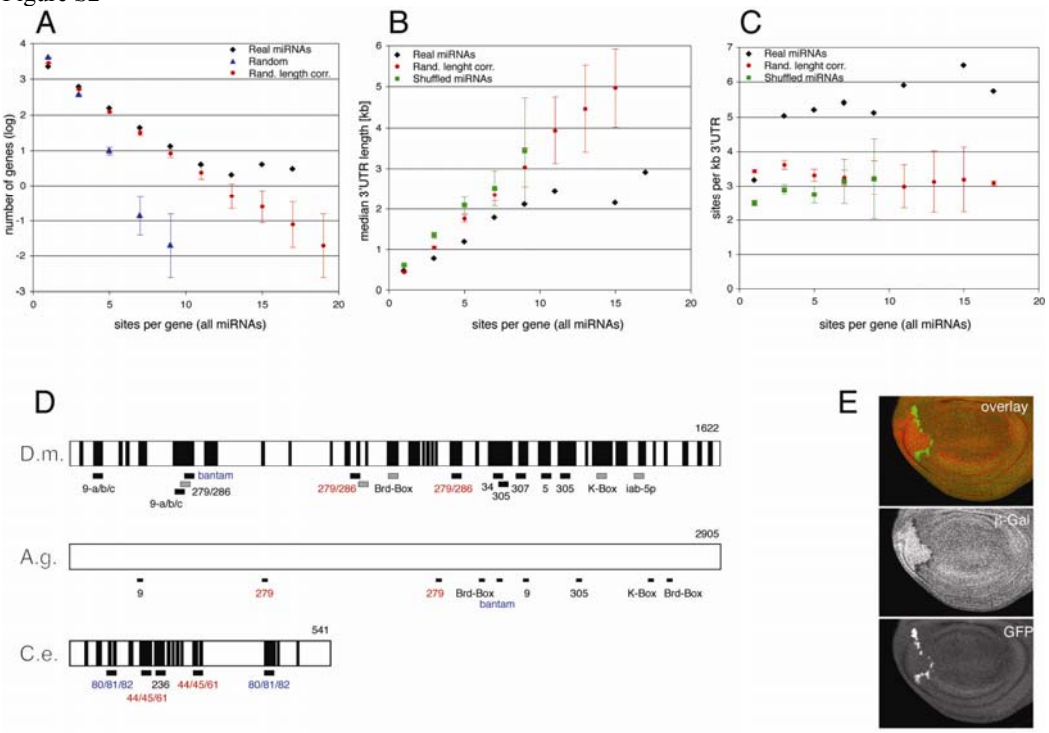
133 miRNA target pairs in *Drosophila* that have been experimentally tested in transgenic flies (W, wing-disc assay as in Brennecke et al., (2003) or in cell-culture (C) by the indicated labs (Brennecke et al., 2003; Brennecke et al., 2005; Burgler and Macdonald, 2005; Lai et al., 2005; Robins et al., 2005; Stark et al., 2003). Functionality of the miRNA target pairs and whether individual methods predict them is indicated (1: functional/predicted; 0: non-functional/not-predicted; n: not considered; *: 3'UTR not in database but site identified when included). Gray shade indicates 3'UTRs that were not annotated and not considered by many methods. Boxed values: Sum and percentage of predicted functional and non-functional pairs for each method and the success-rate (%functional of all predicted and tested). Note that all values in this box disregard the gray shaded and the not considered cases.

Extensive Cooccurrence of Sites for Different miRNAs

Only 5% of all predicted targets contain more than one conserved site for any single miRNA, indicating that stringent target regulation by one miRNA is rare (not shown). In contrast, we observed extensive co-occurrence of sites for different miRNAs in target 3'UTRs (Fig. S2; Enright et al., 2003; Grun et al., 2005; Krek et al., 2005). The 9487 binary interactions correspond to 3125 3'UTRs so that almost 50% of all targets have sites for 2 or more 5'unique miRNAs and some have sites for as many as 12. To assess the significance of site co-occurrence, we compared our results to a random model, in which an identical number of sites is distributed randomly across all genes or randomly across the available 3'UTR sequence by a 'drawing experiment with replacement' (see Methods). We first defined the a priori probability of being a miRNA target for all genes to be identical, arguing that any difference between genes (e.g. in 3'UTR length) might be biologically meaningful. Figure S2A shows that the real site co-occurrence is several orders of magnitude above the level reached by this control, indicating that many genes have more sites than expected. We then corrected the probability above with the genes' 3'UTR lengths, to assess to what extend 3'UTR length differences can explain co-occurrence. This indeed indicates that 3'UTR length differences considerably influence co-occurrence (Fig. S2B). However, for 3 and more sites per gene, the real distribution is significantly above the length-corrected control. This indicates that significantly more genes than expected given the 3'UTR length differences have 3 ore more sites.

This is further emphasized by analyzing 3'UTR lengths and site densities (i.e. the number of sites per kb 3'UTR) for all genes with a given number of sites. Figure S2B and C show that co-occurrence for predicted sites of real miRNAs arises from differences in 3'UTR length and site density: genes with more miRNA sites have on average both, longer 3'UTRs and more sites per kb 3'UTR sequence. In contrast, site co-occurrence for the length corrected control and for shuffled miRNAs comes solely from length differences, while the site densities remain constant (note that – in contrast to the relative contribution of 3'UTR length and site density – the absolute level of site co-occurrence of predictions for shuffled miRNA itself is not informative due to the intrinsically lower number of sites). This shows that many genes have several sites due to increased 3'UTR length and an increased site density. In contrast, genes with few sites have short 3'UTRs and lower site densities. The fact that both measures show identical trends and that the increase in site density is absent in both controls suggests that 3'UTRs have been specifically selected for or against miRNA regulation and argues that miRNAs have a profound impact on 3'UTR evolution.

Figure S2



(A) Co-occurrence of predicted miRNA target sites. Plot of the number of genes (log) versus the number of sites per gene, considering one representative per miRNA family for all cloned *Drosophila* miRNAs (black). To control for the level of random site co-occurrence, we distributed an identical number of sites randomly across all genes by a drawing experiment, which treats all genes equally (i.e. irrespective of 3'UTR lengths; blue) or corrects for 3'UTR length differences (red). Sites complementary to real miRNAs are highly asymmetrically distributed and site co-occurrence is much higher than expected. 3'UTR length differences account for some of the observed co-occurrence but for 3 and more sites per gene, the real curve is significantly above the corrected control. Note that controls based on predictions for shuffled miRNAs are uninformative as – depending on the overall signal-to-noise ratio – fewer sites are predicted overall. Fewer sites distributed across an identical number of genes leads trivially to fewer sites per gene and thus to apparently lower co-occurrence values.

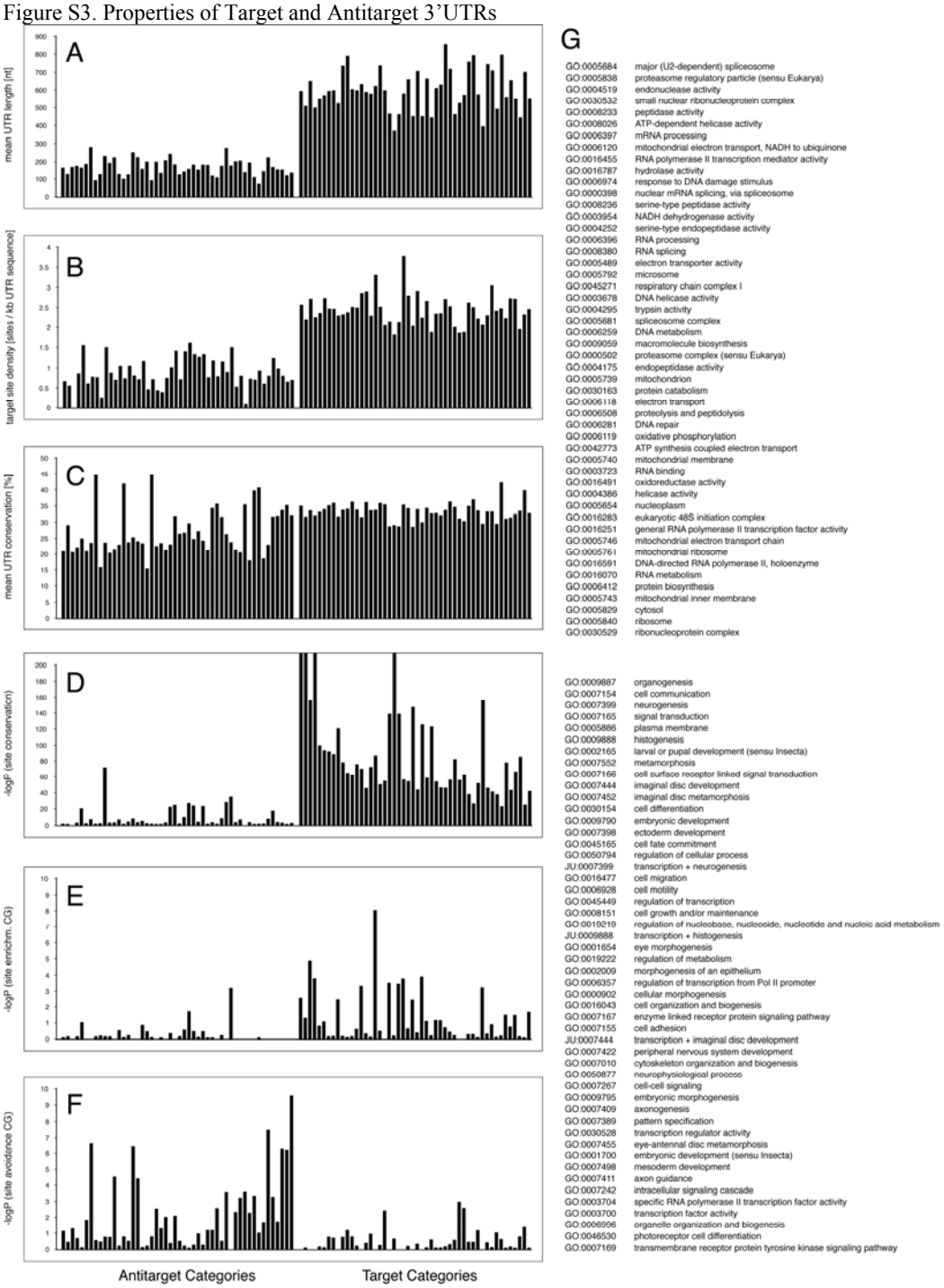
(B) Site co-occurrence and 3'UTR lengths. Median 3'UTR lengths for all genes with a given number of predicted sites for real miRNAs (black), shuffled miRNAs (green) and with a given number of randomly assigned sites in the length corrected control (red, see A). 3'UTR lengths in all cases increase with the number of sites per gene, but the distribution for sites of real miRNAs is significantly below that for the controls. Given the higher degree of site co-occurrence for real miRNAs (A), this indicates that co-occurrence cannot be solely based on 3'UTR length differences.

(C) Site co-occurrence and site density. Plot of the average site density (# sites per kb 3'UTR sequence) for all genes with a given number of sites (specifics as in B). Site density increases with the number of sites per gene for real miRNAs (black), whereas it is constant for both controls (red, green). This indicates that co-occurrence observed for controls (A) is solely based on 3'UTR length differences but that targets for real miRNA acquire sites by both increases in 3'UTR length (see B) and site density.

(D) Predicted target sites in the *nerfin-1* 3'UTR in *Drosophila melanogaster*, *Anopheles gambiae* and *Caenorhabditis elegans* (*egl-46*). Black bars: blocks of sequence conserved between *D. melanogaster* and *D. pseudoobscura* or *C. elegans*, *C. remanei* and *C. briggsae*, respectively. Positions of predicted target sites for miRNAs are shown (K box: miR-2, -6, -11, -13, 308; Brd box: miR-4, -79; (Lai et al., 2005)). Black boxes indicate sites that are conserved in *D. melanogaster*, *D. pseudoobscura*, *D. yakuba*, *D. simulans*, *D. ananassae* and *D. virilis*, sites that are conserved in all but the most distant *D. virilis* are gray. Sites for bantam- and miR-279-related miRNAs were also found in the nematode ortholog *egl-46* (blue and red).

(E) Part of a wing imaginal disc containing a clone of cells mutant for the *dicer1* gene (Lee et al., 2004). Mutant cells were labeled by the absence of β-galactosidase (i.e. absence of red) and showed elevated expression of a Tubulin-EGFP reporter transgene containing the *nerfin-1* 3'UTR (green). Error bars: standard deviation for 10 random controls.

Figure S3. Properties of Target and Antitarget 3'UTRs



Dissection of 3'UTR parameters that correlate with target and anti-target classification as in Figure 2. Shown are values for all genes in the best 50 target and anti-target categories, where obviously redundant GO categories were manually removed. (A) 3'UTR length; (B) target site density; (C) overall 3'UTR conservation considering only conserved nucleotides in blocks of at least 6; (D) selective site conservation (-log p values); (E) specific enrichment of conserved sites (CG: conserved genome analysis; -log p values; statistic intrinsically corrects for 3'UTR length); (F) specific avoidance of conserved sites, specifics as in (E); (G) GO categories shown in (A-F): (top) anti-target categories (left in A-F); (bottom) target categories (right in A-F). Note that the anti-target categories are sorted right to left, so that the best are adjacent to the best target categories.

Detecting Clusters of Functionally Related Target Genes

In light of the recent finding that an average miRNA has > 100 target genes, it seems likely that a given miRNAs regulates groups of genes that are related by a common function or expression pattern rather than a completely arbitrary collection of unrelated genes. Indeed, others and we previously presented evidence that a miRNA controls several genes involved in Notch signaling or many enzymes of a metabolic pathway (Lai et al., 2005; Stark et al., 2003). The set of genes regulated by a particular miRNA might thus contain global common features that are not obvious when considering a single gene. Thus, augmenting the lists of miRNA targets with additional information from other sources (e.g. functional annotations, or expression data), might suggest meaningful associations between them, and reveal more about the biological function of a miRNA then any individual predicted or validated target. For example, as miRNAs are negative regulators of gene expression, it is likely that they regulate genes whose expression would be detrimental for a given cell or tissue. In contrast, they probably avoid genes that are generally required in all cells or specifically required in a miRNA expressing tissue. We thus anticipated patterns of gene presence or absence from the target lists that would correlate with the expression patterns of certain tissues and in turn predict if the miRNA is expressed and/or has an important function in a particular tissue.

We searched our target predictions systematically for over- or under-representation of genes related by a common function or expression pattern (i.e. genes from one category) by three different, though complementary measures. The following paragraphs introduce these measures for category over-representation or enrichment (category under-representation or avoidance works identically, though in an inverted sense). First, we test whether a functional category is over-represented among the predicted target genes, i.e. whether the targets contain more genes of a certain category than is expected given the category's frequency in the 3'UTR database. Second, we assess if the UTRs of a certain category are enriched for certain miRNA binding sites, i.e. whether the UTRs contain more sites than expected given their length and conservation. Third we test if miRNA sites in the genes of a certain category are better conserved than average UTR sequences in that category.

1. Gene Enrichment

If genes that belong to a functional category are predominantly regulated, the genes of that category should be enriched among our predictions. Without specific enrichment each category should appear at the frequency with which it is represented in the 3'UTR database. We assess the significance of enrichment by calculating the probability (p value) that the enrichment is observed at random as the sum of binomial p values for all gene counts that are greater or equal to the observed count. This simple statistical model corresponds to randomly drawing terms associated with the genes from our 3'UTR database. For example, if we observe 15 neurogenesis genes among the 100 predicted targets, the significance (p value) is the likelihood of randomly drawing 100 genes from the database (containing 3.6% neurogenesis genes) and getting 15 or more such genes.

$$P = \sum_{Genes(Category)}^{Genes(All)} \binom{Genes(All)}{Genes(Category)} p^{Genes(Category)} \cdot (1-p)^{Genes(All)-Genes(Category)}$$

$$\text{with } p = \frac{Genes(Category, Database)}{Genes(All, Database)}$$

This statistic corrects for different database category frequencies by, for example, requiring more genes to be found for larger categories for the same significance level. It however assigns an equal *a priori* probability to all genes regardless of differences in UTR-lengths or conservation. Although genes with long and conserved UTRs are obviously more likely to be predicted as miRNA targets by our method (and others), we think it is important not to correct for this: Validated targets often have long and conserved UTRs and we show in this work that UTR length correlates with the number of sites for different miRNAs, which suggests that correcting for these features would be counterproductive. Instead, we first treat all genes equally and then investigate the basis for the presence or absence of target sites, i.e. the effects of UTR length and specific site enrichment.

2. Site Enrichment

We next test whether the 3'UTRs in a functional category are specifically enriched for miRNA target sites over what is expected, given the UTR length and degree of conservation. For this, we create

randomly shuffled miRNA sequences and adjusted their site count on the complete *D. melanogaster* UTR database (i.e. on a single genome) to be equal to the site count of the real miRNA. This allows us to assess differences in conservation between real and shuffled miRNAs (see below), but also differences in the number of sites for real and shuffled miRNAs in each gene category. For example, if 1/2 of all target sites for a given miRNA are found in the 3'UTRs of neurogenesis genes, whereas this ratio is only 1/10 for shuffled miRNAs, we conclude that neurogenesis genes are enriched for target sites of that specific miRNA. We calculated the fraction of the total number of target sites for shuffled miRNAs that are in each of the categories. This basically assesses the differences between the categories in conserved UTR sequence that is sampled by our prediction method, i.e. longer and more conserved categories contain more shuffled site matches. Indeed, the number of sites for shuffled miRNAs correlates linearly with the length of conserved UTR sequence (data not shown). We assess the significance of site enrichment for real miRNAs by calculating the binomial probability (p value) that the observed level of enrichment is random, where the ratios for the random shuffles (R) define the background probability for each category:

$$P = \sum_{Sites(Category)}^{Sites(All)} \binom{Sites(All)}{Sites(Category)} p^{Sites(Category)} \cdot (1-p)^{Sites(All)-Sites(Category)}$$

$$\text{with } p = \frac{R\_Sites(Category)}{R\_Sites(All)}$$

Another interpretation of this statistic is that it assesses the fraction of the (conserved) UTR sequence that corresponds to binding sites for a given miRNA. This statistic corrects for differences in UTR lengths and conservation and other possible dependencies of the prediction method, as it measures the differences in distribution between the functional categories for real and shuffled miRNAs using the same method and database. Also, as the statistics is independent of the overall number of sites and only scores the relative site distribution between the categories, it is applicable for both single genome and conserved genome analysis.

3. Seed Enrichment

We assess whether 3'UTRs in certain functional categories contain more or fewer 6mers complementary to miRNAs than expected given their lengths. For this, we calculated the frequency of all possible 6mers in all 3'UTRs. We then compare the frequency of individual 6mers or certain groups of 6mers (e.g. the 6mers complementary to the 5'ends of all miRNAs) to the corresponding frequencies obtained for 3'UTRs of a specific functional category. We assess the significance of seed enrichment by calculating the binomial probability (p value) that the observed occurance of specific seeds is random, where the frequency in all 3'UTRs define the background probability:

$$P = \sum_{Spec\_Seeds(Category)}^{All\_Seeds(Category)} \binom{All\_Seeds(Category)}{Spec\_Seeds(Category)} p^{Spec\_Seeds(Category)} \cdot (1-p)^{All\_Seeds(Category)-Spec\_Seeds(Category)}$$

$$\text{with } p = \frac{Specific\_Seeds(All\_Categories)}{All\_Seeds(All\_Categories)}$$

4. Selective Conservation

We also assess the conservation of miRNA binding sites, by measuring whether a UTR site that is complementary to a miRNA is better conserved than a randomly chosen UTR site, complementary to a shuffled miRNA (similar to Lewis et al., 2003). For this, we compare the fraction of miRNA-complementary sites that are conserved with the same fraction for shuffled miRNAs. We assess the significance of selective conservation by calculating the binomial probability (p value) that the observed number of conserved sites for real miRNAs can occur randomly, given the overall UTR conservation, where the conservation of sites for randomly shuffled miRNAs (R) defines the background probability (p) for each category:

$$P = \sum_{Sites(Cons)}^{Sites(Dmel)} \binom{Sites(Dmel)}{Sites(Cons)} p^{Sites(Cons)} \cdot (1-p)^{Sites(Dmel)-Sites(Cons)}$$

with $p = \frac{R\_Sites(Cons)}{R\_Sites(Dmel)}$

In this statistic, the conservation ratio for sites of shuffled miRNAs is an approximation of the overall UTR conservation (p=%average UTR conservation; especially as the average for several shuffled miRNAs is used). The comparison then assesses the fraction of conserved UTR sequence that is a miRNA binding site. Note that the average UTR conservation is the denominator in the above ratio and that its value is ≤ 1, where 1 corresponds to a 100% conserved UTR. Conserved sites in overall less-well conserved UTRs are therefore regarded as more significant than those in highly conserved UTRs.

This signal is based solely on the sites that are present in *D. melanogaster* and the extent to which they are conserved in *D. pseudoobscura*. The statistic does not take the overall number of possible sites in a UTR or in all UTRs of a functional category into account. This means that genes without sites do not weaken or dilute the signal and so do not influence the statistics. For the analysis of functional categories, this means that even in large categories, the signal can be derived from only very few genes or sites. For this reason we consider it questionable whether a good conservation signal for a small subset of genes can be used to extrapolate to the entire functional class. For example, if a functional class contains 1000 genes, of which 20 are matched by a specific miRNA in a single genome and (in the best case scenario) all 20 are conserved in additional genome(s), one can only draw conclusions about the relevance of the 20 genes but not extrapolate to all 1000: the 20 might be a very specific subset of the overall class and might be regulated by a miRNA whereas the majority of genes might not. Indeed, in our analysis, the highest signal for selective conservation was often associated with the entire set of genes rather than a specific functional class, but the conclusion that the given miRNA thus regulates all genes is obviously wrong. To assess regulation of an entire set of functionally related genes, we suggest to measure the significant overrepresentation of miRNA-target sites given all putative sites (i.e. the entire length of all UTRs) in that set or the overrepresentation of target genes given all genes in the UTR database.

p Value-Based Statistics to Assess Category Enrichment, Site Enrichment, or Selective Conservation

We chose to use p values to assess the significance of our observations as we think they reflect statistical significance (non-randomness) more accurately than *enrichment factors* or *signal-to-noise ratios*. These have been used before to detect functional clusters of targets (John et al., 2004) or target predictions and signals within certain functional categories (Lewis et al., 2005; Lewis et al., 2003). These factors describe the improvement over random quite intuitively (e.g. "5-fold enrichment"), but do not score non-randomness accurately, because they do not take the number of instances into account. The difference is illustrated by a dice experiment: the probability of throwing a six once is 1/6 and the enrichment factor of throwing only sixes (i.e. 100%) is thus 1/(1/6) = 6. The enrichment factor 6 is independent of whether the number of trials is one or ten, although the likelihood is vastly different with random chances of p=1/6=0.17 for throwing a six once or $p=1.6 \times 10^{-8}$ for throwing a six ten times in a row. In fact, throwing a six 3 out of 10 times, which has about the same probability than throwing a six once, has an enrichment factor of only 1.8. Consequently the enrichment factor correlates poorly with the true non-randomness (i.e. statistical significance) of the observations. It always overestimates the significance of smaller numbers and underestimates that of larger ones. If both cases are ranked within one list, insignificant predictions based on small numbers can outperform significant predictions based on larger numbers. Studies that use these ratios thus often have to introduce additional filters to remove small numbers.

Supplemental References

Aravin A.A., Lagos-Quintana M., Yalcin A., Zavolan M., Marks D., Snyder B., Gaasterland T., Meyer J., Tuschl T. (2003). The small RNA profile during Drosophila melanogaster development. Dev. Cell *5*, 337–350.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell *113*, 25–36.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of MicroRNA-Target Recognition. PLoS Biol. *3*, e85. 10.1371/journal.pbio.0030085.

Burgler, C., and Macdonald, P.M. (2005). Prediction and verification of microRNA targets by MovingTargets, a highly adaptable prediction method. BMC Genomics *6*, 88.

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

Grun, D., Wang, Y.L., Langenberger, D., Gunsalus, K.C., and Rajewsky, N. (2005). microRNA Target Predictions across Seven Drosophila Species and Comparison to Mammalian Targets. PLoS Comput. Biol. *1*, e13. 10.1371/journal.pcbi.0010013.

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

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., Macmenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., and Rajewsky, N. (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495–500.

Lai, E.C. (2004). Predicting and validating microRNA targets. Genome Biol. *5*, 115.

Lai, E.C., Tam, B., and Rubin, G.M. (2005). Pervasive regulation of Drosophila Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. Genes Dev.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell *120*, 15–20.

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

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

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

Robins, H., Li, Y., and Padgett, R.W. (2005). Incorporating structure to predict microRNA targets. Proc. Natl. Acad. Sci. USA *102*, 4006–4009.

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila MicroRNA Targets. PLoS Biol. *1*, e60. 10.1371/journal.pbio.0000060.

RESEARCH COMMUNICATION

A single Hox locus in *Drosophila* produces functional microRNAs from opposite DNA strands

Alexander Stark,[1,2,6,8] Natascha Bushati,[3,6] Calvin H. Jan,[4] Pouya Kheradpour,[1,2] Emily Hodges,[5] Julius Brennecke,[5] David P. Bartel,[4] Stephen M. Cohen,[3,7] and Manolis Kellis[1,9]

[1]Broad Institute of Massachussetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02141, USA; [2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; [3]European Molecular Biology Laboratory, 69117 Heidelberg, Germany; [4]Department of Biology, Howard Hughes Medical Institute and Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology Cambridge, Massachusetts 02139, USA; [5]Watson School of Biological Sciences and Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

MicroRNAs (miRNAs) are ~22-nucleotide RNAs that are processed from characteristic precursor hairpins and pair to sites in messages of protein-coding genes to direct post-transcriptional repression. Here, we report that the miRNA *iab-4* locus in the *Drosophila* Hox cluster is transcribed convergently from both DNA strands, giving rise to two distinct functional miRNAs. Both sense and antisense miRNA products target neighboring Hox genes via highly conserved sites, leading to homeotic transformations when ectopically expressed. We also report sense/antisense miRNAs in mouse and find antisense transcripts close to many miRNAs in both flies and mammals, suggesting that additional sense/antisense pairs exist.

Supplemental material is available at http://www.genesdev.org.

Received September 6, 2007; revised version accepted November 2, 2007.

Hox genes are highly conserved homeobox-containing transcription factors crucial for development in animals (Lewis 1978; for reviews, see McGinnis and Krumlauf 1992; Pearson et al. 2005). Genetic analyses have identified them as determinants of segmental identity that specify morphological diversity along the anteroposterior body axis. A striking conserved feature of Hox complexes is the spatial colinearity between Hox gene transcription in the embryo and the order of the genes along the chromosome (Duboule 1998). Hox clusters also give rise to a variety of noncoding transcripts, including microRNAs (miRNAs) *mir-10* and *mir-iab-4/mir-196*, which derive from analogous positions in Hox clusters in flies and vertebrates (Yekta et al. 2004). miRNAs are ~22-nucleotide (nt) RNAs that regulate gene expression post-transcriptionally (Bartel 2004). They are transcribed as longer precursors and processed from characteristic pre-miRNA hairpins. In particular, Hox miRNAs have been shown to regulate Hox protein-coding genes by mRNA cleavage and inhibition of translation, thereby contributing to the extensive regulatory connections within Hox clusters (Mansfield et al. 2004; Yekta et al. 2004; Hornstein et al. 2005; Ronshaugen et al. 2005). Several Hox transcripts overlap on opposite strands, providing evidence of extensive antisense transcription, including antisense transcripts for *mir-iab-4* in flies (Bae et al. 2002) and its mammalian equivalent *mir-196* (Mainguy et al. 2007). However, the function of these transcripts has been elusive. Here we show that the *iab4* locus in *Drosophila* produces miRNAs from opposite DNA strands that can regulate neighboring Hox genes via highly conserved sites. We provide evidence that such sense/antisense miRNA pairs are likely employed in other contexts and a wide range of species.

[*Keywords*: *Drosophila*; miR-iab-4; Hox; antisense miRNAs]
[6]This authors contributed equally to this work.
[7]Present address: Temasek Life Sciences Laboratory, The National University of Singapore, Singapore 117604.
Corresponding authors.
[8]E-MAIL alex.stark@mit.edu; FAX (617) 253-7512.
[9]E-MAIL manoli@mit.edu; FAX (617) 253-7512.
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.1613108.

Results and Discussion

Our examination of the antisense transcript that overlaps *Drosophila mir-iab-4* revealed that the reverse complement of the *mir-iab-4* hairpin folds into a hairpin reminiscent of miRNA precursors (Fig. 1A). Moreover, 17 sequencing reads from small RNA libraries of *Drosophila* testes and ovaries mapped uniquely to one arm of the *iab-4* antisense hairpin (Fig. 1B). All reads were aligned at their 5′ end, suggesting that the *mir-iab-4* antisense hairpin is processed into a single mature miRNA in vivo, which we refer to as miR-iab-4AS. For comparison, we found six reads consistent with the known miR-iab-4-5p (or miR-iab-4 for short) and one read for its star sequence (miR-iab-4-3p). Interestingly, the relative abundance of mature miRNAs and star sequences for *mir-iab-4AS* (17:0) and *mir-iab-4* (6:1) reflects the thermodynamic asymmetry of the predicted miRNA/miRNA* duplexes (Khvorova et al. 2003; Schwarz et al. 2003). Because they derived from complementary near palindromes, miR-iab-4 and miR-iab-4AS had high sequence similarity, only differing in four positions at the 3′ region (Fig. 1B). However, they differed in their 5′ ends, which largely determine miRNA target spectra (Brennecke et al. 2005; Lewis et al. 2005): miR-iab-4AS was shifted by 2 nt, suggesting targeting properties distinct from those of miR-iab-4 and other known *Drosophila* miRNAs.

We confirmed robust transcription of *mir-iab-4* sense and antisense precursors by in situ hybridization to *Drosophila* embryos (Fig. 1C). Both transcripts were detected in abdominal segments in the posterior part of the embryo, but intriguingly in nonoverlapping domains. As described previously (Bae et al. 2002; Ronshaugen et al. 2005), *mir-iab-4* sense was expressed highly in abdominal segments A5–A7, showing modulation in levels within the segments: *abdominal-A* (*abd-A*)-expressing cells (Fig. 1D; Karch et al. 1990; Macias et al. 1990) ap-

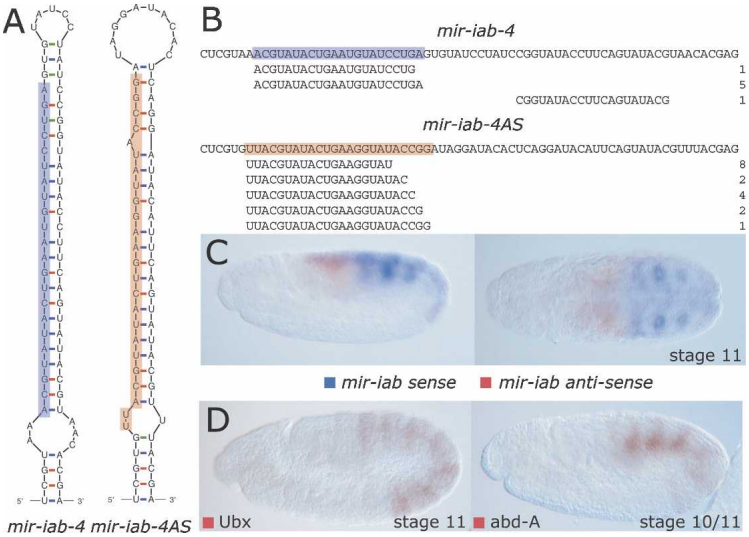


Figure 1. *Drosophila iab-4* contains sense and antisense miRNAs. (*A*) *mir-iab-4* sense and antisense sequences can adopt fold-back stem–loop structures characteristic for miRNA precursors (structure predictions by Mfold [Zuker 2003]; mature miRNAs shaded in blue [miR-iab-4] and red [miR-iab-4AS]). (*B*) Solexa sequencing reads that uniquely align to the *mir-iab-4* hairpin sequence (*top*) or its reverse complement (*bottom*; numbers on the *right* indicate the cloning frequency for each sequence). The mature miRNAs have very similar sequences that are shifted by 2 nt and are different in only four additional positions. (*C*) Expression of primary transcripts for *mir-iab-4* (blue) and *mir-iab-4AS* (red) in nonoverlapping abdominal segments determined by in situ hybridization (lateral [*left* panel] and dorsal [*right* panel] view of embryonic stage 11, anterior is to the left). (*D*) Lateral views of stage 10/11 embryos in which *Ubx* and *abd-A* proteins are visualized (anterior is to the left, and dorsal is upwards).

peared to have more *mir-iab-4*, whereas *Ultrabithorax* (*Ubx*)-positive cells appeared to have little or none (Fig. 1D; Ronshaugen et al. 2005). In contrast, *mir-iab-4AS* transcription was detected in the segments A8 and A9, where *Abdominal-B* (*Abd-B*) is known to be expressed (Fig. 1C; Yoder and Carroll 2006). Primary transcripts for *mir-iab-4* and *mir-iab-4AS* were also detected by strand-specific RT–PCR in larvae, pupae, and male and female adult flies (Supplemental Fig. S1), suggesting that both miRNAs are expressed throughout fly development.

To assess the possible biological roles of the two *iab-4* miRNAs, we examined fly genes for potential target sites by searching for conserved matches to the seed region of the miRNAs (Lewis et al. 2005). We found highly conserved target sites for miR-iab-4AS in the 3′ untranslated regions (UTRs) of several Hox genes that are proximal to the *iab-4* locus and are expressed in the neighboring more anterior embryonic segments: *abd-A*, *Ubx*, and *Antennapedia* (*Antp*) have four, five, and two seed sites, respectively, most of which are conserved across 12 *Drosophila* species that diverged 40 million years ago (Fig. 2A; Supplemental Fig. S2; *Drosophila* 12 Genomes Consortium 2007; Stark et al. 2007a). More than two highly conserved sites for one miRNA is exceptional for fly 3′ UTRs, placing these messages among the most confidently predicted miRNA targets and suggesting that they might be particularly responsive to the presence of the miRNA. The strong predicted targeting of proximal Hox genes was reminiscent of previously characterized miR-iab-4 targeting of *Ubx* in flies and miR-196 targeting of *HoxB8* in vertebrates (Mansfield et al. 2004; Yekta et al. 2004; Hornstein et al. 2005; Ronshaugen et al. 2005).

To test whether miR-iab4AS is functional and can directly target *abd-A* and *Ubx*, we constructed Luciferase reporters carrying the corresponding wild-type 3′ UTRs and control 3′ UTRs in which each seed site was disrupted by point substitutions. *mir-iab-4AS* potently repressed reporter activity for *abd-A* and *Ubx* (Fig. 2B). This repression was specific to the miR-iab-4AS seed sites, as expression of the control reporters with mutated sites was not affected. We also tested whether *mir-iab-4AS* reduced expression of a Luciferase reporter with the *Abd-B* 3′ UTR, which has no seed sites. As expected, *mir-iab-4AS* expression did not affect reporter activity, consistent with a model where miRNAs do not target genes that are coexpressed at high levels (Farh et al. 2005; Stark et al. 2005). In addition to demonstrating specific repression dependent on the predicted target sites, these assays confirmed the processing of the *mir-iab-4AS* hairpin into a functional mature miRNA.

If miR-iab-4AS were able to potently down-regulate

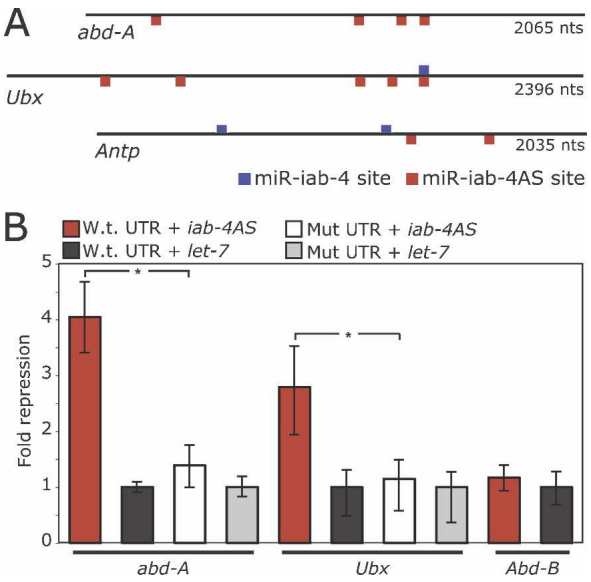


Figure 2. miR-iab-4AS targets neighboring Hox genes. (*A*) miR-iab-4AS has five 3′ UTR seed sites (red) in *Ubx*, four in *abd-A*, and two in *Antp* of which three, four, and one are conserved across 12 *Drosophila* species, respectively (Supplemental Fig. S2). miR-iab-4 has one 3′ UTR seed site (blue) in *Ubx* and two in Antp, while *abd-A* has no such sites. (*B*) miR-iab-4AS mediates repression of luciferase reporters through complementary seed sites in 3′ UTRs from *abd-A* and *Ubx*, but not *Abd-B* (*Antp* was not tested). Luciferase activity in S2 cells cotransfected with plasmid expressing the indicated miRNA with either wild-type luciferase reporters or mutant reporters bearing a single point mutation in the seed. Bars represent geometric means from 16 replicates, normalized to the transfection control and noncognate miRNA control (*let-7*; see Materials and Methods). Error bars represent the fourth largest and smallest values from 16 replicates ([*] $P < 0.0001$, Wilcoxon rank-sum test).

Ubx in the fly, its misexpression should result in a *Ubx* loss-of-function phenotype, a line of reasoning that has often been used to study the functions and regulatory relationships of Hox genes. *Ubx* is expressed throughout the haltere imaginal disc, where it represses wing-specific genes and specifies haltere identity (Weatherbee et al. 1998). When we expressed *mir-iab-4AS* in the haltere imaginal disc under *bx-Gal4* control, a clear homeotic transformation of halteres to wings was observed (Fig. 3). The halteres developed sense organs characteristic of the wing margin and their size increased severalfold, features typical of transformation to wing (Weatherbee et al. 1998). Consistent with the increased number of miR-iab4AS target sites, the transformation was stronger than that reported for expression of *iab-4* (Ronshaugen et al. 2005), for which we confirmed changes in morphology but did not find wing-like growth (Fig. 3D).

We conclude that both strands of the *iab-4* locus are expressed in nonoverlapping embryonic domains and that each transcript produces a functional miRNA in vivo. In particular, the novel *mir-iab-4AS* is able to strongly down-regulate neighboring Hox genes. Interestingly, vertebrate *mir-196*, which lies at an analogous position in the vertebrate Hox clusters, is transcribed in the same direction as *mir-iab-4AS* and most other Hox genes, and targets homologs of both *abd-A* and *Ubx* (Mansfield et al. 2004; Yekta et al. 2004; Hornstein et al. 2005). With its shared transcriptional orientation and homologous targets, *mir-iab-4AS* appears to be the functional equivalent of *mir-196*.

The expression patterns and regulatory connections between Hox genes and the two *iab-4* miRNAs show an intriguing pattern in which the miRNAs appear to reinforce Hox gene-mediated transcriptional regulation (Fig. 4A). In particular, miR-iab-4AS would reinforce the posterior expression boundary of *abd-A*, *Ubx*, and *Antp*, supporting their transcriptional repression by *Abd-B*. *mir-iab-4* appears to support *abd-A*- and *Abd-B*-mediated repression of *Ubx*, reinforcing the *abd-A*/*Ubx* expression domains and the posterior boundary of *Ubx* expression. Furthermore, both *iab-4* miRNAs have conserved target sites in *Antp*, which is also repressed by *Abd-B*, *abd-A*, and *Ubx*. The *iab-4* miRNAs thus appear to support the established regulatory hierarchy among Hox transcription factors, which exhibits "posterior prevalence," in that more posterior Hox genes repress more anterior ones and are dominant in specifying segment identity (for reviews, see McGinnis and Krumlauf 1992; Pearson et al. 2005). Interestingly, *Abd-B* and *mir-iab-4AS* are expressed in the same segments, and the majority of *cis*-regulatory elements controlling *Abd-B* expression are located 3′ of *Abd-B* (Boulet et al. 1991). This places them near the inferred transcription start of *mir-iab-4AS*, where they potentially direct the coexpression of these genes. Similarly, *abd-A* and *mir-iab-4* may be coregulated as both are transcribed divergently, potentially under the control of shared upstream elements.

Our data demonstrate the transcription and processing of sense and antisense *mir-iab-4* into functional miRNAs with highly conserved functional target sites in neighboring Hox genes. In an accompanying study (Bender 2008), genetic and molecular analyses in *mir-iab-4* mutant *Drosophila* revealed that the proposed regulation of *Ubx* by both sense and antisense miRNAs occurs under physiological conditions and, in particular, the regulation by miR-iab-4AS is required for normal development. These lines of evidence establish miR-iab-4AS as a novel Hox gene, being expressed from within the Hox cluster and regulating Hox genes during development.

The genomic arrangement of two miRNAs that are expressed from the same locus but on different strands

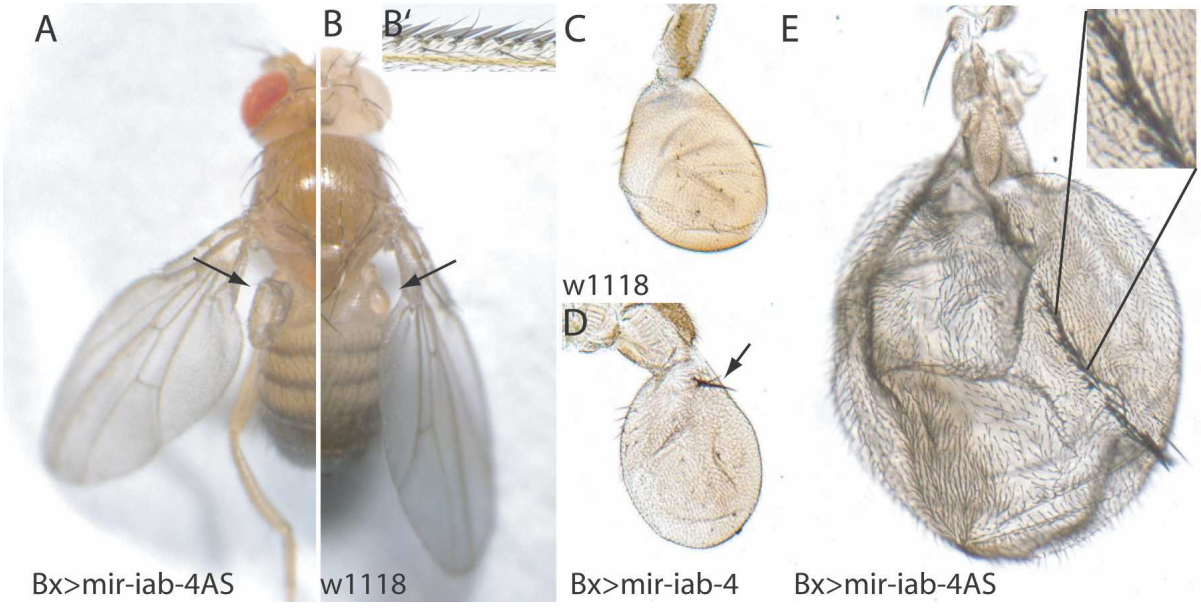


Figure 3. Misexpression of miR-iab-4AS transforms halteres to wings. (*A*,*B*) Overview of an adult wild-type *Drosophila* (*B*) and an adult expressing *mir-iab-4AS* using *bx*-Gal4 (*A*). The halteres, balancing organs of the third thoracic segment, are indicated by arrows. (*C*) Wild-type haltere. (*D*) Expression of *mir-iab-4* using *bx-Gal4* induces a mild haltere-to-wing transformation. Sensory bristles characteristic of wild-type wing margins (shown in *B*′) are indicated by an arrow. (*E*) Expression of *mir-iab-4AS* using *bx-Gal4* induces a strong haltere-to-wing transformation, displaying the triple row of sensory bristles (*inset*) normally seen in wild-type wings (shown in *B*′). Note that *C*–*E* are at the same magnification.

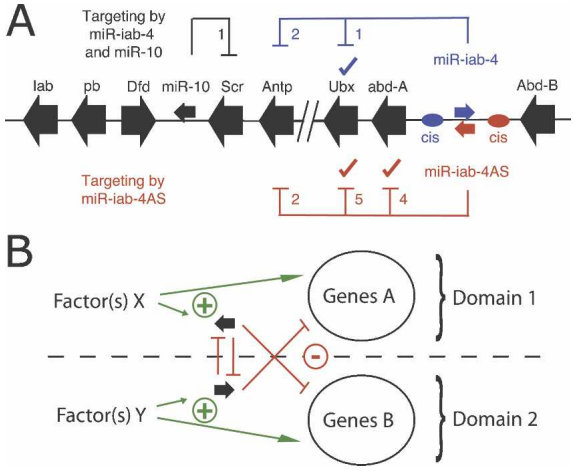


Figure 4. Regulation of gene expression by antisense miRNAs. (*A*) miRNA-mediated control in the *Drosophila* Hox cluster. Schematic representation of the *Drosophila* Hox cluster (Antennapedia and Bithorax complex) with miRNA target interactions (check marks represent experimentally validated targets). miR-iab-4 (blue) and miR-iab-4AS (red) target anterior neighboring Hox genes and miR-10 (black) targets posterior *Sex-combs-reduced* (*Scr*) (Brennecke et al. 2005). *abd-A* and *mir-iab-4* and *Abd-B* and *mir-iab-4AS* might be coregulated from shared control elements (*cis*). Note that *mir-iab-4AS* is expressed in the same direction as most other Hox genes and its mammalian equivalent, *mir-196*. (*B*) General model for defining different expression domains with pairs of antisense miRNAs (black). Different transcription factor(s) activate the transcription of miRNAs and genes in each of the two domains separately (green lines). Both miRNAs might inhibit each other by transcriptional interference or post-transcriptionally (vertical red lines), leading to essentially nonoverlapping expression and activity of both miRNAs. Further, both miRNAs likely target distinct sets of genes (diagonal red lines), potentially re-enforcing the difference between the two expression domains.

might provide a simple and efficient means to create nonoverlapping miRNA expression domains (Fig. 4B). Such sense/antisense miRNAs could restrict each other's transcription, either by direct transcriptional interference, as shown for overlapping convergently transcribed genes (Shearwin et al. 2005; Hongay et al. 2006), or post-transcriptionally, possibly via RNA–RNA duplexes formed by the complementary transcripts. Sense/antisense miRNAs would usually differ at their 5′ ends and thereby target distinct sets of genes, which might help define and establish sharp boundaries between expression domains. Coupled with feedback loops or coregulation of miRNAs and genes in *cis* or *trans*, this arrangement could provide a powerful regulatory switch. The *iab-4* miRNAs might be a special case of tight regulatory integration in which miRNAs and proximal genes appear coregulated transcriptionally in *cis* and repress each other both transcriptionally and post-transcriptionally.

It is perhaps surprising that no antisense miRNA had been found previously, even though, for example, the intriguing expression pattern of the *iab-4* transcripts had been reported nearly two decades ago (Cumberledge et al. 1990; Bae et al. 2002), and *iab-4* lies in one of the most extensively studied regions of the *Drosophila* genome. The frequent occurrence of antisense transcripts (Yelin et al. 2003; Katayama et al. 2005) suggests that more antisense miRNAs might exist. Indeed, up to 13% of known *Drosophila*, 20% of mouse, and 31% of human miRNAs are located in introns of host genes transcribed on the opposite strand or are within 50 nt of antisense ESTs or cDNAs (Supplemental Table S1). These include an antisense transcript overlapping human *mir-196* (see also Mainguy et al. 2007). However, because of the contribution of noncanonical base pairs, particularly G:U pairs that become less favorable A:C in the antisense strand, many miRNA antisense transcripts will not fold into hairpin structures suitable for miRNA biogenesis, which explains the propensity of miRNA gene predictions to identify the correct strand (Lim et al. 2003). Nonetheless, in a recent prediction effort, 22 sequences reverse-complementary to known *Drosophila* miRNAs showed scores seemingly compatible with miRNA processing (Stark et al. 2007b). Deep sequencing of small RNA libraries from *Drosophila* confirmed the processing of small RNAs from four of these high-scoring antisense candidates (Ruby et al. 2007), and the ovary/testes libraries used here showed antisense reads for an additional *Drosophila* miRNA (*mir-312*) (see Supplemental Tables S2, S3). In addition, using high-throughput sequencing of small RNA libraries from mice, we found sequencing reads that uniquely matched the mouse genome in loci antisense to 10 annotated mouse miRNAs. Eight of the inferred antisense miRNAs were supported by multiple independent reads, and two of them had reads from both the mature miRNA and the star sequence (Supplemental Table S2). These results suggest that sense/antisense miRNAs could be more generally employed in diverse contexts and in species as divergent as flies and mammals.

Materials and methods

Plasmids

3′ UTRs were amplified from *Drosophila melanogaster* genomic DNA and cloned in pCR2.1 for site-directed mutagenesis. The following primer pairs were used to amplify the indicated 3′ UTR: abd-A (tctagaGCGGTCAGCAAAGTCAACTC; gtcgacATGGATGGGTTCTCGTTGCAG), Ubx (tctagaATCCTTAGATCCTTAGATCCTTAG; ctcgagATGGTTTGAATTTCCACTGA), and Abd-B (tctagaGCCACCACCTGAACCTTAG; aactcgagCGGAGTAATGCGAAGTAATTG). QuickChange multisite-directed mutagenesis was used to mutate all miR-iab-4AS seed sites from ATACGT to ATAGGT, per the manufacturer's directions (Stratagene). Wild-type and mutated 3′ UTRs were subcloned into pCJ40 between SacI and NotI sites to make *Renilla* luciferase reporters. Plasmid pCJ71 contains the abd-A wild-type 3′ UTR, pCJ72 contains the Ubx wild-type 3′ UTR, pCJ74 contains the Abd-B wild-type 3′ UTR, pCJ75 contains the abd-A mutated 3′ UTR, and pCJ76 contains the Ubx mutated 3′ UTR fused to *Renilla* luciferase. The control *let-7* expression vector was obtained by amplifying *let-7* from genomic DNA with primers 474 base pairs (bp) upstream of and 310 bp downstream from the let-7 hairpin and cloning it into pMT-puro. To express miR-iab-4 and miR-iab-4AS, a 430-bp genomic fragment containing the miR-iab-4 hairpin was cloned, in either direction, downstream from the tubulin promoter as described in Stark et al. (2005). For the UAS-miR-iab-4 and UAS-miR-iab-4AS constructs, the same 430-bp genomic fragment containing the miR-iab-4 hairpin was cloned downstream from pUAST-DSred2 (Stark et al. 2003) in either direction.

Reporter assays

For the luciferase assays, 2 ng of p2129 (firefly luciferase), 4 ng of *Renilla* reporter, 48 ng of miRNA expression plasmid, and 48 ng of p2032 (GFP) were cotransfected with 0.3 µL Fugene HD per well of a 96-well plate. Twenty-four hours after transfection, expression of *Renilla* luciferase was induced by addition of 500 µM $CuSO_4$ to the culture media. Twenty-four hours after induction, reporter activity was measured with the Dual-Glo luciferase kit (Promega), per the manufacturer's instructions on a Tecan Safire II plate reader.

The ratio of *Renilla*:firefly luciferase activity was measured for each well. To calculate fold repression, the ratio of *Renilla*:firefly for reporters cotransfected with *let-7* was set to 1. The Wilcoxon rank-sum test was used to assess the significance of changes in fold repression of wild-type reporters compared with mutant reporters. Geometric means from 16 transfections representing four replicates of four independent transfections are shown. Error bars represent the fourth highest and lowest values of each set.

Drosophila *strains*

UAS-miR-iab-4 and *UAS-miR-iab-4AS* flies were generated by injection of the corresponding plasmids into w^{1118} embryos. bx^{MS1096}*-GAL4* flies were obtained from the Bloomington Stock Center.

In situ hybridization and protein stainings

Double in situ hybridization for the miRNA primary transcripts was performed as described in Stark et al. (2005). Probes were generated using PCR on genomic DNA with primers TCAGAGCATGCAGAGACAT AAAG, TTGTAGATTGAAATCGGACACG for iab-4 sense and ATTT TACTGGGTGTCTGGGAAAG, TAGAAACTGAGACGGAGAAGCAG for iab-4 antisense. Protein stainings were performed as described in Patel (1994). Antibodies used were mouse anti-Ubx (1:30), mouse anti-abd-A (1:5), and HRP-conjugated goat anti-mouse (Dianova, 1:3000).

RT–PCRs

Total RNA was isolated using Trizol (Invitrogen), treated with RQI DNase (Promega), and used for strand-specific cDNA synthesis with SuperScript III (Invitrogen). Primers for cDNA synthesis were CATATAA CAAAGTGCTACGTG (iab-4 sense) and CTTTATCTGCATTTG GATCCG (iab-4 antisense). Both primers were used for subsequent amplification.

Small library sequencing

Drosophila small RNAs were cloned from adult ovaries and testes as described previously (Brennecke et al. 2007) and sequenced using Solexa sequencing. A total of 657,251 sequencing reads uniquely matched known *Drosophila* miRNAs (Rfam release 9.2), and the 69 miRNAs with unique matches had 1011 matches on average (Stark et al. 2007b). Two miRNAs had unique matches to the antisense hairpin (Supplemental Tables S2, S3). Mouse small RNAs were cloned from wild-type and c-kit mutant ovaries (Supplemental Table S4; G. Hannon, pers. comm.) and from Comma-Dbgeo cells, a murine mammary epithelial cell line (Ibarra et al. 2007), and were sequenced using Solexa sequencing. A total of 4,217,883 reads uniquely matched known mouse miRNAs (Rfam release 9.2), and the 286 miRNAs with unique reads showed 256 reads on average. Sequencing reads matching to the plus and minus strand of known mouse miRNAs with antisense reads are listed in Supplemental Table S3.

Multiple sequence alignments and target site prediction

The multiple sequence alignments for the indicated Hox 3′ UTRs were obtained from the University of California at Santa Cruz (UCSC) genome browser (Kent et al. 2002) and were slightly manually adjusted. We predicted target sites according to Lewis et al. (2005) by searching for 3′ UTR seed sites (reverse-complementary to miRNA positions 2–8 or matching to "A" + reverse complement of miRNA positions 2–7).

Antisense transcripts near known miRNAs

To assess the fraction of *Drosophila*, human, and mouse miRNAs that are also putatively transcribed on both strands and might give rise to antisense miRNAs, we determined the number of miRNAs that are near known transcripts on the opposite strand. We obtained the coordinates of all introns of protein-coding genes and all mapped ESTs or cDNAs for the three species from the UCSC genome browser (Kent et al. 2002). We intersected them with the miRNA coordinates from Rfam (release 9.2; Griffiths-Jones et al. 2006), requiring miRNAs and transcripts to be on opposite strands and at a distance of at most 50 nt. For each miRNA, we recorded the number of antisense transcripts and their identifiers. Note that some of the transcripts might have been mapped to more than one place in the genome, such that the intersection represents an upper estimate based on the currently known transcripts.

Acknowledgments

We thank Greg Hannon for providing Solexa sequencing data and support, Juerg Mueller for the anti-Ubx antibody, Thomas Sandmann for *Drosophila* embryos, and Sandra Mueller for preparing transgenic flies. We thank the *Drosophila* genome sequencing centers and the UCSC genome browser for access to the 12 *Drosophila* multiple sequence alignments prior to publication, and Welcome Bender for sharing data prior to publication. A.S. was partly supported by a post-doctoral fellowship from the Schering AG and partly by a post-doctoral fellowship from the Human Frontier Science Program Organization (HFSPO). C.H.J. is an NSF graduate fellow. J.B. thanks the Schering AG for a post-doctoral fellowship. This work was also partially supported by a grant from the NIH.

References

Bae, E., Calhoun, V.C., Levine, M., Lewis, E.B., and Drewell, R.A. 2002. Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci.* **99:** 16847–16852.

Bartel, D.P. 2004. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **116:** 281–297.

Bender, W. 2008. MicroRNAs in the *Drosophila* bithorax complex. *Genes & Dev.* (this issue), doi: 10.1101/gad.1614208.

Boulet, A.M., Lloyd, A., and Sakonju, S. 1991. Molecular definition of the morphogenetic and regulatory functions and the *cis*-regulatory elements of the *Drosophila* Abd-B homeotic gene. *Development* **111:** 393–405.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. 2005. Principles of microRNA-target recognition. *PLoS Biol.* **3:** e85. doi: 10.1371/journal.pbio.0030085.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128:** 1089–1103.

Cumberledge, S., Zaratzian, A., and Sakonju, S. 1990. Characterization of two RNAs transcribed from the *cis*-regulatory region of the abd-A domain within the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci.* **87:** 3259–3263.

Drosophila 12 Genomes Consortium 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450:** 203–218.

Duboule, D. 1998. Vertebrate hox gene regulation: Clustering and/or colinearity? *Curr. Opin. Genet. Dev.* **8:** 514–518.

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

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. 2006. miRBase: MicroRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34 (Database issue):** D140–D144. doi: 10.1093/nar/gkj112.

Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. 2006. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127:** 735–745.

Hornstein, E., Mansfield, J.H., Yekta, S., Hu, J.K., Harfe, B.D., McManus, M.T., Baskerville, S., Bartel, D.P., and Tabin, C.J. 2005. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature* **438:** 671–674.

Ibarra, I., Erlich, Y., Muthuswamy, S.K., Sachidanandam, R., and Hannon, G.J. 2007. A microRNA fingerprint of mammary epithelial stem cells. *Genes & Dev.* **21:** 3238–3243.

Karch, F., Bender, W., and Weiffenbach, B. 1990. abdA expression in *Drosophila* embryos. *Genes & Dev.* **4:** 1573–1587.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al. 2005. Antisense transcription in the mammalian transcriptome. *Science* **309:** 1564–1566.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. 2002. The human genome browser at UCSC. *Genome Res.* **12:** 996–1006.

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

Lewis, E.B. 1978. A gene complex controlling segmentation in *Dro-*

sophila. *Nature* **276:** 565–570.
Lewis, B.P., Burge, C.B., and Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120:** 15–20.
Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. 2003. The microRNAs of *Caenorhabditis elegans*. *Genes & Dev.* **17:** 991–1008.
Macias, A., Casanova, J., and Morata, G. 1990. Expression and regulation of the abd-A gene of *Drosophila*. *Development* **110:** 1197–1207.
Mainguy, G., Koster, J., Woltering, J., Jansen, H., and Durston, A. 2007. Extensive polycistronism and antisense transcription in the mammalian hox clusters. *PLoS ONE* **2:** e356. doi: 10.1371/journal.pone.0000356.
Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G., et al. 2004. MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat. Genet.* **36:** 1079–1083.
McGinnis, W. and Krumlauf, R. 1992. Homeobox genes and axial patterning. *Cell* **68:** 283–302.
Patel, N.H. 1994. Imaging neuronal subsets and other cell types in whole-mount *Drosophila* embryos and larvae using antibody probes. *Methods Cell Biol.* **44:** 445–487.
Pearson, J.C., Lemons, D., and McGinnis, W. 2005. Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **6:** 893–904.
Ronshaugen, M., Biemar, F., Piel, J., Levine, M., and Lai, E.C. 2005. The *Drosophila* microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes & Dev.* **19:** 2947–2952.
Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., and Lai, E.C. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* doi: 10.1101/gr.6597907.
Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115:** 199–208.
Shearwin, K.E., Callen, B.P., and Egan, J.B. 2005. Transcriptional interference—A crash course. *Trends Genet.* **21:** 339–345.
Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. 2003. Identification of *Drosophila* microRNA targets. *PLoS Biol.* **1:** E60. doi: 10.1371/journal.pbio.0000060.
Stark, A., Brennecke, J., Bushati, N., Russell, R.B., and Cohen, S.M. 2005. Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* **123:** 1133–1146.
Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N., et al. 2007a. Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450:** 219–232.
Stark, A., Kheradpour, P., Parts, L., Brennecke, J., Hodges, E., Hannon, G.J., and Kellis, M. 2007b. Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res.* doi: 10.1101/gr.6593807.
Weatherbee, S.D., Halder, G., Kim, J., Hudson, A., and Carroll, S. 1998. Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the *Drosophila* haltere. *Genes & Dev.* **12:** 1474–1482.
Yekta, S., Shih, I.H., and Bartel, D.P. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304:** 594–596.
Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., et al. 2003. Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* **21:** 379–386.
Yoder, J.H. and Carroll, S.B. 2006. The evolution of abdominal reduction and the recent origin of distinct Abdominal-B transcript classes in *Diptera*. *Evol. Dev.* **8:** 241–251.
Zuker, M. 2003. Mfold Web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31:** 3406–3415.

A single Hox locus in Drosophila produces functional microRNAs from opposite DNA strands

Supplementary Data

Figure S1

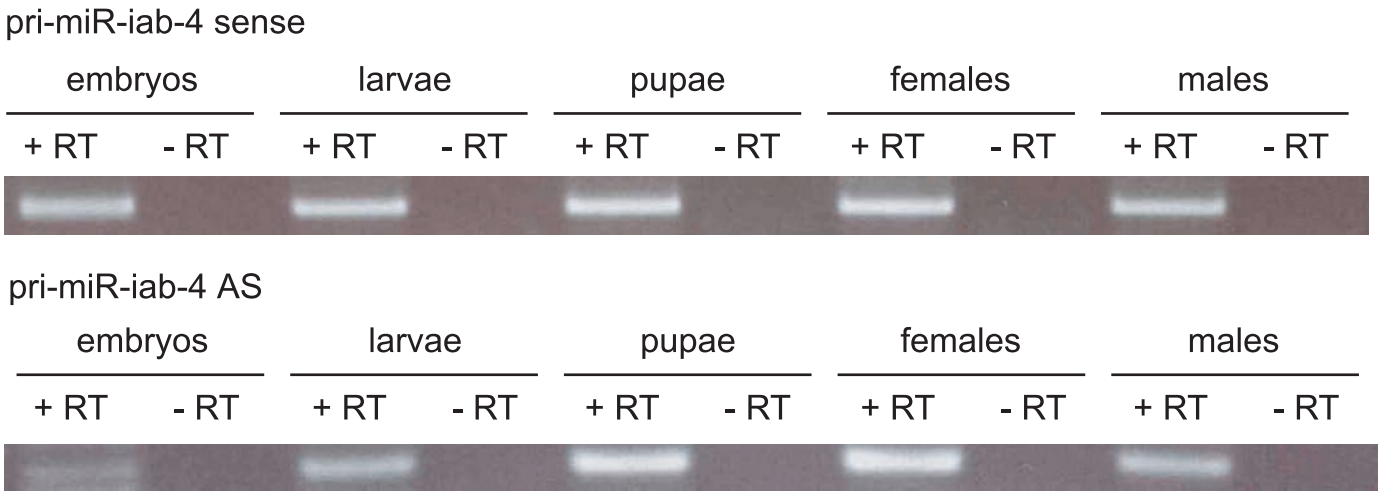


Figure S1: miR-iab-4 sense and anti-sense are expressed throughout fly development. Shown are strand-specific RT-PCR products (neg. control without RT) for the miR-iab-4 sense and anti-sense primary transcripts for different stages of fly development.

Figure S2

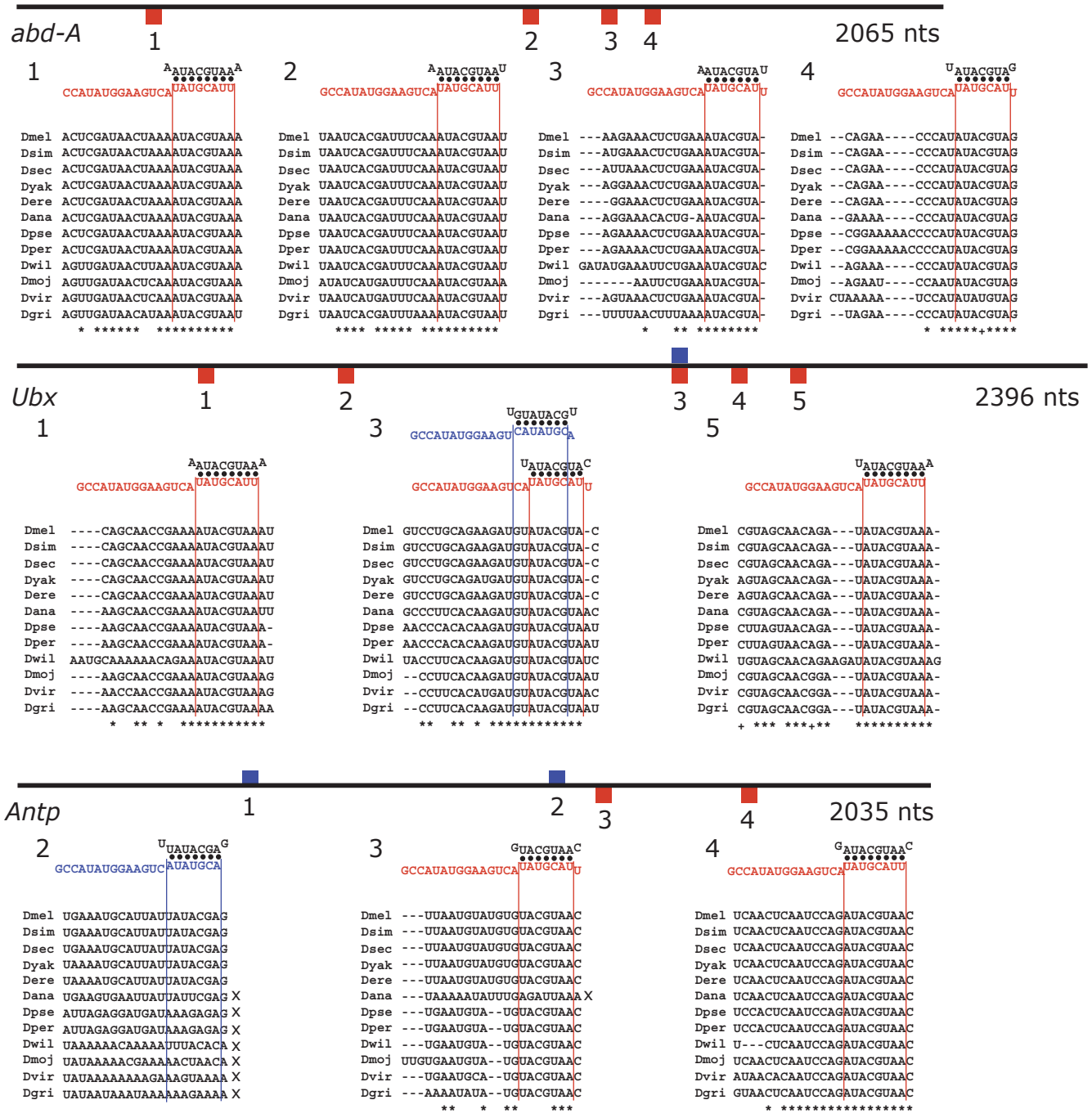


Figure S2. miR-iab-4AS target sites the *Hox* genes *abd-A*, *Ubx*, and *Antp*
miR-iab-4AS has 4 seed sites (nts 2-8 or A+2-7; Lewis et al., 2005; red) in *abd-A, 5 in Ubx,* and 2 in *Antp* of which 4, 3, and 1 are fully conserved across 12 Drosophila species (*Drosophila melanogaster, simulans, sechellia, yakuba, erecta, ananassae, pseudoobscura, persimilis, willistoni, mojavensis, virilis, grimshawi*), respectively. 1 site in the 3'UTR of *Ubx* shows extended complementarity such that it also constitutes a seed site for miR-iab-4 (blue). *Antp* has 2 sites for miR-iab-4 that are distinct from the miR-iab-4AS sites, while *abd-A* has no such sites. * denotes fully conserved sequence positions, and x denotes species where the site is not conserved.

Table S1: Anti-sense transcripts near known miRNAs

Drosophila miRNAs

miRNA	#transcripts	Transcript IDs
FlyBase Noncoding Genes		
dme-mir-281-1	1	CR33594
dme-mir-281-2	1	CR33596
ESTs		
dme-mir-277	2	EC255080,CO344262
dme-mir-2a-1	1	CO341270
dme-mir-284	1	CK135092
dme-mir-33	1	BI631776
dme-mir-2a-2	6	CO341270,EC253459,EC253459,CO341270,CO341270,CO341270
Introns		
dme-mir-289, dme-mir-31b, dme-mir-274		

Table S1: Anti-sense transcripts near known miRNAs (continued)

Human miRNAs

miRNA	#transcripts	Transcript IDs
RefSeq Genes		
hsa-mir-662	2	NM_001025190,NM_001025190
hsa-mir-328	4	NM_024712,NM_024712,NM_024712,NM_024712
hsa-mir-367	4	NM_015454,NM_016648,NM_015454,NM_016648
hsa-mir-191	1	NM_199074
UCSC Genes		
hsa-mir-662	2	uc002cjz,uc002cjz
hsa-mir-328	12	uc002esa,uc002esb,uc002esc,uc002esa,uc002esb,uc002esc,uc002esa,uc002esb,uc002esc,uc002esa,uc002esb,uc002esc
hsa-mir-367	8	uc003iay,uc003iaz,uc003iba,uc003ibb,uc003iay,uc003iaz,uc003iba,uc003ibb
hsa-mir-99b	1	uc002pxg
hsa-let-7e	1	uc002pxg
hsa-mir-636	2	uc002jsz,uc002jta
hsa-mir-125a	1	uc002pxg
hsa-mir-191	1	uc003cvn
ESTs		
hsa-mir-639	11	DA578654,DB036923,BE163339,BX337999,BE173861,AW601441,AW601462,AW602622,AW602643,BE076002,BG978956
hsa-mir-26a-1	1	BG013974
hsa-mir-659	2	N64538,N64538
hsa-mir-632	4	CR982540,BF739929,BF986402,BG005263
hsa-mir-545	1	CF121922
hsa-mir-142	11	AA736584,AI214324,AA480329,AI803986,BM987623,AW075941,BM994627,AA749202,AW075748,AA804593,AA804593
hsa-mir-1-2	2	AI220268,AI265999
hsa-mir-614	8	BG546987,AA494355,BE001688,BE001688,BE001688,BE001688,BE815242,BE825683
hsa-mir-34c	1	CB243753
hsa-mir-423	1	AW001018
hsa-mir-324	2	DB118628,AA428344
hsa-mir-609	5	BF946814,BF946811,BF946814,BF946814,BF946814
hsa-mir-143	3	AW750687,AW750687,BF328624
hsa-mir-612	13	BF371267,BF359112,BQ373809,BQ332593,BQ332596,BF808842,BF359123,BQ373809,BF877468,BG990102,AW939381,BQ327427,BF894409
hsa-mir-124a-	1	BU738550
hsa-mir-570	1	BE162667
hsa-mir-631	10	CV569139,CN361157,BX644907,DA065297,BF827870,CR978078,BF922074,BF827870,BM709090,AA359490
hsa-mir-770	3	AI970289_dup1,AI636778_dup1,AI963928_dup1
hsa-mir-299	2	AW895894,AW895880

hsa-let-7i	1	CR992282
hsa-mir-503	10	R79973,H01302,R79973,R79973,R79973,H01302,R7997 3,R79973,R79973,R79973
hsa-mir-339	1	DB317795
hsa-mir-219-2	1	DA115674
hsa-mir-125b-	1	BG000222
hsa-mir-499	13	BE297149,BE297149,BE297149,BE297149,BE297149,BE 297149,BE297149,BE297149,BE297149,BE297149,BE29 7149,BE297149,BE297149
hsa-mir-517a	1	BF994458
hsa-mir-141	4	AI695443,AA863389,AA863395,AI969882
hsa-mir-223	5	BQ012126,BQ010657,AW192334,AA845292,BQ011993
hsa-mir-371	2	DB508823,DB443882
hsa-mir-24-2	1	BF061990
hsa-mir-193a	1	AI770050
hsa-mir-429	5	AI312008,AI312008,AI312008,AI312008,AI312008
hsa-mir-568	18	CB054179,CA425784,AI383978,AI024792,CK902677,AA 489694,AW024848,AI912487,AA115749,BM668044,AA1 15749,BM668044,AA115749,BM668044,BM668044,AA11 5749,AA115749,AA115749
hsa-mir-565	1	DC404319
hsa-mir-522	2	BI057935,BI057935
hsa-mir-24-1	1	CB852241
hsa-mir-516-3	2	BE466189,BF223893
hsa-mir-624	1	H66135
hsa-mir-224	1	BF367311
hsa-mir-328	20	BM792516,BE868797,BI092635,BI260928,BQ706797,CA 488307,CD243179,BM853445,BM789033,BX390378,BI0 92635,BX390378,BM792516,BQ706797,BE868797,BM85 3445,CA488307,CD243179,BI260928,BM789033
hsa-mir-7-1	1	AW367387
hsa-mir-663	1	BM893562
hsa-mir-643	1	BF356097
hsa-mir-320	1	AI041315
hsa-let-7a-1	1	BF372275
hsa-mir-214	11	AI961147,AI202700,BG055203,BQ012655,AI039268,CF 529290,BQ447950,AI658740,AI242334,AW296598,AI12 6848
hsa-mir-363	1	AA353588
hsa-let-7b	6	BX361219,BX349705,BU732454,AI382133,DB328200,A W028822
hsa-mir-296	1	BE044227
hsa-mir-590	2	AW812109,AW812109
hsa-mir-21	4	BX399365,AA837010,BE932516,CA414094
hsa-mir-186	3	AA258173,BQ271697,AW968323
hsa-mir-622	1	BM542991
hsa-mir-495	3	AA613245,AW593422,BE467211
hsa-mir-144	1	BX644926

hsa-mir-205	60	BX454368,BU607326,BU687118,BU686798,BM976602,BF088470,BF088470,DB313515,AU158081,AU158081,R48249,R73789,BG989285,AA627916,AI969511,AW275818,AW275825,AW662302,BU732467,CN478782,DB313077,W68529,H25944,AI800001,BM980083,BU676846,CB048050,AW628474,R73789,AA631033,AI375939_dup1,CB409600,BQ378988,R48249,AA642987,BF088470,AI611676,R48249,R73789,BE717915,R48249,AW117862,CB409600,R48249,R73789,BE717919,BE717969,BE837877,BE717939,BE645803,R48249,BE717915,AW381579,AW380949,AA730105,R73789,R48249,BE717915,AI220853,AW117862
hsa-mir-125a	1	DB078945
hsa-mir-29b-1	1	BF447465
hsa-mir-641	1	BF805572
hsa-let-7d	8	AI360684,BF511039,BF511477,T85314,BE932147,BE709888,BU736404,AA806216
hsa-mir-194-2	3	AI392832,AW294586,DB331974
hsa-mir-377	5	BF376962,BF376962,BF376962,BF376962,BF376962
hsa-mir-23b	3	BG004545,BG004545,BG004545
hsa-mir-555	1	BF842814
hsa-mir-655	2	BG000132,BG000132
hsa-mir-365-2	1	AA682513
hsa-mir-29a	2	AI261722,BF000417
hsa-mir-564	10	CF126629,BQ082498,BQ082492,BQ896914,BQ082492,BQ082498,AL532442,AL532442,AL532442,AL532442
hsa-mir-675	83	AL571290,AL569789,BF997501,AL514065,AL571212,W72208,BQ366088,AW946259,AA864221,BF590747,AL571290,AI738732,BG000974,BG012341,BQ367462,AL571290,AA928657,R93762,BM678988_dup1,AI146817,AI247386,BF985264,BX453708,AL564130,BG012341,AW196761,AW196780,BG009089,BG013206,BG000974,BG011738,BF985264,BG000974,BF985264,BF770659,BG012341,BX453708,BG000974,BX453708,BF770659,BI044667,BQ367462,BG011738,BF770659,BG012341,AA928657,R93762,BG013206,BG009089,BI044667,AA928657,BQ367462,BG000974,R93762,W72208,AA864221,AL564130,BG012347,AA928657,BQ366088,BQ367462,BG011738,BQ367462,BF770659,BI044667,BQ367462,AI377593,BQ367462,AI377593,AL564130,R93762,BQ367462,W72208,BG012347,BQ367462,BI044667,BQ367462,W72208,AA864221,R93762,AL569789,BQ367462,BG012347
hsa-mir-33b	1	DB361007
hsa-mir-103-2	5	AA477191,BE768941,BE768943,BE768941,BE768943
hsa-mir-130a	1	DA811291
hsa-mir-196b	2	BQ923616,BM549446
hsa-mir-647	7	AL708944,BF857645,DB066850,BI006021,BF857645,DB066850,DB066850
hsa-mir-27a	4	AW797020,AW797020,AW797020,AW797020
hsa-mir-611	4	AA602796,BQ130146,DV461029,AW104310
hsa-mir-190	1	DB288334
hsa-mir-421	1	H63707
hsa-mir-101-1	1	BU786820
hsa-mir-221	1	AW842797
hsa-mir-192	2	AI672369,BG982441

hsa-mir-367	28	BG499366,BQ217140,H43466,BM796975,AA451825,BF790426,BG431297,BG898445,BM853135,BU854431,BX488044,CB146063,CB162796,CN346204,DT218636,W49839,W39691,CB135486,BQ920803,AA160180,AA372829,BF701214,CN419691,BU787911,BE748087,AV714955,BE748087,W39691
hsa-mir-567	1	AA054551
hsa-mir-122a	3	R91986,R91986,R91986
hsa-mir-30c-2	9	BQ022726,AI087792,AI474776,W93081,AW168091,BF439917,BF439062,AI768381,BF594736
hsa-mir-133a-	4	AA211717,DB330925,AA211717,AA211717
hsa-mir-372	2	DB443882,AW833903
hsa-mir-630	10	BX454918,CB854863,BX412844,CF891283,BX412844,BX454918,CB854863,CF891283,BX454918,BX454918
hsa-mir-635	8	AA210899,BQ705942,BQ705942,BQ705942,BQ705942,BQ705942,BQ705942,AA210899
hsa-mir-198	11	AW364521,AW936518,AW936507,AA614015,BX395666,AA897679,BF733997,BF367500,AW364521,BM684744,BX395666
hsa-mir-202	1	AW340536
hsa-mir-106b	2	AA078024,AA078024
hsa-mir-508	1	AW665118
hsa-mir-98	1	BM956031
hsa-mir-130b	1	DV080646
hsa-mir-155	5	BG058661,BG058739,AI863758,AI863758,AI863758
hsa-mir-373	6	AI825624,AI656634,BE466114,AI825746,BE672151,BE046897
hsa-mir-516-2	1	BQ025835
hsa-mir-548a-	1	AW851470
hsa-mir-505	3	BF956603,BF956603,BF956603
hsa-mir-9-2	8	BM676994,DC425874,DB315227,BX412634,BM683678,BX412634,BX412634,BX412634
hsa-mir-135a-	1	BM918632
hsa-mir-636	24	AI890310,BX434220,BU608163,CA306910,AW316782,CB241746,CA420551,CA423953,AL581513,BI012130,BI084978,BU168060,BU628861,CA420567,CD367228,BX453763,BI010983,BU624553,BX434220,BX434220,CB241746,AI890310,BX434220,DB205468
hsa-mir-637	5	AI203861,AW974502,AA768246,AW974502,AW974502
hsa-mir-621	1	BF923438
hsa-mir-25	6	BE077684,BQ377098,BQ377098,BE077684,BE077684,BE077684
hsa-mir-411	2	AW895894,AW895880
hsa-mir-29c	43	AA832487,AA732327,AA814450,AA290626,AI949631,BF431339,AI796930,AI277016,AI139038,BF446797,BF196476,AW015122,AW055230,AI927692,BU686674,BQ009786,BU607478,CA312488,CA312468,AI800208,AI582596,AI634533,BF592005,R48833,AI628443,AI935532,AI435146,BQ018823,BE043082,BE045365,DB302666,AW058564,T59406,T59406,R48833,T59406,R48833,R48833,R48833,T59406,T59406,R48833,R48833
hsa-mir-451	4	H90496,BX644926,H90496,H90496
hsa-mir-145	4	AI825923,AI678858,BF995019,AI659796
hsa-mir-133a-	8	BF508318,BF446240,AA063595,N80784,AI819052,DB327850,BX647033,AA063595
hsa-mir-548d-	2	BQ889295,BQ889295
hsa-mir-301	1	CV414363

hsa-mir-671	98	BQ576140,BF570037,BQ007968,BQ574801,CA945217,CK904076,DB323915,DB323780,BF115617,BX330457,AW664880,BG054996,BQ017347,BE855754,BM680658,BU630312,DB323018,AI978627,AA534198,CA431554,CA450521,BU632964,CB857275,AW195142,BI966035,AW166498,AW166513,BF594067,BU077063,BE206147,AW612913,BE207563,AI819341,AA209319,BM145320,AI139595,AI365580,AW958254,CB856472,CA416896,AW958254,AI620649,AA583627,CF529167,AW005402,BM148417,AA209319,AI022352,AA583619,AW166054,AA209319,AA209319,R51978,AI475193,R51978,AI569467,AI540081,R51978,AA209319,BM144231,BM145320,AI191668,BM149632,R51978,R51978,R51978,R51978,AA209319,AI090886,AA583627,AA534198,AA678590,BE242046,T46979,R51978,T46979,R51978,BU733453,BE242217,AA524450,AA508608,R51978,T46979,R51978,BI524753,T46979,AW451651,R51978,DB334087,AW082159,F09189,DB364152,BQ447944,CF529982,R51978,AA209319,T46979,AA378790
hsa-mir-650	1	BU585017
hsa-mir-616	11	BF839560,BF839556,BF839550,BF839561,BF839551,BF839555,BF839557,BF839558,BF839549,BF839554,BF839560
hsa-mir-30d	1	AA137041
hsa-mir-15a	9	AA748510,BM272203,CK825000,BQ185483,AW976277,AA867999,AI086232,AA748510,AW976277
hsa-mir-29b-2	1	AA766259
hsa-mir-126	24	BX380871,BX397999,BX397999,BX380871,BX397999,BX380439,BX340281,BX397999,BX397999,BX380439,BX381730,BX381730,BX380040,BX381730,BX380871,BX381730,BX380782,BX459190,BX380439,BX380871,BX360778,BX381730,BX397999,BX360778
hsa-mir-191	8	BQ439253,BU594868,BQ645068,BQ652614,BQ646652,BQ647682,DB059215,BQ439253

mRNAs

hsa-mir-302c	1	AK000089
hsa-mir-302b	1	AK000089
hsa-mir-675	2	BC010054,BC004532
hsa-mir-631	2	AK128372,AK026216
hsa-mir-150	1	AK130324
hsa-mir-647	1	AK128082
hsa-mir-636	3	AF015188,AF015189,AF015190
hsa-mir-661	1	AL834492
hsa-mir-335	1	AK055108
hsa-mir-328	12	AK023886,AK056923,BC015524,BC034410,DQ891332,DQ894515,AK023886,AK056923,BC015524,BC034410,DQ891332,DQ894515
hsa-mir-589	1	AY927477
hsa-mir-367	8	AK000089,AF068284,AK000274,AK225900,AL049996,BC066945,BC107709,BX647100
hsa-mir-99b	1	AK125996
hsa-mir-302a	1	AK000089
hsa-let-7e	1	AK125996
hsa-mir-126	1	BC114447
hsa-mir-499	1	AK098707
hsa-mir-302d	1	AK000089
hsa-mir-125a	1	AK125996

Spliced ESTs

hsa-mir-371	1	DB443882
hsa-mir-130b	1	DV080646
hsa-mir-639	11	DA578654,DB036923,BE163339,BX337999,BE173861,AW601441,AW601462,AW602622,AW602643,BE076002,BG978956
hsa-mir-155	1	BG058739
hsa-mir-632	3	BF739929,BF986402,BG005263
hsa-mir-365-2	1	AA682513
hsa-mir-423	1	AW001018
hsa-mir-324	1	AA428344
hsa-mir-565	1	DC404319
hsa-mir-564	4	AL532442,AL532442,AL532442,AL532442
hsa-mir-124a-	1	BU738550
hsa-mir-675	7	AL571290,AL571212,AL571290,AL571290,AL564130,AL564130,AL564130
hsa-mir-196b	2	BQ923616,BM549446
hsa-mir-24-1	1	CB852241
hsa-mir-637	5	AI203861,AW974502,AA768246,AW974502,AW974502
hsa-mir-636	24	AI890310,BX434220,BU608163,CA306910,AW316782,CB241746,CA420551,CA423953,AL581513,BI012130,BI084978,BU168060,BU628861,CA420567,CD367228,BX453763,BI010983,BU624553,BX434220,BX434220,CB241746,AI890310,BX434220,DB205468
hsa-mir-611	4	AA602796,BQ130146,DV461029,AW104310
hsa-mir-25	6	BE077684,BQ377098,BQ377098,BE077684,BE077684,BE077684
hsa-mir-328	20	BM792516,BE868797,BI092635,BI260928,BQ706797,CA488307,CD243179,BM853445,BM789033,BX390378,BI092635,BX390378,BM792516,BQ706797,BE868797,BM853445,CA488307,CD243179,BI260928,BM789033
hsa-mir-367	28	BG499366,BQ217140,H43466,BM796975,AA451825,BF790426,BG431297,BG898445,BM853135,BU854431,BX488044,CB146063,CB162796,CN346204,DT218636,W4983 9,W39691,CB135486,BQ920803,AA160180,AA372829,BF701214,CN419691,BU787911,BE748087,AV714955,BE748087,W39691
hsa-mir-372	1	DB443882
hsa-mir-133a-	4	AA211717,DB330925,AA211717,AA211717
hsa-mir-15a	3	BM272203,AW976277,AW976277
hsa-mir-126	22	BX380871,BX397999,BX397999,BX380871,BX397999,BX380439,BX340281,BX397999,BX397999,BX380439,BX381730,BX381730,BX380040,BX381730,BX380871,BX381730,BX380782,BX459190,BX380439,BX380871,BX381730,BX397999
hsa-mir-141	4	AI695443,AA863389,AA863395,AI969882
hsa-mir-205	15	BG989285,AA627916,AI969511,AW275818,AW275825,AW662302,BU732467,CN478782,DB313077,W68529,H25944,AI800001,BM980083,BU676846,CB048050
hsa-mir-191	5	BU594868,BQ645068,BQ652614,BQ646652,BQ647682

Intronic miRNAs (anti-sense)

hsa-mir-662, hsa-mir-302c, hsa-mir-199a-1, hsa-mir-302b, hsa-mir-181a-2, hsa-mir-1-2, hsa-mir-324, hsa-mir-194-1, hsa-mir-16-2, hsa-mir-599, hsa-mir-486, hsa-mir-570, hsa-mir-631, hsa-mir-610, hsa-mir-328, hsa-mir-181b-2, hsa-mir-367, hsa-mir-133a-1, hsa-mir-215, hsa-mir-548d-2, hsa-mir-15b, hsa-mir-302a, hsa-mir-214, hsa-mir-199a-2, hsa-mir-199b, hsa-mir-302d, hsa-mir-549, hsa-mir-191

Table S1: Anti-sense transcripts near known miRNAs (continued)

Mouse miRNAs

miRNA	#transcripts	Transcript IDs
RefSeq Genes		
mmu-mir-136	3	NM_184109,NM_184109,NM_184109
mmu-mir-127	2	NM_184109,NM_184109
mmu-mir-433	2	NM_184109,NM_184109
mmu-mir-434	2	NM_184109,NM_184109
mmu-mir-367	2	NM_138593,NM_138593
mmu-mir-762	3	NM_009746,NM_009746,NM_009746
mmu-mir-328	4	NM_172760,NM_172760,NM_172760,NM_172760
mmu-mir-135a-1	2	NM_001039586,NM_174846
mmu-mir-431	2	NM_184109,NM_184109
ESTs		
mmu-mir-760	1	BY724041
mmu-mir-219-2	1	BQ174353
mmu-mir-684-1	13	DV073242,BQ174174,AW537749,BE630457,CF578937,AU019690,AW912686_dup1,BF138984,BF138984,EH106532_dup1,EH106867_dup1,BB368041,BE630457
mmu-mir-320	2	AI317636,AU067584
mmu-let-7a-1	2	CA895946,CA895946
mmu-let-7b	6	BX634634,BE990520,BE691213,BP758973,AI481799,AI551238
mmu-mir-410	9	BQ175773,CO039417,CO039417,CO039417,CO039417,CO039417,BQ175773,CO039417,CF586444
mmu-mir-703	69	AW475773,BM569675,BQ829508,AA408019,AW549861,BG061776,BM222193,AW323039,BM202365,CK619245,BM208748,BM209549,BM244993,BQ942052,BQ829508,AW549861,AW323039,BM569675,AA408019,AW475773,BG061776,BM222193,AI037476,BM202365,BM200056,BM235396,BI557490,BQ829508,AI037476,CK619245,AW549861,AI037476,BM202365,AW323039,BM569675,BQ829508,BM200056,BM209549,BM235396,BM244993,BM208748,BQ942052,AI894192,BM241195,BQ829508,BM203696,BM293715,BM293715,BM200056,BM235396,BM202365,BM209549,BM244993,BM569675,AW323039,BM222193,AI037476,BQ942052,CK619245,AA408019,AI894192,AW475773,BG061776,BM203696,BM208748,BM241195,BQ829508,BM293715,AW323039
mmu-mir-719	3	BF319694,BF319694,BF319694
mmu-let-7c-2	1	CX730031
mmu-mir-692-1	1	EH109725_dup1
mmu-mir-762	48	CB590046,BF123867,CB057904,AA219888,BI794412,AA015463,CF583785,BP761456,CB522109,BG247738,BY011013,CB574482,CO806265,CA321672,BQ712754,BQ945556,CB193807,CA464583,BE336140,BX516469,CB057430,BY311097,BY323578,CJ174327,CJ184526,BU935799,BY335256,BU936524,CF550740,CO800690,CO798695,BQ930905,CK792708,BE553370,CB182304,AV445615,BQ928058,BB651583,CA977612,BF123867,AA015463,DV651350,BY022716,BF123867,AA015463,AA929912,BF123867,AA929912

mmu-mir-124a-1	1	EL608186
mmu-mir-133a-2	1	CR517796
mmu-mir-484	3	BI688308,BB628323,BI688308
mmu-mir-370	2	BG870188,BG870188
mmu-mir-127	2	AW060983,BF461489
mmu-mir-299	2	CF578713,BI319615
mmu-mir-195	1	BB469543
mmu-mir-107	1	DT918023
mmu-mir-205	1	BX636072
mmu-mir-715	36	CN677041,CA540888,CA542066,CN675294,AV459680,BI441300,BI789633,CF899888,CD551120,CD553916,CD549707,BI441208,CD565318,CD553951,CA546775,CF902843,CA546648,AV488839,AV458300,CA949213,BI659171,BI690957,BI659626,AV468375,AA538360,BI100293,AA538360,CA949213,BI100293,BG146738,W42347,AV459680,W42347,AV459680,AV459680,AV458300
mmu-mir-130b	2	BF318156,AI550467
mmu-mir-324	1	CK619771
mmu-mir-675	12	BG228779,CR757130,CR757163,CR757130,CR757163,BG228779,BG228779,CR757130,CR758026,CR756565,CR757086,CR758028
mmu-mir-367	12	AA408246,CO795279,BU610513,BG802296,BF160578,CA857540,BM946683,BU525549,CF534412,CO814744,CV561292,DV657845
mmu-mir-682	5	EH111758,BB646391,BF018818,BF018908,BQ033113_dup1
mmu-mir-153	1	BU921583
mmu-mir-101a	2	BX638239,BF318343
mmu-mir-296	1	BE990102
mmu-mir-700	1	AA028433
mmu-mir-129-2	1	CR519766
mmu-mir-704	4	BM519373,BM519462,BM519384,BM519373
mmu-mir-29b-2	2	AA184398,AI591965
mmu-mir-411	5	CF578713,CF578713,CF578713,CF578713,CF578713
mmu-mir-671	23	BG067525,BP766107,AW060945,BF020189,BP764469,C86598,BP763209,AW744648,CD741762,AI551231,AI551231,AW413604,AI551231,BF320330,BE993040,BM197793,BF458705,BE690893,AI551231,BF455885,C86598,BX638041,AW046089
mmu-mir-23b	2	AI848465,AW124037
mmu-mir-804	3	BY707319,BY707319,BX521364
mmu-mir-142	3	BE136127,AW909330,BE136127
mmu-mir-328	23	BY766110,CB950291,BF159242,BE569617,BF123539,BG865727,BI691021,BF159242,BG865727,BF123539,BG865727,BE569617,BF159242,BI691021,BY766110,CB950291,AA266420,AI466854,BG865727,BG865727,BG865727,BG865727,BG865727
mmu-mir-135a-1	5	CB948426,BQ256452,BF780995,BF452753,BY105899
mmu-let-7d	3	BF662590,BQ031149,AI591449
mmu-mir-677	4	CK725393,BE859594,BG794909,EL605651
mmu-mir-713	1	BY019942
mmu-mir-29c	8	AA184398,AI591965,CO042247,BF465648,AW456678,BE944441,EH094197,AW494050
mmu-mir-21	5	BP774280,EH103036,EH105192,DT931951,DT931951
mmu-mir-339	1	BX629854

mmu-mir-692-2	97	BG090122,AI876713,AI120525,C77615,BQ265350,BP771591,BM570058,AW212806,AW208734,C76709,EG563506,AU042028,BQ264923,C81314,C85950,AU041056,BM198032,AU018935,AI042821,C85307,AW259615,AW742780,AA066642,AU021287,BG090001,BF319101_dup1,BM199052,BF148348,C80431,BF148348,BF148348,AI503783,AA066642,BP769370,AA066642,AI837703,AA066642,AA066642,BM199579,BM210661,AA086583,AI386321,AA086583,AW212806,CF583084,AA086583,AA086583,AA066642,AI839982,AA086583,BP769370,BM116361,AA086583,AA086583,BF148625,BM196958,BF146594,AA542429,AA561457,BM199549,BM199547,BQ264923,BM198032,BM198037,AA760480,BM199059,BM199182,BF319100,BP774123,BM199052,BG146514,AA542429,AA542429,BM202444,AA542429,AA760480,AA561457,AA542429,AI837703,BM199579,BM210661,CD741947,AA086583,AA561457,AI386321,AA086583,EH109725,AA086583,BQ128462,AA086583,AA542429,AI839982,AA086583,BM116361,AA086583,BP774123,AA086583
mmu-mir-689-1	5	AW211456,AW211456,AW211456,AW211456,AW211456
mmu-mir-138-2	1	AI854067
mmu-mir-685	9	AA176001,AA177668,BE335968,BE335968,AA176001,AA177668,AA176001,AA177668,BE335968
mmu-mir-137	3	AI852436,EL607439,BX638046
mmu-mir-678	30	AI504737,AI851123,AI845257,BE954206,CD774044,AW321042,AW541993,BM114883,BE993523,AI852462,CA889352,CA882196,BM227453,CA881388,BM021608,BM226267,C87892,CF582154,EH094417,EH101496,EH102235,EH106884,AA152773,AA152782,AA152773,AA152773,EH094419,AA152773,AA152773,AA152782
mmu-mir-763	1	CF578614
mmu-mir-214	7	BX631348,AI414009,CA774438,AI414009,BQ418332,AI414009,CA947318
mmu-mir-705	3	BX636473,AW046224,CF585469
mmu-mir-702	2	AI647501,AI647501
mmu-mir-431	2	AA048217,BQ127396
mmu-mir-181b-2	4	BI319432,CJ065621,CJ065471,BI319432
mmu-mir-686	32	AW539968,AI848457,AW543310,AW557943,AW541849,BE200146,BM203159,AU045989,BF722037,AW492012,BQ126981,BQ268530,BE952271,BG228853,BG093897,AU021597,AW539968,AU021597,AW541849,AU021597,AU021597,AU045989,AU021597,AU045989,AU021597,BE952271,BE952271,BE952271,BE952271,BE952271,BG093897,BG093897

mRNAs

mmu-mir-29c	1	AK081202
mmu-mir-207	1	AK139286
mmu-mir-29b-2	1	AK081202
mmu-mir-671	1	AK158019
mmu-mir-715	2	BC116307,BC131795
mmu-mir-762	6	Y11905,BC058530,AK131935,BC005673,AK049064,Y11905
mmu-mir-685	6	BC128280,BC128469,BC128470,BC128280,BC128469,BC128470
mmu-mir-804	1	AK007228
mmu-mir-328	10	AK037067,AK157875,BC018516,BC026617,BC058752,AK037067,AK157875,BC018516,BC026617,BC058752

mmu-mir-135a-1	5	AK052709,BC025935,BC036136,BC033063,AK049425
mmu-mir-484	1	AK035704
mmu-mir-181b-2	1	AK082091

spliced ESTs

mmu-mir-704	4	BM519373,BM519462,BM519384,BM519373
mmu-mir-320	2	AI317636,AU067584
mmu-mir-410	9	BQ175773,CO039417,CO039417,CO039417,CO039417, CO039417,BQ175773,CO039417,CF586444
mmu-mir-719	3	BF319694,BF319694,BF319694
mmu-let-7c-2	1	CX730031
mmu-mir-762	46	CB590046,BF123867,CB057904,AA219888,BI794412,AA 015463,CF583785,BP761456,CB522109,BG247738,BY01 1013,CB574482,CO806265,CA321672,BQ712754,BQ945 556,CB193807,CA464583,BE336140,BX516469,CB0574 30,BY311097,BY323578,CJ174327,CJ184526,BU935799, BY335256,BU936524,CF550740,CO800690,CO798695,B Q930905,CK792708,BE553370,CB182304,AV445615,BQ 928058,BB651583,CA977612,BF123867,AA015463,BF12 3867,AA015463,AA929912,BF123867,AA929912
mmu-mir-804	3	BY707319,BY707319,BX521364
mmu-mir-328	23	BY766110,CB950291,BF159242,BE569617,BF123539,BG 865727,BI691021,BF159242,BG865727,BF123539,BG86 5727,BE569617,BF159242,BI691021,BY766110,CB9502 91,AA266420,AI466854,BG865727,BG865727,BG86572 7,BG865727,BG865727
mmu-mir-484	3	BI688308,BB628323,BI688308
mmu-mir-677	4	CK725393,BE859594,BG794909,EL605651
mmu-mir-713	1	BY019942
mmu-mir-195	1	BB469543
mmu-mir-367	12	AA408246,CO795279,BU610513,BG802296,BF160578,C A857540,BM946683,BU525549,CF534412,CO814744,CV 561292,DV657845
mmu-mir-675	5	CR757130,CR757163,CR757130,CR757163,CR757130
mmu-mir-678	7	CD774044,CA881388,EH094417,EH101496,EH102235,E H106884,EH094419
mmu-mir-137	3	AI852436,EL607439,BX638046
mmu-mir-700	1	AA028433
mmu-mir-702	2	AI647501,AI647501
mmu-mir-686	26	AW539968,AI848457,AW543310,AW557943,AW541849, BE200146,BM203159,AU045989,BF722037,AW492012,B Q126981,BQ268530,BE952271,BG228853,BG093897,A W539968,AW541849,AU045989,AU045989,BE952271,B E952271,BE952271,BE952271,BE952271,BG093897,BG 093897

Intronic miRNAs (anti-sense)

mmu-mir-684-1, mmu-mir-199a-1, mmu-mir-133a-1, mmu-mir-486, mmu-mir-703, mmu-mir-1-2, mmu-mir-181a-2, mmu-mir-297-2, mmu-mir-215, mmu-mir-804, mmu-mir-328, mmu-mir-199a-2, mmu-mir-484, mmu-mir-194-1, mmu-mir-324, mmu-mir-199b, mmu-mir-689-2, mmu-mir-214, mmu-mir-695, mmu-mir-181b-2

Table S1: Anti-sense transcripts near known miRNAs. Shown are known miRNAs from human, mouse, and Drosophila (col. 1), the number of anti-sense transcripts near these miRNAs (within 50nts) and their sequence identifiers. The transcripts are separated into genes, ESTs, and spliced ESTs as indicated. The table also contains a list of known miRNAs that are located in introns of host genes, which are transcribed in the opposite direction.

Table S2: Solexa reads matching sense/anti-sense miRNA pairs

Drosophila sense/anti-sense miRNAs

Mouse sense/anti-sense miRNAs

```
mmu-mir-34b chr9 50855792 50855875 -
GTGCTCGGTTTGTAGGCAGTGTAATTAGCTGATTGTAGTGCGGTGCTGACAATCACTAACTCCACTGCCATCAAAACAAGGCAC
(((((..((((...(((((((...((((.((((((((((.....))).)))))))))).)))))))....)))).)))))) (-34.10)
(((((((.((((..(((((((.((((((.(((((((((.......)).)))))))))).)))))))..)))).))).)))) (-34.50)
                                            ----------------------          ATGGCAGTGGAGTTAGTGATTGT 1 1
                                            ----------------------          TGGCAGTGGAGTTAGTGATTGT 5 1
                                               ++++++++++++++++++++++       ATCACTAACTCCACTGCCATCAA 2 1
                                               ++++++++++++++++++++++       ATCACTAACTCCACTGCCATCA 51 1
                                              ++++++++++++++++++++++        AATCACTAACTCCACTGCCATCA 17 1
                                                ++++++++++++++++++++        TCACTAACTCCACTGCCATC 1 1
                                               ++++++++++++++++++++         ATCACTAACTCCACTGCCATC 24 1
                                              ++++++++++++++++++++++        AATCACTAACTCCACTGCCATC 259 1
                                               ++++++++++++++++++++         ATCACTAACTCCACTGCCAT 1 1
                                              ++++++++++++++++++++          AATCACTAACTCCACTGCCAT 22 1
                                               ++++++++++++++++++           ATCACTAACTCCACTGCCA 2 1
                                              ++++++++++++++++++            AATCACTAACTCCACTGCCA 28 1
                                             ++++++++++++++++++++           CAATCACTAACTCCACTGCCA 1 1
                                             ++++++++++++++++++             CAATCACTAACTCCACTGCC 1 1
                  ++++++++++++++++++++++                                    GGCAGTGTAATTAGCTGATTGTA 1 1
                ++++++++++++++++++++++++                                    TAGGCAGTGTAATTAGCTGATTGTA 2 1
                       ++++++++++++++++                                     GTAATTAGCTGATTGT 1 1
                     ++++++++++++++++++                                     GTGTAATTAGCTGATTGT 3 1
                    ++++++++++++++++++++                                    AGTGTAATTAGCTGATTGT 6 1
                 ++++++++++++++++++++++                                     AGGCAGTGTAATTAGCTGATTGT 638 1
                ++++++++++++++++++++++++                                    TAGGCAGTGTAATTAGCTGATTGT 26 1
                     ++++++++++++++++++                                     AGTGTAATTAGCTGATTG 4 1
                 ++++++++++++++++++++++                                     AGGCAGTGTAATTAGCTGATTG 23 1
                ++++++++++++++++++++++                                      TAGGCAGTGTAATTAGCTGATTG 117 1
                ++++++++++++++++++++++                                      TAGGCAGTGTAATTAGCTGATT 14 1
                ++++++++++++++++++++                                        TAGGCAGTGTAATTAGCTGAT 1 1
                ++++++++++++++++++                                          TAGGCAGTGTAATTAGCT 4 2

mmu-mir-182 chr6 30115928 30116002 -
ACCATTTTTGGCAATGGTAGAACTCACACCGGTAAGGTAATGGGACCCGGTGGTTCTAGACTTGCCAACTATGGT
(((((..(((((((.(.(((((((...(((((((.............)))))))))))..)))))))..)))))) (-29.14)
(((((..(((((((...(((((((...(((((((.............)))))))))))))...)))))))..)))))) (-31.64)
                                            ----------------------      GTTGGCAAGTCTAGAACCACCG 2 1
                              ++++++++++++++++++                        GTAAGGTAATGGGACCCG 43 1
      ++++++++++++++++++++++++++++                                      TTTGGCAATGGTAGAACTCACACCGGT 8 1
      ++++++++++++++++++++++++++++                                      TTTGGCAATGGTAGAACTCACACCGG 95 1
            ++++++++++++++++++++                                        AATGGTAGAACTCACACCG 40 1
         ++++++++++++++++++++++                                         GGCAATGGTAGAACTCACACCG 4 1
        ++++++++++++++++++++++++                                        TGGCAATGGTAGAACTCACACCG 5 1
       ++++++++++++++++++++++++++                                       TTGGCAATGGTAGAACTCACACCG 48 1
      ++++++++++++++++++++++++++++                                      TTTGGCAATGGTAGAACTCACACCG 11377 1
     ++++++++++++++++++++++++++++++                                     TTTTGGCAATGGTAGAACTCACACCG 27 1
        ++++++++++++++++++++++++                                        TGGCAATGGTAGAACTCACACC 3 1
       +++++++++++++++++++++++++                                        TTTGGCAATGGTAGAACTCACACC 1358 1
        ++++++++++++++++++++++                                          TTGGCAATGGTAGAACTCACAC 11 1
       ++++++++++++++++++++++                                           TTTGGCAATGGTAGAACTCACAC 2295 1
        ++++++++++++++++++++                                            TTGGCAATGGTAGAACTCACA 4 1
       ++++++++++++++++++++                                             TTTGGCAATGGTAGAACTCACA 1734 1
        ++++++++++++++++++                                              TTGGCAATGGTAGAACTCAC 1 1
       ++++++++++++++++++                                               TTTGGCAATGGTAGAACTCAC 1024 1
       +++++++++++++++++++                                              TTTGGCAATGGTAGAACTCA 250 1
       ++++++++++++++++++                                               TTTGGCAATGGTAGAACTC 113 1
       ++++++++++++++++++                                               TTTGGCAATGGTAGAACT 20 1
       ++++++++++++++++++                                               TTTGGCAATGGTAGAAC 4 2

mmu-mir-203 chr12 112578688 112578763 +
GCCTGGTCCAGTGGTTCTTGACAGTTCAACAGTTCTGTAGCACAATTGTGAAATGTTTAGGACCACTAGACCCGGC
(((.((((.((((((((((.((((.((((.(((((.((...)).)))))))))).))))).)))))))))).))))) (-35.60)
(((.(((((((((((.((..((.((((.((((.((((.((...)).)))))))))).)))).)))))))))))))) (-32.40)
        ----------------------                                          TTGAACTGTCAAGAACCACTGG 1 1
          --------------------                                          TTGAACTGTCAAGAACCACT 1 1
                                          +++++++++++++++++++++++       TGTGAAATGTTTAGGACCACTAG 29 1
                                           ++++++++++++++++             GTGAAATGTTTAGGAC 1 3
                                           +++++++++++++++++            GTGAAATGTTTAGGACCA 1 1
                                           ++++++++++++++++++           GTGAAATGTTTAGGACCAC 1 1
                                           ++++++++++++++++++++         GTGAAATGTTTAGGACCACT 16 1
                                           ++++++++++++++++++++++       GTGAAATGTTTAGGACCACTA 3 1
                                           ++++++++++++++++++++++++     GTGAAATGTTTAGGACCACTAG 223 1
                                           +++++++++++++++++++++++++    GTGAAATGTTTAGGACCACTAGA 11 1
                                            ++++++++++++++++            TGAAATGTTTAGGACCA 1 3
                                            ++++++++++++++++++          TGAAATGTTTAGGACCACTA 7 1
                                            ++++++++++++++++++++        TGAAATGTTTAGGACCACTAG 130 1
                                            ++++++++++++++++++++++      TGAAATGTTTAGGACCACTAGA 17 1
                                             ++++++++++++++++++++       GAAATGTTTAGGACCACTAG 1 1

mmu-mir-194-1 chr1 187014107 187014173 +
ATCGGGTGTAACAGCAACTCCATGTGGACTGTGCTCGGATTCCAGTGGAGCTGCTGTTACTTCTGAT
((((((.(((((((((.((((((.(((((((.....)))..)))))))))).)))))))))).)))))) (-34.10)
.......(((((((((((((((..(((((.............)))))..)))))))))))))))..... (-28.90)
     ----------------------                                             CACATGGAGTTGCTGTTACAC 3 1
      ++++++++++++++++++++++                                            TGTAACAGCAACTCCATGTGGA 104 2
      ++++++++++++++++++++++++                                          TGTAACAGCAACTCCATGTGGAC 1 1
```

```
mmu-mir-27b chr13 63309934 63310006 +
AGGTGCAGAGCTTAGCTGATTGGTGAACAGTGATTGGTTTCCGCTTTGTTCACAGTGGCTAAGTTCTGCACCT
(((((((((((((((((....(((((((.....(((...)))..)))))))..)))))))))))))))) (-40.20)
(((((((((.((((((.....((((((.......(....).....)))))))....))))))..))))))))) (-31.30)
    ----------------------                                                  TTCACCAATCAGCTAAGCTCTGC 1 1
     ++++++++++++++++++++++++                                               CAGAGCTTAGCTGATTGGTGAACA 19 1
      ++++++++++++++++++                                                    AGAGCTTAGCTGATTGGT 5 1
      ++++++++++++++++++++                                                  AGAGCTTAGCTGATTGGTGA 8 1
      ++++++++++++++++++++                                                  AGAGCTTAGCTGATTGGTGAA 65 1
      ++++++++++++++++++++++                                                AGAGCTTAGCTGATTGGTGAAC 87 1
      ++++++++++++++++++++++                                                AGAGCTTAGCTGATTGGTGAACA 274 1
      ++++++++++++++++++++++++                                              AGAGCTTAGCTGATTGGTGAACAG 19 1
      ++++++++++++++++++++++++                                              AGAGCTTAGCTGATTGGTGAACAGT 1 1
                                          ++++++++++++++++++                GTTCACAGTGGCTAAGTT 2 2
                                          ++++++++++++++++++++              GTTCACAGTGGCTAAGTTCT 64 1
                                          ++++++++++++++++++++++            GTTCACAGTGGCTAAGTTCTG 135 1
                                          ++++++++++++++++++++++++          GTTCACAGTGGCTAAGTTCTGC 188 1
                                           ++++++++++++++++                 TTCACAGTGGCTAAG 2 20
                                           +++++++++++++++++                TTCACAGTGGCTAAGT 4 3
                                           +++++++++++++++++                TTCACAGTGGCTAAGTT 109 3
                                           ++++++++++++++++++               TTCACAGTGGCTAAGTTC 1294 2
                                           ++++++++++++++++++++             TTCACAGTGGCTAAGTTCT 3140 1
                                           ++++++++++++++++++++             TTCACAGTGGCTAAGTTCTG 12686 1
                                           ++++++++++++++++++++++           TTCACAGTGGCTAAGTTCTGC 32597 1
                                           ++++++++++++++++++++++++         TTCACAGTGGCTAAGTTCTGCA 617 1
                                           ++++++++++++++++++++++++++       TTCACAGTGGCTAAGTTCTGCAC 25 1
                                            ++++++++++++++++++              TCACAGTGGCTAAGTTC 1 4
                                            ++++++++++++++++++++            TCACAGTGGCTAAGTTCTG 54 1
                                            ++++++++++++++++++++            TCACAGTGGCTAAGTTCTGC 16 1
                                            ++++++++++++++++++++++          TCACAGTGGCTAAGTTCTGCA 2 1
                                             ++++++++++++++++++             CACAGTGGCTAAGTTCTG 5 1
                                             ++++++++++++++++++             CACAGTGGCTAAGTTCTGC 4 1
                                             ++++++++++++++++++++           CACAGTGGCTAAGTTCTGCA 1 1
                                              ++++++++++++++++              ACAGTGGCTAAGTTC 2 13
                                              ++++++++++++++++              ACAGTGGCTAAGTTCT 7 4
                                              ++++++++++++++++++            ACAGTGGCTAAGTTCTG 25 3
                                              ++++++++++++++++++            ACAGTGGCTAAGTTCTGC 51 2
                                               ++++++++++++++++++           CAGTGGCTAAGTTCTGC 3 2
                                                 ++++++++++++++             AGTGGCTAAGTTCTG 3 9
                                                  ++++++++++++++++          AGTGGCTAAGTTCTGCA 9 2

mmu-mir-126 chr2 26413366 26413438 +
TGACAGCACATTATTACTTTTGGTACGCGCTGTGACACTTCAAACTCGTACCGTGAGTAATAATGCGCGGTCA
.(((.((.((((((((((.(((((((.(---(((----)))..).)))))))).)))))))))).)).)))). (-32.30)
.(((.(((((((((((...((((((.((((....)))))).)))))))...)))))))))).)))). (-34.40)
       ----------------------                                               CGCGTACCAAAAGTAATAATGT 1 1
        ++++++++++++++++++                                                  CATTATTACTTTTGGTAC 3 1
        ++++++++++++++++++++                                                CATTATTACTTTTGGTACG 41 1
        ++++++++++++++++++++                                                CATTATTACTTTTGGTACGC 781 1
        ++++++++++++++++++++++                                              CATTATTACTTTTGGTACGCG 786 1
        ++++++++++++++++++++++                                              CATTATTACTTTTGGTACGCGC 44 1
        ++++++++++++++++++++++++                                            CATTATTACTTTTGGTACGCGCT 5 1
        +++++++++++++++++++++++++++                                         CATTATTACTTTTGGTACGCGCTGTGA 1 1
        +++++++++++++++++++++++++++++                                       CATTATTACTTTTGGTACGCGCTGTGACA 1 1
         ++++++++++++++++++++++                                             ATTATTACTTTTGGTACGCGC 1 1
          ++++++++++++++++++                                                TTATTACTTTTGGTACGC 2 1
           ++++++++++++++++++                                               TATTACTTTTGGTACGCG 1 1
                                          ++++++++++++++                    CTCGTACCGTGAGTA 1 1
                                          ++++++++++++++++                  CTCGTACCGTGAGTAAT 1 1
                                          ++++++++++++++++++++              CTCGTACCGTGAGTAATAATG 4 1
                                          ++++++++++++++++++++++            CTCGTACCGTGAGTAATAATGC 26 1
                                          ++++++++++++++++++++++++          CTCGTACCGTGAGTAATAATGCG 11 1
                                           ++++++++++++++++++               TCGTACCGTGAGTAATAA 1 1
                                           ++++++++++++++++++               TCGTACCGTGAGTAATAAT 25 1
                                           ++++++++++++++++++++             TCGTACCGTGAGTAATAATG 101 1
                                           ++++++++++++++++++++             TCGTACCGTGAGTAATAATGC 658 1
                                           ++++++++++++++++++++++           TCGTACCGTGAGTAATAATGCG 1165 1
                                           ++++++++++++++++++++++++         TCGTACCGTGAGTAATAATGCGC 2 1
                                            ++++++++++++++++++              CGTACCGTGAGTAATAAT 3 1
                                            ++++++++++++++++++              CGTACCGTGAGTAATAATG 28 1
                                            ++++++++++++++++++++            CGTACCGTGAGTAATAATGC 240 1
                                            ++++++++++++++++++++++          CGTACCGTGAGTAATAATGCG 561 1
                                            ++++++++++++++++++++++++        CGTACCGTGAGTAATAATGCGC 3 1
                                             ++++++++++++++++++++           GTACCGTGAGTAATAATGC 1 1
                                              ++++++++++++++++++            TACCGTGAGTAATAATGCG 1 1
                                               ++++++++++++++++             ACCGTGAGTAATAATGC 1 1
                                                ++++++++++++++              CCGTGAGTAATAATGC 1 1
                                                ++++++++++++++++            CCGTGAGTAATAATGCG 2 1
                                                  ++++++++++++              GTGAGTAATAATGCG 1 1
```

```
mmu-let-7b chr15 85535085 85535169 +
GCAGGGTGAGGTAGTAGGTTGTGTGGTTTCAGGGCAGTGATGTTGCCCCTCCGAAGATAACTATACAACCTACTGCCTTCCCTGA
.(((((.(((((((((((((((((((((((.((((((.....))))))....)).....)))))))))))))))))))))))). (-49.00)
.(((((..(((.((((((((((...(((..((((((.......))))))((....)).)))...))))))))))).)))).))))). (-37.50)
      ++++++++++++++                                                                    TGAGGTAGTAGGTTG 6 10
      ++++++++++++++++                                                                  TGAGGTAGTAGGTTGT 53 8
      +++++++++++++++++                                                                 TGAGGTAGTAGGTTGTG 14 1
      ++++++++++++++++++                                                                TGAGGTAGTAGGTTGTGT 94 1
      +++++++++++++++++++                                                               TGAGGTAGTAGGTTGTGTG 1041 1
      ++++++++++++++++++++                                                              TGAGGTAGTAGGTTGTGTGG 1351 1
      +++++++++++++++++++++                                                             TGAGGTAGTAGGTTGTGTGGT 34985 1
      ++++++++++++++++++++++                                                            TGAGGTAGTAGGTTGTGTGGTT 136624 1
      +++++++++++++++++++++++                                                           TGAGGTAGTAGGTTGTGTGGTTT 33456 1
      ++++++++++++++++++++++++                                                          TGAGGTAGTAGGTTGTGTGGTTTC 299 1
      +++++++++++++++++++++++++                                                         TGAGGTAGTAGGTTGTGTGGTTTCA 26 1
      ++++++++++++++++++++++++++                                                        TGAGGTAGTAGGTTGTGTGGTTTCAG 5 1
      +++++++++++++++++++++++++++                                                       TGAGGTAGTAGGTTGTGTGGTTTCAGGG 1 1
      ++++++++++++++++++++++++++++++                                                    TGAGGTAGTAGGTTGTGTGGTTTCAGGGCA 7 1
       ++++++++++++++++++++                                                             GAGGTAGTAGGTTGTGTGGT 104 1
       ++++++++++++++++++++                                                             GAGGTAGTAGGTTGTGTGGTT 139 1
       ++++++++++++++++++++++                                                           GAGGTAGTAGGTTGTGTGGTTT 11 1
        ++++++++++++++++++                                                              AGGTAGTAGGTTGTGTGGT 6 1
        ++++++++++++++++++++                                                            AGGTAGTAGGTTGTGTGGTT 44 1
        ++++++++++++++++++++++                                                          AGGTAGTAGGTTGTGTGGTTT 3 1
         ++++++++++++++++++++                                                           GGTAGTAGGTTGTGTGGTT 4 1
         ++++++++++++++++++++++                                                         GGTAGTAGGTTGTGTGGTTT 1 1
          ++++++++++++++++++                                                            GTAGTAGGTTGTGTGGT 1 1
          ++++++++++++++++++++                                                          GTAGTAGGTTGTGTGGTT 22 1
          ++++++++++++++++++++                                                          GTAGTAGGTTGTGTGGTTT 1 1
           ++++++++++++++++++                                                           TAGTAGGTTGTGTGGT 22 1
           ++++++++++++++++++                                                           TAGTAGGTTGTGTGGTT 5 1
           ++++++++++++++++++                                                           TAGTAGGTTGTGTGGTTT 4 1
            ++++++++++++++++                                                            AGTAGGTTGTGTGGTT 2 1
            ++++++++++++++++                                                            AGTAGGTTGTGTGGTTT 20 1
             ++++++++++++++++                                                           GTAGGTTGTGTGGTT 2 3
              ++++++++++++++++++++++++++                                                TAGGTTGTGTGGTTTCAGGGCAGTGA 1 1
                     ++++++++++++++++++++++++++++++++                                   TCAGGGCAGTGATGTTGCCCCTCCGAAGAT 1 1
                                          ++++++++++++++++++++++++++                    AGATAACTATACAACCTACTGCCTTC 1 1
                                               ++++++++++++++++++++                     CTATACAACCTACTGCCTTC 15 1
                                               ++++++++++++++++++++++                   CTATACAACCTACTGCCTTCC 43 1
                                               ++++++++++++++++++++++++                 CTATACAACCTACTGCCTTCCC 1 1
                                             ----------------                           GTAGGTTGTATAGTTA 7 3
                                             -----------------                          AGTAGGTTGTATAGTTA 4 1
                                              ----------------                          AGTAGGTTGTATAGTT 41 3
                                              -----------------                         CAGTAGGTTGTATAGTT 1 1
                                               ---------------                          AGTAGGTTGTATAGT 1 9


(mmu-mir-374 chrX 99775779 99775873 - NO uniquely matching anti-sense reads)
GAAGAAATCCTACTCGGGTGGATATAATACAACCTGCTAAGTGTTCTAGCACTTAGCAGGTTGTATTATCATTGTCCGAGGTTATGGCTCTCGTC
...((...((..((((((..(..(((((((((((((((((((((....)))))))))))))))))))))))..)..))))))......))...)).... (-46.30)
...((((.(((((((.(......(((((((((((((((((((((((.....)))))))))))))))))))))).....)))))..))).)).)))).... (-42.60)
                                          --------------------                          ATAATACAACCTGCTAAGTG 1 2
                                          -------------------                           TAATACAACCTGCTAAGTG 1 2
                                          ------------------                            ATACAACCTGCTAAGTG 1 5
                                             ++++++++++++++++++++                       TTAGCAGGTTGTATTATCATT 1 1
                                           ++++++++++++++++++++++                       CTTAGCAGGTTGTATTATCATT 114 1
                                          ++++++++++++++++++++++                        ACTTAGCAGGTTGTATTATCATT 1 1
                                     ++++++++++++++++++++++++++++++                     TTCTAGCACTTAGCAGGTTGTATTATCATT 1 1
                                            ++++++++++++++++++++                        CTTAGCAGGTTGTATTATCAT 5 1
                                           ++++++++++++++++++++++                       ACTTAGCAGGTTGTATTATCAT 3 1
                                            ++++++++++++++++++                          CTTAGCAGGTTGTATTA 3 3
                        ------------------                                              CTTAGCAGGTTGTATTA 3 3
                        ++++++++++++++++++++                                            TAATACAACCTGCTAAGTGT 1 1
                       ++++++++++++++++++++                                             ATAATACAACCTGCTAAGTGT 1 1
                     ++++++++++++++++++++++                                             ATATAATACAACCTGCTAAGTGT 91 1
                          ++++++++++++++++                                              ATACAACCTGCTAAGTG 1 5
                        ++++++++++++++++++                                              TAATACAACCTGCTAAGTG 1 2
                       ++++++++++++++++++++                                             ATAATACAACCTGCTAAGTG 1 2
                      ++++++++++++++++++++                                              TATAATACAACCTGCTAAGTG 3 1
                     ++++++++++++++++++++                                               ATATAATACAACCTGCTAAGTG 316 1
                      ++++++++++++++++++                                                TATAATACAACCTGCTAAGT 4 1
                     ++++++++++++++++++++                                               ATATAATACAACCTGCTAAGT 603 1
                     +++++++++++++++++++                                                ATATAATACAACCTGCTAAG 28 1
                     ++++++++++++++++++                                                 ATATAATACAACCTGCTAA 5 1
```

Table S3: Solexa reads matching sense/anti-sense miRNA pairs

Drosophila miRNAs

Read Sequence	#cloned/sequenced	#matches to genome	miRNA matches	miRNA anti-sense matches
TTACGTATACTGAAGGTAT	8	1		dme-mir-iab-4
TGCACTTGAGACGGCCTGA	2	1	dme-mir-312	
TTACGTATACTGAAGGTATACCG	2	1		dme-mir-iab-4
TATTGCACTTGAGACGGCCTGA	3027	1	dme-mir-312	
TTGCACTTGAGACGGCCTGA	4	1	dme-mir-312	
TTGCACTTGAGACGGCCTGAT	1	1	dme-mir-312	
ACGTATACTGAATGTATCCTG	1	1	dme-mir-iab-4	
ATTGCCCTTGTGACGAACCAAAT	2	1		dme-mir-312
TATTGCACTTGAGACG	1	1	dme-mir-312	
TATTGCACTTGAGACGGCCTGATT	4	1	dme-mir-312	
TGGTTCGTCACAAGGGCAATTCT	151	1	dme-mir-312	
TTACGTATACTGAAGGTATAC	2	1		dme-mir-iab-4
TTGCACTTGAGACGGCCTG	1	1	dme-mir-312	
CGGTATACCTTCAGTATACG	1	1	dme-mir-iab-4	
TATTGCACTTGAGACGGCC	34	1	dme-mir-312	
TGGTTCGTCACAAGGGC	1	1	dme-mir-312	
TGGTTCGTCACAAGGGCAA	17	1	dme-mir-312	
TTACGTATACTGAAGGTATACC	4	1		dme-mir-iab-4
TGGTTCGTCACAAGGGCA	6	1	dme-mir-312	
TATTGCACTTGAGACGGCCTGAT	94	1	dme-mir-312	
TATTGCACTTGAGACGGC	27	1	dme-mir-312	
ACGTATACTGAATGTATCCTGA	5	1	dme-mir-iab-4	
TTACGTATACTGAAGGTATACCGG	1	1		dme-mir-iab-4
TGGTTCGTCACAAGGGCAATT	90	1	dme-mir-312	
TATTGCACTTGAGACGGCCT	225	1	dme-mir-312	
CACTTGAGACGGCCTGA	1	1	dme-mir-312	
TGGTTCGTCACAAGGGCAATTC	66	1	dme-mir-312	
TATTGCACTTGAGACGGCCTG	1165	1	dme-mir-312	
ATTGCACTTGAGACGGCCTG	3	1	dme-mir-312	
ATTGCACTTGAGACGGCCTGAT	6	1	dme-mir-312	
TGCACTTGAGACGGCCT	2	1	dme-mir-312	
TGGTTCGTCACAAGGGCAAT	44	1	dme-mir-312	
ATTGCACTTGAGACGGCCTGA	18	1	dme-mir-312	

Table S3: Solexa reads matching sense/anti-sense miRNA pairs

Mouse miRNAs

Read Sequence	#cloned/sequenced	#matches to genome	miRNA matches	miRNA anti-sense matches
GTAGGTTGTATGGTT	4	5	mmu-let-7c-2	
ACAGTGGCTAAGTTCT	7	4	mmu-mir-27b	
CAATCACTAACTCCACTGCC	1	1	mmu-mir-34b	
GCACCATTTGAAATCAGTGTT	39	2	mmu-mir-29b-2	
ACTTAGCAGGTTGTATTATCATT	1	1	mmu-mir-374	
GTAGTAGGTTGTATAGTT	21	2	mmu-let-7a-1	
AGTAGGTTGTATAGTTA	4	1		mmu-let-7b
AGTAGGTTGTATAGTTT	1	2	mmu-let-7a-1	
CGTACCGTGAGTAATAATGCG	561	1	mmu-mir-126	
AGTAGGTTGTATGGTT	8	3	mmu-let-7c-2	
TGGAATGTAAAGAAGTATGTAT	45	2	mmu-mir-1-2	
TCGTACCGTGAGTAATAAT	25	1	mmu-mir-126	
CAGAGCTTAGCTGATTGGTGAACA	19	1	mmu-mir-27b	
TGAGGTAGTAGGTTGTATGGTT	198548	2	mmu-let-7c-2	
AGGTAGTAGGTTGTGTGGTT	44	1	mmu-let-7b	
TAGCACCATTTGAAATC	6	3	mmu-mir-29b-2	
CACCATTTGAAATCAGTG	4	3	mmu-mir-29b-2	
GTAGTAGGTTGTGTGGT	1	1	mmu-let-7b	
CGCGTACCAAAAGTAATAATGT	1	1		mmu-mir-126
CTCGTACCGTGAGTAATAATGC	26	1	mmu-mir-126	
TGAGGTAGTAGGTTGTGTG	1041	1	mmu-let-7b	
GGCAATGGTAGAACTCACACCG	4	1	mmu-mir-182	
GCTGGTTTCACATGGTGGCTTAGA	2	1	mmu-mir-29b-2	
TGAGGTAGTAGGTTGTATGG	1404	2	mmu-let-7c-2	
AGAGCTTAGCTGATTGGTGAACA	274	1	mmu-mir-27b	
TATACAATCTACTGTCTTTC	2	2	mmu-let-7a-1;mmu-let-7c-2	
TTTTGGCAATGGTAGAACTCACACCG	27	1	mmu-mir-182	
GTGAAATGTTTAGGACCACTA	3	1	mmu-mir-203	
TTTGGCAATGGTAGAACTCA	250	1	mmu-mir-182	
CTATACAATCTACTGTCTTT	218	2	mmu-let-7a-1;mmu-let-7c-2	
TGGCAGTGGAGTTAGTGATTGT	5	1		mmu-mir-34b
CACCATTTGAAATCAGTGTT	13	2	mmu-mir-29b-2	
TCGTACCGTGAGTAATAATG	101	1	mmu-mir-126	
CTCGTACCGTGAGTA	1	1	mmu-mir-126	
AGTAGGTTGTATGGTTT	2	2	mmu-let-7c-2	
TCCTGTACTGAGCTGCCCCG	11	2	mmu-mir-486	mmu-mir-486
TAGCACCATTTGAAAT	6	6	mmu-mir-29b-2	
CGGGGCAGCTCAGTACAGGAT	46	1		mmu-mir-486
ATTTGAAATCAGTGTTT	1	4	mmu-mir-29b-2	
CTCGTACCGTGAGTAAT	1	1	mmu-mir-126	
ACAGTGGCTAAGTTCTG	25	3	mmu-mir-27b	
TGAGGTAGTAGGTTGTATAGTTTT	54	1	mmu-let-7a-1	
AGGCAGTGTAATTAGCTGATTG	23	1	mmu-mir-34b	
TGTGAAATGTTTAGGACCACTAG	29	1	mmu-mir-203	
GGTAGTAGGTTGTATAGT	5	2	mmu-let-7a-1	
GTAATTAGCTGATTGT	1	1	mmu-mir-34b	
GTGAAATGTTTAGGACCACTAGA	11	1	mmu-mir-203	
ACACGGACACCGCAGGG	6	1	mmu-let-7b	
TTTGGCAATGGTAGAACTCACACC	1358	1	mmu-mir-182	
TGAAATGTTTAGGACCACTA	7	1	mmu-mir-203	
AATCACTAACTCCACTGCCAT	22	1	mmu-mir-34b	
TTCACAGTGGCTAAGTTC	1294	2	mmu-mir-27b	
TTTGGCAATGGTAGAACTCAC	1024	1	mmu-mir-182	
ATGAGGTAGTAGGTTGTATAGTT	53	1	mmu-let-7a-1	
TTGGCAATGGTAGAACTCAC	1	1	mmu-mir-182	
ACCATTTGAAATCAGTGTT	24	2	mmu-mir-29b-2	
TGAGGTAGTAGGTTGTATGGT	30009	2	mmu-let-7c-2	
CGTACCGTGAGTAATAATGCGC	3	1	mmu-mir-126	
GTAGTAGGTTGTGTGGTT	22	1	mmu-let-7b	
GAGGTAGTAGGTTGTATGGT	65	2	mmu-let-7c-2	
TTTGGCAATGGTAGAAC	4	2	mmu-mir-182	
AGTAGGTTGTGTGGTTT	20	1	mmu-let-7b	
ATATAATACAACCTGCTAAGTGT	91	1	mmu-mir-374	
TCAGCACCAGGATATTGTTGGGG	1	1		mmu-mir-338
TTAGGGTCACACCCACCACTGGG	1	1	mmu-let-7a-1	
TGAGGTAGTAGGTTGTGTGGT	34985	1	mmu-let-7b	
GAGGTAGTAGGTTGTGTGGT	104	1	mmu-let-7b	
CAGTGGCTAAGTTCTGC	3	2	mmu-mir-27b	
CTATACAATCTACTGTCTTTCCT	8	2	mmu-let-7a-1;mmu-let-7c-2	
CTAGCACCATTTGAAATCAGTGT	335	2	mmu-mir-29b-2	
CATCAGTGATTTTGTTG	1	1	mmu-mir-338	
AGAGCTTAGCTGATTGGTGAAC	87	1	mmu-mir-27b	
TGGAATGTAAAGAAGTATGT	1	2	mmu-mir-1-2	
GGTAGTAGGTTGTATAGTTT	2	2	mmu-let-7a-1	
ACAGTGGCTAAGTTC	2	13	mmu-mir-27b	
TGAGGTAGTAGGTTGTGTGGTTTCAGGG	1	1	mmu-let-7b	
GTAAGGTAATGGGACCCG	43	1	mmu-mir-182	
TTCACAAAGCCCATACACTTT	35	1	mmu-mir-350	
CATTATTACTTTTGGTACGC	781	1	mmu-mir-126	
TCACAGTGGCTAAGTTCTGCA	2	1	mmu-mir-27b	
TAGTAGGTTGTATAG	1	4	mmu-let-7a-1	
TCCAGCATCAGTGATTTTGTTG	15	1	mmu-mir-338	

CCATTTGAAATCAGTGTT	15	2	mmu-mir-29b-2	
TTGAACTGTCAAGAACCACT	1	1		mmu-mir-203
ATCTAGCACCATTTGAAATCAGTGTT	4	1	mmu-mir-29b-2	
CACAGTGGCTAAGTTCTGCA	1	1	mmu-mir-27b	
CTAGCACCATTTGAAATCAGTGTT	126	2	mmu-mir-29b-2	
TAGCACCATTTGAAATCAGTGTTTTA	5	1	mmu-mir-29b-2	
CACATGGAGTTGCTGTTACAC	3	1		mmu-mir-194-1
AATGGTAGAACTCACACCG	40	1	mmu-mir-182	
TGAAATGTTTAGGACCA	1	3	mmu-mir-203	
AGTAGGTTGTATAGT	1	9	mmu-let-7a-1	mmu-let-7b
TAGCACCATTTGAAATCAGTGTT	80418	2	mmu-mir-29b-2	
CCGTGAGTAATAATGC	1	1	mmu-mir-126	
GGTAGTAGGTTGTATGGTTT	1	2	mmu-let-7c-2	
TAGCACCATTTGAAATCAGTGT	20749	2	mmu-mir-29b-2	
ATCACTAACTCCACTGCCATCA	51	1	mmu-mir-34b	
GCACCATTTGAAATCAGTGT	14	2	mmu-mir-29b-2	
GGCAGTGTAATTAGCTGATTGTA	1	1	mmu-mir-34b	
TTCACAAAGCCCATACACTTTCAC	145	1	mmu-mir-350	
TCCAGCATCAGTGATTTTG	1	1	mmu-mir-338	
TGAGGTAGTAGGTTGTATGGTTTTGGG	1	1	mmu-let-7c-2	
AAAGTGCATGCGCTTTGGG	8	1	mmu-mir-350	
TGAGGTAGTAGGTTG	6	10	mmu-let-7a-1;mmu-let-7c-2;mmu-let-7b	
GGTAGTAGGTTGTATAGTT	35	2	mmu-let-7a-1	
CTATACAACCTACTGCCTTC	15	1	mmu-let-7b	
TAGTAGGTTGTATAGTTT	6	2	mmu-let-7a-1	
TAATACAACCTGCTAAGTG	1	2	mmu-mir-374	mmu-mir-374
AAAGTGCATGCGCTTTGGGA	66	1	mmu-mir-350	
GAGGTAGTAGGTTGTATGGTT	681	2	mmu-let-7c-2	
TTCACAGTGGCTAAGTT	109	3	mmu-mir-27b	
AACACGGACACCGCAGGG	1	1	mmu-let-7b	
TTAGGGTCACACCCACCACTGGGAGATAA	28	1	mmu-let-7a-1	
GCTGGTTTCACATGGTGGCTTAGATTT	1	1	mmu-mir-29b-2	
TCAACAAAATCACTGATGCTGGAGT	11	1		mmu-mir-338
TGTAACAGCAACTCCATGTGGA	104	2	mmu-mir-194-1	
TATAATACAACCTGCTAAGT	4	1	mmu-mir-374	
ACCATTTGAAATCAGTG	1	5	mmu-mir-29b-2	
CTGGTTTCACATGGTGGCTTAG	4	1	mmu-mir-29b-2	
CGTACCGTGAGTAATAAT	3	1	mmu-mir-126	
AGAGCTTAGCTGATTGGTGAA	65	1	mmu-mir-27b	
TAGTAGGTTGTATAGTT	150	2	mmu-let-7a-1	
CATTATTACTTTTGGTACGCGCTGTGA	1	1	mmu-mir-126	
TGAGGTAGTAGGTTGTATAGTTTTAGG	1	1	mmu-let-7a-1	
GGTAGTAGGTTGTGTGGTTT	1	1	mmu-let-7b	
TGGCAATGGTAGAACTCACACCG	5	1	mmu-mir-182	
GTAGGTTGTATAGTTA	7	3		mmu-let-7b
AGTTCCAGGACAGCCAGGGCTATACAGAGA	1	19184	mmu-mir-706	
TAGTAGGTTGTATGGTTT	2	2	mmu-let-7c-2	
CTAGCACCATTTGAAATCAGT	15	2	mmu-mir-29b-2	
TGAGGTAGTAGGTTGTATGGTTTTGG	1	1	mmu-let-7c-2	
TTCACCAATCAGCTAAGCTCTGC	1	1		mmu-mir-27b
TGGGTACATAAAGAAGTATGTGC	5	1		mmu-mir-1-2
AGGTAGTAGGTTGTATGGTT	25	2	mmu-let-7c-2	
TTCACAAAGCCCATACA	1	1	mmu-mir-350	
CAATCACTAACTCCACTGCCA	1	1	mmu-mir-34b	
ACAGTGGCTAAGTTCTGC	51	2	mmu-mir-27b	
TAGGTTGTGTGGTTTCAGGGCAGTGA	1	1	mmu-let-7b	
TTATTACTTTTGGTACGC	2	1	mmu-mir-126	
TAGGCAGTGTAATTAGCTGATTG	117	1	mmu-mir-34b	
CTTAGCAGGTTGTATTATCATT	114	1	mmu-mir-374	
TTTGGCAATGGTAGAACTCACACCG	11377	1	mmu-mir-182	
TAGTAGGTTGTATGG	1	4	mmu-let-7c-2	
CACAGTGGCTAAGTTCTGC	4	1	mmu-mir-27b	
AGATAACTATACAACCTACTGCCTTC	1	1	mmu-let-7b	
CATTATTACTTTTGGTACGCGCTGTGACA	1	1	mmu-mir-126	
CTGGTTTCACATGGTGGCTTAGA	15	1	mmu-mir-29b-2	
ATACAACCTGCTAAGTG	1	5	mmu-mir-374	mmu-mir-374
CCATTTGAAATCAGTGT	6	2	mmu-mir-29b-2	
GGTAGTAGGTTGTGTGGTT	4	1	mmu-let-7b	
TCGTACCGTGAGTAATAATGC	658	1	mmu-mir-126	
TCACAGTGGCTAAGTTCTG	54	1	mmu-mir-27b	
CATTATTACTTTTGGTACGCGCT	5	1	mmu-mir-126	
CGTACCGTGAGTAATAATGC	240	1	mmu-mir-126	
TCTCTGTATAGCCCTGGCTGTC	1	25381		mmu-mir-706
GTGAAATGTTTAGGACCAC	1	1	mmu-mir-203	
TAGTAGGTTGTGTGGTTT	4	1	mmu-let-7b	
CAGTAGGTTGTATAGTT	1	1		mmu-let-7b
TCGTACCGTGAGTAATAATGCGC	2	1	mmu-mir-126	
AGTGTAATTAGCTGATTGT	6	1	mmu-mir-34b	
TCCTGTACTGAGCTGCCCC	1	2	mmu-mir-486	mmu-mir-486
TTTGGCAATGGTAGAACTC	113	1	mmu-mir-182	
TTGGCAATGGTAGAACTCACAC	11	1	mmu-mir-182	
CTATACAATCTACTGTCTTTCC	102	2	mmu-let-7a-1;mmu-let-7c-2	
TAGTAGGTTGTGTGGTT	5	1	mmu-let-7b	
TAGTAGGTTGTATGGT	24	4	mmu-let-7c-2	
GAGGTAGTAGGTTGTATGGTTT	120	2	mmu-let-7c-2	
TCCTGTACTGAGCTGCCCCGAGC	3	1	mmu-mir-486	
TCCTGTACTGAGCTGCCCCGAGG	1	1		mmu-mir-486
GTAGTAGGTTGTATGGTT	5	2	mmu-let-7c-2	
TCACTAACTCCACTGCCATC	1	1	mmu-mir-34b	

CATTATTACTTTTGGTACGCGC	44	1	mmu-mir-126	
TGAAATGTTTAGGACCACTAGA	17	1	mmu-mir-203	
TGAGGTAGTAGGTTGTATA	95	2	mmu-let-7a-1	
TGAGGTAGTAGGTTGTATG	381	2	mmu-let-7c-2	
TGGTTTCACATGGTGGCTTAGA	2	1	mmu-mir-29b-2	
TTGGCAATGGTAGAACTCACACCG	48	1	mmu-mir-182	
CATTTGAAATCAGTGT	5	6	mmu-mir-29b-2	
GTACCGTGAGTAATAATGC	1	1	mmu-mir-126	
CTAGCACCATTTGAAATCAGTG	30	2	mmu-mir-29b-2	
CTAGCACCATTTGAAATCAG	2	2	mmu-mir-29b-2	
GTTCACAGTGGCTAAGTTCT	64	1	mmu-mir-27b	
TTTGGCAATGGTAGAACTCACACCGGT	8	1	mmu-mir-182	
AGGTAGTAGGTTGTATGGT	16	2	mmu-let-7c-2	
GGTAGTAGGTTGTATA	1	2	mmu-let-7a-1	
TAGCACCATTTGAAATCAGT	265	2	mmu-mir-29b-2	
TGAGGTAGTAGGTTGTGTGG	1351	1	mmu-let-7b	
TATTACTTTTGGTACGCG	1	1	mmu-mir-126	
AGTAGGTTGTATAGTT	41	3	mmu-let-7a-1	mmu-let-7b
ATGAGGTAGTAGGTTGTATAGT	5	1	mmu-let-7a-1	
TGAGGTAGTAGGTTGTATAGTT	66680	2	mmu-let-7a-1	
TAGTAGGTTGTGTGGT	22	1	mmu-let-7b	
CATTATTACTTTTGGTACG	41	1	mmu-mir-126	
TGAGGTAGTAGGTTGT	53	8	mmu-let-7a-1;mmu-let-7c-2;mmu-let-7b	
ATCTATCTATCTATCTATC	2	212505		mmu-mir-350
CATTTGAAATCAGTGTT	44	3	mmu-mir-29b-2	
AATCACTAACTCCACTGCCA	28	1	mmu-mir-34b	
CGGGGCAGCTCAGTACAGGATG	56	1		mmu-mir-486
AGTAGATTGTATAGTT	11	5		mmu-let-7a-1;mmu-let-7c-2
ATCACTAACTCCACTGCCATCAA	2	1	mmu-mir-34b	
TACCGTGAGTAATAATGCG	1	1	mmu-mir-126	
AAACCAGCTTCCAGAAG	1	3		mmu-mir-29b-2
AGGTAGTAGGTTGTATGGTTT	7	2	mmu-let-7c-2	
AGTGTAATTAGCTGATTG	4	1	mmu-mir-34b	
TGGAATGTAAAGAAGTATGTA	2	2	mmu-mir-1-2	
GAGGTAGTAGGTTGTGTGGTTT	11	1	mmu-let-7b	
GTGTAATTAGCTGATTGT	3	1	mmu-mir-34b	
CGGGGCAGCTCAGTACAGGA	1	2	mmu-mir-486	mmu-mir-486
AGCACCATTTGAAATCAGTGT	3	2	mmu-mir-29b-2	
TAGCACCATTTGAAATCAG	45	2	mmu-mir-29b-2	
TTCACAAAGCCCATACACTTTC	190	1	mmu-mir-350	
AGGTAGTAGGTTGTGTGGT	6	1	mmu-let-7b	
GGTGAGGTAGTAGGTTGTAT	1	1	mmu-let-7c-2	
GAGGTAGTAGGTTGTATGGTTTT	2	1	mmu-let-7c-2	
TCCTGTACTGAGCTGCCCCGA	66	2	mmu-mir-486	mmu-mir-486
TAATACAACCTGCTAAGTGT	1	1	mmu-mir-374	
TGAGGTAGTAGGTTGTATAGT	7590	2	mmu-let-7a-1	
GAGGTAGTAGGTTGTATAGT	14	2	mmu-let-7a-1	
TTTGAAATCAGTGTT	1	21	mmu-mir-29b-2	
TTCACAAAGCCCATACACTT	4	1	mmu-mir-350	
GAACACGGACACCGCAGGG	2	1	mmu-let-7b	
CTGGTTTCACATGGTGGCTTAGATTTT	2	1	mmu-mir-29b-2	
TTGGGCTCTGCCCCGCTCTGCGGTAA	21	1	mmu-let-7c-2	
ATTTGAAATCAGTGTT	12	8	mmu-mir-29b-2	
TAGGCAGTGTAATTAGCTGATT	14	1	mmu-mir-34b	
ATATAATACAACCTGCTAAGT	603	1	mmu-mir-374	
GTGAGGTAGTAGGTTGTATGGTT	1	1	mmu-let-7c-2	
GGTAGTAGGTTGTATGGT	2	2	mmu-let-7c-2	
CATTATTACTTTTGGTACGCG	786	1	mmu-mir-126	
ATGGCAGTGGAGTTAGTGATTGT	1	1		mmu-mir-34b
TTTGGCAATGGTAGAACTCACA	1734	1	mmu-mir-182	
GTAGGTTGTGTGGTT	2	3	mmu-let-7b	
ACCATTTGAAATCAGTGT	9	2	mmu-mir-29b-2	
TGGCAATGGTAGAACTCACACC	3	1	mmu-mir-182	
CTCGTACCGTGAGTAATAATGCG	11	1	mmu-mir-126	
TTCACAGTGGCTAAG	2	20	mmu-mir-27b	
AGGTAGTAGGTTGTGTGGTTT	3	1	mmu-let-7b	
AGTGGCTAAGTTCTG	3	9	mmu-mir-27b	
GTAGTAGGTTGTATGGTTT	1	2	mmu-let-7c-2	
AGTAGGTTGTGTGGTT	2	1	mmu-let-7b	
AGAGCTTAGCTGATTGGTGAACAG	19	1	mmu-mir-27b	
TGAGGTAGTAGGTTGTGTGGTT	136624	1	mmu-let-7b	
CACCATTTGAAATCAGTGT	14	2	mmu-mir-29b-2	
ATAGATAGATAGATAGATA	4	210178	mmu-mir-350	
TAGGCAGTGTAATTAGCTGAT	1	1	mmu-mir-34b	
TGAGGTAGTAGGTTGTAT	327	4	mmu-let-7a-1;mmu-let-7c-2	
TCGTACCGTGAGTAATAA	1	1	mmu-mir-126	
GTTCACAGTGGCTAAGTTCTGC	188	1	mmu-mir-27b	
GTGAAATGTTTAGGACCACT	16	1	mmu-mir-203	
CCGTGAGTAATAATGCG	2	1	mmu-mir-126	
AGAGCTTAGCTGATTGGTGA	8	1	mmu-mir-27b	
TCCTGTACTGAGCTGCCC	2	2	mmu-mir-486	mmu-mir-486
AACAATATCCTGGTGCTGAGT	11	1	mmu-mir-338	
CTGGTTTCACATGGTGGCTTAGAT	7	1	mmu-mir-29b-2	
TTTGGCAATGGTAGAACTCACACCGG	95	1	mmu-mir-182	
GAGGTAGTAGGTTGTATGG	4	2	mmu-let-7c-2	
CTATACAACCTACTGCCTTCCC	1	1	mmu-let-7b	
TGAGGTAGTAGGTTGTA	34	4	mmu-let-7a-1;mmu-let-7c-2	
TGAGGTAGTAGGTTGTG	14	1	mmu-let-7b	
TGGTTTCACATGGTGGCTTAGAT	1	1	mmu-mir-29b-2	

TAGCACCATTTGAAATCAGTGTTTT	68	1	mmu-mir-29b-2	
ATATAATACAACCTGCTAAG	28	1	mmu-mir-374	
CATCTTTGCCGGTGACAGCA	7	1	mmu-mir-126	
GTTCACAGTGGCTAAGTTCTG	135	1	mmu-mir-27b	
TTGAACTGTCAAGAACCACTGG	1	1		mmu-mir-203
GTTGGCAAGTCTAGAACCACCG	2	1		mmu-mir-182
TTCACAAAGCCCATACACTTTCA	188	1	mmu-mir-350	
TGAGGTAGTAGGTTGTATAGTTTTAGGG	1	1	mmu-let-7a-1	
TCGTACCGTGAGTAATAATGCG	1165	1	mmu-mir-126	
TCCAGCATCAGTGATTTTGT	12	1	mmu-mir-338	
CGTACCGTGAGTAATAATG	28	1	mmu-mir-126	
AGTGGCTAAGTTCTGCA	9	2	mmu-mir-27b	
ATTTGAAATCAGTGT	1	29	mmu-mir-29b-2	
TGAGGTAGTAGGTTGTATGGTTTT	1395	1	mmu-let-7c-2	
AATCACTAACTCCACTGCCATC	259	1	mmu-mir-34b	
GAGGTAGTAGGTTGTATGGTTTTGG	2	1	mmu-let-7c-2	
ATCTATCTATCTATC	1	246025		mmu-mir-350
TAGGCAGTGTAATTAGCT	4	2	mmu-mir-34b	
ACTTAGCAGGTTGTATTATCAT	3	1	mmu-mir-374	
GTAGTAGGTTGTGTGGTTT	1	1	mmu-let-7b	
TGAGGTAGTAGGTTGTGTGGTTTCA	26	1	mmu-let-7b	
GTGAGTTCCAGGACAGCCAGGGCTATACA	1	19153	mmu-mir-706	
TTCACAGTGGCTAAGTTCTGC	32597	1	mmu-mir-27b	
TAGCACCATTTGAAATCA	15	2	mmu-mir-29b-2	
TGAGGTAGTAGGTTGTATAGTTT	2680	2	mmu-let-7a-1	
TCACAGTGGCTAAGTTCTGC	16	1	mmu-mir-27b	
TTCACAGTGGCTAAGTTCTGCAC	25	1	mmu-mir-27b	
CACAGTGGCTAAGTTCTG	5	1	mmu-mir-27b	
AAAGTGCATGCGCTTTGGGACA	49	1	mmu-mir-350	
GTGAAATGTTTAGGACCACTAG	223	1	mmu-mir-203	
ATCTTTGCCGGTGACAGCA	4	1	mmu-mir-126	
TTCACAGTGGCTAAGTTCTG	12686	1	mmu-mir-27b	
TTTGGCAATGGTAGAACT	20	1	mmu-mir-182	
CTAGCACCATTTGAAATCA	1	2	mmu-mir-29b-2	
CCATTTGAAATCAGTGTTT	4	1	mmu-mir-29b-2	
GGTAGTAGGTTGTATGGTT	10	2	mmu-let-7c-2	
TGAGGTAGTAGGTTGTATAGTTTTA	1	1	mmu-let-7a-1	
ATCACTAACTCCACTGCCATC	24	1	mmu-mir-34b	
TTTGGCAATGGTAGAACTCACAC	2295	1	mmu-mir-182	
AGTAGGTTGTATGGT	1	6	mmu-let-7c-2	
TTCTAGCACTTAGCAGGTTGTATTATCATT	1	1	mmu-mir-374	
TCAGGGCAGTGATGTTGCCCCTCCGAAGAT	1	1	mmu-let-7b	
TGAGGTAGTAGGTTGTGTGGTTTCAGGGCA	7	1	mmu-let-7b	
CTTAGCAGGTTGTATTATCAT	5	1	mmu-mir-374	
GTGAAATGTTTAGGAC	1	3	mmu-mir-203	
TCCTGTACTGAGCTGCCCCGAG	152	2	mmu-mir-486	mmu-mir-486
TGAGGTAGTAGGTTGTATGGTTTTGGGCT	1	1	mmu-let-7c-2	
CTGGTTTCACATGGTGGCTTAGATT	7	1	mmu-mir-29b-2	
TTAGCAGGTTGTATTATCATT	1	1	mmu-mir-374	
TGAGGTAGTAGGTTGTATGGTTT	27795	2	mmu-let-7c-2	
CCAACAATATCCTGGTGCTGA	2	1	mmu-mir-338	
TCACAGTGGCTAAGTTC	1	4	mmu-mir-27b	
ATCACTAACTCCACTGCCAT	1	1	mmu-mir-34b	
AGAGCTTAGCTGATTGGT	5	1	mmu-mir-27b	
ACTATACAATCTACTGTCTTTC	4	2	mmu-let-7a-1;mmu-let-7c-2	
CTGGTTTCACATGGTGGCTTAGATTT	6	1	mmu-mir-29b-2	
GAGGTAGTAGGTTGTATAGTT	103	2	mmu-let-7a-1	
TAGTAGGTTGTATGGTT	24	3	mmu-let-7c-2	
AGGCAGTGTAATTAGCTGATTGT	638	1	mmu-mir-34b	
ATATAATACAACCTGCTAAGTG	316	1	mmu-mir-374	
TGAGGTAGTAGGTTGTGTGGTTTCAG	5	1	mmu-let-7b	
TGAGGTAGTAGGTTGTATAG	1312	2	mmu-let-7a-1	
TGAAATGTTTAGGACCACTAG	130	1	mmu-mir-203	
TGAGGTAGTAGGTTGTGT	94	1	mmu-let-7b	
GAAATGTTTAGGACCACTAG	1	1	mmu-mir-203	
AATCACTAACTCCACTGCCATCA	17	1	mmu-mir-34b	
TTCAACAAAATCACTGATGCTGGAG	1	1		mmu-mir-338
CTTAGCAGGTTGTATTA	3	3	mmu-mir-374	mmu-mir-374
TAGGCAGTGTAATTAGCTGATTGTA	2	1	mmu-mir-34b	
TAGCACCATTTGAAATCAGTGTTT	844	1	mmu-mir-29b-2	
TCCAGCATCAGTGATTTTGTT	25	1	mmu-mir-338	
TGAGGTAGTAGGTTGTATGGTTTTG	4	1	mmu-let-7c-2	
CTGACTTCGGCCCCCATGTCAGCAGATGC	1	143		mmu-mir-680-1;mmu-mir-680-2
TCTTTGCCGGTGACAGCA	6	1	mmu-mir-126	
TTAGGGTCACACCCACCACTGGGAGAT	6	1	mmu-let-7a-1	
CTATACAATCTACTGTCTTTC	241	2	mmu-let-7a-1;mmu-let-7c-2	
ATATAATACAACCTGCTAA	5	1	mmu-mir-374	
ATAATACAACCTGCTAAGTGT	1	1	mmu-mir-374	
TCCAGCATCAGTGATTTTGTTGA	10	1	mmu-mir-338	
AGGTAGTAGGTTGTATAGTT	20	2	mmu-let-7a-1	
TTCACAGTGGCTAAGTTCT	3140	1	mmu-mir-27b	
TGTAACAGCAACTCCATGTGGAC	1	1	mmu-mir-194-1	
TATAATACAACCTGCTAAGTG	3	1	mmu-mir-374	
TGAGGTAGTAGGTTGTGTGGTTT	33456	1	mmu-let-7b	
ATTATTACTTTTGGTACGCGC	1	1	mmu-mir-126	
TGAGGTAGTAGGTTGTATAGTTTTAGGGT	9	1	mmu-let-7a-1	
CTCGTACCGTGAGTAATAATG	4	1	mmu-mir-126	
ATAATACAACCTGCTAAGTG	1	2	mmu-mir-374	mmu-mir-374
ACCGTGAGTAATAATGC	1	1	mmu-mir-126	

TAGCACCATTTGAAATCAGTG	352	2	mmu-mir-29b-2
CATTTGAAATCAGTG	1	22	mmu-mir-29b-2
GTGAAATGTTTAGGACCA	1	1	mmu-mir-203
TTGGCAATGGTAGAACTCACA	4	1	mmu-mir-182
TAGTAGGTTGTATAGTTTTAGGG	2	1	mmu-let-7a-1
TTCACAGTGGCTAAGT	4	3	mmu-mir-27b
AGCACCATTTGAAATCAGTGTT	102	2	mmu-mir-29b-2
CGGTCAGCAGCCCAGCGCCA	1	1	mmu-mir-126
TTCACAGTGGCTAAGTTCTGCA	617	1	mmu-mir-27b
TAGTAGGTTGTATAGT	13	2	mmu-let-7a-1
CATTATTACTTTTGGTAC	3	1	mmu-mir-126
ATCACTAACTCCACTGCCA	2	1	mmu-mir-34b
AGAGCTTAGCTGATTGGTGAACAGT	1	1	mmu-mir-27b
GAGGTAGTAGGTTGTGTGGTT	139	1	mmu-let-7b
AAAGTGCATGCGCTTTGGGAC	21	1	mmu-mir-350
TTAGGGTCACACCCACCACTGGGA	2	1	mmu-let-7a-1
GTAGTAGGTTGTATGGT	1	2	mmu-let-7c-2
CTATACAACCTACTGCCTTCC	43	1	mmu-let-7b
GTGAGTAATAATGCG	1	1	mmu-mir-126
GTTCACAGTGGCTAAGTT	2	2	mmu-mir-27b
TAGGCAGTGTAATTAGCTGATTGT	26	1	mmu-mir-34b
TTGGGCTCTGCCCCGCTCTGCGGT	2	1	mmu-let-7c-2
AGGTAGTAGGTTGTATAGT	4	2	mmu-let-7a-1
TGAGGTAGTAGGTTGTGTGGTTTC	299	1	mmu-let-7b

Table S4: Solexa reads from wt and c-kit mutant ovaries matching mouse miRNAs

Read Sequence	#cloned
TAGTAGTTTGTACAGT	2
GTTCACAGTGGCTAAGTTCCGC	3
GCTGTCTCGGAGCCTGGGGCA	1
TAAAGTGCTGACAGTGCA	4
CCTGAACTAGGGGTCTGGAGA	1
TCTACAGTGCACGTGTCT	2
AGCTCGACTCATGGTTTGAACCATTACTTA	1
AGCAGCATTGTACAGGGCTATGAAG	12
ATAGTAGACCGTATAGCGTAC	2
CTAGCACCATTTGAAATCGGTT	1
CAATCACTAACTCCACTGCC	1
TTTCCTCATATCCATTCAGGAGT	1
ACTTAGCAGGTTGTATTATCATT	1
TGAGGGGCAGAGAGCGAGACTT	3
AGCGAGGTTGCCCTTTGTATATT	5
TAGGTAGTTTCATGTTGTTG	44
TTGCACTTGTCCCGGCCTGT	4
AGTAGGTTGTATAGTTA	1
AAACATTCGCGGTGCACTTCTTTT	1
AAGAGGTAGTAGGTTGCATAGT	28
CGTACCGTGAGTAATAATGCG	552
TCGGTTATCATGGTACCGATG	1
AACATGAAGCGCTGCAACAC	1
AGTAGGTTGTATGGTT	2
TGGAATGTAAAGAAGTATGTAT	37
TGTGCAAATCTATGCAAAA	11
CTCCCACATGCAGGGTTTGCA	10
CTGCCAATTCCATAGGTCA	1
ACTGATTTCTTTTGGTGTTCAGAG	10
TTCAAGTAATTCAGGATAGGTTG	87
AGTAGTCTGCACATTGGT	2
CACCATTTGAAATCAGTG	4
AACATTCAACCTGTCGGTGAGT	179
TGTAAACATCCTCGA	3
AATCACATTGCCAGGGATTACCA	3
AACCGTGGCTTTCGATTGTTA	4
TGAGGTAGTAGGTTGTGTG	115
CAGCTACATCTGGCTACTGGGTC	2
TGAGGTAGTAGGTTGTATGG	818
TTTGTGACCTGGTCCACTAACC	2
AATGCACCCGGGCAAGGATTT	1
AAACCGTTACCATTACTGAGT	6
ATTCAAGTAATCCAGGATAG	1
GTCCAGTTTTCCCAGGAATCC	955
CAAAGCACAGGGCCTGCAGAG	2
ATTGGGAACATTTTGCATGC	4
ACCGATTTCTCCTGGTGTTCA	1
CTCCATTTGTTTTGATGATGGA	1
AAGGTGCATCTAGTGCAGATAG	1
ATCACAAGTCAGGTTCTTGGG	6
AAGGAGCTCACAGTCTATTGAG	21
CCCTGAGACCCTAACTTGTGA	343
ACAGATTCGATTCTAGGGGAA	1
CAGTGTAGTTAGCTGATTGC	1
TGTAAACATCCCCGACTGGA	22
CTTGGTACATCTTTGAGTG	1
ACAGAGTTGTCATGTGTCT	1
TTCACAGTGGCTAAGTTCCGC	6342
CCCCCGCCGGGCGTCGGGACCGGGGTCCGG	1
ACCAATATTATTGTGCTGC	2
ATTCAACGCTGTCGGTGAGTTT	1
GCAGGCACAGACAGGCAG	1
CGTGTATTTGACAAGCTGAGTTGGACACT	9
ATAAATTCACTCTAGTGCTT	1
CTGCTGGGGTGGAACCTGGTT	1
TGACCGATTTCTCCTGGTGTT	18
CACGCGGGAACCGAGTCCA	1
GCAGTGTCTTAGCTGGTTGT	1
AGCAGCATTGTACAGGGCTATCAA	32
CCCAGTGTTTAGACTACCTGTTC	166
ACAGGGTAGAACCACGGA	1
TATGGCTTTTCATTCCTATGTGA	13
AGGTAGTTTCATGTTGTTGGG	1
CAACGGAATCCCAAAAGCAG	54
TAACACTGTCTGGTAAAGATG	1296
TGAGGTAGTAAGTTGTATTG	1
TGTTCTGGCAATACCTGTG	1
GCCCCTGGGCCTATCCTAGAA	11
CGTCCTGAGGTTGTTGAGCTG	1
CTGTACAGCCTCCTAGCTTTCC	3
GCTGACAGTGCAGATA	1
AAGGGATTCTGATGTTGGTCACACTC	91
GGCTTCTTTACAGTGCTGCCTTG	12
TCACATTGCCAGGGATTTCC	8
TCCAGTTTTCCCAGGAATCC	4
AATCACTAACTCCACTGCCAT	12
AGCTACATCTGGCTACTGGG	6
AACATTCATTGTTGTCGGTGGGTT	89
ATACTGCCGGGTAATGATG	3
GTGTTCAGACTACCTGTTC	2
GTATGTGCCTTTGGACTACAT	28
TAGCAGCACATAATGGTTTGT	228
AAGAGAGGTTATCCTTTGTGTGT	1
TCCTTCATTCCACCGGAGTCTG	5
CCTCCCACATGCAGGGTTTGC	1
AGCTTCTTTACAGTGTTGCCTT	4
TAGTAGGTTGCATAGT	5
ATCACAAGTCAGGTTCTTGGGAC	12
TAGGTAGTTTCATGTTGTT	5
AATCCTTTGTCCCTGGGTGAAAATG	7
TAGCAGCGGGAACAGTACTGC	89
ATGACCTATGATTTGACAGAC	1
CCGTAGATCCGATCTTGTG	2
AAAAGCTGGGTTGAGAGGG	2
AATGACACGATCACTCCCGTTGAGTG	1
AGGCAAGATGCTGGCATAGCT	13
ACTGCCCTGCTGATGGAT	1
CAAAGAATTCTCCTTTTGGGCTT	20
TTGGTCCCCTTCAACCAGCTGTA	2
CAGCAGCACACTGTGGTTTG	1
CCAGTGTTCAGACTACCTGTTC	1
TCTCTGGGCCTGTGTCTTAGGCT	3
ATCATAGAGGAACATCCACTTT	5
AACATTCATTGCTGT	1
CTGAGATGAAGCACTGTAGCTC	17
CTCGGGGATCATCATGTCACGAGA	2
ATCATGATGGGCTCCTCGGTGTT	3
TAGCAGCACGTAAATAT	1
TAACAGTCTACAGCCATGGT	17
AATACTGCCGGGTAATGATGG	1
CTGTACAGGCCACTGCCTTGCCA	8
TTTTCCCAGGAATCCCT	17
TAGCTTATCAGACTGATGT	106
GGAGCTTACAATCTAGCTGG	2
TGCCTGTCTACACTTGCTGTG	80
AGGAAGCCCTGGAGGGGCTGGAGGTGA	1
CACTAGATTGTGAGCTGCTGGA	17
TATGGCTTTTTATTCCTATGTG	233
TCCTGACTCCAGGTCCTGTGT	27
TAGCAGCACGTAAATATTGGC	1765
TGGCTCAGTTCAGCAGGAACAGG	2
GCCTGTCTACACTTGCTGTG	11
TGTGACAGATTGATAACTGA	26
TCACAGTGAACCGGTCTCTTTTT	1
TATGTGCCTTTGGACTACAT	18
AGCAGCACGTAAATATTGGC	3
TAACACTGTCTGGTAAC	6
TAGCACCATCTGAAATCGGTTA	691
AGCTGGTGTTGTGAATCAGGCCGTTG	9
AATGACACGATCACTCCCGTTGAGT	80
AACATTCGCGGTGCACTTCTTT	1
ACAAGCTTGTGTCTATAGGTAT	11
AAGCCCTTACCCCAAAAAGC	12
AATACTGCCTGGTAATGATGA	26
AGTTGTGTGTGCATGTTCAT	1
TATGGCTTTTCATTCCTATG	1
TCAAGTAATCCAGGATAGG	6
CTTTCAGTCGGATGTTTGCAGC	319
TAGTAGTTTGTGCTGTT	4
CGGCTACTTCACAACACCAGGGC	1
AGGGGTGCTATCTGTGATTGAGGGACAT	4
TTTCTTGGGACTCGGAAGC	1
TGTCTGGTAATGCCG	1
AGTAGTTTGTACAGTT	1
AAGTGCTGTTCGTGCAGGTAGT	1
GCACCATTTGAAATCAGTGTT	30
CCCCCAGGTGTGATTCTGATTTGT	9
AGAGGTAGTAGGTTGCA	3
CCCTGAGACCCTAACTTGTGAG	1
TTGTGCTTGATCTAACCATGTGGT	1
TTTGGCACTAGCACATTTTTGCTT	1
AGTAGGTTGTATAGTTT	1
TACGGTGAGCCTGTCATTATTC	1
TGGTAGACTATGGAACGTAGG	16
ACATTCATTGTTGTCGGTGGGT	1
CAAAGTGCTTACAGTGCAGGT	385
AACATTCATTGCTGTCGGTGGGT	193
ATGGCACTGGTAGAATTCACT	1
AACACTGTCTGGTAAAGATGGC	6
TGTAAACATCCTACACTCTCA	546
TGGCTCAGTTCAGCAGGA	2
ACAGGTGAGGTTCTTGGGAGCCT	1
TGAGGTAGTAGGTTGTATGGTT	35877
TAGCACCATTTGAAATC	3
GCTCTTTTCACATTGTGCTAC	1
ACATTCAACGCTGTCGGTGAGTT	1
TGTAAACATCCTACACTCAG	50
CGAAGGAGCCTCGGTTGGCCCCGGATAG	1
TATACGACCTGCTGCCTTTCT	1
CTCGTACCGTGAGTAATAATGC	26
CACAGCTCCCATCTCAGAACA	5
TTGGTGTACTGTGTGTGA	1
TTCCTATGCATATACTTC	2
CCACAGGGTAGAACCACGGA	10
TATTGGGAACATTTTGCATGC	18
AAACCGTTACCATTACTGAGTTTAGTA	5
CAGTGCAATGGTATTGTCAAAG	6
AACTGGCCTACAAAGTCCC	9
CCCTGGGCTGTAACCATGTTGTAGTGTG	2
CAGCTTCTTTACAGTGTTGCCTT	3
CTGAGGTAGTAGTTTGTACAG	1
CCGTTACCATTACTGAGTT	1
GGTCCGGTGCGGAGAGCCGTTCGTCTT	1
AGAGGTATAGCGCATGGGAAGAT	2
CCACAGGGTAGAACCACGGACA	22
GTAGTAGTTTGTGCTGT	1
TAGCACCATTTGAAATCGGT	23
GGGTGCTATCTGTGATTGAGGGAC	1
AGCAATGTCAGCGGTGCCT	5
GTGGATATTCCTTCTATGGTT	1
TAATCCAGGATAGGC	3
AACATTCAACGCTGTCGGTGAGTTTG	8
TGTAAACATCCTCGACTG	4
ATTTGAAATCAGTGTTT	1
TGAGAACTGAATTCCATAGGCTG	28
GCTACATCTGGCTACTGGGT	1
TCACATTGCCAGGGATTTC	7
TGCGGGGCTAGGGCTAACAGCA	59
CCCAGTGTTTAGACTACCTGT	39
CCTTGGCTCTAGACTGCTTACT	1
AGCAGCATTGTACAGGGCTA	102
TAGGTAGTTTCATGTTGTTGGGAT	12
CATCACTCGACTCCT	1
GTCCAGTTTTCCCAGGAATCCCTTG	1
AGCTACATTGTCTGCTGGGTTT	93
CAAAACATGAAGCGCTGCAA	6
CACATTACACGGTCGACCTCT	3
AGAGGTAGTAGGTTGCATAGTTTTA	5
TGAGGTAGTAAGTTGTATT	1
TGGGGAGAGCACCTCGCAGGCGACTACTTA	1
TAGAAGCAAAGTGATACTGGGGAGG	1
AATCCTTTGTCCCTGGGTGAAAATGC	64
TGGGTCTTTGCGGGCAAGATGAGA	1
CCTCATCTGTCTGTTGGGCT	1
AGAGGTAGTAGGTTGCATAGTTTTAG	1
AATCACTAACCACACAGCCAGGTA	3
CGTAGATCCGATCTTGTG	2
GTAAGTGCCTGCATGTATATGCGT	2
CACAGGGTAGAACCACGGA	1
CTAGCACCATCTGAAATCGGTTA	32
TCCCCCAGGTGTGATTCTGATTTGTT	1
AATCACTAACCACACAGCCAGGTAA	1
CTCTAAGCCAAGTTTCAGTTCA	1
TCTACAGTGCACGTGTCTCC	26
TGTGCTTGATCTAACCATGTG	2
TGTAAACATCCTACACTCTCAG	284
TAAGGCTCCTTCCTGTGCTTGCA	3
TAGGTAGTTTCCTGTTGTTGGG	280
TACAGTATAGATGATGTA	3
CCATCTTCCAGTGCAGTGTTGG	1
CCGGAGTCGGGAAAAGCTGGGTTGAGAGGG	1
TGAGGGTCTATGATACCACCCGGTACAGG	1
CTTTCAGTCGGATGTTTACA	110
ACCATTTGAAATCAGTGTT	7
CATCCCTTGCATGGTGGA	1
TCACATTGCCAGGGATTTCCAA	9
CGCATCCCCTAGGGCATTGGTGTA	8
TGACCTATGAATTGACAGCC	3
GCAGTGCAATGTTAAAAGGGCAT	1
AACATTCAACGCTGTCGGTG	49
TATGTAGTATGGTCCACATCTT	2
TCTCCCAACCCTTGTACCAGTGCTGTGC	1
AGTAGGTTGTGTGGTTT	1
TGAGGTAGTAAGTTGTATTGT	24
CAACAAATCACAGTCTGCCA	1
CTGTGCTGACGTTTGATTCTTGAGC	5
GAACAGTGCCTTTCTGTGTAGG	1
CACATCATGGTTTACA	1
AATGCCCCTAAAAATCCTTAT	2
CCTGTGCTGACGTTTGATTCTTGAGCAG	1
AAGCTGCCAGTTGAAGA	1
GTAGTGTTTCCTACTTTAT	8
TAGCAGCACATCATGGTTTACATACT	3
AGGTAGTAGGTTGCATAGT	1
AGCGCCTCGGCGACAGAGCCG	1
GTAAGTGCCTGCATGTATAT	1
AAGGAGCTTACAATCTAGC	9
TTGGGAACATTTTGCATGCAT	2
TGAGGTAGGAGGTTGTATAGTTGAGGAAGA	1
TAAAGTGCTTATAGTG	3
AGGCAAGATGCTGGCATAG	2
CCAGTGGGGCTGCTGTTAT	2
AGTAGTTTGTGCTGTT	2
ACCGCACTGTGGGTACTTGCT	1
TAGCAGCGGGAACAGTACTGCA	185
CCCATAAAGTAGAAAGCACTACTA	1
ATCCCTGAGTGTATGTGGTGAA	2
GGAGCCTCGGTTGGCCCCGGATAGCCGGGT	1
AGTGTTTCCTACTTTATGGATGA	13
ATCATAGAGGAACATCCACTT	3
CTGTACAGGCCACTGCCTTGC	42
ACAGCAGGCACAGACAGGCAG	19
TGATATGTTTGATATATTAGGTTG	38
TTTGGCAGTGTCTTAGCTGGTTGT	2
TTCTGCCATGGAAAACGTCAGCTGAAGA	2
CGGTGCTGGTGGAGCAGTGAGCACGC	1
AAGCCCTTACCCCAAAAAGT	1
ACATCCTACACTCTCAGC	1
TCGAGGAGCTCACAGT	1
TGAGAACTGAATTCCATGGGTT	7
ATCACATTGCCAGGGATTTCCA	5799
TTCAAGTAATCCAGGAT	3
AATTCGTATCTAGGGGAAT	1
GCAAAGCACAGGGCCTGCAGAGAGGTA	1
CATCGGGAATGTCGTGT	2
TGTAAACATCCCCGACTG	1
AGTAATCCAGGATAGGC	4
AGCACATCATGGTTT	2
TATTGCACTTGTCCCGGCCTG	41
GCAGCGCCGAAGGAGCCTCGGTTGGCCCC	1
ACAAGTCAGGTTCTTGGGAC	47
GCTGGTGTTGTGAATCAGGCCGT	3
CCCTGAGACCCTAACTTGTGAT	1
ATCCCTGAGTGTATGTGGTGA	6
GTAGTAGGTTGTATAGTT	5
CTCGTGTCTTGTGTTGCAG	1
TAATTGAATGGCGCCACTAGGGTTGTG	1
TTGGCAGTGTCTTAGCTGGTTG	1
GGCAGTGTCTTAGCTGGTTGT	1
TCCCTGAGACCCTAACTTGTG	3514
GCTCGACTCATGGTTTGAAC	3
TGAGATGAAGCACTGTAGC	53
TCGTACCGTGAGTAATAAT	23
CAGCTTCTTTACAGTGTTGCCTTGT	16
TGAATTACCGAAGGGCCATAA	2
AATGCCCCTAAAAATCC	1
AGGTAGTAGGTTGTGTGGTT	3
ATACTGCCTGGTAATGATGAC	7
TGTCAGTTTGTCAAATACCCCA	24
CTATACAATCTATTGCCTTCCC	1
TTGAGGTAGTAGGTTGTATGG	5
TTATCAGAATCTCCAGGGGTA	18
GCCTGCTGGGGTGGAAC	1
TAGCTTATCAGACTGATGTTGACTGTTG	6
ATCACTAACCACACAGCCAGG	1
TGAGGTTGGTGTACTGTGTGTGAG	6
AGAGCTTAGCTGATTGGTGAACA	2
TATTGCTTAAGAATACGCGTAGT	1
CATTGCACTTGTCTCGGTCTGA	1343
CAGCTTCTTTACAGTGTTGCCT	4
AATGTTGCTCGGTGAACCCCT	7
TAATCCAGGATAGGCT	16
CTATACAATCTACTGTCTTT	12
CAGTAGTCTGCACATTGGT	76
TCACATTGCCAGGGATTTCCA	56
CTGTCTCGGAGCCTGGGGCA	1
GTCTTAGCTGGTTGT	1
AGTTTTCCCAGGAATCC	3
CTTGTACTGCCCTGCTGATGGAT	1
TGAGGTAGTAGATTGTA	1
CACCATTTGAAATCAGTGTT	7
TCGTACCGTGAGTAATAATG	100
ATCACTAACCACACAGCCAG	1
TAACACTGTCTGGTAACGA	31
CATTGCACTTGTCTCGGTCT	250
TGTGCCTTTGGACTACATCGTG	1
CGGGGCAGCTCAGTACAGGAT	4
AAGTTGCCCGCGTGTTTTTC	1
ATCACATTGCCAGGGATTAC	654
AGAGGTAGTAGGTTGCATAGT	1770
AAACCGTTACCATTACTGAGTTTAG	24
TAAGGTGCATCTAGTG	1
TAGCTTATCAGACTGATGTTGACTG	1
TACCACAGGGTAGAACCACG	29
AGGGGTGCTATCTGTGATTGAGGGACA	69
TCTGGCTCCGTGTCTTCACT	23
TGAGGTAGTAGGTTGTATAGTTTA	7
AATGCACCTGGGCAAGGGTTC	43
CACTAGATTGTGAGCTGCTGG	2
CCACTGCCCCAGGTGCTGCTGGG	1
AAGGAGCTCACAGTCTATTGA	7
AGCAGCACAGAAATATTGGCA	2
TGAGGTAGTAGTTTGTA	2
ATCACTAACCACACAGCCAGGT	9
CTAACACTGTCTGGTAACGATGTT	3
CCCTGAGACCCTTTAACCTGTG	2
ACAAACATGGTGCACTTCTTT	1
GTGAAATGTTTAGGACCACTAGA	5
CAAAGTGCTGTTCGTGCAGGTAG	2902
CATCTTACCGGACAGTGCTGG	3
CAAAGTGCTGTTCGTG	1
TTTGGCAATGGTAGAACTCACACC	2
TAACACTGTCTGGTAACGATG	963
ATCGTAGAGGAAAATCCACGTT	6
AGCTACATCTGGCTACTGGGTC	18
CCGTCTCAGTTACTTTATAGCC	1
TTCACAGTGGCTAAGTTC	284
AACACTGTCTGGTAACGATG	2
TCTCACACAGAAATCGCACCCGT	43
GGCTCAGTAGCCAGTGTAGATCC	1
AAGAGGTAGTAGGTTGCATAGTTT	1
AAAAGCTGGGTTGAGAGGGCGAAAA	39
ACATTGCCAGGGATTTCCA	2
ATTCCTGGAAATACTGTTCTT	38
TCGACAGCACGACACTGCCTTC	1
TTTCAGTCGGATGTTTGCAG	38
TGGCTCAGTTCAGCAGGAACAG	362
CAGCATTGTACAGGGCTATCA	1
TGAGTATTACATGGCCAATCT	1
GAAAAGCTGGGTTGAGAGGGCGA	3
GAGGTAGTAGGTTGTATGGT	3
GTAAACATCCTCGACTGGAAGCT	8
TGGAGAGAAAGGCAGTTCCTGA	15
ATGAGGTAGTAGATTGTATAGTT	6
CGTGAGTGGCACACGCAGGGCAACTCGA	1
TAGCAGCACATAATGGTTTG	1
TTATCAGAATCTCCAGGGGTAC	187
GCTACTTCACAACACCAGGGC	19
CTGAAGCTCAGAGGGCTCTGAT	11
CTAACACTGTCTGGTAAAGATGG	1
TACTCAGTAAGGCATTGTTC	4
CAGCTTCTTTACAGTGTTGC	2
GCTCGACTCATGGTTTGAA	2
CACGTAAATATTGGCG	12
CTCGGGAGGGGCGGGAG	5
CCAGTGTTTAGACTACCTGTTC	1
CGAATCATTATTTGCTGCTC	1
CCCCTGGGCCTATCCTAGAA	4
TACCCTGTAGATCCGAATTTGTG	204
GCACATCATGGTTTAC	3
TACAGTACTGTGATAACT	1
CAAGAGGGCCTTTCTG	1
TGGCTTTTTATTCCTATGTG	3
TTTTGCGATGTGTTCCTAATATGTG	2
ATATAATACAACCTGCTAAGTGT	40
TCAGTGCATCACAGAACTTTG	37
GCAGCATTGTACAGGGCTATGA	6
GTTCACAGTGGCTAAGTTCCG	1
AGGCAGTGCATTGCTAGCTGGCTGT	2
TTTGGCACTAGCACATTTTTG	2
ACCAATATTATTGTGCTGCT	3
GTACAGTACTGTGATAACTGA	8
TTCCCTTTGTCATCCTATGCC	89
GCACCATCTGAAATCGGTTA	2
CTGTTGCCACTAACCTCAACC	26
TAATCCTTGCTATCTGGGTGCTTAGTGC	3
AGGTAGTAGATTGTATAGTTT	1
AGCAGCATTGTACAGGGCTATCAAAG	1
TAGGTAGTTTCCTGTTGTTGGGA	56
ACCACAGGGTAGAACCACGG	51
AAGCTGCCAGTTGAAGAACTG	91
AACATTCAACGCTGTCGGTGAGTTT	2732
TGTCTGGTAATGCCGT	3
CTATGCCAACATATTGCCATC	1
CATCCCTTGCATGGTGGAGGG	1
AACATTCATTGTTGTCGGTGGGTTGT	27
TGTGCAAATCTATGCAAAACTG	140
TAAAATGGGTTCCTAGGA	2
CCGGTGCGGAGAGCCGTTCGTCTTGGGAA	1
CTAGATTGTGAGCTGCTGGAG	2
TGAAAGGTGCCATACTATGTATA	1
CATTCAACGCTGTCGGTGAGTTT	2
TGAGGTAGTAGTTTGTACAGTT	524
TAGGTAGTTTCATGTTGTTGG	179
AACATTCAACCTGTCGGTGAG	16
CAGTGCAATGTTAAAAGGGCATC	5
TATTGCACTTGTCCCGGCCTGA	24
TCCCTGAGGAGCCCTTTGAGCCT	823
CGTGTATTTGACAAGCTGAGTTG	2
CCCAGTGTTCAGACTACCTGTT	918
TATGTGCCTTTGGACTAC	1
AGCAGCATTGTACAGGG	4
TTTTGCGATGTGTTCCTAATATGTGCTATA	1
ACAGTCTACAGCCATGGTC	2
GCTATTTCACGACACCAGGGT	1
GCGCCGAAGGAGCCTCGGTTGGCCCCGGA	2
GCACAGAAATATTGGC	2
TTGTGCTTGATCTAACCATGTGCTT	7
CTATACGGCCTCCTAGCTTT	3
TTTGGCACTAGCACATTTTTGC	2
CCCCCAGGTGTGATTCTGATTT	2
CGTTACCATTACTGAGTT	1
TCTCACACAGAAATCGCACCC	1
AACATTCATTGCTGTCGGT	10
CTTGGTACATCTTTGAGTGA	3
CATGCCTTGAGTGTAGGACCGTT	28
CTTGCGAGGTATGAGAAAAA	1
TATGCAAGGGCAAGCTCTCTT	9
GCTGGTTTCACATGGTGGCTTAGA	2
CAACCCTAGGAGGGGGTGCCATTCACATAG	2
TAATACTGCCTGGTAATGAT	513
CTGTGCGTGTGACAGCGGCTGA	370
TTTGAACCATCACTCGACTCCT	24
TACCACAGGGTAGAACCACGGA	191
AACAAACATGGTGCACTTCTTT	11
GATTCCTGGAAATACTGTTCTT	1
CAAATTCGTATCTAGGGGA	1
AAACATCCTCGACTGGAAGC	1
GCGCCACTAGGGTTGTG	1
CTGGCTCAGTTCAGCAGGAACAG	1
AAACCGTTACCATTACTGAGTTTAGTAA	1
TGCTGTTCGTGCAGGTAG	1
TGAGAACTGAATTCCATAGGCTGT	113
AATCACACAAAGGCAACTTTTGT	2
GCTTCTTTACAGTGTTGCCTTG	1
CTCGTACCGTGAGTA	1
AGCAGCATTGTACAGGGCTATGAAAGA	2
TCCTGTACTGAGCTGCCCCG	5
CCCGTAGATCCGATCTTGTG	5
GCACTGAGATGGGAGTGGTGTAA	38
TAGGTTATCCGTGTTGCCTTCGCTTTAT	2
CAATGTTAAAAGGGCA	1
CTGTGGGCCACCTAGTCACC	1
TCGGTTGGCCCCGGATAGCCGGGT	1
TTGTGCTTGATCTAACCATGT	1276
CAAATTCGTATCTAGGGGAA	2
CTATGCCAACATATTGCCATCT	1
TGCTTATAGTGCAGGTAGT	1
TCCCTGAGGAGCCCTTTGA	133
AGCTGGTGTTGTGAATCAGGCCGTT	1371
TGAGGTAGTAGGTTGTATAGTTTT	20
AGCTTCTTTACAGTGCTGCCTTG	19
GACCGATTTCTCCTGGTGTT	7
AATGACACGATCACTCCCGTTGAGTGG	2
CCAAAGTGCTGTTCGTGCAGGTAG	2
TCCCTGAGGAGCCCTTTGAGCC	239
TGAGGTAGTAGTTTGTG	30
CAGTGCAATGGTATTGTC	1
CAAGGGTCACCCTCTGACTCTGT	1
CCGCACTGTGGGTACTTGCTG	15
CAAAGTGCTTACAGTG	3
TTTTGCGATGTGTTCCT	4
CCGGGGTCCGGTGCGGAGAGCCGTTC	1
ACACACCCAGCTAACCTTTTT	1
CTGGGAGAAGGCTGTTTACTC	1
ACTGCTGAGCTAGCACTTCCCG	2
AGGTGCATCTAGTGCAGATAG	1
ATTGGGAACATTTTGCATGCAT	27
GCCTGTCTACACTTGCTGT	1
TAAAGTGCTTATAGTGCA	4
AGTGTTTCCTACTTTATG	3
TGGCTCAGTTCAGCAGGAAC	111
ATGAGGTAGTAGGTTGTATAGTT	11
TCTGGCTCCGTGTCTTCACTCC	77
TAATCCTTGCTATCTGGGTGCTTAGT	2
CAGTGCAATGATGAAAGGG	1
CTGTAGATCCGAATTTGTG	2
GTCCAGTTTTCCCAGG	6
TAGGTAGTTTCATGTTGT	3
TAAGTCACTAGTGGTTCCGTTT	6
TAATCCTTGCTATCTGGGTGCTT	1
TAGCTTATCAGACTGA	23
CGTACCGTGAGTAATAATGCGC	3
CCTGAACTAGGGGTCTGGAG	1
CTACATCTGGCTACTGGGTCTC	1
TCCGGGGCTGAGTTCTGTGCAC	1
CCCCCAGGTGTGATTCTGATT	1
TAGCAGCACATCATGGTTT	104
TTGCACTTGTCTCGGTCTGA	1
TCAGTGCATGACAGAACTTGGG	101
AAGGAGCTTACAATCTAG	17
AATGACACGATCACTCCCGTTGAG	5
AGAGGCTGGCCGTGATGAATTC	2
TGAGGTTGGTGTACTGTGTGTGAGT	19
CTGGGTTGAGAGGGCGA	1
TATCCCTGTGCTAGTTCCATGTT	1
AGCACAGAAATATTGGCA	1
TGAGCGCCTCGGCGACAGA	1
CAGTGCAATGTTAAAAGG	18
TCTCGCTGGGGCCTCCA	83
CAGCAGCAATTCATGTTTTG	260
CTGCACTTGGATTTCGTTCCC	3
CAGTGCAATGGTATTGTCAAAGCAT	29
AATTGCACGGTATCCATCTGTAA	1
CAGCGCCGAAGGAGCCTCGGTTGGCCCC	6
CGAATGTTGCTCGGTGAACCCCTTT	1
AGTTTTGCAGGTTTGCATCCAGCT	1
AAGGGATTCTGATGTTGGTCACAC	3
AATCCTTGGAACCTAGGTGTGAAT	27
TCGGATCCGTCTGAGCTTGG	3
CTGCAGTGAGGGCACTTGTAGCATT	1
TGCTTATAGTGCAGGTAG	1
GAGTATTGTTTCCACTGCCTGG	4
AAAGTGCTGTTCGTGCAGGTAG	5
CAGCATTGTACAGGGCT	1
TGAGGTAGTAGTTTGTGCTGTT	2258
TACCACAGGGTAGAACCACGG	93
CCATCCCACTCAGTCGTGCCC	8
TGAGGTAGTAGGTTGTGTGGT	3202
TAATGCCCCTAAAAATC	6
TAATACTGTCTGGTAATGCCG	1846
GAATGGCGCCACTAGGGTTGTG	1
GAGGTAGTAGGTTGTGTGGT	1
TTGTACAGGGCTATGA	3
CAGTGGCTAAGTTCTGC	3
TCCAGTTTTCCCAGGAATCCCT	91
TCCCTGAGACCCTAACT	1
TATTGGGAACATTTTGCATAAA	1
AATCAGCAAGTATACTGCCCTA	4
GACCGATTTCTCCTGGTGTTCAGAGTC	1
TCACAGCTCTGGTCCTTGGAGC	1
TGTAACGTTGTCTATATA	1
AGCAGCATTGTACAGGGCTATGAAA	177
CAGTGCAATGGTATTGTCAAAGC	10
AACATTCAACCTGTCGGTGAGTTTG	12
TAAGGTGCATCTAGTGCAGA	82
CAAACCACACTGTGGTGTTAG	2
TCTGCCTGTCTACACTTGCTGT	2
CACCCGTAGAACCGACCTTGCG	241
AATCATACACGGTTGACCTAT	1
TATTGCACTTGTCCCGGCCTGT	822
CCCAGTGTTCAGACTACCTGTTCA	4
CATCGGGAATGTCGTGTCC	1
AGGGACTTTCAGGGGCAGCTGTGT	2
TTGCACTTGTCCCGGCCTGTT	1
TGAGGTAGTAGATTGTATAGTTGT	2
TTCAGCTCCTATATGATGCCTTT	1
TTCAAGTAATTCAGG	2
TCGGTTGGCCCCGGATAGCCG	1
ACCCTGTAGAACCGAATTTGT	7
CCCTGAGACCCTAACTTGTG	60
CAGTGCAATAGTATTGTCAAAGCA	15
CAACGGAATCCCAAAAGCAGCT	263
TGATCTAGCCAAAGCCTGACTG	2
CTAACACTGTCTGGTAAAGAT	2
TGTGAAAAGTAAGAAAGC	1
GGCTTCTTTACAGTGCTGCCTT	1
ATGTATGTGTGCATGTGCATGT	31
TAGCTTATCAGACTGAT	68
GCACATTACACGGTCGACCTC	1
AGTGCACGTGTCTCCAGT	1
TGTGCAAATCCATGCAAAA	238
ACCGATTTCTCCTGGTGTTCAGA	12
CTGAAGCTCAGAGGGCTCTGATT	5
CTAGACTGAGGCTCCTTG	1
GTGCATTGTAGTTGCATTGCA	1
TAAACATCCTCGACTGGAAGC	3
TAGCTTATCAGACTGATGTTGACT	1218
GTGAAATGTTTAGGACCACTA	3
TTTGGCAATGGTAGAACTCA	1
AACACACCCAGCTAACCTTTTT	7
TGTAAACATCCTCGACTGGA	56
GCTCGACTCATGGTTTGAACCA	3
TGAGAGATGCCATTCTATGTAG	1
CACATTGCCAGGGATTTCCA	3
CAAGCTCGTGTCTGTGGGTC	13
TAAAGTGCTTATAGTGCAGG	35
AAGTGCTGACAGTGCAGATA	1
TCCGAGCCTGGGTCTCCCTCTT	27
TAAAGTGCTGACAGTGCAGATA	364
CTGAGACCCTAACTTGTG	3
TAGCACCATTTGAAAT	1
AAGGGATTCTGATGTTGGTC	7
GCGCCGAAGGAGCCTCGGTTGGCCCCGGAT	3
AGGTTACCCGAGCAACTTTGC	2
CTCGTACCGTGAGTAAT	1
ATGCAGTCCACGGGCATATA	1
TCTTTGGTTATCTAGCTGTATGA	2
CCCAGTGTTCAGACTACCTGTTC	818
CAGTTATCACAGTGCTGATGCTGTCC	3
CCTGAGGAGCCCTTTGAGCCT	1
ACACTGTCTGGTAAAGATGGC	7
AGCTACATTGTCTGCTGGGT	2
TCGAGGAGCTCACAGTCTAGTA	76
TTAGCCGCTGAAATAGATGG	1
ATGTTGAAAAGGTGCAGGCCATACTGTG	1
AACATCCTACACTCTCAGC	4
CCTGAGACCCTAACTTGTG	3
AGGCAGTGTAATTAGCTGATTG	3
ACATTCAACCTGTCGGTGAGT	1
CGGCGTCATGCAGGAGTTGATT	7
GGTAGTAGGTTGTATAGT	4
GCAGCAGCAATTCATGTTTTG	1
AACACTGTCTGGTAAAGATG	4
TGAAACATACACGGGAAACCT	4
AAAAGCTGGGTTGAGAGGGCG	9
TGGCTCAGTTCAGCAGGAACA	103
AATCCTTGGAACCTAGGTGTGAATGCT	98
TGAATTACCGAAGGGCCATAAA	1
ACAGCAGGCACAGACAGGC	2
TAACACTGTCTGGTAACGATGTT	7558
CTAAAGGTACAGTACTGTGATAACTGAA	1
TGAAGCTCAGAGGGCTCTGATT	1
TTATTGCTTAAGAATACGCGTAG	1
AACATTCAACGCTGTCGGT	31
AATCCTTGCTATCTGGGTGCTTAGTGCT	7
AAACAAACATGGTGCACTTCT	1
TCGGGGATCATCATGTCACGAG	1
TGTCTCGGAGCCTGGGGCA	4
AGTTCTTCAGTGGCAAGCTT	14
AGCTGGTGTTGTGAATC	3
TAAGGTGCATCTAGTGCAGATAGTGAAGTA	1
CTGTTGCCACTAACCTCAACCT	9
ACTGTACAGGCCACTGCCTTGC	3
TTCAAGTAATTCAGGATAGGTTGT	11
TGAGGTAGTAGGTTGTATGGT	7726
GTAGTAGGTTGTGTGGTT	2
TAGCAGCGGGAACAGTACTGCAGT	61
AATACTGCCTGGTAATGATGACG	2
TTGGTCCCCTTCAACCAGCTGTAG	1
CAGCACACTGTGGTTTGT	2
CCAAGATAAATTCACTCTAGTGCTT	1
ACATTCAACGCTGTCGGTGA	2
CATCTTACTGGGCAGCATT	1
GCTACATTGTCTGCTGGGTTT	1
TGCTGTTCGTGCAGGTAGT	2
AGATCAGAAGGTGACTGTGGC	2
AAAACATGAAGCGCTGCAA	1
AATGGCGCCACTAGGGTT	2
CCGCACTGTGGGTACTTGC	3
TTATCAGAATCTCCAGGGGTACT	18
CAAGCTCGTGTCTGTGGGTCCGT	15
TCCTTCATTCCACCGGAG	1
GAGGTTGGTGTACTGTGTGTGA	12
TCGGTTGGCCCCGGATAGCCGGGTCCCCG	13
ATCACAAGTCAGGTTCTTGG	4
GCTGACCCCTAGTCCAGTGCTT	4
ACCAATATTATTGTGCTGCTTT	12
CGTCTTACCCAGCAGTGTTTGG	1
AATCCTTGGAACCTAGGTGTGAA	9
ATCATCGTCTCAAATGAGTCT	1
TGCTATGCCAACATATTGCCA	71
GCAAAGCACAGGGCCTGCAGAGAGG	11
TCAGACTGATGTTGAC	1
TTGTGACAGATTGATAACTGA	2
TGTCACTCGGCTCGGCCCACTA	1
TACAGCAGGCACAGACAGGCAGT	264
TCCTTCATTCCACCGGAGTCT	2
AATGCCCCTAAAAATCCTTATT	1
TAACAGTCTCCAGTCACGGC	3
CCAAAGTGCTGTTCGTGCAGGTAGT	1
TAACACTGTCTGGTAAAGAT	111
ATGAGGTAGTAGATTGTATAGT	1
CTGTACCACCTTGTCGGA	7
CGCCGAAGGAGCCTCGGTTGGCCCCGGA	2
TTGTGTCAATATGCGATGATGT	24
TGGCAGTGTCTTAGCTGGTTGTTGT	17
AAGCTGCCAGTTGAAGAACT	147
AAGGAGCTCACAGTCTATTGAGT	3
TGAGGTAGTAGTTTGTGCTGTTGGTCGG	1

GCTGGTTTCATATGGTGGTTTAGATTT	2
TACAGCAGGCACAGACA	1
CTCGGGGATCATCATGTCAC	1
CAGTGCACTACAGAACTTTG	4
ACAAGTCAGGTTCTTGG	1
CAACGGAATCCCAAAAGCAGC	108
TACTCAGTAAGGCATTGT	1
AGTTCTTCAGTGGCAAGCTTTA	314
CAGTGCACTACAGAACTTTGTC	1
TATGTAACATGGTCCACTAA	1
AACAAATCACAGTCTGCCA	1
TGGAATGTAAAGAAGTATGT	1
TCAACGCTGTCGGTGAGTT	1
ATAGCTTATCAGACTGATGTT	2
CCCAGTGTTTAGACTAC	1
TCCCCCAGGTGTGATTCTGATTTGTTT	6
TACTGCCGGGTAATGATG	1
TAGCAGCACAGAAATATTGG	3
ATACTGCATCAGGAACTGACTGGA	1
TAGCAGCACGTAAATATTGGCGTTA	17
TCAGCTCCTATATGATGCCT	2
CGCATCCCCTAGGGCATTGGTGTAA	7
TGTAGGGATGGAAGCCATGAAA	5
TTAGGGTCATACCCCATCTTGGAGATAA	2
AAGAGGTATAGCGCATGGGA	4
TAGTAGACCGTATAGCGTACGCTT	1
AACATTCAACGCTGTCGGTGAGTTTGGGA	1
CTTTCAGTCAGATGTTTGCTG	8
GGTTTTGTTTGGGTTTGTT	1
TCCGTCTCAGTTACTTTATAG	2
CCGTCCTGAGGTTGTTGAGCTGT	38
AAGAGGTAGTAGGTTGCATAG	3
TGAGGTAGTAGATTGTATA	6
TAGCAGCACATCATGGTT	8
AAAACATGAAGCGCTGCAACAC	3
CACATTGCCAGGGATTACCAC	1
AAGCTCGGTCTGAGGCCCCTCAGT	2
CTTTCAGTCGGATGTTTAC	21
TTTCATTGATCTTGGTGTCCTCAAATTGA	1
AAGCTGCCAGTTGAAGAAC	52
CCCAGTGTTTAGACTACCTG	3
GCACTGAGATGGGAGTGGTGTA	97
CGGCGTCATGCAGGAGTTGAT	4
GGCTTCTTTACAGTGCTGCCT	1
TATTGCACTTGTCCCGG	1
TTCAAGTAATCCAGGATAGGCT	31185
AGTGTTTCCTACTTTATGGAT	116
CAAAGTGCTTACAGTGCAGGTA	134
GAATGGCGCCACTAGGGTTGT	1
CAGCAATTCATGTTTTGGA	4
TTTGCAGTATGTTCCTGAATAC	1
CATTGCACTTGTCTCGGTC	18
TAATACTGCCTGGTAATG	10
TACAGTATAGATGATGTACTA	2
TCAAGAGCAATAACGAAAAATG	16
CTGTCTGGTAAAGATG	10
GGCAGTGTAATTAGCTGATTGTA	1
AGGCAGTGCATTGCTAGCTGGCTGTT	4
TCCAGTTTTCCCAGGAATCCCTT	5
CTGAGACCCTAACTTGTGA	23
TAATACTGCCTGGTAATGATGAC	2602
TGAGGTAGTAGTTTGTGCTG	31
ATCCTTGGAACCTAGGTGTGA	1
GTATGTGCCTTTGGACTACATC	1
GCTTACAGTGCAGGTAG	1
TGACCGATTTCTCCTGGTGTTCA	1
AAAGTGCTTATAGTGCAGGTAG	6
TCAGGCTCAGTCCCCTCCCGATAA	1
CCTTGGATGCTAAGATGGGGAT	1
TACAGTACTGTGATAGCTGA	5
ATAGCTTATCAGACTGATGTTG	130
AATCCCAAAAGCAGCTG	1
AAACATGAAGCGCTGCAACAC	85
TTCAAGTAATCCAGGATA	44
AATCATTCACGGACAACACTT	2
GTAGAACCGACCTTGC	1
GTGTTGTGAATCAGGCCG	1
TGCTGACCCCTAGTCCAGTG	2
GCTGGTGTTGTGAATCAGGC	3
TTGTGCAAATCTATGCAAAACTG	15
TAATACTGCCTGGTAATGATG	3241
TTAATGCTAATTGTGATAGGG	7
ACCATCGACCGTTGATTGTAC	2
GCTGGTTTCACATGGTGGCTTAGATTT	1
AAGCAGTGCAATGATGAAAGGGCAT	1
TCCCTGAGACCCTAACTTGTGAG	1
TAATGCCCCTAAAAATCC	2
AGGTTGTCTGTGATGAGTTCG	2
ATATACATACACACACCTACAC	6
AAGGGATTCTGATGTTGGTCAC	4
TCAACAAAATCACTGATGCTGGAGT	11
TGTAACAGCAACTCCATGTGGA	4
ACCACAGGGTAGAACCACGGA	224
TTTGGCAGTGTCTTAGCTGGTTG	2
GGTGCAGTGCTGCATCTCTGGTCAGTTGGG	7
CCCATAAAGTAGAAAGCACTA	6
TGTAAACATCCTTGACTGGAAGCTGT	7
CTATACGACCTGCTGC	1
CACTGGCTCAGTTCAGCAGGAACAG	1
TAGCAGCACATAATGGTTTGTGG	3
TCCCTGAGGAGCCCTTTGAGCCTGG	10
CAAGGAGCTCACAGTCTATTG	1
AACATTCATTGTTGTCGGTGGGTTG	1
CATTATTACTTTTGGTACGCGCTGTGA	1
CAACGGAATCCCAAAAGCAGCTG	572
TATGTATGTGTGCATGTGCATGT	1
ATAGTAGACCGTATAGCGTACG	7
AAGAGGTATAGCGCATGGGAAGA	2
AAGCTGCCAGTTGAAGAACTGT	1250
AGGGACTTTTGGGGGCAGATGTG	2
CAATCATGTGTAGTGCCAATAT	2
TCTGGCTCCGTGTCTTCACTCCC	59
CAAAGTGCTGTTCGTGCAGGT	529
TTTTTCATTATTGCTCCTGAC	1
GGCAGTGTCTTAGCTGGTTGTT	1
CTAGCACCATTTGAAATCAGT	9
ACTGTGCGTGTGACAGCGGCTGAT	2
CAAAGCACAGGGCCTGCAGAGAGG	8
TGATCTAGCCAAAGCCTGACTGT	7
AACATAGAGGAAATTTCACGT	4
ACTGCAGTGAGGGCACTTGTA	7
TTGCCCTTGGTGAATTC	1
CAGTGCAATAGTATTGTCAAAG	289
TACTCAGTAAGGCATTGTT	1
ACGTAAATATTGGCG	2
CCCTGAGACCCTAACTTGTGAGG	3
TCACAGTGAACCGGTCTCTTTT	10
TTCACAAAGCCCATACA	1
TGTGCATGTGTGTATAGTTGTGT	1
AATCGTACAGGGTCATCCACTT	17
CGCATCCCCTAGGGCATTGGT	9
GCAGTCCACGGGCATATACACT	30
TTTGTGACGAAACAAACATGGTGCACTT	1
CATTGTCTGCTGGGTTT	1
GTGCCTTTGGACTACATCGT	1
TAATGCCCCTAAAAATCCT	13
TAGCACCATTTGAAATCGGTT	231
TGCTGACCCCTAGTCCAGTGCT	31
ATCACATTGCCAGGGATTTCC	1071
CTGTCTGGTAAAGATGG	9
GTAGTCTGCACATTGGTT	2
GTAGTGTTTCCTACTTTATG	72
ACAAGTCAGGTTCTTGGG	2
AGAGATGCCATTCTATGTAG	1
CATCTGAAATCGGTTA	2
CTGTGCGTGTGACAGCGGCT	44
TAGCTTATCAGACTGATGTTGACTGTTGAA	3
GTAGATCCGATCTTGTG	1
TAGGTAGTTTCATGTTGTTGGGC	1
AATACTGTCTGGTAATGCCG	2
AAACCCGTAGATCCGATCTTGT	1
CAGTGCACTACAGAACTTTGT	1
CTGTGAGGTGACTCTTGGTGT	1
TTTGGTCCCCTTCAACCAGCT	6
CATCAGTGATTTTGTTG	1
ATCGGGAATGTCGTGTCCGCC	1
TGCTGTGCGTGCGGAATCGACATCAAG	1
TTAGCCGCTGAAATAGATGGA	20
AATCCTTTGTCCCTGGGTGAAAATGCTATT	1
CAGTTCAGCAGGAAC	1
ACAAGCTTGTGTCTATAGGTA	1
ACCAAGATAAATTCACTCTAGTGCTT	1
CAAAGTGCTGTTCGTGCAGG	49
GAGGGTTGGGTGGAGGCTCTCC	2
TTTCAGTCGGATGTTTAC	1
CAAAGTGCTGTTCGTGCAG	6
CAAGTAATCCAGGATAGGC	7
TAATACTGTCTGGTAATGCCGTC	2
TGAGGTAGTAGTTTGTGCT	12
TTGTGCTTGATCTAACCAT	7
ACACTGTCTGGTAAAGATGG	2
TGGAGAGAAAGGCAGTTCCTG	6
GTAAACATCCTTGACTGGAAGC	6
TTCCCTTTGTCATCCTATGCCTGA	7
TAGCAGCACATAATGGTTTGTG	90
CATCTTACCGGACAGTGCTGGAT	2
TCCAGCATCAGTGATTTTGTTG	15
GTGTAGTTAGCTGATTGC	2
TAGCTTATCAGACTGATGTTGACTGT	8
TTCACAGTGGCTAAGTTCCG	309
ACAGTCTACAGCCATGGTCG	1
ACTGCAGTGCCAGCACTTCTTA	1
CTAGCACCATTTGAAATCAGTGTT	24
CTGGCCCTCTCTGCCCTTCCG	5
TAACAGTCTCCAGTCACGGCC	7
TGAGGTAGGAGGTTGTAT	12
CTGACCTATGAATTGACAGCCA	57
CCAATATTGGCTGTGCTGCTCCA	1
CAGCATTGTACAGGGCTATGA	47
TGAAATGTTTAGGACCA	1
AACCCGTAGATCCGATCTTGT	315
TTTCAGTCAGATGTTTGCTGC	13
CACGCAGGGCAACTCGATTGCTG	1
GTATGTGCCTTTGGACTACATCGT	1
TTTGAACCATCACTCGACTCC	16
TATTGCACATTACTAAGTTGCA	13
GTTCAGACTACCTGTT	1
GGTAGTAGGTTGCATAGT	4
GTTCTTCAGTGGCAAGCTTTA	2
TGAGGTAGGAGGTTGTATAGTTGA	16
GACCTTAGTCATGGGGGCTC	1
CAGCAGCAATTCATGTTTTGG	355
TAATCTCAGCTGGCAACTGTG	1
ATTGGGAACATTTTGCATAAA	1
TGTAAACATCCTCGACTGGAAGCT	362
GCTGGTGTTGTGAATCAGG	2
CTAGGTATGGTCCCAGG	1
AACATTCATTGCTGTCGGTGGGTTGA	6
TGTAAACATCCTCGACTGGAAGC	1208
CCTCCCACACCCAAGGCTTGCA	20
ACAGTAGTCTGCACATTGGTTAG	18
CACTGAGATGGGAGTGGTGT	2
ACAGCAGGCACAGACAGG	1
AAGCTGGGTTGAGAGGGCGA	3
CGGGGTCCGGTGCGGAGAGCCGTTCGTCTT	2
TGAGGTAGTAGGTTGTATGGTTTTGGG	1
TGAGGTAGTAGGTTG	3
AGCTCGACTCATGGTTTGA	1
CATTGCACTTGTCTCGGT	54
TGAGGAGCCCTTTGAGCCT	1
TTTGGTCCCCTTCAACCAG	4
AAGGAGCCTCGGTTGGCCCCGGATAGCCGG	6
AGCTCGACTCATGGTTTGAAC	1
GAGGGACTTTCAGGGGCAGCTGTG	1
CATGCCTTGAGTGTAGGACC	9
TATGTAACATGGTCCACTAAC	4
TTCACAGTGGCTAAGTT	44
ATACATACACGCACACATAAGA	2
AATGGCGCCACTAGGGTTGTGC	43
ATCATGATGGGCTCCTCGGTGT	14
AAGGGATTCTGATGTTGGTCACA	8
ATGGCTTTTTATTCCTATGTGA	2
CTGTAGAACCGAATTTGTGT	2
AGGTGCAGTGCTGCATCTCTGG	7
AATCACATTGCCAGGGATTT	5
TTCAAGTAATCCAGGATAGG	2778
TCCAGTTTTCCCAGGAATC	2
TCAGTTATCACAGTGCTGATGC	173
TTTCCCAGGAATCCCTT	1
TATTGCACATTACTAAGTTG	10
TGAGGGGCAGAGAGCGAGACTTTT	4
CCATCTGAAATCGGTTA	1
AGAGGTAGTAGGTTGCATA	27
AACATTCATTGCTGTCGGTGGGTTG	3
CAGTCGGATGTTTACAGC	1
CCAATATTGGCTGTGCTGCTC	2
TAACACTGTCTGGTAAAGATGGC	3961
TAATACTGCCGGGTAATGA	59
GTAAGTGCCTGCATGTATATGC	16
AGCTACATTGTCTGCTGGGTT	12
TAAGTTGCATGTTGTCACGGCCTCAAT	1
CATCCCACTCAGTCGTGCCC	52
AAACATGAAGCGCTGCAACACC	5
GTAATCCAGGATAGGC	4
TAACACTGTCTGGTAACG	14
GTGCTGTTCGTGCAGGTAG	1
CCTGACTCCAGGTCCTGTGTGT	1
AAGTGCTTATAGTGCAGGTAG	12
GGGTCCGGTGCGGAGAGCCGTTCGTCTT	1
GGCTTCTTTACAGTGCTGCCTTGTT	2
TCAGGCTCAGTCCCCTCCCGA	8
TGATCTAGCCAAAGCCTGA	1
AGTTCCAGGACAGCCAGGGCTATACAGAGA	1
TAGTAGGTTGTATGGTTT	2
AATGCACCTGGGCAAGGGTTCA	8
GCTTATCAGACTGATGTTGA	68
CTATACGGCCTCCTAGCTTTC	8
CCTGTTCTTGATTACTTGTTTC	3
CCGTGTATTTGACAAGCTGAGTTGGACA	1
AAGTGCTTACAGTGCAGGTAGT	1
AGAGGTATAGCGCATGGGAA	3
TTCAAGTAATTCAGGATAGGTT	7618
TAAGGTGCATCTAGTGCAGATAG	573
GGTGCAGTGCTGCATCTCTGG	17
TGTCTGCCCGAGTGCCTGCCTCT	4
TCCCTTAGGGTTAGAGTGCACTGAAG	1
CACCCGTAGAACCGACCTTGC	193
AACAAATCACAGTCTGCCATA	3
AGTGGGGAACCCTTCCATGAGGA	1
AATCACATTGCCAGGGATTTCCAA	5
CAATCACTAACTCCACTGCCA	1
CAGTGCAATGATGAAAGGGCA	7
TAATACTGCCTGGTAATGA	101
TAGCAATGTCAGCGGTGCCT	1
GGTGCAGTGCTGCATCTCTG	1
CCTGAGACCCTTTAACCTGTGA	3
TTTCAGTCAGATGTTTGCTG	1
AGCTTCTTTACAGTGTTGCCTTGT	5
ACAGTGGCTAAGTTCTGC	1
TGAGAACTGAATTCCATGGGTTA	5
TCTCCCAACCCTTGTACCAGTG	153
GCTTATAGTGCAGGTAGT	4
TTATTACTTTTGGTACGC	2
GGTCAAGAGGCGCCTGGGAA	8
ATCCCACTCAGTCGTGCCC	70
GTTTTCCCAGGAATCC	1
TTTGGCAATGGTAGAACTCACACCG	17
GCAAAGCACAGGGCCTGCAGAGA	8
TCCTGACTCCAGGTCCTGTGTG	3
GCACACGCAGGGCAACTCGATTGCTGT	2
ACATTCATTGCTGTCGGTGGGTTGA	1
TGCACTGAGATGGGAGTGGTGTAA	1
GAAGTTGTTCGTGGTGGATT	1
ACTCCATTTGTTTTGATGATG	4
CTAGACTGAGGCTCCTTGA	6
TACAGTACTGTGATAGCTG	1
AATCACATTGCCAGGGATTACCAC	1
ACATTCAACGCTGTCGGTGAGT	13
CAAGAGGGCCTTTCTGG	1
AGTATTACATGGCCAATCTC	1
AGGCAAGATGCTGGCATAGC	2
TAAGGTGCATCTAGTGCAGATA	316
AACCCGTAGATCCGAACTTGT	54
TTTCCTCATATCCATTCAGGAG	2
CAGTGGTTTTACCCTATGGTAGG	2
CAGCGGGAACAGTACTGCAG	3
TAACACTGTCTGGTAAAGATGG	2615
AATCCTTGGAACCTAGGTGTGA	112
TCAGCTGATCTGTGGCT	1
CTGTGCGTGTGACAGC	2
CATTATTACTTTTGGTACGC	781
TCAGTTATCACAGTGCTGATG	59
TGCTATGCCAACATATTGCCATC	33
CCAGTGTTTAGACTACCTGT	1
TCAGTGCACTACAGA	2
GTACAGTACTGTGATAGCTGA	4
TGTCTTGCAGGCCGTCATG	4
AGGTTGGTGTACTGTGTGTGA	1
ACCATCTGAAATCGGTT	1
ACTGGCCTACAAAGTCCCA	1
AGCAGCACAGAAATATTGGC	1
TAGCACCATTTGAAATCAGTGTTTTA	4
GGTCAAGAGGCGCCTGGGA	3
TGAGGTAGTAGATTGTATAGT	388
CAGTGGTTTTACCCTATGGTAGGT	2
AATCCTTTGTCCCTGGGTG	1
TCCCTGAGGAGCCCTTTGAGC	801
TGGAGAGAAAGGCAGTTCCTGAT	3
TATTGCACATTACTAAGTTGCAT	2
CTGGGAGAGGGTTGTTTACTCC	1
GCCGAAGGAGCCTCGGTTGGCCCCGGATAG	1
TCTGGCTCCGTGTCTTCAC	3
CTATGCATATACTTCTT	1
TCAGTTATCACAGTGCTGA	1
AAAAGCTGGGTTGAGAGGGCGAA	30
CAACAGCAGTCGATGGGCTGT	1
TAGTAGGTTGCATAGTT	12
TATTGCACTTGTCCCGGCCTGTT	63
GGTAGTAGGTTGTATGGTTT	1
CCCAGTGTTCAGACTACCT	14
TAGCACCATTTGAAATCAGTGT	13176
TGTAAACATCCTACACTCAGCTGTCATAC	1
CTAGCACCATCTGAAATCGGTTAT	4
CACGGGTTAGGCTCTTGGGA	1
TCCCTGAGGAGCCCTTTGAGCCTGGAG	2
TCTCCCAACCCTTGTACCAGT	20
CAACGGAATCCCAAAAGCA	5
TTTCCCTGAGGAGCCCTTTGAGC	1
TCCCTGTCCTCCAGGAGCTCAC	2
AGCTGGTGTTGTGAATCAGGCCG	2005
AGCGAGGTTGCCCTTTGTATAT	4
ACTGGACTTGGAGTCAGAAG	19
CAGCACGTAAATATTGGC	1
TAGCACCATCTGAAATCGGTTAT	65
AAAAGCTGGGTTGAGAGGGCGAAAAA	8
TATGTGCCTTTGGACTACA	1
ATTCCTAGAAATTGTTCACAAT	1
TGGCAGTGTCTTAGCTGGTTG	227
TAGCACCATTTGAAATCGGTTAT	13
TAGCAGCGGGAACAGT	2
TCAGCCCATCCCACTCAGTCGTGCCC	1
TAGTAGGTTGTATAGTTT	5
TCGTGTCTTGTGTTGCAGC	1
TAATACAACCTGCTAAGTG	1
TCGAGGAGCTCACAGTCTAGT	250
GCTTATCAGACTGATGTTGACT	3
AACACGGACACCGCAGGG	1
TCAGTGCATGACAGAACTTGG	955
TTAGGGTCACACCCACCACTGGGAGATAA	28
CGGTGCGGAGAGCCG	1
TCCCTGAGACCCTTTAACCTGT	102
TCCATCTTCCAGTGCAGTGTT	3
TCCCTGAGACCCTAACTTG	39
TCCCTGAGACCCTAACTTGTGAT	33
AATACTGCCTGGTAATGATG	9
TGTAAACATCCTCGACTGGAAGCTGT	2
AATGCACCCGGGCAAGGATTTGG	2
AAGCCCTTACCCCAAAAAGTAT	17
CCGTCCTGAGGTTGTTGA	1
AACATGAAGCGCTGCAACA	1
GCAGCACAGAAATATTGGCAT	1
ACCATTTGAAATCAGTG	1
AACATTCATTGCTGTCGGTGGGTTT	453
AATCACATTGCCAGGGATTTCC	4
CTGCGCAAGCTACTGCCTTGCT	17
CGGTTATCATGGTACCGATGCTG	1
ATGACCTATGATTTGACAGACC	1
CATGCCTTGAGTGTAGGACCGT	246
GCAATGTTAAAAGGGCAT	1
ACTGATTTCTTTTGGTGTTCAG	27
TCCCCCAGGTGTGATTCTGATTT	8
TAGCAGCACATCATGGTTTACATACTACAG	3
TGAGGTAGTAGGTTGTATAGTTTTAGG	1
AATGTTGCTCGGTGAACC	1
ACAGTAGTCTGCACATTGGTTA	88
AATCACTAACCACACAGCCAG	10
CAGTTTTCCCAGGAATCCCTT	4
ACCCGTAGATCCGATCTTGTG	6
ACTCAGCCCATCCCACTCAGTCGTGCCC	3
CTGTAGATCCGAATTTGT	2
AAGTGCTGACAGTGCAGAT	1
AATCACATTGCCAGGGATTAC	5
AGCTACATCTGGCTACTGGGT	40
CGGTGCTGGTGGAGCAGTGAGCACGCCAT	1
TGTAGTGTTTCCTACTTT	2
CCTGGTACAGGCCTGGGGGAT	1
AGGGGTGCTATCTGTGATTGAGG	16
TATGTGCCTTTGGACTACATC	8
AACATTCATTGCTGTCGGTGGGTTGAAC	2
TCTACAGTGCACGTGTCTCCAG	47
ATCATGATGGGCTCCTCGGTG	1
AGGTAGTAGGTTGTATGGTT	8
TGAAACATACACGGGAAACCTCTT	4
TAGCAGCACATCATGGTTTAC	2534
AACCCGTAGATCCGATCTTG	22
TTGTGCAAATCTATGCAAAACTGA	8
AGTACTGTGATAGCTGA	1
AGCGCCGAAGGAGCCTCGGTTGGCCCCGGA	69
TGAGGGGCAGAGAGCGAGACT	2
TAGGTAGTTTCCTGTTGTTGGGATCC	2
CTGCCGGGTAATGATGG	2
AACCCGTAGATCCGATCTTGTGG	1
TGTAGTGTTTCCTACTTTATG	792
TAGTGCTGGTGATGCTCAGCGCCGCGGAGG	2
TTCAAGTAATTCAGGA	1
TTTGGCACTAGCACATTTTTGCT	15
AATCCTTGCTATCTGGGTGCTTAGTGC	79
AAGGAGCTTACAATCTAGCTGG	799
CTCCCACCTTGTCCTCACG	2
TGAAACATACACGGGAAACCTCT	19
TAGGTTGTGTGGTTTCAGGGCAGTGA	1
CATAAAGTAGAAAGCACTAC	9
TCCGGTTCTCAGGGCTCCACC	1
AACATTCATTGCTGTCGGTGGGTTGAA	1
TAGGCAGTGTAATTAGCTGATTG	6
TGTGCAAATCCATGCAAAACTG	4261
CAGGCACAGACAGGCAG	2
TCAGTGCACTACAGAACTTTG	551
CTCGTGTCTTGTGTTGCAGCCG	4
TGCCATATCAACTCCTCTCTTCCCAAAAC	1
CTAGCACCATTTGAAATCAGTGT	206
TGAGGTAGTAGTTTGTACAG	14
AACCCGTAGATCCGAACTTG	4
CTACAGTATAGATGATGTACT	5
AGAGCTTAGCTGATTGGTGAAC	2
CTACTTCACAACACCAGGGC	1
CAAAGTGCTCATAGTGCAGGTAG	36
TTGGTCCCCTTCAACCAGCT	1
TACGTAGTATAGTGCTTTTCA	1
TGAGGTAGTAGTTTGTGCTGT	358
CCTGAGACCCTAACTTGTGA	32
GCACTGAGATGGGAGTGGTGT	404
CCGCACTGTGGGTACTTGCTGCT	2
GGTAGTAGGTTGCATAGTT	7
TACAGTAGTCTGCACATTGGTT	1267
TGAGGTAGTAGGTTGTGTGGTTTCAGGG	1
CAGTAGTCTGCACATTGGTT	597
TGAGGTAGTAGTTTGTGCTGTTG	43
GCACCATCTGAAATCGGTT	2
ATACTGCATCAGGAACTGACTGG	2
CATTGCCAGGGATTTCCA	4
CAAAGCACAGGGCCTGCAGAGAGGT	48
CCTATGCATATACTTCTTT	1
TCTGGCTCCGTGTCTTCACTC	39
TTTTCCCAGGAATCCC	3
TTTTGCGATGTGTTCCTAATAT	1592
CAGTTATCACAGTGCTGATG	10
TGTGTGTGTACATGTACAT	1
TGTGCAAATCTATGCAAAACTGA	157
ATAGTTGTGTGTGGATGTGTGTAT	59
CCCAGTGTTCAGACTACCTGTTCAGGA	1
GGTAGTTTCCTGTTGTTG	3
CAAGCTTGTGTCTATAGGTAT	20
TCCCTGAGGAGCCCTTTGAGCCTG	465
TATGGCTTTTTATTCCTATGT	51
CGCATCCCCTAGGGCATTGGTGT	439
GTAGAACCGAATTTGT	2
CAAAGTGCTTACAGTGCAG	22
AGTAGGTTGTATAGT	1
TGAGACCCTAACTTGTGA	5
AGGCAAGATGCTGGCATAGCTGT	10
TAGCACCATTTGAAATCAGTGTT	21723
CTCGGTTGGCCCCGGATAGCCGGGTCCCCG	14
TAAAGTGCTGACAGTGCAGATAGT	9
CATAAAGTAGAAAGCACTACTA	6
TAGCAGCACAGAAATATTG	3
AGGGTAGAACCACGGAC	1
AAGAGGTAGTAGGTTGCATAGTT	22
GCTGGTTTCATATGGTGGTTTAGAT	3
TAAGTGCCTGCATGTATATGCG	527
TGAGGTAGTAGGTTGTATGGTTTAG	20
TGTAAACATCCCCGACTGGAA	24
CAAGCTCGTTTCTATGGGTCT	2
CTTTCAGTCGGATGTTTGCAGCT	4
TTGGACTGTGAGGTGACTCTTGG	1
TTCACAAAGCCCATACACTTTCAC	33
CTCCCACATGCAGGGTTTGC	3
AGTTCTTCAGTGGCAAGCT	4
TCAGCTCCTATATGATGCC	1
AATCCTTTGTCCCTGGGTGA	5
CTCCCAACCCTTGTACCAGT	1
ATTGCACTTGTCCCGGCCTGT	13
AATGGCGCCACTAGGGTTGTGCA	1
TGATATGTTTGATATATTAGGTTGTT	1
CCTAAGGTGAATTTTTTGGGA	1
CTAATACTGCCGGGTAATGATG	22
TTGAGGTAGTAGGTTGTATAGT	5
AAGGAGCTTACAATCTAGCT	3
TACTGCCTGGTAATGATGA	3
TCCCTGAGACCCTTTAACC	3
AATCGTACAGGGTCATCCACT	3
AACGGAATCCCAAAAGCAGCTGT	9
CGTCCTGAGGTTGTTGAGCTGT	3
AAAGTGCATGCGCTTTGGGA	5
ACAGTAGTCTGCACATT	2
CTAACACTGTCTGGTAAAG	1
CATTGTACAGGGCTATG	7
TAGGTTATCCGTGTTGCCTTC	28
ACAAACATGGTGCACTTCTTTT	1
TGGCAGTGTATTGTTAGCTGGT	37
ACACGAGACTTTGCCTAG	1
ACTGATTTCTTTTGGTGTTCAGA	59
CAGTGCATGACAGAACTTGG	3
CTAACACTGTCTGGTAAAGATGGC	5
CACTGCCCCAGGTGCTGCTGG	2
AGCTACATCTGGCTACTGGGTCTCTG	1
CGCATCCCCTAGGGCATTGGTGTAAA	6
AGCGCCGAAGGAGCCTCGGTTGGCCCCG	3
CACAGCTCCCATCTCAGAACAA	9
AATGGCGCCACTAGGGTTGT	200
CGGTGCGGAGAGCCGTTCGTCTTGGGAA	1
TTCTTCAGTGGCAAGCTTTA	1
TTTCAGTCGGATGTTTACAGC	17
CGTACCGTGAGTAATAAT	1
CTATACGGCCTCCTAGCTTTCC	26
CACTGAGATGGGAGTGGTGTA	7
AATACTGTCTGGTAATGCC	1
TGTCAGTTTGTCAAATACC	2
ACCCTGTAGAACCGAATTTGTGT	11
TAACACTGTCTGGTAACGAT	132
CAGCTACATCTGGCTACTGGGTCT	1
ATCCCTGAGTGTATGTGGTGAACCTGAA	1
AAGCTGCCAGTTGAAGAACTGTT	45
GAGGTAGTAGGTTGCATAG	4
CGGGACCGGGGTCCGGTGCGGAGAGCCGT	2
TTCCTATGCATATACTTCTTTGT	3
GGTAGTAGGTTGTGTGGTTT	1
CATCTTCCAGTGCAGTGTTGGA	26
TAGCTTATCAGACTGATGTTG	18978
AGCTACATCTGGCTACTGGGTCT	35
TCAGTGCATCACAGAACTTTGT	56
TCTCTGGGCCTGTGTCTTAGG	8
TACAGTATAGATGATGTACT	19
TGAGGTAGTAGGTTGTATGGTTTTGG	1
TTCACCAATCAGCTAAGCTCTGC	1
TAGCTTATCAGACTGATGTTGA	46488
ATTGGGGATGCTTTGCAT	3
AATACTGCCGGGTAATGATGGA	2
TGAGGTAGTAGGTTGTATGGTTTAGAGTTA	1
TCCATCTTCCAGTGCAGT	1
TGTCTTGCAGGCCGTCATGC	4
AAACCGTTACCATTACTGAGTTT	25
CCTCTGGGCCCTTCCTCCAG	2
CTTTTTGCGGTCTGGGCTTGC	62
ATACTGTCTGGTAATGCCGT	2
GTTTCCTACTTTATGGA	1
CAAGCTCGTTTCTATGGGTCTGT	1
TAATACTGTCTGGTAATGCCGT	3023
AGCAGCACATCATGGTTTAC	1
TCTCCCAACCCTTGTACCA	2
AAGGTGCAGGCCATACTGTGCTGCCTCA	1
CTCGAGGAGCTCACAGTCTAGTAT	1
CTGAACCTACCTGCTTGGGAC	1
ACTGCCTGGTAATGATGAC	2
TGAGAACTGAATTCCATAGGCT	1
GTAAACATCCTACACTCTCAGCT	1
CATTGCTGTCGGTGGGTT	1
ATGTATGTGTGCATGTGCAT	2
GTACAGTACTGTGATAGCTGAAG	1
GCTCGACTCATGGTTTGA	1
TTTCAGTCGGATGTTTGC	1
GTGTTTGTCCGAGGAGG	1
AGGCAGTGTAGTTAGCTGATTG	19
ATGTATGTGTGCATGAACATGTG	1
CTTCCAGTGCAGTGTTGG	1
AGCTACATCTGGCTACTGGGTCTC	18
GCAGCAGCAATTCATGTTTTGG	1
AAGTAATCCAGGATAGGC	5
CTATACAATCTACTGTCTTTCCT	2
AGGTAGGAGGTTGTATAGT	1
TCTTTACAGTGTTGCCT	1
TAATACTGCCTGGTAAT	4
TGAGGTAGTAGTTTGTGC	7
CTACATTGTCTGCTGGGTTT	2
CAGTGCAATAGTATTGT	1
AAAACATGAAGCGCTGCAACA	7
ACTGCAGTGAGGGCACTTGTAGC	18
TCTACAGTGCACGTGTCTC	6
TAACACTGTCTGGTAAAGA	2
TTCAGCTCCTATATGATGCCT	2
CCCACTCAGTCGTGCCC	2
AGCTTCTTTACAGTGTTGC	1
GAATGTTGCTCGGTGAACCCC	1
AGCAGCATTGTACAGGGCTATGAAAG	1
TCAGTGCATCACAGAACTTTGTC	1
TGTAAACATCCTACACTCTCAGCTG	1
TTCACAAAGCCCATACACTTT	25
GGGGTTCCTGGGGATGGGA	1
CTGAGGTAGTAGTTTGTACAGT	9
AGGAGGTTGTATAGTTG	1
TCGTGTCTTGTGTTGCAGCCGGAG	2
TAGCAGCACAGAAATATTGGC	658
TAGTAGGTTGTATAG	1
CCATTTGAAATCAGTGTT	12
TAACAGTCTACAGCCATGGTCG	82
CCGTCTCAGTTACTTTATAGC	2
ATATACAAGGGCAAGCTCTCTGT	3
TAGATGTTGTCTCTTGCTATCCCTGTG	2
CACAGTGGCTAAGTTCTGCA	1
TACAGTACTGTGATAACTGA	21
CTGGACTTGGAGTCAGAAGGCC	2
AGCTGGTGTTGTGAATCAG	1
AACCCGTAGATCCGATCTT	2
AACATCCTACACTCTCAG	1
GACCGATTTCTCCTGGTGTTCAGAGT	1
AGGTTGCCCTTGGTGAATTCGCTTTATTGA	1
TTAATGCTAATTGTGATAGG	1
TTTGTGACCTGGTCCACTAACCC	1
CCCTGAGGAGCCCTTTGAGC	2
GCAGTCCACGGGCATATACACTT	2
AAACATGAAGCGCTGCAA	2
AGTAATTCAGGATAGGT	2
TTTCCTATGCATATACTTCTTT	203
AATGGCGCCACTAGGGTTGTG	610
CCGTGAGTAATAATGC	1
TTTTGCGATGTGTTCCTAAT	133
CATTCAACGCTGTCGGTGAGT	1
ACCGTGGCTTTCGATTGTT	2
ATCACTAACTCCACTGCCATCA	26
TAATACTGCCTGGTAATGATGACG	13
GCACCATTTGAAATCAGTGT	14
CTATGCCAACATATTGCCAT	1
ACTGATTTCTTTTGGTGTTCA	12
CCATCTTCCAGTGCAGTGTTGGA	1
CACATAGTGACGCAGGCAGTGCGACC	1
TCGAGGAGCTCACAGTCTAG	39
TCCAGCATCAGTGATTTTG	1
TTTGGTCCCCTTCAACCAGCTGTA	1
TAACACTGTCTGGTAAAG	7
CAACGGAATCCCAAAAGCAGCTGT	341
AGTAGTCTGCACATTGGTT	11
CTGGTACAGGCCTGGGGGATA	1
AACCCGTAGATCCGATCTTGTG	5068
GGTAGTAGGTTGTATAGTT	4
TATGGCTTTTTATTCCTATG	16
CCACTGCCCCAGGTGCTGCTG	5
TACTCAGTAAGGCATTGTTCT	6
CTATACAACCTACTGCCTTC	13
GGTGCAGTGCTGCATCTCTGGT	11
TGTCTTGCAGGCCGTCAT	3
ATACTGCATCAGGAACTGACTG	1
CTAATACTGTCTGGTAATGCCGT	2
GAGGTAGTAGGTTGTATGGTT	37
TGTAAACATCCTACACTCAGCTG	17
GAGACCTCTAGCATGGTGTTGTGGGAC	1
CAAGTAATCCAGGATAGG	1
TACCCTGTAGATCCGAATTTGT	152
TTTTGCGATGTGTTCCTAATATG	11
CCCTGAGACCCTTTAACCTGTGA	3
TAAAGTGCTTATAGTGCAGGTAGTGTGTAG	2
AATCACATTGCCAGGGATTA	1
ATCTTACTGGGCAGCATTGGA	1
CTGTTGCCACTAACCTCAACCTT	3
ATGTGCCTTTGGACTACATCGT	1
CCCCCAGGTGTGATTCTGATTTG	1
CAGTGCAATGTTAAAAGGGC	188
TAGCTTATCAGACTG	2
TGTAAACATCCTACACTCAGCT	1817
AGTTTTGTGTGCATGTGCAT	2
AACATTCGCGGTGCACTTCTTTT	1
CTGGTTTCACATGGTGGCTTAG	3
CTGTGCAAATCCATGCAAAACTG	1
CCGGTGCGGAGAGCCGTTCGTCTTGGG	1
TGTCTCTCTCTGTGTCCTGC	1
CTCGCTGGGGCCTCCA	7
TAGTAGGTTGTATAGTT	37
ATGTCACTCGGCTCGGCCCACTACC	1
TTGAGGTAGTAGGTTGTATGGTT	25
CACACGAGACTTTGCCTAG	3
TTTCCCAGGAATCCCT	19
TGCAAATCTATGCAAAACTG	1
AATCCAGGATAGGCT	2
TAAGGCTCCTTCCTGTGCTT	5
TAAGGCTCCTTCCTGTGCTTG	1
TTTCAGTCGGATGTTTGCAGCT	7
TGTGCAAATCCATGCAAA	3
ACAAATTCGTATCTAGGGGAAT	1
AGCTGCCAGTTGAAGAAC	1
CAGCAGCAATTCATGTTTTGGAG	2
ACAAATTCGTATCTAGGGGAA	2
TAGTAGATTGTATAGTT	2
ATCTCGCTGGGGCCTCCA	131
CGCTCGGCGGGGTCCCCGCGTCGTC	1
AGCTTATCAGACTGATGTTG	5
TGTAAACATCCCCGACTGGAAG	69
CTATACAGTCTACTGTCTTTCC	1
TTTTTCATTATTGCTCCTGA	1
AGGAGCTTACAATCTAGCTGGG	3
ATTGGGGATGCTTTGCATTCA	14
TTTTGCAGGTTTGCATCCAGC	1
GCTCCTGACTCCAGGTCCTGTGT	1
AGTCTGCACATTGGTTA	1
ACGGGTTAGGCTCTTGGGAGC	1
TATACATACACACACCTACAC	3
TAATCTCAGCTGGCAACTGTGA	1
GCAGCAGCAATTCATGTTTTGGA	2
CTTATAGTGCAGGTAG	4
CAACGGAATCCCAAAAGC	1
GTTCTTCAGTGGCAAGCTTT	1
TTTCAGTCGGATGTTTGCAGC	79
CAAAACATGAAGCGCTGCAAC	31
TCAGTGCACTACAGAACTTTGTCT	18
TATGTATGTGTGCATGTGCAT	2
GCTTTCAGTCGGATGTTTACAG	2
AGTGCAATGTTAAAAGGGC	1
CATTGTACAGGGCTATCA	1
TACAGTAGTCTGCACATTG	1
CCAGTGTTCAGACTACCTGTT	2
TGAGCGCCTCGGCGACAGAGCCG	2
TCGAGGAGCTCACAGTCTAGTAT	30
TGTCTTGCAGGCCGTCATGCAG	4
AGCTTCTTTACAGTGCTGCCTTGT	3
CTTAGCAGGTTGTATTATCATT	19
TAGGTAGTTTCATGTTGTTGGGA	132
TAGTAGACCGTATAGCGTACG	26

TGCCATCAGAAGACTGAAGGAACTGAAAGA	1
CTAGCACCATCTGAAATCGGTT	242
GGCTCAGTTCAGCAGGAACAG	1
AATAATACATGGTTGATCTT	1
CTGCCTGGTAATGATGA	3
TGCGGGGCTAGGGCTAACAGC	5
CTTTGTCCCTGGGTGAAAATGC	1
CATTATTACTTTTGGTACGCGCTGTGACA	1
CATTGCACTTGTCTCG	1
CCATTTGAAATCAGTGT	5
TGGCAGTGTCTTAGCTGGTTGTTG	8
TATGTAACATGGTCCACTAACT	1
ATTCAACGCTGTCGGTGAGT	2
TACAGTAGTCTGCACATTGGTTAG	1
AACTGGCCTACAAAGTC	1
GTAGTGTTTCCTACTTTATGGATG	2
CTTTCAGTCGGATGTTTGC	21
ATCACATTGCCAGGGATTTCCAAC	3
CATTATTACTTTTGGTACGCGCT	5
TGTGCGTGTGACAGCGGCTGA	1
ATCGTACAGGGTCATCCACTT	1
TGCCTGTCTACACTTGCTGT	10
TGAGAACTGAATTCCATAGGC	2
TTCAAGTAATCCAGGATAG	507
TATGGCACTGGTAGAATTCACTG	1
CTATACAATCTATTGCCTTC	7
ACCGTGGCTTTCGATTGTTACT	58
CGTGTCTTGTGTTGCAGCCGGA	3
TCGTACCGTGAGTAATAATGCGC	2
CTGTAACCATGTTGTAGTGTG	1
TAGCAGCGGGAACAGTACTGCAGTGAGT	1
CAGTGTCTTAGCTGGTTGT	1
CTAATACTGCCGGGTAATGAT	10
TGAGCGCCTCGGCGACAGAG	2
TCACAGTGAACCGGTCTCTT	5
TTCCTATGCATATACTT	7
ATCACATTGCCAGGGATTACCACG	11
CTGGCCCTCTCTGCCCTTCCGT	22
TAGGTAGTTTCCTGTTGTTG	74
TCCCCCAGGTGTGATTCTGATTTG	4
TGAACAGTGCCTTTCTGTGTAGG	2
TGCCTGTCTACACTTGCTGTGC	412
CTGTACAACCTTCTAGCTTTCC	5
TAGTTAGCTGATTGC	1
GCAGCACGTAAATATTGGCGT	4
ACCCTGTAGAACCGAATTTGTG	50
CCCCCAGGTGTGATTCTGATTTGTT	4
GAGGTAGTAGGTTGTATGGTTT	1
CAGCAGCACACTGTGGTTTGTA	43
TGCTGACCCCTAGTCCAGTGCTT	7
TCCTGTACTGAGCTGCCCCGAGC	1
TGTAGTGTTTCCTACTTTAT	123
CAGCATTGTACAGGGCTATGAA	1
TCACTAACTCCACTGCCATC	1
TGTCACTCGGCTCGGCCCACT	1
ATTCAAGTAATCCAGGATAGGC	4
CAAAGTGCTAACAGTGCAG	1
TGAGGTAGTAGGTTGTATG	112
TAAGTGCCTGCATGTATATGCGT	72
CACAGCTCCCATCTCAGAAC	7
CATTTGAAATCAGTGT	2
CATCTTACTGGGCAGCATTGG	11
TACAGTACTGTGATAGCTGAA	3
TCGGATCCGTCTGAGCTTGGCT	21
TTCCTATGCATATACTTCT	90
ATCTTACTGGGCAGCATTGGAT	1
CACCATTTGAAATCGGTT	1
ACAGGTGAGGTTCTTGGGAGCC	6
ACAGTAGTCTGCACATTGGTT	6823
TCGGATCCGTCTGAGCTTGGC	4
GTGGGGTAGTGATTTTACCCTGTTTAGGAG	2
AATGACACGATCACTCCCGTTG	25
ACACTGTCTGGTAACGATG	2
CAGGCCATACTGTGCTGCCT	2
CCGAAGGAGCCTCGGTTGGCCCCGGAT	2
AACCCGTAGATCCGAACTTGTGCTGATT	2
CAAAGAATTCTCCTTTTGGGCT	13
ATGAGGTAGTAGGTTGTATAGT	5
TTGCGATGTGTTCCTAATAT	2
TAGCAGCACGTAAATATTGGCGTTAA	2
TCCATTCTAAAGGTACAGTACTGTGATAA	2
CTAGGTATGGTCCCAGGGATCCCAGATCAA	2
CATCCCTTGCATGGTGGAGGGTGA	1
ACTGCCTGGTAATGATG	2
CACGCGGGAACCGAGTCCAC	1
TAAGTGCCTGCATGTATAT	10
TGAAGAGAGGTTATCCTTTGTGT	1
AATCGTACAGGGTCATCCACTTT	10
TAACAGTCTCCAGTCACGGCCAC	1
CATTCAACGCTGTCGGTGA	1
TTCAAGTAATCCAGGATAGGC	13035
CACTGTCTGGTAACGATGTT	3
AACATTCATTGCTGTCGGTGGGTT	359
CGAATGTTGCTCGGTGAACCC	1
TACCGTGAGTAATAATGCG	1
TGAGGTAGTAGTTTGTGCTGTTGGT	2
GCAGCGCCGAAGGAGCCTCGGTTGGCCCCG	1
AGTGTTTCCTACTTTAT	1
CATAAAGTAGAAAGCACTA	1
TAAGGCTCCTTCCTGTGCT	1
TCCTATGCATATACTTCTTT	1
TCCCTGAGGAGCCCTTTG	14
ATATGTTTGATATATTAGGTT	1
ATAGCTTATCAGACTGATGTTGACT	5
CAGTTATCACAGTGCTGATGCTGTCCATT	1
AATCACATTGCCAGGGATTTCCA	32
AGCACCATTTGAAATCAGTGT	3
TGAGCGCCTCGGCGACAGAGC	3
GGTGAGGTAGTAGGTTGTAT	1
TTGTGCTTGATCTAACCATG	55
CAGTGCAATGTTAAAAGGGCA	341
TGCACTGAGATGGGAGTGGTGTA	2
TGTCAGTTTGTCAAATACCCC	2
TATACAAGGGCAAGCTCTCTG	5
TCCTGTACTGAGCTGCCCCGA	52
TAGGTTATCCGTGTTGCCTTCGCTT	1
TATGGCTTTTCATTCCTATGTG	2
CAAAGTGCTGTTCGTGCAGGTAGTG	4
AAAGCTGGGTTGAGAGGGCGAAAA	2
AACATTCATTGTTGTCGGTGGGT	89
GAGGGTTGGGTGGAGGCTCTC	1
TGTATGCCCTAACCGCTCAGT	1
AGTGCATGACAGAACTTGGG	1
ATAGTTGTGTGTGGATGTGTGTATTT	5
CAGTGCACGTGTCTCCAG	1
AGGCAGTGTAGTTAGCTGATTGCT	1
TGAGCGCCTCGGCGACAGAGCC	3
TTCACAAAGCCCATACACTT	4
AAGGGATTCTGATGTTGGT	1
TTCCCTTTGTCATCCTATGCCT	1357
GAGCCTCGGTTGGCCCCGGATAGCCGGGT	8
GTAGTAGGTTGCATAGT	2
GGTGTTGTGAATCAGGCCG	2
CATCTTACTGGGCAGCATTG	1
ACTGCAGTGAGGGCACTTGTAG	2
ACCCTGTAGATCCGAATTTGT	11
GCTCTGACTAGGTTGCACTAC	1
TTTGGTCCCCTTCAACC	2
AGTTCAGCAGGAACAG	1
ACTGTCTGGTAACGATGTT	2
GTAATTCAGGATAGGT	2
ATTTGAAATCAGTGTT	10
TATTGCACATTACTAAGTTGC	264
TTATAAAGCAATGAGACTGAT	4
AGAGGTAGTAGGTTG	2
GTGAGGTAGTAGGTTGTATGGTT	1
CATGCCTTGAGTGTAGGACCG	25
TACCCTGTAGATCCGAATTTG	50
CAGCTTCTTTACAGTGTTGCCTTG	4
CAGTAGTCTGCACATTGGTTA	135
AATATAACACAGATGGCCTGT	10
TTTGGTCCCCTTCAACCAGCTG	29
CCCGGAGTCGGGAAAAGCTGGGTTGAGAG	1
TAGTAGGTTGTATGG	1
TAAGTGCCTGCATGTATATGC	159
AAGGGATTCTGATGTTGGTCACACT	211
CCGAAGGAGCCTCGGTTGGCCC	1
CAAAGTGCTTACAGTGCA	3
ATGAAGCACTGTAGCTCAGGA	1
ATAGGTTATCCGTGTTGCCTTCG	1
TGGCCATGCACCCGTGTCTCGGTGC	1
GTAGGTTGCATAGTT	1
TGAGGTAGTAGATTGTATAGTTG	26
CTCGGGGATCATCATGTCACGA	237
TTCAAGTAATTCAGGATAG	42
CTACACTGCCGGCCTCTGGG	2
AGCTTCTTTACAGTGTTGCCT	7
TGTGCAAATCCATGCAAAACTGATT	6
TGAGGTAGTAGATTGTATAG	188
GTTGAGGTAGTAGGTTGTATGGT	3
GTAGAACCGAATTTGTG	1
ATCACATTGCCAGGGATTACCA	2852
TAGTAGGTTGTGTGGTTT	1
ATTGCACTTGTCCCGGC	1
GTCCGGTGCGGAGAGCCGTTCGTCTTGGGA	5
AGTAGGTTGCATAGTT	8
CCGTCTCAGTTACTTTATAGCCA	1
TGACCGATTTCTCCTGGTGTTC	3
ACTGCAGTGCCAGCACTTCTT	1
TATGCAAGGGCAAGCTCTCT	2
GGGGGTCCCCGGTGCTCGGATCT	1
TTGTCTCCAGAGCATTCCA	1
TACCCTGTAGAACCGAATTT	1
TCCTGTACTGAGCTGCCCC	1
GTACAGTACTGTGATAACT	1
AAACATCCCCGACTGGAAGC	1
GTGCTTATAGTGCAGGTAG	1
CTCTACAACCTTAGGACTTGCA	1
ATCACACAAAGGCAACTTTTGT	4
AGATCAGAAGGTGACTGTGGCT	3
TCACATTGCCAGGGATTAC	22
CTTTCAGTCGGATGTTTACAG	118
AGAGTTGTCATGTGTCT	2
TAGTAGGTTGTATGGT	2
GCTCAGTTCAGCAGGAACAG	1
TGAGGTAGTAGATTGT	1
CATTCATTGCTGTCGGTGGGT	1
TAGGTTATCCGTGTTGCCTT	5
CAGTGCAATGTTAAAAGGG	95
TCCTGTACTGAGCTGCCCCGAGG	1
TAGCAGCACATCATGGTTTACA	1735
TGTAAACATCCTTGACTGGAAGCTG	4
CATTATTACTTTTGGTACGCGC	42
AGCTGGTGTTGTGAATCA	2
ACTATATATCAAGCATATTCCT	1
TCAGGCTCAGTCCCCTCCCGATAAA	2
TGCCTGGTAATGATGA	1
TGTGACTGGTTGACCAGAGG	12
GCCCTAAGGTGAATTTTTTGGGA	1
AGTTGTGTGTGCATGTTCATGTC	6
GTACCGTGAGTAATAATGC	1
TGTGACTGGTTGACCAGAG	4
CAGTGTCTTAGCTGGTTGTT	1
TGACCCCTAGTCCAGTGCTTGTGG	1
GTTCACAGTGGCTAAGTTCC	1
TCCGAGCCTGGGTCTCCCTC	4
CTGCGCAAGCTACTGCCTTGC	12
TATGTCACTCGGCTCGGCCCACTACCC	1
AGTTTTCCCAGGAATCCCTT	1
CAGTGGTTTTACCCTATGGT	5
TGAGGTAGTAGGTTGTGTGG	517
ACTCTACAACCTTAGGACTTGC	7
TAAACATCCTCGACTGGAAGCT	1
AAAAGCTGGGTTGAGAGGGCGA	129
TAAGGCACGCGGTGAATGCCA	4
AACTGGCCTACAAAGTCC	6
AGTTGTGTGTGCATGTTCATG	3
AGTTTTGTGTGCATGTGCATGTG	12
GCAGCACATAATGGTTTGT	1
TGTCAGTTTGTCAAATACCCCAA	10
ATCACAAGTCAGGTTCTTGGGA	22
ATTGCCAGGGATTTCCA	2
CATTATTACTTTTGGTACG	41
TCACAAGTCAGGTTCTTGGGAC	22
TGTAAACATCCTACACTCA	18
AATCACTAACTCCACTGCCA	23
TAACACTGTCTGGTAA	1
CGGGGCAGCTCAGTACAGGATG	18
TTCAAGGCTATCACCATCCCCCTTCAG	1
GAGCTTACAATCTAGCTGG	1
ACAGTCTACAGCCATGGTCGC	3
TGAGGTAGTAGATTGTATAGTT	3051
ATCACTAACTCCACTGCCATCAA	2
TGTAAACATCCTCGAC	2
CGGTGCGGAGAGCCGTTCGTC	1
CTCACACAGAAATCGCACCCGTC	1
AGGGGTGCTATCTGTGATTGA	6
CAAGTAATCCAGGATAGGCT	11
CAAAGTGCTGTTCGTGCAGGTAGT	1023
CGCCACTAGGGTTGTG	1
ACCACAGGGTAGAACCACGGAC	342
CACCATCTGAAATCGGTTA	2
AGGTTACCCGAGCAACTTTGCAT	3
ATACCCTGTAGATCCGAATTTG	1
CATCTAGTGCAGATA	1
AGTAGTCTGCACATTGGTTA	1
AGTGTTTCCTACTTTATGGA	54
CTAATACTGCCGGGTAATG	1
TCCTTCATTCCACCGGAGTCTGT	3
CATCTGAAATCGGTTAT	1
TGAGGTAGTAGATTGTAT	10
AAAGCTGGGTTGAGAGGGCGA	8
AGTTCTTCAGTGGCAAGC	1
TGTGCAAATCCATGCAAAACT	385
TAATACAACCTGCTAAGTGT	1
TAATACTGTCTGGTAATG	6
GTTGTGTGTGCATGTTCATGT	1
TGAGGTAGTAGGTTGTATAGT	3054
AACATTCATTGCTGTCGGTGGGTTTGA	4
ACCTTGGCTCTAGACTGCTTA	2
GAGGTAGTAGGTTGTATAGT	3
CCACAGGGTAGAACCACGG	2
TAGTAGATTGTATAGT	1
CGACAGCACGACACTGCCTTCA	1
AGCACGTAAATATTGGC	1
TTGTGCAAATCTATGCAAAACTGAT	2
CAACAAGTCCCAGTCTGCCACA	5
GTTACCATTACTGAGTTT	1
ACTGTCTGGTAAAGATGG	2
ACCGGGGTCCGGTGCGGAGAGCCG	5
CTGGTTTCACATGGTGGCTTAGATTTT	2
CAAAGAATTCTCCTTTTGGGCTTT	1
CCCTTTAACCTGTGA	1
ACAGTAGTCTGCACATTGG	16
AGCTACATTGTCTGCTGGGTTTCA	7
AGGTTGTCCGTGTTGTCTTCTCTTT	1
TGTTTCTAATGTACTTCACCTGGTCCACTA	1
TAGCAGCACGTAAATATTGG	26
TAAAGTGCTTATAGTGCAGGTAG	7188
AGCTCGGTCTGAGGCCCCTCAG	3
TCTCGCTGGCTGTTCCTTT	1
CTCCTGACTCCAGGTCCTGTG	18
TGTGCAAATCCATGCAAAAC	111
ATATAATACAACCTGCTAAGT	32
CATCCTTGACTGGAAGC	1
TAGGTTATCCGTGTTGCCT	2
TGCAGTCCACGGGCATATACA	1
GAAAAAAAAAAACAGGTCAAGAGGCGC	2
AAACAAACATGGTGCACTTCTT	16
TGAGGTAGTAGATTGTATAGTTTT	1
TGAGGTAGTAGATTGTATAGTTGTGGGGTA	1
CCCTGAGACCCTAACTTGT	2
TCCCTGAGACCCTTTAACCTGTG	607
ATGGTTGACCATAGAACATGC	1
CCTGCCGCGCCCCCCGGGC	1
CTGGTTTCACATGGTGGCTTAGA	12
GTACAGTACTGTGATAGCTGAA	2
AAGGAGCCTCGGTTGGCCCCGGATAGCC	2
ACATTCATTGCTGTCGGTGGGT	7
TCCCTGTCCTCCAGGAGCT	3
TGGAAGACTAGTGATTTTGTTG	1
TGAGGTAGTAGATTGTATAGTTT	104
TAATCCTTGCTATCTGGGTGCTTAGTG	1
AATAATACATGGTTGATCTTTT	3
CAACAAATCACAGTCTGCCATAT	3
TGTGACAGATTGATAACTGAAAG	16
CGTACCGTGAGTAATAATGC	235
ACACTGTCTGGTAACGATGT	1
TGCAAATCCATGCAAAACTGA	1
AGTAGGTTGCATAGTTT	3
GTTGAAAAGGTGCAGGCCATACTGTGCTG	1
AACACTGTCTGGTAAAGATGGCC	3
AATCATTCACGGACAACACTTT	4
TAAGGCACGCGGTGAATGC	1
GTATGTGCCTTTGGACTACATCG	10
CTGTGCTGACGTTTGATTCTTGAGCAG	3
GTTGGCCCCGGATAGCCGGGTCCCCGTCCG	1
TGTGACAGATTGATAACTGAAA	7
TAAACATCCCCGACTGGAAGC	1
GCAGTCCACGGGCATATACAC	4
CCGTCCTGAGGTTGTTG	1
TGCTATGCCAACATATTG	1
AGGTAGTAGATTGTATAGTTGTGGGGTAGT	1
AACATTCATTGCTGTCGGTGG	48
AAACAAACATGGTGCACTTC	1
TTTCCTCATATCCATTCAGGA	1
CATCTTCCAGTGCAGTGTTGGAT	1
CTATACAATCTACTGTCTTTCC	17
GCTGTTCGTGCAGGTAGT	2
AACAAACATGGTGCACTTCTT	4
CCTCCCACACCCAAGGCTTGC	47
ACAGGTGAGGTTCTTGGGAGC	7
AATGGCGCCACTAGGGT	1
CACTCTTTCCCTGTTGCACTAC	1
TCCCTGTCCTCCAGGAGCTC	5
CTGACCTATGAATTGACAGC	50
AATGACACGATCACTCCCGTTGA	17
ACTGCCCCAGGTGCTGCTGG	30
CCAGTTTTCCCAGGAATCCCT	19
GTAGTAGGTTGTATGGTT	2
GCAGCACGTAAATATTGGCG	10
CCCAGTGTTCAGACTACCTGT	161
CAAAGCACAGGGCCTGCAGAGAGGTA	7
CATGCAGTCCACGGGCATATACA	4
CATCTTACCGGACAGTGCTGGA	18
AGTGTTTCCTACTTTATGGATG	144
GACCGATTTCTCCTGGTGT	1
TAATGCCCCTAAAAATCCTTATT	109
CTGCCTGGTAATGATGAC	1
TAGCTTATCAGACTGATG	70
AAGGAGCTCACAGTCTATTG	11
TCACATTGCCAGGGATTACCA	116
GTGGTCCTCTCTGTGCTA	3
CACATTGCCAGGGATTACCA	1
CGGTTGGCCCCGGATAGCCGGGTCCCCG	2
ATAGTTGTGTGTGGATGTGTGTA	1
GCAGCATTGTACAGGGCTAT	2
TCCCACTCAGTCGTGCCC	15
CATCTTCCAGTGCAGTGTTGG	2
GGGGGTCCCCGGTGCTCGGATCTCGAGGGT	1
ATTGGGGATGCTTTGCATTCAT	15
TCGGCAACAAGAAACTGCCTGA	1
GGTCCGGTGCGGAGAGCCGTTCGTCTTGGG	1
TTTGCGATGTGTTCCTAATAT	6
TAGCAGCACAGAAATATTGGCATGGGG	1
AGTTCTTCAGTGGCAAG	1
AACAGTCTACAGCCATGGTCG	5
TCAGTGCACTACAGAACTTTGTC	55
TTAATGCTAATTGTGATAGGGGT	24
TCAGTTATCACAGTGCTGAT	4
ATCACATTGCCAGGGATTACCAC	683
TAACACTGTCTGGTAACGATGTTC	4
TGTAAACATCCTACACTCTCAGCT	1331
TGTAAACATCCTACACTCT	54
TTTCCCAGGAATCCC	2
CAGCAATTCATGTTTTGG	1
TGAAACATACACGGGAAACCTCTTTT	10
TGACCTATGAATTGACAGC	1
TAATACTGCCGGGTAATG	6
TCAAGTAATCCAGGATAGGCT	77
TGTAAACATCCTTGACTGGAAGC	436
GCTGGTGTTGTGAATCAGGCCGTT	2
TTATCAGAATCTCCAGGGGT	26
GGGAGCCAGGAAGTATTGATGTT	3
TCAAGTAATCCAGGATAGGC	23
CAGTGGTTTTACCCTATGGTAG	61
CCCAGTGTTTAGACTACCTGTT	99
TTCACAGTGGCTAAGTTCCGCC	1
CTGAGATGGGAGTGGTGTAA	1
TAGTGTTTCCTACTTTATGGA	112
AAGTGCCTGCATGTATATGCG	1
GTCTTACCCAGCAGTGTTTGG	1
TATGCAAGGGCAAGCTCT	1
TGGCAGTGTCTTAGCTGGTTGTTGTGAGT	2
TAGCACCATTTGAAATCAG	15
TGTAAACATCCTTGACTGGAA	15
TGTAGTTAGCTGATTGC	1
AAAGTGCTGTTCGTGCAGGTAGT	1
TTCTGCTGTGCAAATCCATGCAAAACTG	2
ACTGCTGAGCTAGCACTTCCCGA	3
ACATTGCCAGGGATTTCCAA	1
TATTGGGAACATTTTGCATGCA	18
CCCAGTGTTTAGACTACCT	3
GCAGCACGTAAATATTGGC	6
ATAGCTTATCAGACTGATGTTGAC	22
AATATAACACAGATGGCCT	1
TGTAAACATCCTCGACTGGAA	117
TTGTGCAAATCTATGCAAA	1
TGTGTGCATGTGCATGTGT	1
TCGTGTCTTGTGTTGCAGCCG	13
TCCGTCTCAGTTACTTTATAGCCA	3
TGTTGTGAATCAGGC	1
TATGGCTTTTTATTCCTATGTGAT	2
TAAGGTGCATCTAGTGCAG	57
ACTGCCGGGTAATGATGG	1
TGCCTGTCTACACTTG	4
CAGCGCCGAAGGAGCCTCGGTTGGCCCCGG	21
AAGTGCTGACAGTGCAG	1
TGTAAACATCCTTGACTGGA	10
ACAGTCTCCAGTCACGGCCA	1
TCTACAGTGCACGTGTCTCCA	1
AACATTCATTGCTGTCGGTGGG	21
TCCGGTTCTCAGGGCTCCACCT	1
TAGCAGCACATCATGGTTTA	211
CCCTGAGACCCTTTAACCTGT	2
TCAAGAGCAATAACGAAAAATGT	11
CGTAGATCCGAACTTGT	1
TAATGCCCCTAAAAATCCTT	28
TAGCAGCGGGAACAGTA	1
ACTGTGCGTGTGACAGCGGCTGA	9
GTAAACATCCTACACTCAGCT	1
GAAGGAGCCTCGGTTGGCCCCGGATAGCCG	2
AAACAAACATGGTGCACTTCTTT	15
TTTCCTCATATCCATTCAGGAGTGT	1
CTATACGACCTGCTGCCTTTCT	20
TGAGGTAGTAGTTTGTACAGT	121
TCTCTGGGCCTGTGTCTTAGGC	2
AGCATTGTACAGGGCTATGA	3
AGCTGCCAGTTGAAGAACTGT	7
ACAAGAGGGCCTTTCTGG	8
AATGACACGATCACTCCCGTTGAGTGGG	1
TAGCAGCGGGAACAGTACTGCAG	197
CACAGTGGCTAAGTTCTGC	1
TAAGGTGCATCTAGTGCAGATAGTGA	2
TCCATCTTCCAGTGCAGTGTTGGA	4
ATACAACCTGCTAAGTG	1
GTAAGTGCCTGCATGTATATG	2
CAAAGTGCTTACAGTGCAGGTAGTG	1
TTCCCTTTGTCATCCTATGC	88
GTACTGCCCTGCTGATGGAT	2
AGGGGTGCTATCTGTGATTGAGGG	17
TGGCTCCGTGTCTTCACTCCC	1
TTTGTTCGTTCGGCTCGCGTGA	117
AACATTCAACGCTGTCGGTGA	366
AACACACCCAGCTAACCTTT	1
TAAAGTGCTGACAGTGC	1
CTAGACTGAGGCTCCTTGAGGA	15
GTTTTGTGTGCATGTGCATG	1
TAGCACCATCTGAAATCGG	6
GTGAAATGTTTAGGACCAC	1
GTAAGTCTGGGTCTTCCACATGGGCA	1
TAAAGTGCTGACAGTGCAGA	71
CCGCACTGTGGGTACTTGCTGC	44
TGTCACTCGGCTCGGCCCACTAC	1
TGAGCGCCTCGGCGACA	1
TAGGAGGTTGTATAGTT	2
CACTGAGATGGGAGTGGTGTAA	5
ATACTGCCGGGTAATGA	1
TGTAAACATCCTCGACTGG	57
TCTCCCAACCCTTGTACCAG	3
TAGCACCATCTGAAATC	4
TCCCTGAGACCCTTTAACCTGTGA	418
AAGGAGCTTACAATCTAGCTGGGG	2
TGCCGGGTAATGATGG	1
TATGCAAGGGCAAGCTCTCTTC	24
AAGTAATTCAGGATAGGT	2
TTGCACTTGTCCCGGCCT	1
CAGTTATCACAGTGCTGATGCT	25
CTTATCAGACTGATGTTG	2
ATACCCTGTAGAACCGAATTTGTG	1
TAAAGTGCTTATAGTGCAGGT	370
ACCAATATTATTGTGCTGCTT	1
AGAATCTCCAGGGGTAC	1
GAAGTTGCCCATGTTATTTTTCGCTT	1
GCTCAGTTCAGCAGGAAC	1
GCAGCACATCATGGTTTACA	3
AGGGGTGCTATCTGTGATTGAGGGA	48
CTAGGTATGGTCCCAGGGATCC	1
TTATAAAGCAATGAGACTGATT	17
TAAGTGCCTGCATGTATATG	24
TTTTGCAGTATGTTCCTGAATACA	4
TGAAATGTTTAGGACCACTAGA	9
CCTGAGACCCTTTAACCTGTG	2
TCCCTGAGACCCTAAC	6
AGGCAGTGTAGTTAGCTGATTGC	75
TGGTTTCACATGGTGGCTTAGA	2
TTGGTCCCCTTCAACCAGCTGT	131
CAAGCTCGTGTCTGTGGGTCC	8
CTAGCACCATTTGAAATCAGTG	28
CTAGCACCATTTGAAATCAG	2
AGTGTTTCCTACTTTATGG	7
CGTCTCAGTTACTTTATAGCC	1
TGAATGGCGCCACTAGGGTTGTG	2
TAGGTAGTTTCATGTTGTTGGGATT	2
GCCCTAAGGTGAATTTTTGG	1
CAGTGCAATGATGAAAGGGCAT	125
GGTAGTAGGTTGTATA	1
TAGCACCATTTGAAATCAGT	89
AAAGTGCTGTTCGTGCAGGT	1
AATGACACGATCACTCCCGTTGAGTGGGC	1
TATGGCACTGGTAGAATTCA	2
AGTAGGTTGTATAGTT	9
ACTGCCCCAGGTGCTGCTG	5
TGAGAGATGCCATTCTATGTAGA	3
AACACACCTGTTCAAGGATTC	1
AGTTGTGTGTGCATGTTCATGT	17
CTGAGAACTGAATTCCATGGGT	1
TAAGTGCCTGCATGTATA	2
TGTAAACATCCTTGACTGG	9
CCCTGTAGAACCGAATTTGTG	2
TCCCTGAGACCCTAA	1
TAGGTAGTTTCCTGTTGTT	4
AGCACATAATGGTTTGT	1
TGAAGCTCAGAGGGCTCTGAT	1
TTCCCTTTGTCATCCTATG	5
GAATGTTGCTCGGTGAACCCCTTT	2
TAACACTGTCTGGTAAAGATGGCCC	18
AGGTTCTGTGATACACTCCGACT	2
CTGAAGCTCAGAGGGCTCTGA	1
AACCCGTAGATCCGATCTTGTGGT	1
TTCCTATGCATATACTTCTTT	5699
TAAAGTGCTTATAGTGCAGGTAGTGTGT	4
CAGTGGTTTTACCCTATGG	1
ATATTGCACATTACTAAGTTGC	5
CTTTTTGCGGTCTGGGCTTGCT	2
ACATCCTCGACTGGAAGC	1
TATTGCACTTGTCCCGGCC	2
ACAGCAGGCACAGACAG	3
CAAAGTGCTAACAGTGCAGGTAG	13
TTGTGCAAATCTATGCAAAACT	6
TGGAATGTAAAGAAGTATGTA	2
ATGCAGTCCACGGGCATATACA	5
CGGGGCAGCTCAGTACAGGA	1
TGACCCCTAGTCCAGTGCTTGTGGTGGCTA	1
CAGGTGAGGTTCTTGGGAGCC	4
AGCTCGGTCTGAGGCCCCTCAGT	3
CGGCGTCATGCAGGAGTTG	1
AGTGGCTAAGTTCCGC	1
CCCACGTGCGGTGCTTGAAGA	1
CAACTAGACTGTGAGCTTCTAGA	1
CATCCCTTGCATGGTGGAGGGT	2
CAAGCTCGTTTCTATGGGTC	3
TTGTGCTTGATCTAACC	1
TGAGGTTGGTGTACTGTGTGTGAGTAT	3
CAGTTATCACAGTGCTGATGCTG	5
TAACACTGTCTGGTAAAGATGGCC	145
TCACAAGTCAGGTTCTTGGGA	13
GAACGGCTACTTCACAACACCAGGGC	1
CCATCTGAAATCGGTT	2
CAACCCTAGGAGGGGGTGCCATTC	5
TTTGAAATCAGTGTT	1
AGCTTATCAGACTGATGTTGAC	8
GAAAAGCTGGGTTGAGAGGGCGAAA	2
AGCTGGTGTTGTGAATCAGGC	228
AAAAGCTGGGTTGAGAGGGCGAAAAAG	1
AACAGTCTACAGCCATGGTC	1
TAACAGTCTCCAGTCACGGCCA	3
AGGAGCTTACAATCTAGCTG	1
GGACAGAGTTGTCATGTGTCT	1
TGTAAACATCCTACACTCAGC	1296
TGTAAACATCCTACACT	6
TCCCTGTCCTCCAGGAGCTCACGTAT	3
TCAAGAGCAATAACGAAAAAT	10
TGCGGGGCTAGGGCTAA	1
ACATTGCCAGGGATTA	1
TAATACTGCCGGGTAATGAT	192
CAACTAGACTGTGAGCTTCTA	2
TAAGGTGCATCTAGTGCAGAT	466
AGAGGTAGTAGGTTGCATAGTTTT	58
ACCACCGACCGTTGACTGTACC	2
TCCTATGCATATACTTCTT	1
TAAGTCACTAGTGGTTCCGTTTAG	14
TTTTCCCAGGAATCCCTT	2
CCACAGGGTAGAACCACGGAC	17
TTGGTCCCCTTCAACCAG	2
CTAGCACCATCTGAAATCGGT	16
CATTGCACTTGTCTCGGTCTGAC	2
CGGTGCTGGTGGAGCAGTGAGCACG	3
CAAATTCGTATCTAGGGGAAT	41
GCTGGGGTGGAACCTGGTTTGT	2
TGTACTGCCCTGCTGATGGAT	2
AGATAACTATACAACCTACTGCCTTC	1
AACGGAATCCCAAAAGCAGCTG	4
TGATATGTTTGATATATTAGGT	78
AAGTGCCTGCATGTATATGC	1
GGTAGTAGGTTGTGTGGTT	2
ATACTGCCGGGTAATGATGGA	10
TATTGCACTTGTCCCGGCCTGTTGA	1
TCGTACCGTGAGTAATAATGC	656
TCTTTGGTTATCTAGCTGTAT	5
GCGCCGAAGGAGCCTCGGTTGGCCCCGG	1
GGGTGCTATCTGTGATTGAGGGACA	1
TCCGTCTCAGTTACTTTATAGCC	42
TAACACTGTCTGGTAACGATGT	3654
TAGGTAGTTTCCTGTTGTTGG	189
GAGGTAGTAGGTTGCATA	1
ATGAAGCGCTGCAAC	1
TATGTGGGACGGTAAACCGCT	1
TCCATCTTCCAGTGCAGTGTTGG	36
GTCCAGTTTTCCCAGGAAT	73
AGGTTGCCCATGGTGTGTTCATTTTATTTA	1
TCAAGTAATCCAGGATAGGCTG	1
AGAGGTAGTAGGTTGCATAGTTT	568
TAGCAGCGGGAACAGTACTGCAGTG	1
TGAGGTTGGTGTACTGTGTGTG	129
TCCCTGAGGAGCCCTTTGAG	397
CGCCGAAGGAGCCTCGGTTGGCCCCGGATA	7
TTAAGACTTGCAGTGATGTT	1
AGGAGCTTACAATCTAGCTGG	1
TGGGTCTTTGCGGGCAAGATG	1
CACATCTGCACATGCTGGGG	1
AATTCAGGATAGGTT	1
CAGTCCACGGGCATATACACT	3
TAAAGTGCTTATAGTGCAG	22
GAGGTAGTAGTTTGTGCTGTT	2
CCCCCAGGTGTGATTCTGATTTGTTT	1
TAGTAGGTTGTGTGGTT	4
ATGTATGTGTGCATGTGCATGTG	1
CTCGGGGATCATCATGTCACG	31
AAGCCCTTACCCCAAAAAGTA	1
TGAGGTTGGTGTACTGTGTGTGA	1505
TGTAGTGTTTCCTACTTTATGG	880
GTAGTGTTTCCTACTTTATGGA	661
AGCTACATTGTCTGCT	1
TAGGTAGTTTCCTGTTGT	1
TACCCTGTAGAACCGAATTTGTGT	33
TAAAGTGCTGACAGTGCAG	110
TTGCATAGTCACAAAAGTGATC	1
TTGGTCCCCTTCAACCAGC	1
AATCCTTGCTATCTGGGTGCTTAGT	28
TAGCAGCACGTAAATATTGGCG	3440
TCGTGTCTTGTGTTGCAGCC	1
TGAGGTAGTAGGTTGTATA	41
CAAGTAATTCAGGATAGGT	1
AGCGCCGAAGGAGCCTCGGTTGGCCCCGG	1
TCTCTGGGCCTGTGTCTTAG	4
ACCTGGCATACAATGTAGATTT	1
TACACTGCCGGCCTCTGGG	5
CTGCCAGTTGAAGAACTGT	1
AATCCTTGCTATCTGGGTGCTTA	1
CTATACAATCTATTGCCTTCC	6
AACAAATCACAGTCTGCCATAT	5
ATCTTACCGGACAGTGCTGGATT	1
GTATGTGCCTTTGGACTACATCGTGAACG	3
TTTGGCAATGGTAGAACTCACACCGGT	4
TGCTATGCCAACATATTGC	9
GCCCCTGGGCCTATCCTAGA	12
TATTACTTTTGGTACGCG	1
CATTGTACAGGGCTATGA	2
CAGTGCAATGTTAAAAGGGCAT	8552
GTAATCCAGGATAGGCT	8
CATAAAGTAGAAAGCACTACT	39
AGCTACATTGTCTGCTGGGTTTC	119
CAGCAGCACACTGTGGTTTGTAC	1
TGAGGTAGTAGGTTGTATAGTT	9311
CGGAATCCCAAAAGCAGCT	1
AGGAGCCTCGGTTGGCCCCGGATAGCCGGG	21
CAACGGAATCCCAAAAGCAGCTGTTG	1
CAGTTCAGCAGGAACAG	3
TGAGGTAGTAGGTTGT	11
TGGCAGTGTATTGTTAGCTGGTT	3
CATTTGAAATCAGTGTT	8
AAACATTCGCGGTGCACTTCT	2
GCGACCCATACTTGGTTTCAGT	1
AATCCTTGGAACCTAGGTGTGAATGC	682
TGAGATGAAGCACTGTAGCTC	483
ATGTAAACATCCTACACTCAGC	3
TGACCGATTTCTCCTGGTGTTCAGA	1
AAAACATGAAGCGCTGCAAC	3
ACCACAGGGTAGAACCA	1
TTATGTGTGTGTACATGTACA	1
CAGTGCAATGTTAAAAG	2
GTGTTTCCTACTTTATGGATG	1
CAAAGTGCTGTTCGTGCAGGTA	652
GCACCATCTGAAATCGGT	1
TGATTTCTTTTGGTGTTCAG	1
AACCGTGGCTTTCGATTGT	1
TCTACAGTGCACGTGTCTCCAGT	226
CTGGACTTGGAGTCAGAAGG	2
TATTGGCTGGCTCTGGGTC	1
GTGTAATTAGCTGATTGT	1
TCACATTGCCAGGGATTACC	15
TTCACAAAGCCCATACACTTTC	69
TTCAAGTAATCCAGGA	8
GTAGTGTTTCCTACTTTATGGAT	54
GCTATGCCAACATATTGCC	1
CTTCATTTATGGCGCACATTACACGGTCGA	2
CTGAGGAGCCCTTTGAGCCTG	1
TCAGCTCCTATATGATGCCTT	3
GGGCGTCGGGACCGGGGTCCGGTG	1
TTAGCTGCTTGTGAGCA	1
TCAAAGTGCTTACAGTGCAGGTA	8
CCCCTTCAACCAGCTGT	3
GCCTGTCTACACTTGCTGTGCA	29
TCAAAGTGCTTACAGTGCAG	7
TAGCAGCACAGAAATATTGGCAT	40
TTTTGCAGTATGTTCCTGAATAC	11
GGGGTGCTATCTGTGATTGAGGGACAT	6
TCGGTCGATCGGTCGGTCGGT	1
GCAGCAGCAATTCATGTTTT	2
AATCCATGCAAAACTGA	1
TAACAGTCTACAGCCATGGTC	89
CACAAGTCAGGTTCTTGGGA	1
ACCACCGACCGTTGACTGTAC	1
AGGTGCAGTGCTGCATCTCTGGTC	1
AGAGGTTTTCTGGGTCTCTGTTTC	1
AGCAGCACACTGTGGTTTGTAC	2
AGGTAGTAGTTTGTGCTGT	1
CCCATAAAGTAGAAAGCACT	6
AGTTTTGCAGGTTTGCATCCA	6
CCCCTTCAACCAGCTG	1
CCTGTCTACACTTGCTGTGCAG	1
TATGTATGTGTGCATGTGCATG	1
ACATTGCCAGGGATTT	1
ATTCAAGTAATCCAGGATAGG	1
CAGTGCAATGGTATTGTCAA	2
TGTAAACATCCTACACTCTC	107
CCCAGTGTTTAGACTACCTGTTCAGGAC	1
CAGTTATCACAGTGCTGATGC	25
CAATGTTTCCACAGTGCATCA	5
CCTCGGTTGGCCCCGGATAGCCGGGTCCC	2
AAGTAATCCAGGATAGGCT	21
GCCCCTGGGCCTATCCTAGAAC	2
GCTCTTTTCACATTGTGCTACT	5
ACATAACATACACACACACGTAT	1

GGTAGTAGGTTGTATGGT	2	CTGCAGTGAGGGCACTTGTAGC	1	ATCTGAAATCGGTTA	2	TTTGTTCGTTCGGCTCGCGTG	7	CATTATTACTTTTGGTACGCG	767
TCACACAGAAATCGCACCCGTC	3	TACAGTAGTCTGCACATTGGT	275	ATTCCTGGAAATACTGTT	2	TAGCAGCACGTAAATATTGGCGTAG	1	AGGTTACCCGAGCAACTTTGCATCT	1
AGCAGCATTGTACAGGGCTATCA	503	CAGCAGCAATTCATGTTT	1	CGTGTATTTGACAAGCTGAGTTGGACAC	2	TAAGGTGCATCTAGTGCAGATAGT	75	AGCTCGACTCATGGTTTGAACCA	1
CGGTTATCATGGTACCGATGCTGTAGCT	2	ACTCCATTTGTTTTGATGATGGA	10	ACTGCCTGGTAATGATGA	2	TAGCAGCACATAATGGT	1	TCACATTGCCAGGGATT	1
CAAACCACACTGTGGTGTTAGA	2	CACTGTCTGGTAAAGATGG	1	TGTAAACATCCTACACTCTCAGCTGT	45	AGGTAGTAGATTGTATAGT	1	CAGTGCAATAGTATTGTCAAAGCATC	14
AGTCTGCACATTGGT	1	GTGCAATGTTAAAAGGGCAT	1	CTGAGATGAAGCACTGTAGCTCA	2	ACCCTGTAGATCCGAATTTG	4	TTTGGCAATGGTAGAACTCACA	2
GCACATCATGGTTTACA	1	GCAGTCCACGGGCATATACA	9	CAGGTGAGGTTCTTGGG	1	TATACATACACGCACACATAAGA	2	ACAAGTCAGGTTCTTGGGACCTA	17
AGAGGTATAGCGCATGGGA	2	TGAAGAGAGGTTATCCTTTGTGTGT	7	CAGTGCAATAGTATTGTCAA	29	TTCAAGTAATTCAGGATAGG	342	ACCATTTGAAATCAGTGT	9
ACCACAGGGTAGAACCACGGACA	173	ACAGTAGTCTGCACATTGGT	678	TTTCAGTCGGATGTTTACAG	3	CCTGGCTGGACAGAGTTGTCATGTGTCT	1	GTAATTCAGGATAGGTT	3
TGAGGTAGGAGGTTGTATAGTT	593	GCAAAGCACAGGGCCTGCAGAGAGGT	55	CAAAGTGCTTACAGTGCAGGTAG	4674	CTTTCAGTCAGATGTTTGCTGCT	2	CTGTTGCCACTAACCTCAAC	1
AAGCCCTTACCCCAAAAAGCAT	6	TCAGTGCACTACAGAACTT	3	TGGCTACTGGGCCCTGAACTAGGGGTCTGG	1	CTGACCTATGAATTGACAG	6	CCACAGCACTGCCTGGTCAGA	1
TATGGCTTTTTATTCCTATGTGA	736	CTCGTACCGTGAGTAATAATGCG	11	AGCTTCTTTACAGTGTTGCCTTG	23	TGGAGAGAAAGGCAGTTCC	1	ACCTTGGCTCTAGACTGCTTACTG	2
TAGCAGCACGTAAATATTGGCGTA	1	TAGCAGCACGTAAATATTGGCGTT	16	TGTAAACATCCTCGACTGGAAG	474	GACCGATTTCTCCTGGTGTTCAGA	2	GCACGTAAATATTGGCG	3
AAGTGCTTATAGTGCAGGTA	1	TACTGCATCAGGAACTGACTGG	6	TTGTGACAGATTGATAACTGAA	1	CAGTTTTCCCAGGAATCCCT	46	GTACAGCAGGCACAGACAGGCA	2
TGACCTATGAATTGACAGCCA	8	AGAGGTTTTCTGGGTCTCTGTTTCT	2	CGAATCATTATTTGCTGCTCTA	11	TGGCAGTGTCTTAGCTGG	4	TAAAGTGCTTATAGTGCAGGTA	570
GCAGCACAGAAATATTGGC	3	CCACTGCCCCAGGTGCTGCT	3	TTGCATAGTCACAAAAGTGAT	1	TGTAAACATCCCCGA	2	TTCACAGTGGCTAAG	2
TCAGGCTCAGTCCCCTCC	1	AGGTAGTAGGTTGTGTGGTTT	1	CAGTCCACGGGCATATACACTT	11	CAAAGTGCTGTTCGTGCA	3	AGTGGCTAAGTTCTG	1
CTCAGTCCCCTCCCGAT	1	TGTAGTGTTTCCTACT	1	TATGTGCCTTTGGACTACATCGT	130	TAAAGTGCTTATAGTGCAGGTAGT	1190	AGGAAGGGAGAAGATGTCCTGCAGCCGTCA	1
TGAGGTAGGAGGTTGTA	2	CTAGACTGAGGCTCCTTGAG	10	TACAAGGGCAAGCTCTCTGT	1	CAGTGCATCACAGAACTTTGT	1	CAACAAGTCCCAGTCTGCCACAT	1
AAGGGATTCTGATGTTGGTCA	20	CCAGTGTTCAGACTACCTGT	1	TATACAAGGGCAAGCTCTCTGT	69	CAAAGTGCTAACAGTGCAGGT	4	GAGGTAGGAGGTTGTATAGTTG	1
GCTTATCAGACTGATGTTGAC	46	GTAGTAGGTTGTATGGTTT	1	AATCATACACGGTTGACCTATT	37	AAACCGTTACCATTACTGAGTTTA	4	TTCCCAGGAATCCCT	2
TAAAGTGCTGACAGTG	2	CTGTACAGGCCACTGCCTTG	5	CCCAGTGTTCAGACTAC	10	AGGAGCTTTCAGTCGGATGTTTACAGC	2	AACATTCAACCTGTCGGTGA	12
AGCTTATCAGACTGATGTTGA	26	CTCCTGACTCCAGGTCCTGTGTGT	2	ACCTTGGCTCTAGACTGCTTACT	11	ACACTGGTGCTCTGTGGGA	3	TGAGGTAGTAGGTTGTGTGGTT	11073
TAAGGTGCATCTAGTGCA	1	TAAAGTGCTGACAGTGCAGATAG	14	CAGGCCATACTGTGCTGCCTC	1	TAATACTGTCTGGTAATGC	34	TTTTCCTATGCATATACTTCTT	2
TGAGGTAGTAGTTTGT	1	CACCATTTGAAATCAGTGT	10	ATGACACGATCACTCCCGTTGAGT	1	AAGAGGTATAGCGCATGGGAA	4	ATCACATTGCCAGGGATTTC	615
AGGGTTGGGTGGAGGCTCTC	1	CTTTCAGTCGGATGTTTGCA	46	ACATTCAACGCTGTCGGTGAGTTT	8	TAGCAGCACGTAAATATTG	6	TCAAAGTGCTTACAGTGCAGG	8
AATCCTTGCTATCTGGGTGCTTAG	2	CCGTCCTGAGGTTGTTGAGC	30	TAAACATCCCCGACTGGAA	1	TCAAGTAATTCAGGATAGGTT	3	TAGGCAGTGTAATTAGCTGAT	1
AGCAGCATTGTACAGGGCTATCAAA	62	TGAGGTAGTAGGTTGTAT	121	TCGTACCGTGAGTAATAA	1	TCCCTGTCCTCCAGGAGCTCA	8	ACAAGTCAGGTTCTTGGGA	15
CAAAGTGCTTACAGTGCAGG	29	GAAAGGTGCCATACTATGTATA	1	TTGTGGTGGCTACTGGGC	1	TGTAGATCCGAATTTGT	1	GAAGTTGTTCGTGGTGGATTCG	1
AGCTGGTGTTGTGAATCAGGCCGA	184	AGCTGGTGTTGTGAATCAGGCCGT	667	CTGCAGTGAGGGCACTTGTAG	1	CTGTGCAAATCCATGCAAAACTGA	1	TAAACATCCTACACTCTCAGC	2
TTCCTCATATCCATTCAGGAGT	1	GGTCAAGAGGCGCCTGGGAAC	50	AATCCTTGGAACCTAGGTGTG	42	TGTGTGCATGTGCATGTGTGT	1	TGTAGGGATGGAAGCCATGA	3
GTTCACAGTGGCTAAGTTCTGC	32	TAATGCCCCTAAAAATCCTTA	150	GTGAAATGTTTAGGACCACT	13	CCGTGAGTAATAATGCG	2	CTTGGATGCTAAGATGGGG	2
AGAGCTTAGCTGATTGGTGA	2	TCAAGCAGCATTGTACAGGGCTATGA	1	ACTCTTTCCCTGTTGCACTACT	6	CAGTGTTTAGACTACCTGTTCAGGACTC	1	CAAGTAATTCAGGATAGGTT	4
AGCAGCATTGTACAGGGCTATC	154	AGCAGCATTGTACAGGGCTATG	151	AATCACTAACCACACAGCCAGGT	25	CCTGCTGGGGTGGAACCTGGTT	1	TGCACTTGTCTCGGTCTG	1
ATAGTTGTGTGTGGATGTGTGT	29	CTAATACTGCCGGGTAATGATGGA	18	TACAGTAGTCTGCACATTGGTTA	9	CCCTGAGGAGCCCTTTGAGCCT	2	AACCCGTAGATCCGATC	1
AATACTGTCTGGTAATGCCGT	4	TTTTTTGGTAATTGGGCAGCTGGTGA	1	TCCTGACTCCAGGTCCTGTGTGT	3	TCTTTGGTTATCTAGCTGTATG	7	CTGCGCAAGCTACTGCCTTG	3
CTGACCTATGAATTGACAGCCAGTACTCT	1	AAAGCTGGGTTGAGAGGGCGAA	2	TAATACTGTCTGGTAATGCC	65	AATGCCCCTAAAAATCCTTA	1	CAGCCCATCCCACTCAGTCGTGCCC	1
TAAGTCACTAGTGGTTCCGTTTAGT	18	TCCTGTACTGAGCTGCCC	2	TGTGCAAATCCATGCAAAACTGAC	1	TGTGCAAATCCATGCAAAACTGAT	63	CAAAGTGCTTACAGTGCAGGTAGTGA	3
CCCATAAAGTAGAAAGCAC	1	CAATGTTAAAAGGGCAT	1	CTTTCAGTCAGATGTTTGCT	4	AGTTCTTCAGTGGCAAGCTTT	95	GTGTCTTAGCTGGTTGTT	1
AACAATATCCTGGTGCTGAGT	11	ACTGCCCTAAGTGCTCCTTCT	2	CAAGCTCGTTTCTATGGGTCTG	4	GAATGTTGCTCGGTGAAC	1	GTACTGTGATAGCTGAA	1
TAGTCTGCACATTGGTTA	11	TCAGCTCCTATATGATGCCTTT	42	TAAGGCACGCGGTGAATGCC	1	CAGTGCAATAGTATTGTC	7	TGGTTTACCGTCCCACATAC	2
TAAACATCCTACACTCAGCT	1	CTGTGGGCCACCTAGTCACCA	1	TGTGCAAATCCATGCAAAACTGA	4075	TGGCAGTGTCTTAGCTGGTTGTT	875	AGGTGCAGTGCTGCATCTCTG	3
CTGGTTTCACATGGTGGCTTAGAT	7	GCTGACCCCTAGTCCAGTGCT	9	TGTCACTCGGCTCGGCCCACTACCC	15	AGAGGTAGTAGGTTGCATAGTT	4925	TTAGGGTCATACCCCATCTTGGA	1
GCAGTGTCTTAGCTGGTTG	2	TGTTGAAAAGGTGCAGGCCATACTGTG	3	CATTGCCAGGGATTACCA	2	GGTAGTAGTTTGTGCTGTTG	1	AACATTCATTGCTGTCGGTGGGTTTG	6
TGCGGGGCTAGGGCTAACAG	13	GCTTATAGTGCAGGTA	1	TGAGGTAGTAGTTTGTACA	2	ATAAAGTAGAAAGCACTACT	2	TTGGCTGGCTCTGGGTC	1
AATCCTTGCTATCTGGGTGCTTAGTG	13	TTCCTACTTTATGGATG	1	CGTGTATTTGACAAGCTGAGTTGGAC	2	CTCCTGACTCCAGGTCCTG	1	TTGTGCTTGATCTAACCATGTGC	16
TTGTGCTTGATCTAACCATGTGG	16	CTAACACTGTCTGGTAAAGATG	10	AAAATGGGTTCCTAGGA	1	CAGGTCGTCTTGCAGGGCTTC	1	AACATTCAACGCTGTCGG	1
GTGCTTATAGTGCAGGTAGT	1	TAGCAGCACATCATGGTTTACATACTAC	5	TGTGCAAATCTATGCAAAACTGAT	14	ACAACTGGCCTGAGGAGC	1	AGTTCTTCAGTGGCAAGCTTTAT	8
TTTTTGCGGTCTGGGCTTGCTGT	1	CTGTGCGTGTGACAGCGGCTGAT	64	TTAATGCTAATTGTGATAGGGGTTTT	12	TTAATGCTAATTGTGATAGGGG	9	TGATATGTTTGATATATTAGG	14
CCGTCCTGAGGTTGTTGAGCTG	59	ATGGTTTACCGTCCCACATACA	1	TTTGGCAATGGTAGAACTCACACCGG	3	CCGTCCCCGCTCGGCGGGGTCCCCGCGTCG	1	GACCGATTTCTCCTGGTGTTCAG	1
TAATACTGCCGGGTAATGATGGA	4379	CTGGACTTGGAGTCAGAAG	1	CTCGGTGCTGGTGGAGCAGTGAGCACG	1	AACATTCAACGCTGTCGGTGAGTT	220	TAATACTGCCGGGTAATGATG	1144
TATTGCACTTGTCCCGGCCT	9	ACCTTGGCTCTAGACTGCTTAC	2	AGGGACTTTCAGGGGCAGCTGTG	1	TAGTTGTGTGTGCATGTTCATGTCT	2	AATACTGCCTGGTAATGAT	1
AATGACACGATCACTCCCGTTGAGTGGGCA	1	TTCCCTTTGTCATCCTATGCCTG	160	TGAGGTAGTAGGTTGTA	17	TGAGGTAGTAGGTTGTG	9	CCTGTAGAACCGAATTTGTG	1
CAACCCTAGGAGGGGGTGCCA	2	TGTAAACATCCCCGACTGGAAGC	295	CTATACGGCCTCCTAGCTT	1	TGGTTTCACATGGTGGCTTAGAT	1	TAGTAGACCGTATAGCGTAC	5
TAGCACCATTTGAAATCAGTGTTTT	27	TAGCAGCACATAATGGTTTGTGGA	4	ATATAATACAACCTGCTAAG	1	CAGTAGTCTGCACATTGGTTAG	2	ACTGTGCGTGTGACAGCGGCT	1
GGATTCCTGGAAATACTGTTCT	7	TCTCACACAGAAATCGCACCCGTC	65	GCTGGTGTTGTGAATCAGGCCG	12	AGTTTTGCAGGTTTGCATCCAGCTGT	2	TCCCTGAGACCCTTTAAC	16
CATCTTTGCCGGTGACAGCA	7	AATGGCGCCACTAGGGTTG	24	CAAAGAATTCTCCTTTTGGGC	3	TCTCCCAACCCTTGTACCAGTGC	54	AATACTGCCGGGTAATGATG	5
TATACATACACGCACACATAAG	2	GTTCACAGTGGCTAAGTTCTG	2	TACAGTACTGTGATAACTGAAG	13	TTGAACTGTCAAGAACCACTGG	1	TTGGTCCCCTTCAACCAGCTG	20
TACTGTCTGGTAATGCCG	1	GGACTTGGAGTCAGAAGGC	1	TGAGATGAAGCACTGTAGCTCA	145	ATAGTTGTGTGTGGATGTGTGTATT	15	AAAGTGCTTACAGTGCAGGTAG	2
TTCACAAAGCCCATACACTTTCA	43	GCTACTTCACAACACCAGGG	1	ACCAGGAGGCTGAGGTCCCTTA	1	TGTGACTGGTTGACCAGAGGGG	23	CAAGCTCGTGTCTGTGGGT	9
ATCATACACGGTTGACCTATT	1	TGAGGTAGTAGGTTGTATAGTTTTAGGG	1	GGTCAAGAGGCGCCTGGGAACT	42	AAACCGTTACCATTACTGAGTTTAGT	40	TGTAAACATCCTTGA	1
CTATACAGTCTACTGTCTTT	2	ATCACATTGCCAGGGATTTCCAA	712	AGTTTTCCCAGGAATCCCT	34	TACGTAGATTGGTACCTATCATG	1	GTTTCATGTTGTTGGGA	1
GTAAACATCCTTGACTGGAAGCT	2	TTGCTGAAGCCAGATGCCGTTCCTGAGAAG	1	TTTCAGTCAGATGTTTGCTGCT	2	AGTGTTTAGACTACCTGTTC	1	TCGTACCGTGAGTAATAATGCG	1068
CGTATGTGCCTTTGGACTACAT	5	TCCAGCATCAGTGATTTTGT	12	TGGCAGTGTCTTAGCTGGTTGT	2253	TACCCTGTAGAACCGAATTTGT	348	TGTAGTGTTTCCTACTTTATGGAT	15
GGCTGCAGCGTGATCGCCTGCTC	1	TCAAAGTGCTTACAGTGCAGGT	76	CTTATCAGACTGATGTTGAC	4	TACCCTGTAGATCCGAATTTGTGTA	1	TGTAAACATCCCCGACTGGAAGCT	78
TACAGTACTGTGATAACTG	1	TAGCAGCACATAATGGTTT	2	CGTACCGTGAGTAATAATG	28	TATGGCACTGGTAGAATTCACT	7	CAGTAGTCTGCACATTGG	1
ACTGATTTCTTTTGGTGTTC	7	AGGGCTTCTGTCGTGAGTG	1	AAGAGGTATAGCGCATGGGAAGAT	2	TGTCTTGCAGGCCGTCATGCA	20	CCAGTATTGACTGTGCTGCTGAA	2
ACGGAATCCCAAAAGCAGCTGT	1	AGCAGCATTGTACAGGGCT	21	CCTCTGGGCCCTTCCTCCAGT	60	ATTTGAAATCAGTGT	1	TTATCAGACTGATGTTGA	3
TGAGATGAAGCACTGTAG	3	TGAGGTAGTAGGTTGTATGGTTTA	96	TTGTTCGTTCGGCTCGCGTGA	2	TGAGGTAGTAGGTTGTATGGTTTT	233	TTGTGCTTGATCTAAC	1
CGAATCATTATTTGCTGCT	2	TACTGCCTGGTAATGATGAC	2	ACTGCATTACGAGCACTTAAAG	9	TTTCAGTCGGATGTTTGCA	13	AACATTCAACCTGTCGGTGAGTTT	59
ACAGTAGTCTGCACATTG	7	TAAAGTGCTTATAGTGCAGGTAGTG	2	ACCTGGCATACAATGTAGATTTCTGT	3	ATGAGGTAGTAGATTGTATAG	1	ATACTGCATCAGGAACTGACTGGAT	1
TATGTAACACGGTCCACTAAC	53	AACACTGTCTGGTAACGATGT	10	AATCACTAACTCCACTGCCATC	157	CTGTCTGGTAACGATGTT	1	ATCACATTGCCAGGGATTA	132
TCAGGCTCAGTCCCCTCCCGATA	19	ATCACATTGCCAGGGATTT	356	CCATCCCACTCAGTCGTGCC	1	ATACTGCCTGGTAATGATG	7	CAAAACATGAAGCGCTGCA	5
TACAGCAGGCACAGACAGGCA	9	ATTGGGGATGCTTTGCATT	4	TAAGTCACTAGTGGTTCCGTTTA	1	TAGCAGCACATCATGGTTTACATAC	1	TGGTGTACTGTGTGTGA	1
ACAAGCTTGTGTCTATAGGT	1	GCAGCATTGTACAGGGCTATC	1	AGTTTTGTGTGCATGTGCATGTGTGT	1	AGAGGTTTTCTGGGTCTCTGTT	1	GTCAAGAGGCGCCTGGGAAC	8
TGACCCCTAGTCCAGTGCTTGTGGT	1	AGCAGCATTGTACAGGGC	7	AGTAATCCAGGATAGGCT	8	ACAGCAGGCACAGACAGGCA	3	GAGGTAGTAGGTTGTATGGTTTTGG	2
CCAATATTGGCTGTGCTGCT	2	TCAGTTATCACAGTGCTGATGCTG	1	TCAGGCTCAGTCCCCTCCCG	1	CCAATATTGGCTGTGCTGCTCC	4	TGTCTTGCAGGCCGTCA	1
TGAGGGGCAGAGAGCGAGA	1	AGTTTTGTGTGCATGTGCATG	1	CTTTCAGTCGGATGTTTA	6	CTTTCAGTCGGATGTTTG	3	AATCACATTGCCAGGGATT	4
AGGTTACCCGAGCAACTTTGCA	1	TAGCACCATTTGAAATCGGTTA	143	TTCAAGTAATCCAGGATAGGCTGT	20	GGCAGCGCCGAAGGAGCCTCGGTTGGCC	1	ACCGTGGCTTTCGATTGTTA	2
AAATCACATTGCCAGGGATTACCAC	1	TAGTCTGCACATTGGTT	2	GTCTGCACATTGGTTA	1	AAGTAATTCAGGATAGGTT	1	GGCAGAGGAGGGCTGTTCTTCCCTT	1
TTTCATTGATCTTGGTGTCCTCAAATT	1	TTTGGCAGTGTCTTAGCTGGTTGTT	1	ACTTAGCAGGTTGTATTATCAT	1	GTAGTAGGTTGTGTGGTTT	1	AGTGCAATGTTAAAAGGGCA	1
TGTAGAACCGAATTTGTG	1	TAGTAGGTTGCATAGTTT	2	AGCGCCGAAGGAGCCTCGGTTGGCC	1	TAGCAGCACGTAAATATTGGCGT	378	TCAGTGCATGACAGAA	1
TTGTGCTTGATCTAACCA	1	ACCACAGGGTAGAACCACG	9	CTATACGACCTGCTGCCTTTC	9	CAAAGTGCTCATAGTGCAGGT	5	CTTTGGTTATCTAGCTGTATGA	3
GTTTGATATATTAGGTT	1	GACCGATTTCTCCTGGTGTTC	2	GAGGTTGGTGTACTGTGTGTG	2	TGAGGTAGTAGGTTGTGTGGTTTCA	12	GTGAGTTCCAGGACAGCCAGGGCTATACA	1
CAGTACTGTGATAGCTGAAGA	1	TCAGGCTCAGTCCCCTCCCGAT	193	TCTCACACAGAAATCGCACCCGTCA	7	TGGTAGACTATGGAACGTAG	2	TGAAGCACTGTAGCTC	5
TATTGCACTTGTCCCGGC	29	TTCAGTGATGATTAGCTTCTGACTTT	1	CAGTTTTCCCAGGAATCC	4	CTTATCAGACTGATGTTGA	6	GAGGTAGTAGGTTGCATAGTT	18
AAGAGGCGCCTGGGAAC	1	AAGAGGTATAGCGCATGGGAAG	9	AACATTCATTGTTGTCGGTG	2	TTCACAGTGGCTAAGTTCTGC	11757	ATGCAGTCCACGGGCATATACAC	1
TGAGGTAGTAAGTTGTATTGTT	80	GGAGTATTGTTTCCACTGCCT	1	TCCCTGAGACCCTAACTTGT	75	TGGTTTACCGTCCCACATACA	3	TTTCAGTCAGATGTTTGCT	1
TAGCACCATTTGAAATCA	2	TAGCACCATTTGAAATCG	1	AGCTGCCAGTTGAAGAA	1	AGCTGGTGTTGTGAATCAGGCC	133	TGAGGTAGTAGGTTGTATAGTTT	507
TTTTGTTCGTTCGGCTCGCGTGA	2	ATCGTAGAGGAAAATCCACGT	7	TGTAGTGTTTCCTACTTTATGGATGA	1	GCTGTTCGTGCAGGTAG	2	TAGCAGCACATCATGGTTTACAT	12
TGTGACTGGTTGACCAGAGGG	9	GCTGGTTTCATATGGTGGTTTAGATTTAAA	1	CAGTTGCTAGCTGCACTACCCT	1	CTTTTTGCGGTCTGGGCTTG	3	GCAGTGCAATGATGAAAGGGC	1
TATGTCACTCGGCTCGGCCCACT	1	CTGACCTATGAATTGACAGCCAGT	4	AGCAGCATTGTACAGGGCTATGA	2741	AATCCTTTGTCCCTGGGTGAAA	2	TAATCCTTGCTATCTGGGTGCTTAGTGCT	1
TGACCCCTAGTCCAGTGCTTGT	1	TGGAAGACTAGTGATTTTGTTGTTG	1	TAGCACCATTTGAAATCGG	2	CTGAGATGAAGCACTGTAGCT	4	CATCGGGAATGTCGTGTCCGC	11
TATGGCTTTTTATTCCTAT	4	TCACAGTGGCTAAGTTCTGC	5	GGGGTGCTATCTGTGATTGAGGG	1	TTTGCTGAAGCCAGATGCCGTTCCTGAGAA	1	TTCACAGTGGCTAAGTTCTGCAC	4
CACAGTGGCTAAGTTCTG	1	AACATTCATTGCTGTCGGTG	14	TCACAAGTCAGGTTCTTGGG	2	TCCCTGAGACCCTAACTTGTGA	8378	TGGTTTACCGTCCCACATACAT	7
AGGGCCCCCCCTCAATCCTGT	1	TCCCTGAGGAGCCCTTT	1	TTTTTCATTATTGCTCCTGACC	5	CTGTAGAACCGAATTTGTG	1	GCCTGTCTACACTTGCTGTGC	94
CAACGACATCAAACCACCTGAT	1	TGGAGTGTGACAATGGTGTTT	1	AACATTCATTGCTGTCG	1	TATTGGGAACATTTTGCATA	1	TATTGGGAACATTTTGCATG	1
AGCTGCCAGTTGAAGAACTGTT	1	GTGAAATGTTTAGGACCACTAG	95	ATCTTTGCCGGTGACAGCA	4	TTGGATGTTGGCCTAGTTCTGTG	1	TGGAAGACTAGTGATTTTGTTGTT	7
AGAGGTATAGCGCATGGGAAGA	19	CTTCAGTGATGATTAGCTTCTGAC	1	TTCACAGTGGCTAAGTTCTG	1201	CATCATCGTCTCAAATGAGTCT	3	ATCCCACTCAGTCGTGCC	2
TAGCACCATCTGAAATCG	2	ATTGTGTCAATATGCGATGATGT	2	TGACCTATGAATTGACAGCCAGT	10	CTGTGCGTGTGACAGCGGCTG	27	AGCAGCATTGTACAGGGCTATGAA	227
GGGGTTCCTGGGGATGGGATT	1	TCCTGACTCCAGGTCCTGTGTGTT	1	AAAGACACCAAGCTGAGTAG	1	CACTGGTGCTCTGTGGGA	1	AAACATTCGCGGTGCACTTCTTTT	1
TTTGGCAATGGTAGAACT	1	CCAGGAGGCTGAGGTCCCTTA	1	CCAGTTTTCCCAGGAATCC	4	AGGGGTGCTATCTGTGATTGAG	9	ACATGGTGCACTTCTTT	1
ATGTAGGGATGGAAGCCATGAA	10	GCTTATCAGACTGATGTTG	28	CTCCTGACTCCAGGTCCTGTGT	160	TACAGTATAGATGATGTAC	5	CCGTCCTGAGGTTGTTGAG	17
AAGAGGTAGTAGGTTGCAT	1	TCAAGTAATCCAGGATAG	1	GAGGTAGTAGGTTGCATAGT	12	GTAGTGTTTCCTACTTTATGG	76	CTAGCAGCACGTAAATATTGGC	1
AAGAGCAATAACGAAAAATGTT	1	AAACCGTTACCATTACTGAGTTTAGTAAT	2	AAACATTCGCGGTGCACTTCTTT	17	ACTGCATTACGAGCACTTAAA	4	TTCAAGTAATCCAGGATAGGCTG	71
TTTTTATTCCTATGTGA	1	ATTGGGAACATTTTGCATAAAT	4	TCGTGTCTTGTGTTGCAGCCGGA	9	AACATTCATTGTTGTCGGTGG	35	AATGGCGCCACTAGGG	1
AGTTTTGTGTGCATGTGCATGT	17	TTGGACTGTGAGGTGACTCTTGGT	1	TCAAAGTGCTTACAGTGCAGGTAG	51	ATCCCTGAGTGTATGTG	1	TTGTGACAGATTGATAACTGAAA	1
CAAGCTCGTGTCTGTGGGTCCG	24	TGGCAGTGTCTTAGCTGGTTGTTGTG	1	CAGCAGCAATTCATGTTTTGGA	1450	TGGAGACGCGGCCCTGTTGGAGT	1	TACCACAGGGTAGAACCACGGACA	48
CTAGCACCATTTGAAATCA	1	CCATTTGAAATCAGTGTTT	2	GATTCCTGGAAATACTGTTCT	4	GGTAGTAGGTTGTATGGTT	6	ATCACATTGCCAGGG	4
CATTGCACTTGTCTCGGTCTG	206	TCCTGACTCCAGGTCCTGTG	3	GGCTCAGTTCAGCAGGAAC	1	AACACTGTCTGGTAACGATGTT	48	AACCGTGGCTTTCGATTGTTAC	3
CACAGGGTAGAACCACGGAC	1	AGTTGTGTGTGCATGTTCATGTCT	29	GATAGAAGTCAGTGCACTACAGAACTTTGT	1	CCCATCCCACTCAGTCGTGCCC	1	GGAGTATTGTTTCCACTGCCTGG	1
TGGCAGTGTATTGTTAGC	2	TAGTGTTTCCTACTTTATG	7	AGCAGCACGTAAATATTGGCG	1	CCTGTAGAACCGAATTTGT	1	GTACAGCAGGCACAGACAGGCAGT	9
CAAAGCACAGGGCCTGCAGAGAG	7	CAAAGTGCTCATAGTGCAGGTAGT	2	CGCATCCCCTAGGGCATTGG	2	ACTGCATTATCGAGCACTTAAAGT	24	TAGCACCATCTGAAATCGGT	127
TGAGGTAGTAGGTTGTATAGTTTTA	1	ATCACTAACTCCACTGCCATC	20	TGGTGTCCTCAAATTGAAAGCCAAGG	1	CCCAACCCTTGTACCAGTGC	1	TTTCCCTGAGGAGCCCTTTGAGCCTG	1
TAGCTTATCAGACTGATGTTGACTGTTGA	2	CTAATACTGTCTGGTAATGCCG	1	CAAAGTGCTGTTCGTGCAGGTAGTGT	17	TGAACAGTGCCTTTCTGTGTAG	1	TTCAGTCGGATGTTTGCAG	1
CCCTAAGGTGAATTTTTGGGAA	1	AGGTTGGGATTTGTCGCAATGC	3	ACCCTGTAGATCCGAATTTGTGT	28	TCCGTCTCAGTTACTTTAT	2	AAAAGCTGGGTTGAGAGGGCGAAA	107
GGTAGTTTCATGTTGTTGGG	1	TTAATGCTAATTGTGATAGGGGTTT	18	TACAGTACTGTGATAGCTGAAG	25	GTCAAGAGGCGCCTGGGAA	2	CTAATACTGCCGGGTAATGATGG	18
TATGTGCCTTTGGACTACATCGTG	6	ACTGCTGAGCTAGCACTTCCC	1	TATTGCTTAAGAATACGCGTAG	1	GTCAAGAGGCGCCTGGGAACT	3	TCGATCGGTCGGTCGGTCAGT	5
GCAGCACAGAAATATTGGCA	2	TGGTCCCCTTCAACCAGCTGT	1	AACATTCAACGCTGTCGGTGAGT	1427	TAGGTAGTTTCATGTTGTTGGG	819	AGTAGGTTGTATGGT	1
TCACATTGCCAGGGATTACCACG	2	TCCCTGTCCTCCAGGAGCTCACG	7	AACATTCAACGCTGTCGGTGAG	287	TAGCAGCACAGAAATATTGGCA	459	GCAGCACATCATGGTTTAC	9
TGAAACATACACGGGAAACCTC	3	AATCCTTTGTCCCTGGGTGAAAATGCT	9	CAGCATTGTACAGGGCTATC	1	CAGCATTGTACAGGGCTATG	4	CTGAGACCCTTTAACCTGTGA	3
CGAAGGAGCCTCGGTTGGCCCCGGA	2	TCACATTGCCAGGGATTTCCAACT	1	CGGATCCGTCTGAGCTTGGCT	1	CAGTGCACGTGTCTCCA	2	AGCGCCTCGGCGACAGAGCCGG	1
TTCTAGCACTTAGCAGGTTGTATTATCATT	1	TCAGTTATCACAGTGCTGATGCT	43	CAGGCCATACTGTGCTGCCTCA	24	TCAGGGCAGTGATGTTGCCCCTCCGAAGAT	1	GCAGTCCACGGGCATATA	1
GTAGGTTGCATAGTTT	1	TGTGCTGACGTTTGATTCTTGAGCAG	8	TGAGGTAGTAGGTTGTGTGGTTTCAGGGCA	7	CTTAGCAGGTTGTATTATCAT	5	ACTGCAGTGAGGGCACTTGT	3
GTCCAGTTTTCCCAGGAATCCCT	15921	AATCCTTGCTATCTGGGTGCTT	1	GGCTTCTTTACAGTGCTGCCTTGTTGC	1	CACCCGTAGAACCGACCTTG	19	GTCCAGTTTTCCCAGGAATCCC	1474
TAGTAGTTTGTACAGTT	1	AATCACATTGCCAGGGATTACC	5	ATATACATACACACACCTACA	1	TAGCATGGTGTTGTGGGAC	1	CACATTGCCAGGGATTAC	1
CAAAGCACAGGGCCTGCAGAGA	12	TCGAGGAGCTCACAGTC	1	TCCTGTACTGAGCTGCCCCGAG	95	CTTACAGTGCAGGTAG	1	ACATTCATTGCTGTCGGTGGGTTT	5
ATATTCTGCTGTGCAAATCCATGCAAAA	1	AGCACATCATGGTTTAC	2	TACCCTGTAGAACCGAATTTGTG	314	AAAGCTGGGTTGAGAGGGCG	2	TCCCTGTCCTCCAGGAGCTCACGT	2
CTGGTTTCACATGGTGGCTTAGATT	7	ATCACATTGCCAGGGA	6	TGAGGTAGTAGATTG	1	CATCGGGAATGTCGTGTCCGCC	11	TTAGCAGGTTGTATTATCATT	1
AGCAGCATTGTACAGGGCTAT	651	TGTAGTGTTTCCTACTTTATGGA	1174	TGAGGTAGTAGGTTGTATGGTTT	2740	GCACCATTTGAAATCGGTT	3	AATGACACGATCACTCCCGTT	4
TAGCAGCACAGAAATATTGGCATGGGGAAG	1	CCAACAATATCCTGGTGCTGA	2	GCTGGTGTTGTGAATCAGGCC	2	TGATATGTTTGATATATTAGGTT	206	TGTGCAAATCCATGCA	2
TGAGGTAGGAGGTTGTATAGT	309	ATACTGCCTGGTAATGAT	3	ATTCCTGGAAATACTGTTC	3	GTAGTAGTTTGTACAGTTT	1	AGAGGTATAGCGCATGGGAAG	21
GAGGGACTTTCAGGGGCAGCTGTGT	3	TAGCACCATCTGAAATCGGTT	1268	TGAGGTAGTAGGTTGTATGGTTTAGAGTT	2	CAAAACATGAAGCGCTGCAACA	30	TCTAGGTATGGTCCCAGGGATCC	1
TATGTGCCTTTGGACTACATCGTGAACGCA	1	AGCTCCTATATGATGCCTTTCT	1	AGGTAGTAGGTTGCATAGTT	5	TCACAGTGGCTAAGTTC	1	ACCGATTTCTCCTGGTGTTC	1
ATCACTAACTCCACTGCCAT	1	CAGTGCAATAGTATTGTCAAAGC	689	ATTATGCTGACAGCTGCCTCGGTGGGA	1	TAACAGTCTACAGCCATGGTCGC	4	TGTAAACATCCTACACTC	3
TGAGGTAGGAGGTTGTATAG	61	CATCGGGAATGTCGTGTCCGCCC	2	GGGTCCGGTGCGGAGAGCCGTTCGTCTTGG	2	CTGGTTTCACATGGTGGCTTAGATTT	6	TGTCACTCGGCTCGGCCCACTACC	1
GCGCCGAAGGAGCCTCGGTTGGCCC	1	AATCCTTGGAACCTAGGTGT	10	TGGCTCAGTTCAGCAGGAA	15	TTCAAGTAATTCAGGATAGGT	3219	GCATTGTACAGGGCTATGA	9
TAGTGTTTCCTACTTTATGGATG	2	ACTGTCTGGTAACGATGT	2	TGCTATGCCAACATATTGCCATCT	3	TATGTGTGTGTACATGTACAT	1	TGTAAACATCCTACACTCAGCTGT	4
GTTTTCCCAGGAATCCCT	30	CAAAGTGCTGTTCGTGC	3	CGAATGTTGCTCGGTGAACCCC	1	ACTGGACTTGGAGTCAGAAGGC	316	CACGGGTTAGGCTCTTGGGAG	1
TAAAGTGCTTATAGTGCAGGTAGTGT	18	CTAATACTGCCTGGTAATGATGA	2	TTTGGTCCCCTTCAACCAGC	6	GAGGTAGTAGGTTGTATAGTT	6	TTGTGCTTGATCTAACCATGTG	1261
CAGTGCAATGATGAAAGGGC	2	ACCCGTAGATCCGAACTTGTGC	2	AGAACGGCTACTTCACAACACCAGGGC	1	AGAGGCTGGCCGTGATGAAT	1	TTTGGTCCCCTTCAACCAGCTGT	75
TTTTGCGATGTGTTCCTAATATGTGCT	2	TGCAGTGCTGCATCTCTGG	3	AACAAACATGGTGCACTTCT	1	TTCCTGGAAATACTGTTCTT	1	ACACTGCCGGCCTCTGGG	3
TTTTGCAGTATGTTCCTGAAT	12	TTGAGGTAGTAGGTTGTATG	1	ACTCCATTTGTTTTGATGATGG	30	ACTGGACTTGGAGTCAGAAGGCC	1	TAGTAGGTTGTATGGTT	9
TAGGTTATCCGTGTTGCC	1	TCAAGTAATTCAGGATAGGT	1	CGCATCCCCTAGGGCATTGGTG	4	CCTGTCTACACTTGCTGTGC	1	GGGGTGCTATCTGTGATTGAG	1
TGTTTAGACTACCTGTTC	1	GAATGTTGCTCGGTGAACCCCTTTT	2	TGTAGATCCGAATTTGTGT	1	AACTGGCCTACAAAGTCCCA	36	AGGCAGTGTAATTAGCTGATTGT	36
ATTGTACAGGGCTATGA	2	TGAGGTAGTAGGTTGTATGGTTTAGAGT	2	CAACAAATCACAGTCTGCCAT	12	TGAGGTAGTAGGTTGTATGGTTTAGA	16	TCCGGTGCGGAGAGCCGTTCGTCTTGGGAA	2
ATCACATTGCCAGGGAT	8	AGAGGCTGGCCGTGATGAATTCG	3	GGGAGCCAGGAAGTATTGATGTTT	6	GCCGAAGGAGCCTCGGTTGGCCCCGGAT	1	ACGACTAGGAGGCCTCT	1
TGTAGGGATGGAAGCCATGAA	4	AGGAAGCCCTGGAGGGGCTGGAG	1	AGCTCCTATATGATGCCTTTCTT	1	ATATAATACAACCTGCTAAGTG	237	TTCAACGCTGTCGGTGAGT	1
TGAGGTAGTAGGTTGTGTGGTTTCAG	5	TGACCGATTTCTCCTGGTGT	2	TTGAGGTAGTAGGTTGTATGGT	28	ACCCTGTAGATCCGAATTTGTG	98	TCACATTGCCAGGGATTA	1

TCACATTGCCAGGGATTT	3
ATCCCTGAGTGTATGTGGTGAACC	1
TTTCCTATGCATATACTTCT	6
TAGGTTATCCGTGTTGCCTTCGCT	1
GAGGTAGTAGATTGTATAGTT	1
CCGTCCTGAGGTTGTTGAGCT	74
ATTCCTGGAAATACTGTTCT	34
TGAGGTTGGTGTACTGTGTGT	109
AGTTGTGTGTGCATGTTCATGTCTAT	5
TCCCTGAGACCCTTTAACCT	1
TGAAATGTTTAGGACCACTAG	84
TGAGGTAGTAGGTTGTGT	39
ATCTCGCTGGGGCCTCC	2
AAAAAAACAGGTCAAGAGGCGCCTGGGAA	1
AATCACTAACTCCACTGCCATCA	17
TAATGCCCCTAAAAATCCTTAT	847
TACAGCAGGCACAGACAG	5
CCTAAGGTGAATTTTTTGGGAA	3
ATACTGCCGGGTAATGATGG	5
TACCCTGTAGAACCGAATTTG	53
GAGGTAGTAGGTTGCATAGTTT	3
GCTTATAGTGCAGGTAG	10
ATCCCCGACTGGAAG	1
CTAATACTGCCTGGTAATGAT	1
TCTTCCTGCCAGAGCATACCTGGGCCT	1
AGCATTGTACAGGGCTAT	1
AGTTTTGCAGGTTTGCATCCAG	12
AAACCCGTAGATCCGATCTTGTG	2
CTATACAATCTACTGTCTTTC	10
GTAAGTGCCTGCATGTATATGCG	6
ATAATACAACCTGCTAAGTGT	1
TGAGGTTGGTGTACTGTGTG	6
ACTCTTTCCCTGTTGCACTAC	4
GCTGGTTTCATATGGTGGTTTA	1
AGAGGTAGTAGGTTGCAT	16
CGAATGTTGCTCGGTGAACCCCT	4
TTCACAGTGGCTAAGTTCC	101
CCAGTTTTCCCAGGAATCCC	2
TATAATACAACCTGCTAAGTG	1
TATTCAGATTGGTGCCTGTCAT	1
TAGCTTATCAGACTGATGTTGACTGTT	6
ATTATTACTTTTGGTACGCGC	1
CTCGTACCGTGAGTAATAATG	4
ATACTGCCTGGTAATGATGA	9
CAGTGCAATAGTATTGTCAAA	42
GCCCTAAGGTGAATTTTTTGGG	2
ATAATACAACCTGCTAAGTG	1
AGGTGCATCTAGTGCAGATAGT	2
GCTATGCCAACATATTGCCATC	7
TGTGCAAATCCATGCAAAACTGACT	1
CATTGCACTTGTCTCGGTCTGACA	1
AATCCTTTGTCCCTGGGTGAAAAT	7
GTATGTGCCTTTGGACTACA	3
ACTGCATTACGAGCACTTA	3
GATGTTGAAAAGGTGCAGGCCATACTGTG	1
TGCCTGTCTACACTTGCTGTGCA	12
AATCACTAACCACACAGCCAGGTAAA	2
TGCTCGCTTCAGAGGGTTGGGTGGAGGCT	1
ACCACAGGGTAGAACCACGGACAG	1
CTCAGCTGATCTGTGGCT	1
AACCCGTAGATCCGAACTTGTGCTGA	1
GCTGACCCCTAGTCCAGTGC	3
ACCATCGACCGTTGAGTGGAC	1
ACTGGCCTACAAAGTCCCAGT	4
TGAGAACTGAATTCCATGGGT	5
CAGTGCACTGGCTCAGTTCAGCAGGAA	1
TTTTGCGATGTGTTCCTA	10
TGTGACAGATTGATAACTGAA	10
AATGTTGCTCGGTGAACCCCTTT	1
ACTGCCCTAAGTGCTCCTTCTG	1
TGTAAACATCCTACACTCTCAGCTGTGA	1
TCCGAGCCTGGGTCTCCCT	3
CTGAGGTAGTAGTTTGTACAGTTT	2
TGAATTTGCTGAAGCCAGATGCCGTTC	1
AGGTAGTAGGTTGCATAGTTT	1
ATAGCTTATCAGACTGATGTTGA	147
CTGTGCGTGTGACAGCGGC	5
AAAGTGCTTACAGTGCAGGTAGT	2
GAGGTAGTAGGTTGTGTGGTT	2
AAGGAGCTTACAATCTAGCTGGG	617
ACAGGTGAGGTTCTTGGGAG	2
GAATGTTGCTCGGTGAACCCCT	10
AAACATGAAGCGCTGCAAC	11
TTCCTATGCATATACTTCTTTG	23
GGGGTGCTATCTGTGATTGAGGGA	3
CAGCTACATCTGGCTACTGGGT	1
TTTTGCGATGTGTTCCTAA	14
AATCCTTGGAACCTAGGTGTGAATG	37
CGAATCATTATTTGCTGCTCT	10
CAATTTAGTGTGTGTGATATT	7
TGAGGGTCTATGATACCACCCGGTACAGGA	13
TGCCTACTGAGCTGATATCAGT	2
TGGCAGTGTCTTAGCTGGT	7
TGAATGGCGCCACTAGGGTTGT	3
CAAAGTGCTCATAGTGCAGGTA	12
TTTGCGATGTGTTCCTAATA	1
CGTGTCTTGTGTTGCA	1
CAGCAGCAATTCATGTTTT	51
GTAAGGTGTTGAGAGGAG	2
TGATTGTCCAAACGCAATTC	1
ACCGTGGCTTTCGATTGTTAC	9
AGCTGGTGTTGTGAATCAGGCCGTTGC	1
ATTCAAGTAATCCAGGATAGGCT	8
AATAATACATGGTTGATCTTT	5
AGCACATCATGGTTTACA	1
TATTGGGAACATTTTGCATAA	3
ACAGCAGGCACAGACAGGCAGT	761
ACTGTCTGGTAAAGATGGC	2
TAGTTGTGTGTGCATGTTCATGT	2
TACCACAGGGTAGAACCA	1
AGTTGTGTGTGCATGTTCA	1
CAAGAAGCCATCGGGAATGTCGTGTCCGC	1
TGTAAACATCCTTGACTGGAAGCT	112
ACTGGACTTGGAGTCAGAAGG	31
CTTGGTACATCTTTGAGTGAG	15
TAGCACCATTTGAAATCAGTGTTC	9
AGCATTGTACAGGGCTATG	3
AACATACACGGGAAACCTCT	1
TTTGTCATCCTATGCCTG	1
CGGGGTCCGGTGCGGAGA	1
CAGCAGCAATTCATGTT	1
AGTAGGTAGTTTCATGTTGTTG	1
TGAGGTAGGAGGTTGTATA	5
CTGTACAGCCTCCTAGCTTTC	2
ATTGGGAACATTTTGCATGCA	4
TGAGGTAGGAGGTTGTATAGTTGAGG	1
TGGCAGTGTCTTAGCTGGTT	27
TGGCAGTGTCTTAGCTGGTTGTTGTGAGTA	1
TACCGCACTGTGGGTACTTGCT	1
GCTTTCAGTCGGATGTTTGCAG	1
TATTGCACATTACTAAGTT	1
ACATTCATTGCTGTCGGTGGGTT	4
ATTTGCTGAAGCCAGATGCCGTTCCTGAGA	12
TGATTGTCCAAACGCAATTCTCG	2
TTCACAGTGGCTAAGTTCT	485
TGTAACAGCAACTCCATGTGGAA	2
AGGCAGTGTAGTTAGCTGATT	2
CCTGGTACAGGCCTGGGGGATA	1
AAGTGCTTATAGTGCAGGT	1
TGATATGTTTGATATATTAGGTTGT	30
ACCATCGACCGTTGAGTGGACC	1
TTTTGTTCGTTCGGCTCGCGT	1
CTCACAGCTCTGGTCCTTGGA	2
AGCGCCGAAGGAGCCTCGGTTGGCCC	1
TAGCAGCGGGAACAGTACT	9
GTCTGCACATTGGTT	1
ATGTATGTGTGCATGTGCATG	1
ACCGTGAGTAATAATGC	1
ACTGATTTCTTTTGGTGTT	7
AGGTAGTAGATTGTATAGTT	1
ACAAGTCAGGTTCTTGGGACCT	82
ATCGTAGAGGAAAATCCACG	1
TGCTATGCCAACATATTGCC	90
TGAGATGAAGCACTGTAGCT	130
CTGAGGTAGTAGTTTGTACAGTT	19
CATCTTACTGGGCAGCATTGGA	22
CAACGGAATCCCAAAAGCAGCTGTT	25
TCGACAGCACGACACTGCCTTCA	2
AGTGCAATGTTAAAAGGGCAT	10
GGCTTCTTTACAGTGCTGCCTTGTTGCAT	2
TGCACTGAGATGGGAGTGGTGT	27
CAGTGGTTTTACCCTATGGTA	7
AGTTTTCCCAGGAATCCC	1
TTGAAAGGCTGTTTCTTGGTCA	1
GAGGTAGGAGGTTGTATAGTT	1
TCGTGTCTTGTGTTGCAGCCGG	14
AAAGTGCTTATAGTGCAGGTA	1
TGAGATGAAGCACTGTA	1
GGTAGTAGTTTGTGCTGTT	3
AGGGGTGCTATCTGTGATTGAGGGAC	114
TCCGTCTCAGTTACTTTATAGC	56
ATGTATGTGTGCATGAACATGT	3
TCGGGAATGTCGTGTCCGCCC	1
TACGGTGAGCCTGTCATTATT	1
TGTAAACATCCTTGACTGGAAG	40
GTTGAGGTAGTAGGTTGTATGGTT	4
AATCACATTGCCAGGGATTTC	8
TTAATGCTAATTGTGATAGGGGTT	8
AAAGTGCATGCGCTTTGGGAC	3
GCTGGTTTCATATGGTGGTT	1
AACATTCATTGCTGTCGGTGGGTTTGAA	1
TATGTGCCTTTGGACTACATCGTGA	3
CACAAATTCGTATCTAGGGGAA	1
CACAGACAGGCAGTCACATGA	1
TACAGCAGGCACAGACAGG	11
GACCGATTTCTCCTGGTGTTCA	1
CAAGTCAGGTTCTTGGGACCT	5
GTAGTAGGTTGTATGGT	1
CAAGCTTGTGTCTATAGGTATGT	5
ACAAGTCAGGTTCTTGGGACC	38
GCATTGTACAGGGCTATG	4
CTGACCTATGAATTGACAGCC	115
GTAGTCTGCACATTGGTTA	5
AGGTAGTAGGTTGTATAGT	2
CCCTAAGGTGAATTTTTTGGG	1
TGAGGTAGTAGGTTGTGTGGTTTC	28
CCTGTTCTCCATTACTTGGCTC	2
TGCCACAAAGATGCCATATCAACTCCTCT	2
CAACCCTAGGAGGGGGTGCCATTCA	6
TAGTGTTTCCTACTTTATGG	6
ATCACATTGCCAGGGATT	160
TAGCAGCACATCATGGT	1
TGAGGTAGTAGGTTGTATAG	526
GGCACAGACAGGCAGT	1
TTAGGGCAGAGATTTTGCCCACAAGGAGTT	2
ACTATATATCAAGCATATTCC	1
CACTAGATTGTGAGCTGCTGGAG	1
CTGCCCGGGGCCCTCAGGCCAGTGCCCAGG	1
GAAATGTTTAGGACCACTAG	1
TAATACTGCCGGGTAATGATGGAG	1
TTCAACAAAATCACTGATGCTGGAG	1
AACCGTGGCTTTCGATTGTT	1
TAGGCAGTGTAATTAGCTGATTGTA	2
TAGCACCATTTGAAATCAGTGTTT	272
TTCAAGTAATTCAGGATA	3
GGGGTGCTATCTGTGATTGAGGGAC	12
TGTAAACATCCTACACTCAGCTGTCATA	2
TGAGGTAGTAGGTTGTATGGTTTTG	3
AGAGGTAGTAGGTTGC	4
CATCGGGAATGTCGTGTCCG	10
AAACCGTTACCATTACTGAGTT	23
ATTCAACCTGTCGGTGAGTTT	1
TCTTTGCCGGTGACAGCA	6
TCCGGTGCGGAGAGCCGTTCGTCTTGGGA	2
TGCGGGGCTAGGGCTAACA	2
ATATAATACAACCTGCTAA	1
TCCAGTTTTCCCAGGAATCCC	5
CCCAGTGTTTAGACTACC	3
CTGGACTTGGAGTCAGAAGGC	49
TCCAGCATCAGTGATTTTGTTGA	10
AGGTAGTAGGTTGTATAGTT	4
TACCACAGGGTAGAACCACGGAC	122
TGTAAACATCCTTGACTG	1
ATCGGGAATGTCGTGTCCGCCC	2
TACAGCAGGCACAGACAGGCAG	11
TCACAAGTCAGGTTCTTGGGACC	5
TGAGGGGCAGAGAGCGAGACTTT	3
CAGCAGGCACAGACAGGCAG	1
CTGCCTGGTAATGATG	1
TCAGTGCATGACAGAACTTG	50
CATCTGGCTACTGGGTCTCT	1
CCTCCCACACCCAAGGCTTGCAG	1
ATGACACGATCACTCCCGTTGA	1
AAGCCCTTACCCCAAAAAG	1
CCTGTGGGCCACCTAGTCACC	1
GGCTTCTTTACAGTGCTGCC	1
CAGCAGGCACAGACAGGCAGT	7
TCCCTGAGGAGCCCTTTGAGCCTGGA	8
TGTAAACATCCTACACTCTCAGC	5354
CATTTGAAATCAGTG	1
CTGCCGGGTAATGATGGA	2
TACCCTGTAGATCCGAATTT	3
TGCTGACCCCTAGTCCAGTGC	64
AAGCTGGGTTGAGAGGGCGAAAA	1
TAGCTTATCAGACTGATGTT	594
TGCTATGCCAACATATTGCCAT	44
GTTTTCCCAGGAATCCCTT	2
TCCGGTGCGGAGAGCCGTTCGT	1
GTAGTAGGTTGCATAGTT	4
ACTAGATTGTGAGCTGCTGGAG	2
TGTTGTCGGTGGGTT	1
TACCCTGTAGATCCGAATTTGTGT	20
TTCACAGTGGCTAAGT	3
TGTGCAAATCTATGCAAAAC	3
TATGTAACACGGTCCACTAA	5
GTCCAGTTTTCCCAGGAATC	155
CTCGGGGATCATCATGTCA	4
TAAAGTGCTGACAGTGCAGAT	1031
AACTGGCCTACAAAGTCCCAG	318
AAGCCCTTACCCCAAAAAGCA	3
ATTGCACTTGTCCCGGCCTGTT	2
ATCACTAACTCCACTGCCA	2
TTAGCCGCTGAAATAGATGGAT	1
TCACATTGCCAGGGATTACCAC	26
TCTCACACAGAAATCGCACCCG	6
GTAAACATCCTCGACTGGAAGC	5
CTGTCTGGTAAAGATGGC	1
TTCCTATGCATATACTTCTT	645
ATCGTACAGGGTCATCCACTTT	7
CTGTCTGGTAACGAT	2
AATACTGCCTGGTAATGATGAC	13
TCGGTTGGCCCCGGATAGCCGGGTCCCCGT	4
TGAGGTAGTAGTTTGTACAGTTT	47
ATTGCCAGGGATTTCCAA	1
GTGTTTCCTACTTTATGGAT	1
TATGTCACTCGGCTCGGCCCACTACC	1
TGAAACATACACGGGAAACCTCTTT	1
AACATTCAACCTGTCGGTGAGTT	85
CGAAGGAGCCTCGGTTGGCCCCGGATAGC	2
AATCACTAACCACACAGCCAGG	25
TCCGAGCCTGGGTCTCCCTCT	27
TGAGGTAGGAGGTTGTATAGTTG	109
CAACAAATCACAGTCTGCCATA	55
ACATTCAACGCTGTCGGTGAGTTTG	2
GTTGTGAATCAGGCCG	1
GCAGCAATTCATGTTTTGG	1
TATGTGCCTTTGGACTACATCG	190
TAGGTTATCCGTGTTGCCTTCG	161
GGGGTGCTATCTGTGATTGAGGGACA	17
CCCATAAAGTAGAAAGCACTAC	22
TACAGTACTGTGATAA	1
TGACCTATGAATTGACAGCCAG	12
TGTGACTGGTTGACCAGA	1
CTGTACAGGCCACTGCCTTGCC	4
TTATCAGAATCTCCAGGGGTACTT	1
CTTTCAGTCAGATGTTTGCTGC	51
TAAGGTGCATCTAGTGCAGATAGTG	2
TTGAGGTAGTAGGTTGTATAGTT	8
TCACAGTGAACCGGTCTCTTT	26
TGTGCAAATCTATGCAAAACT	17
ATCACATTGCCAGGGATTACC	678
TACAGTACTGTGATAACTGAA	10
AACTGGCCTACAAAGTCCCAGT	1185
TAATACTGCCGGGTAATGATGG	2668
GGCTTCTTTACAGTGCTGCCTTGTTG	1
TTGCTGAAGCCAGATGCCGTTCCTGAGA	1
CTGACTTCGGCCCCCATGTCAGCAGATGC	1
CGAAGGAGCCTCGGTTGGCCCCGGAT	1
TTGTGCTTGATCTAACCATGTGCT	9
CAGCACGTAAATATTGGCG	5
CTATATATCAAGCATATTCCTA	1
AAACATTCGCGGTGCACTTCTT	18
AGTTTTGCAGGTTTGCATCCAGC	27
AACATTCATTGTTGTCGGTGGG	4
ACTAGATTGTGAGCTG	1
CCGCACTGTGGGTACTTGCT	7
AGTGCAATAGTATTGTCAAAGC	2
TACTCAGTAAGGCATTGTTCTT	7
CCTTCAACCAGCTGT	1
AGGTGCAGTGCTGCATCTCTGGT	2
ATGCACCCGGGCAAGGATTTG	1
CACATAATGGTTTGTG	2
CTGTGCGTGTGACAGCGGCTGATC	2
TGAGGTAGTAGGTTGTGTGGTTT	2194
TTGCCAGGGATTTCCA	1
TTAATGCTAATTGTGATAG	1
CCCAGTGTTCAGACTACC	6
TGTAAACATCCCCGACTGG	6
CTAGCACCATTTGAAATCGGT	1
AAGGAGCTTACAATCTAGCTG	221
TCCCCCAGGTGTGATTCTGATTTGT	35
CCCCTGGGCCTATCCTAGA	2
AGTTTTGCAGGTTTGCATCC	3
AGTTTTGTGTGCATGTGCATGTGT	32
CTCGAGGAGCTCACAGTCTAGT	6
TTATGTGTGTGTACATGTAC	1
CAATCAGCAAGTATACTGCCCTA	1
CTGGGAGAAGGCTGTTTACT	10
GTGAAATGTTTAGGACCA	1
CTAATACTGCCTGGTAATGATG	1
ACCCGTAGATCCGAACTTGTG	2
ACTGTCTGGTAAAGATG	4
AAACATGAAGCGCTGCAACA	78
AAGTGCTTATAGTGCAGGTAGT	4
TGTGCAAATCCATGCAAAACTGACTGTGG	1
ATCCAGGATAGGCTG	1
CCACTGCCCCAGGTGCTGCTGG	52
TTGTGCAAATCTATGCAAAAC	1
CTTTCAGTCGGATGTTTACAGC	498
CAAAGTGCTTACAGTGCAGGTAGT	1816
TGAGAACTGAATTCCATAGG	3
GTTTTGTTTGGGTTTGTT	1
ACTGTCTGGTAACGATG	2
TTCACAGTGGCTAAGTTCTGCA	116
TAGTAGGTTGTATAGT	9
CTTTGTCATCCTATGCCT	1
TACAGTAGTCTGCACATTGG	9
AGTTGTTCGTGGTGGATTCGCTTTA	2
GAGGTAGTAGGTTGCATAGTTTT	1
TTTCCTATGCATATACTTCTT	34
ACTGATTTCTTTTGGTGTTCAGAGT	45
GAGACCCTAACTTGTGA	4
AGGCAGTGTAGTTAGCTGATTGCTAA	1
CTGGCTCAGTTCAGCAGGAAC	1
AACCTGGCATACAATGTAGATTTCTGT	1
TAGCAGCACATAATGGTTTGTGGAT	2
AAGACGCCGCTTTGATCACTGTCTCCAGC	1
GCAAAGCACAGGGCCTGCAGAG	2
ATGTAGGGATGGAAGCCATGA	1
CACGCGGGAACCGAGTCCACC	5
ACCAATATTATTGTGCTGCTTTA	2
TAGGTAGTTTCCTGTTGTTGGGAT	5
CCTCATATCCATTCAGGAGT	1
CTATACAACCTACTGCCTTCC	12
GTGAGTAATAATGCG	1
ATGGCGCCACTAGGGTTGTG	6
TAGGCAGTGTAATTAGCTGATTGT	7
TATGGCTTTTCATTCCTATGT	1
GCAGTGCAATAGTATTGTCAAAG	1
TAAACATCCTCGACTGGAA	1
CTGGTGTTGTGAATCAGG	1
CCCAGTGTTCAGACTACCTG	23
TTGAGGTAGTAGGTTGTATAG	1
TCCCCCAGGTGTGATTCTGATT	10
TACTGCCTGGTAATGATG	1
TGAGGTAGGAGGTTGTATAGTTGAGGA	2
ACACTGTCTGGTAACGATGTT	8
TGGCGCCACTAGGGTTGT	1
TCAAAGTGCTTACAGTGCAGGTAGT	118
TAAGGCTCCTTCCTGTGCTTGC	2
AGCACGTAAATATTGGCG	1
ACATTCAACGCTGTCGGTGAG	1
ACAGGGTAGAACCACGGAC	1
GGTAGTTTCCTGTTGTTGG	1
CAACCCTAGGAGGGGGTGCCAT	2
TAATACTGCCGGGTAAT	1
AACCCGTAGATCCGAACTTGTG	243
TTTTGCGATGTGTTCCTAATATGTGC	1
CAGCAGCACACTGTGGTTTGT	202
GCTGGTTTCATATGGTGGTTTAGA	5
CAGGACAATGCCGTTGTA	1
ACAGTACTGTGATAACTGAA	1
TCCAGCATCAGTGATTTTGTT	25
AGCTGGTGTTGTGAATCAGG	251
TGAGGGGCAGAGAGCGAGAC	2
AGTATTACATGGCCAATCTCC	1
TATGTAACACGGTCCACTAACC	3
ACAGGGTAGAACCACGG	1
TTAGGGTCACACCCACCACTGGGAGAT	6
TGAGGGACTTTTGGGGGCAGATGTG	1
TCCATCTTCCAGTGCAGTGT	1
CAATGGCTGAGGTGAGGCACCTCCCGGGC	1
CTAGACTGAGGCTCCTTGAGG	46
CCAAGGTTCAAAACATGAAGCGCTGCAACA	1
GTCCAGTTTTCCCAGGAATCCCTT	571
CAAGCTTGTGTCTATAGGTATG	10
ATTTCTGGTATGGTTGCTGC	1
CCATCCCACTCAGTCGTGC	1
CTCCATTTGTTTTGATGATGG	1
CTATACGACCTGCTGCCTTT	22
ATCTTACCGGACAGTGCTGGA	1
CCCTGTAGATCCGAATTTGTG	5
AGCTTATCAGACTGATGTTGACT	1
TGAGGTAGTAGGTTGTATAGTTTTAGGGT	9
CCCTAAGGTGAATTTTTTGGGA	6
CAGTGCAATAGTATTGTCAAAGCAT	1521
TAATACTGCCTGGTAATGATGA	5646
TTTGTTCGTTCGGCTCGCGT	3
ATACCCTGTAGATCCGAATTTGT	1
GTTGAGGTAGTAGGTTGTATAGT	1
TCCTTCATTCCACCGGAGT	2
TAGCACCATTTGAAATCAGTG	341
TCAGTGCACTACAGAACTTTGT	639
TTCCTATGCATATACT	1
CTCACAGCTCTGGTCCTTGGAGC	1
TCAGTGCATGACAGA	1
ACGCTGTCGGTGAGT	2
AGGCAAGATGCTGGCATAGCTG	31
AGTGCAATGATGAAAGGGCAT	1
ATCGGGAATGTCGTGTCCGC	4
TCCATCTTCCAGTGCAGTGTTG	47
TAGCTTATCAGACTGATGTTGAC	32747
CAATTCATGTTTTGG	1
AATCCTTGGAACCTAGGTG	2
AGCCTCGGTTGGCCCCGGATAGCCGGGTC	9
CACCATCTGAAATCGGTT	3
CAGCGCCGAAGGAGCCTCGGTTGGCC	2
TCAGTGCACTACAGAACTTT	4
AGCACCATTTGAAATCAGTGTT	19
AACGCTGTCGGTGAGT	1
CGGTCAGCAGCCCAGCGCCA	1
GCACAGAAATATTGGCA	1
TCCTTGGAACCTAGGTGTGAATGC	1
AGAGGTAGTAGGTTGCATAG	281
CATTATTACTTTTGGTAC	3
CTGAGATGAAGCACTGTAGC	6
GTCAAGAGGCGCCTGGGA	1
CAGCTACATCTGGCTACTGGG	1
TTGAGGTAGTAGGTTGTATGGTTT	2
CAGTTTTCCCAGGAATCCC	7
TGTCAGTTTGTCAAATACCC	1
CTCAGCCCATCCCACTCAGTCGTGCCC	2
AGCTACATCTGGCTACTGGGTCTCT	12
AGCGCCTCGGCGACAGAGCC	2
TATACGACCTGCTGCCTTT	7
CTTTCAGTCGGATGTTTGCAG	104
TGAGGTAGTAGATTGTATAGTTGTGGGG	1
AAGCTGCCAGTTGAAGAA	16
TCCCTGAGACCCTTTAACCTG	29
AGCGAGGTTGCCCTTTGTAT	1
AAAAGCTGGGTTGAGAGGGC	1
AAGGAGCTTACAATCTA	1
TAGTGTTTCCTACTTTATGGAT	20
CCCAGTGTTCAGACTACCTGTTCAGGAC	2
TTTTGCGATGTGTTCCTAATA	106
TGGCAGTGTCTTAGCT	4
GCTGACCCCTAGTCCAGTG	1
AGTAATCCAGGATAGG	1
CAGCGCCGAAGGAGCCTCGGTTGGCCCCG	1
CTCGGTTGGCCCCGGATAGCCGGGTCCCC	1

Temporal reciprocity of microRNAs and their targets during the maternal to zygotic transition in Drosophila

Natascha Bushati[1], Alexander Stark[2], Julius Brennecke[3] and Stephen M. Cohen [1,4]

[1]European Molecular Biology Laboratory, Developmental Biology Programme, Meyerhofstr. 1, 69117 Heidelberg, Germany

[2]Broad Institute, MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

[3]Cold Spring Harbor Laboratory, Watson School of Biological Sciences and Howard Hughes Medical Institute, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

[4]Temasek Life Sciences Laboratory and Department of Biological Sciences, 1 Research Link, The National University of Singapore, Singapore 117604

Summary

During oogenesis, female animals load their eggs with messenger RNAs (mRNAs) that will be translated to produce new proteins in the developing embryo. Some of these maternally provided mRNAs are stable and continue to contribute to development long after the onset of transcription of the embryonic (zygotic) genome. However, a subset of maternal mRNAs are degraded during the transition from purely maternal to mixed maternal-zygotic gene expression. In Drosophila, two independent RNA degradation pathways are used to promote turnover of maternal transcripts during the maternal to zygotic transition [1]. The first is driven by maternally encoded factors, including SMAUG [2], whereas the second is activated about 2 hours after fertilization, coinciding with the onset of zygotic transcription. Here we report that a cluster of zygotically expressed microRNAs (miRNAs) targets maternal mRNAs for turnover, as part of the zygotic degradation pathway. miRNAs are small non-coding RNAs that silence gene expression by repressing translation of their target mRNAs and by promoting mRNA turnover. Intriguingly, use of miRNAs to promote mRNA turnover during the maternal to zygotic transition appears to be a conserved phenomenon, as a comparable role was reported for miR-430 in zebrafish [3]. The finding that unrelated miRNAs regulate the maternal to zygotic transition in different animals suggests convergent evolution.

Results and Discussion

A Mutant deleting the miR-309 microRNA cluster is viable and fertile

The Drosophila miR-309 cluster contains 8 miRNA genes, which encode 6 different miRNAs (Fig. 1A). Nucleotides 2 to 8 at the miRNA 5' end comprise the 'seed' region, which serves as the primary determinant of target specificity [4, 5]. The cluster encodes miRNAs with 5 distinct seed sequences, and so has the potential to regulate a broad spectrum of target mRNAs (reviewed in [6]).

Using homologous recombination [7], we generated a mutant in which the 1.1kb comprising the miR-309 cluster was deleted and replaced by GFP.

Northern blot analysis was used to verify that the first and last miRNAs in the cluster, miR-309 and miR-6, were not produced in the mutant (Fig. 1A). Homozygous mutant animals completed embryogenesis with no apparent defects in patterning (not shown and Supplemental Fig 1F), but ~20% died as larvae at different larval stages (Fig. 1B, C). Some individuals stopped growing at the size of L2 larva and arrested at this developmental stage for a few days before dying. Approximately 80% of mutants survived to adulthood and were viable and fertile. Introducing a transgene containing a 2.6 kb fragment of genomic DNA spanning the miRNA cluster (Fig. 1A) restored survival of the mutants to normal levels (Fig. 1B). The mutant animals showed a developmental delay during larval stages. This delay was suppressed in simultaneously collected and staged mutant larvae carrying the rescue transgene (Fig. 1C). The phenotypes that result from complete deletion of the three miR-6 miRNA genes (together with the rest of the cluster mRNAs) contrast with the severe embryonic defects that were reported using anti-sense 2'-O-methyl antisense oligonucleotide injection to deplete miR-6 or miR-286 [8].

RNA samples from precisely staged embryos [9] were used to examine the expression of the miR-309 cluster during early embryogenesis. We compared the levels of mature miR-6 and miR-309 in these samples by quantitative real-time PCR (qPCR). Samples were normalized to two reference miRNAs, miR-310 and miR-184, which we found to be expressed at constant levels when normalized to total RNA. miR-6 and miR-309 were expressed at barely detectable levels in RNA collected from embryos during a 30 minute period before the onset of zygotic transcription (Fig. 2A, time-point T0; [9]). The miRNAs were then strongly induced coincident with the onset of zygotic transcription (Fig. 2A, T1). In situ hybridization analysis at this stage, showed expression of the miR-309 cluster primary transcript throughout the embryo, except in pole cells ([10] [11], Fig. 2B). This transcript was not detectable in miR-309 cluster mutant embryos.

Although the mature miRNA products persist for some time, the expression of the primary transcript shows a dynamic spatial pattern by in situ hybridization

(Supplemental Fig. S1A-E; see also [11]). At midpoint of cellularization, expression of the cluster is turned off at the posterior pole and in a stripe in the anterior region of the embryo. During gastrulation expression is lost ventrally and laterally, resulting in transient stripes in the dorsal ectoderm. By the onset of germ-band elongation, the primary transcript was essentially undetectable, but in northern blots the mature miRNAs are detectable until larval stages [8].

The miR-309 cluster is predicted to target many mRNAs, including those of several genes implicated in embryo patterning. However, immunolabelling to detect these proteins did not reveal alterations in their expression levels or patterns in the miR-309 cluster mutant. For example, Supplemental Figure S1F compares the expression of the predicted miR-3/miR-309 target Ftz with Even Skipped (which is not a predicted target). There was no striking difference between mutant and control embryos, consistent with the observation that miR-309 cluster mutant embryos did not show discernable embryonic patterning defects. The significance of the dynamics of spatial expression of the cluster miRNAs and the implied potential to regulate genes involved in embryonic patterning remains unclear.

miRNA-mediated downregulation of maternal mRNAs

Given that the early onset of cluster miRNA expression does not appear to play a role in regulating zygotic mRNAs involved in patterning, we turned our attention to their potential to regulate the maternal to zygotic transition. We compared expression of the miR-309 cluster to a high-resolution temporal gene expression profile of early embryonic development [9]. The miRNA expression analysis was performed on the RNA samples used by Pilot et al ([9]; RNA samples kindly provided by Thomas Lecuit), in order to ensure comparability in the staging of the samples. mRNAs with a temporal expression profile most similar to that of the miR-309 cluster contained significantly fewer 7mers complementary to miR-309 cluster miRNAs in their 3'UTRs than would be expected by chance (Fig 3A correlation coefficient bin 0.9 to 1 at far right; Table S1, $p<0.01$). This suggests that these mRNAs have been under selection to reduce their regulation by the cluster miRNAs with

which they are coexpressed. Reciprocally, 7-mer seed matches complementary to cluster miRNAs were enriched in the 3'UTRs of maternal transcripts that were strongly downregulated as miRNA expression increases (Fig 3A, correlation coefficient bin –1 to –0.9; Table S1). The same trends hold true for 6-mer seed matches to cluster miRNAs. For the 6mer set, the correlation data are more significant because of overall larger numbers of miRNA targets in each bin (Table S2).

To investigate whether early zygotic miR-309 cluster miRNA expression might contribute to this down-regulation, we performed microarray analyses of control and mutant embryos at 0-1h and 2-3h of embryonic development. During the first hour, miR-309 cluster miRNAs are expressed at barely detectable levels, whereas they are strongly induced during the 2-3 hour interval (Fig. 2A). We compared messenger RNA levels in control and miRNA mutant embryos. Messenger RNAs whose expression was upregulated in the absence of the cluster miRNAs were examined with reference to two sets of maternal mRNAs that had previously been classified as being moderately or strongly down-regulated during the maternal-zygotic transition [12]. 42 of the 291 mRNAs (14%) that normally decrease by >3 fold between 2-3h of embryonic development were upregulated by over 1.5-fold in mutant embryos at this stage. This represents a 5-fold enrichment among the upregulated mRNAs, and is statistically significant (P<e-16; Fig. 3B, Table S3). The effect of removing the miRNAs was stronger in the group of the 32 maternal transcripts annotated to decrease by >10-fold at this stage. 35% of these were upregulated in the mutant (12/32), a 12.5 fold enrichment (P<1e-10; Fig. 3B, Table S3).

The degree of enrichment of these annotated gene sets among our upregulated transcripts is likely to underestimate the true degree of correlation, because only 30% of the genome was included in the original classification of moderately or strongly down-regulated maternal gene sets [12]. To get a more complete picture we performed a similar analysis on the larger set of maternal mRNAs recently reported by Tadros et al., [2]. 1065 mRNAs were classified as unstable maternal transcripts based on expression

profiling RNA from unfertilized wild-type embryos and assessing the degree of their destabilization over time. 138 of the 1065 unstable maternal mRNAs were among the 410 mRNAs upregulated in cluster mutant embryos at 2-3 hours (Table S4). This represents >4 fold enrichment and is statistically highly significant (P<e-52). There was no significant enrichment in 0-1 hour embryos (before the miRNAs are expressed). Much less enrichment was seen in the stable maternal class (1.2 fold; P=0.0035), which contains both stable transcripts and transcripts that are stable in unfertilized embryos, but likely degraded by the zygotic pathway in fertilized embryos. For example, some of the stable maternal class mRNAs [2] were classified as 3x down or 10x down by Arbeitman et al. [12]. 16 of these mRNAs were upregulated in the miRNA mutant, and probably contribute to the 1.2 fold enrichment of mRNAs classified as maternal stable in this set. This analysis indicates that down-regulation of maternal transcripts is impaired in the miRNA cluster mutant, suggesting that these miRNAs play a role in the zygotic pathway of maternal mRNA turnover (as hypothesized by Tadros et al., [2]).

Temporal reciprocity in miRNA/target relationships

The foregoing observations suggest that the miRNA cluster and its targets have largely reciprocal temporal expression patterns, a situation analogous to the spatially reciprocal relationship between many miRNAs and their targets at later stages of embryogenesis [13] and to the temporal relationship between the C elegans heterochronic miRNAs and their targets [14-17]. To assess the significance of these observations, we compared the occurrence of miRNA cluster target sites among the regulated mRNAs with what would be expected to occur by chance. Among the 410 transcripts upregulated in the miRNA cluster mutant, 96 contained 7mers complementary to the seed of one or more cluster miRNAs (Table S3, S4). This represents a statistically significant enrichment of 1.8-fold (P<1e-8).

Among the mRNAs upregulated in cluster mutant embryos at 2-3 hours, mRNAs from the set of maternal mRNAs examined by Arbeitman et al. (maternal_all), which contained such 7mer sites were enriched 3.6-fold, [12] (Table S3, P<1e-5). The enrichment was 6.4-fold in the class of maternal

mRNAs 3x down-regulated containing such 7mers (P<1e-6, Fig. 3B lower panel, Table S3) and 48 fold in 10x down-regulated set containing miR-309 cluster 7mer sites (P<e-5, Fig. 3B lower panel, Table S3). Importantly, no significant enrichment of 7mers was observed in 0-1h embryos, prior to the onset of miRNA cluster expression (Table S3, Fig 3B lower panel). Comparable analysis for the larger set of mRNAs [2] produced similar results (Table S4; Fig. 3B lower panel). Maternal mRNAs containing target sites were enriched 2.5 fold (P<e-8) and the set of unstable maternal mRNAs carrying target sites 6 fold among the mRNAs upregulated in cluster mutant embryos at 2-3 hours (P<e-18). Again no significant enrichment was seen in the 0-1 hour samples.

These statistical relationships suggest that the regulation of these mRNAs depends on the presence of the miRNA sites. To confirm that such sites are indeed functional, we prepared luciferase reporter constructs containing the 3'UTRs of 32 of the affected maternal mRNAs from the different functional categories mentioned above and expressed them together with the miR-309 cluster in Drosophila S2 cells (Fig. 3C). 29 of the 32 reporters were statistically significantly down-regulated upon miR-309 cluster expression, indicating that they carry functional miR-309 cluster target sites.

The cluster encodes miRNAs with 5 different seed sequences, reflecting the capacity to regulate different sets of target mRNAs. To assess the contribution of individual miRNAs to the effects of the cluster as a whole, we examined 7mer seed matches complementary to individual miR-309 cluster miRNAs. 4 of the 5 unique seeds (miR-3 and 309 have the same seed sequence) were significantly enriched among the upregulated mRNAs at 2-3 hours, but not at 0-1 hours (Table S5). The magnitude of the enrichment and the statistical significance were stronger for miR-6, suggesting that it may contribute disproportionately to the effects of the cluster. This might be in part because miR-6 is present in 3 copies and so might be expressed at a higher level than the others. These data suggest that, with the possible exception of miR-286, the 5 distinct miRNAs encoded in the cluster act in concert to regulate a broad spectrum of mRNAs during the maternal-zygotic transition.

Comparison of the maternal and zygotic systems for mRNA turnover

SMAUG has been identified as a key component of the maternal system for maternal mRNA turnover in the embryo [2], whereas the evidence presented above suggests that the miR-309 cluster acts zygotically to promote turnover of maternal mRNAs. A priori, these systems might be functionally related, acting in concert. Alternatively they might represent independent systems. To explore these possibilities we examined the degree to which the sets of targets regulated by these two systems overlap.

Of the 1065 unstable maternal transcripts identified by Tadros et al [2], 710 were identified as SMAUG targets by expression profiling RNA from unfertilized embryos laid by *smaug* mutant flies (note: SMAUG is deposited maternally and acts on maternally deposited mRNAs). As mentioned before, 138 of the transcripts upregulated in the miR-309 cluster mutant at 2-3 hours were classified as unstable maternal transcripts, which represents >4 fold enrichment (Table S3; Fig. 3B upper panel; P<e-52). 92 of these transcripts were also targeted by SMAUG, which represents >4 fold enrichment (Table S3, Fig. 3B upper panel; P<e-33). Of these, 20 (21.7%) had 7mer seed matches complementary to cluster miRNAs in their 3'UTRs and so might represent a set of mRNAs potentially co-regulated by the maternal and zygotic systems. Other mRNAs among the SMAUG targets were not affected in the miRNA cluster mutants, for example Hsp83, whose down-regulation depends strongly on the SMAUG system [18]. Of the 355 unstable transcripts that had been reported to be SMAUG-independent [2], 46 were among the 410 mRNAs upregulated in the miR-309 cluster mutant embryos. This represents a >4 fold enrichment (Table S3, Fig. 3B upper panel; P<e-16). 18 (39%) of these carry 7mers complementary to miR-309 cluster miRNAs; an 8-fold enrichment (Table S3, Fig. 3B lower panel; P<e-11). This set includes mRNAs such as orb, oskar and exuperantia, and may represent the set of mRNAs regulated mainly by the zygotic system. Together, these data suggest that the maternal and zygotic systems regulate distinct, but overlapping sets of maternal mRNAs.

Conclusions

These findings indicate that the early zygotic onset of miR-309 cluster miRNA expression acts to promote the turnover of many maternally deposited mRNAs. Failure to down-regulate maternal mRNAs by this zygotic mechanism has knock-on effects on zygotic gene expression (not shown) and may result in a late onset phenotype reflected by reduced survival and delayed larval development for many of the surviving animals. Elimination of the early zygotic expression of the miR-430 miRNA gene family also led to substantial misregulation of maternal mRNAs and to a late onset zygotic defect in Zebrafish [3]. Although miRNAs have been shown to act to ensure a proper transition between maternal and zygotic gene expression programs in flies and fish, the miRNAs involved are not conserved. Perhaps the fact that miRNAs act in part by leading to mRNA deadenylation, and subsequent destabilization, provided a means to promote turnover of a selected set of maternally deposited mRNAs. miRNAs may have been co-opted independently during evolution to fulfill a comparable function in different animals. The mechanistic basis for their action and the biological output are both conserved, but the miRNAs themselves and the identity of their targets are not. This may be an example of convergent evolution.

Experimental Procedures

Fly strains and genetics

w^{1118} flies were used as the control strain in all experiments. *miR-309-6*$^{\Delta 1}$ mutants were generated using ends-out homologous recombination essentially as described in [7] For the genomic rescue, a 2.6kb genomic fragment was amplified from w^{1118} genomic DNA using the primers GGAGCCCATAGTGACTTCAATTA and GCCACTCGGTTTCCTCTATCCT, cloned into pCasper4 and injected into w^{1118} to create transgenic flies.

miRNA Northern Blotting

Northern blotting was performed as described in [19].

In Situ hybridization

In situ analysis was performed as described in [13]. Primers used to generate the pri-miR-309 probe were CAGTCGCCACCTATACAGTTTAAGG and TGCCACAACGAACTTCAATGG.

miRNA Q-PCR

Wild-type total RNA from precisely staged early embryos was generously provided by Thomas Lecuit [9]). Primer sets designed to amplify mature miRNAs (miR-6, miR-309, miR-184, miR-310) were obtained from Applied Biosystems. Products were amplified from 10ng of total RNA samples with the ‘‘TaqMan MicroRNA Assay’’, Quantitative-PCR machine, and software from Applied Biosystems. The fold inductions of miR-6 and mir-309 were calculated relative to miR-184. Comparable results were obtained normalizing to miR-310.

Microarray Experiments

One-hour egg collections of w^{1118} and *miR-309-6*$^{\Delta 1}$ mutant flies at 25°C were either aged (2-3h sample) at 25°C or directly processed (0-1h sample). Total RNA was extracted using Trizol reagent (GibcoBRL) according to the manufacturer’s instructions. cDNA was synthesized from 3μg of total RNA of six independent samples for each time-point according to Affymetrix One-

Cycle protocol. Labeling and hybridization on Affymetrix Release 2 microarrays were performed according to Affymetrix protocols. Raw data was normalized using GCRMA [20] and significantly regulated genes were identified by Significance Analysis of Microarrays (SAM) [21]; [22] (q-value <0.05 and fold change >1.5).

Larval Survival

First instar larvae were seeded in fly vials and surviving larvae and pupae counted on the indicated days by floating them out of the food with a 20% sucrose solution and transferring them into fresh vials.

Luciferase Assays

Luciferase assays were performed as described in [13].

Statistics

To assess whether targets of the cluster miRNAs were preferentially co-expressed or anti-correlated in early wild-type embryos, we calculated the Pearson correlation coefficient for each gene's expression against miRNA expression across all 5 time-points [9]. We binned all genes according to this coefficient (bin-size 0.1) and determined the enrichment or depletion of 6, or 7mers complementary to miRNA 5'ends per bin by a hypergeometric P-value. To assess mis-regulation in miR-309 cluster knock-out flies, we selected genes that were 1.5x up-regulated compared to wild-type (q-value < 5%). We then determined the enrichment of maternal genes [2], [12], genes containing 7mers complementary to miRNA 5'ends and genes falling into both categories by a hypergeometric P-value.

References

1. Bashirullah, A., Halsell, S.R., Cooperstock, R.L., Kloc, M., Karaiskakis, A., Fisher, W.W., Fu, W., Hamilton, J.K., Etkin, L.D., and Lipshitz, H.D. (1999). Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in Drosophila melanogaster. Embo J *18*, 2610-2620.
2. Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Dev Cell *12*, 143-155.
3. Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal mRNAs. Science *312*, 75-79.
4. Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol *3*, e85.
5. Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.
6. Bushati, N., and Cohen, S.M. (2007). microRNA functions. Annu Rev Cell Dev Biol *23*, 175–205.
7. Gong, W.J., and Golic, K.G. (2003). Ends-out, or replacement, gene targeting in Drosophila. Proc Natl Acad Sci U S A *100*, 2556-2561.
8. Leaman, D., Chen, P.Y., Fak, J., Yalcin, A., Pearce, M., Unnerstall, U., Marks, D.S., Sander, C., Tuschl, T., and Gaul, U. (2005). Antisense-mediated depletion reveals essential and specific functions of microRNAs in Drosophila development. Cell *121*, 1097-1108.
9. Pilot, F., Philippe, J.M., Lemmers, C., Chauvin, J.P., and Lecuit, T. (2006). Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of Drosophila cellularisation. Development *133*, 711-723.
10. Biemar, F., Zinzen, R., Ronshaugen, M., Sementchenko, V., Manak, J.R., and Levine, M.S. (2005). Spatial regulation of microRNA gene expression in the Drosophila embryo. Proc Natl Acad Sci U S A *102*, 15907-15911.
11. Aboobaker, A.A., Tomancak, P., Patel, N., Rubin, G.M., and Lai, E.C. (2005). Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. Proc Natl Acad Sci U S A *102*, 18017-18022.
12. Arbeitman, M.N., Furlong, E.E., Imam, F., Johnson, E., Null, B.H., Baker, B.S., Krasnow, M.A., Scott, M.P., Davis, R.W., and White, K.P. (2002). Gene expression during the life cycle of Drosophila melanogaster. Science *297*, 2270-2275.
13. Stark, A., Brennecke, J., Bushati, N., Russell, R.B., and Cohen, S.M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133-1146.
14. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

15. Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.
16. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.
17. Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell *5*, 659-669.
18. Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. Curr Biol *15*, 284-294.
19. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the pro-apoptotic gene hid in Drosophila. Cell *113*, 25-36.
20. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res *31*, e15.
21. Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A *98*, 5116-5121.
22. Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. BioTechniques *34*, 374-378.

Figure legends

Figure 1. A Mutant deleting the miR-309 miRNA cluster.

(A) Schematic representation of the genomic organization of the miRNA cluster. The region deleted in *miR-309-6$^{\Delta 1}$* flies, and the genomic rescue construct are indicated. Middle: Colors indicate distinct seed sequences. Bottom: RNA blots showing miR-309 and mir-6 in control and *miR-309-6$^{\Delta 1}$* 0-22h embryos.

(B, C) Histograms comparing survival of control, *miR-309-6$^{\Delta 1}$* and rescued larvae. Samples indicate the average of 3 independent collections. Error bars indicate standard deviation. **(B)** 100 first instar larvae were seeded per vial and surviving pupae counted. The reduced survival of homozygous mutant *miR-309-6$^{\Delta 1}$* larvae was rescued by introduction of a ~2.6kb genomic DNA fragment (see panel A) as a transgene. **(C)** 50 first instar larvae were seeded per vial and surviving larvae and pupae counted on the indicated days.

Figure 2. miR-309 cluster expression

(A) Quantitative-PCR showing levels of mature miR-309 and miR-6, normalized to miR-184. Comparable results were obtained normalizing to miR-310. Time-points are defined in [9]. Note that the x-axis is not a linear time scale, but indicates temporal bins (the larger gap between the TO and T1 bins is indicated by //).

(B) In situ hybridization to detect the miR-309 cluster primary transcript. Lateral views are shown of control (WT) and *miR-309-6$^{\Delta 1}$* embryos in the process of cellularization. Anterior is to the left, dorsal is up.

Figure 3. Effects of miR-309 cluster miRNAs on target expression

(A) Histogram showing the proportion of 7mer seed matches for all miRNAs and for the miR-309 cluster in a set of mRNAs sorted according to the degree of correlation between their temporal expression pattern (determined by expression profiling [9]) and that of the miRNAs in the miR-309 cluster. Note that both expression profiles were done on the very same RNA samples to ensure maximal comparability.

(B) Histograms showing the fold enrichment of maternal mRNAs (upper

panel) and maternal mRNAs containing miR-309 cluster 7mer sites (lower panel) among the transcripts upregulated by >1.5-fold (q<0.05) in *miR-309-6*$^{\Delta 1}$ 0-1h and 2-3h embryos. Gene classes maternal 3xdown and maternal 10xdown are defined in [12], the remaining gene classes are defined in [2]. (*) denotes p<1e-05.

(C) Target validation using 3'UTR reporters in a luciferase assay. 3'UTRs of transcripts up-regulated in the miR-309 mutant and carrying at least one 7mer complementary to a miR-309 cluster miRNA in their 3'UTR were tested. The control 3'UTRs dpld and CG13850_sh lack miR-309 cluster 7mers and are not repressed upon miR-309 cluster expression. The classifications by Arbeitman et al [12] (maternal all, 10x and 3x down indicated as all, 10x and 3x) and by Tadros et al [2] (maternal stable, maternal unstable and SMAUG dependent vs independent), are indicated below. Error bars: standard deviation. (*) p<0.05, n=3; double-sided t-test.

Supplemental Figure S1. (A-E) Fluorescent pri-miR-309 in situ hybridization of wild-type embryos during early development. Anterior is to the left, posterior to the right. (A-D) Lateral views. (A, B) Stage 4. (C, D) Stage 5. (E) Dorsal view of a late stage 5 embryo. **(F, G)** Embryos labeled with antibodies to detect Ftz (green) and Eve (red) proteins. Genotypes as indicated.

Supplemental Tables

Table S1. Temporal expression of genes containing 7mers complementary to miR-309 cluster seeds.

All genes with significant signals for at least one time-point (defined in Pilot, et al. 2006) were binned according to their degree of co-regulation with the miRNA (Pearson correlation, column 1). Positive numbers indicate genes whose temporal expression pattern is similar to that of the miRNAs. Negative numbers indicate anti-correlation, with miRNAs in the –1 to –0.9 group being least similar to the miRNA pattern. Column 2 shows the number of target genes with 7mer seed matches to the cluster miRNAs in each class. Column 3 shows the abundance of 7mers for all miRNAs in each class. Column 4 shows the overall abundance of 7mers for cluster

miRNAs in all genes. Column 5 shows the overall abundance of 7mers for all miRNAs in all genes. Columns 6 and 7 show hypergeometric P-values indicating significant enrichment or depletion. miR-309 cluster target sites are overrepresented in the most strongly anti-correlated mRNAs (yellow, not statistically significant for 7mers) and underrepresented in the mRNAs expressed most similarly to the miRNAs (green).

Table S2. Temporal expression of genes containing 6mers complementary to miR-309 cluster seeds.
Same as S1 but for 6mers, where trends are more significant due to overall higher numbers. Classes showing P-values < 0.05 are highlighted.

Table S3. Up-Regulation of gene classes defined by Arbeitman et al. (2002) in the miR-309 cluster mutant.
Target genes stands for transcripts containing 7mers complementary to miR-309 cluster miRNAs. Column 3 shows the number of genes for different classes. Column 2 shows the intersection of these classes with genes upregulated >1.5 fold (q<0.05) in the miR-309 cluster mutant. Column 4 shows the number of genes upregulated >1.5 fold (q<0.05) in the miR-309 cluster mutant compared to control embryos of the stage indicated. Column 5 shows the total number of genes included in the microarray analysis. Columns 6 and 7 show the resulting enrichment (enrichment factor and P-value). Classes showing enrichment P-values < 1e-5 are highlighted in yellow.

Table S4. Up-Regulation of maternal mRNA classes defined by Tadros et al. (2007) in the miR-309 cluster mutant.
Same as S3 but using the gene classes defined by Tadros et al. (2007). Classes showing enrichment P-values < 1e-8 are highlighted in yellow.

Table S5. Occurrence of 7mer seed matches for individual miRNAs in maternal mRNAs up-regulated in the miR-309 cluster mutant.
Genes were grouped by occurrence of 7mer seed matches for all cluster

miRNAs (row 1) or individual miR-309 cluster miRNAs (column 1). Column 2 shows the total number of genes carrying 7mers complementary to the indicated miRNAs. Column 3 shows the total number of genes included in the analysis. Columns 4-7 show the results obtained for the genes upregulated by >1.5 fold in 0-1h mutant embryos, columns 8-11 show those obtained in 2-3h mutant embryos. Columns 4, 8 show the number of upregulated genes carrying 7mers complementary to the indicated miRNAs. Columns 5, 9 show the total number of upregulated genes. Columns 6, 7, 10, 11 show the resulting enrichment (enrichment factor and P-value). Enrichment P-values <1e-2 are highlighted in yellow.

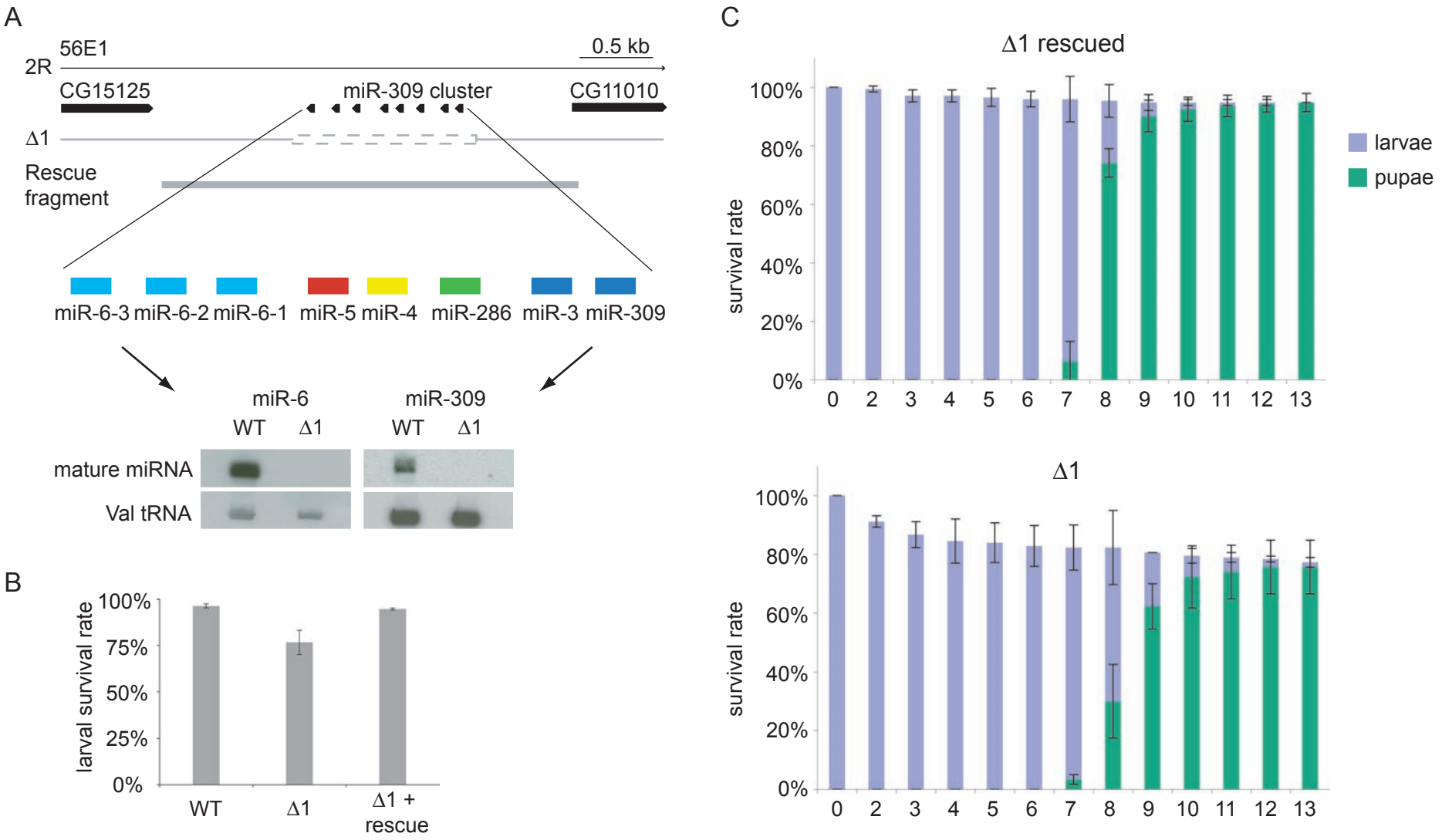

A
56E1
0.5 kb
2R
CG15125
miR-309 cluster
CG11010
Δ1
Rescue
fragment
miR-6-3 miR-6-2 miR-6-1 miR-5 miR-4 miR-286 miR-3 miR-309
miR-6
miR-309
WT Δ1 WT Δ1
mature miRNA
Val tRNA
B
larval survival rate
100%
75%
50%
25%
0%
WT
Δ1
Δ1 +
rescue
C
Δ1 rescued
survival rate
100%
80%
60%
40%
20%
0%
0 2 3 4 5 6 7 8 9 10 11 12 13
larvae
pupae
Δ1
survival rate
100%
80%
60%
40%
20%
0%
0 2 3 4 5 6 7 8 9 10 11 12 13

Figure 1

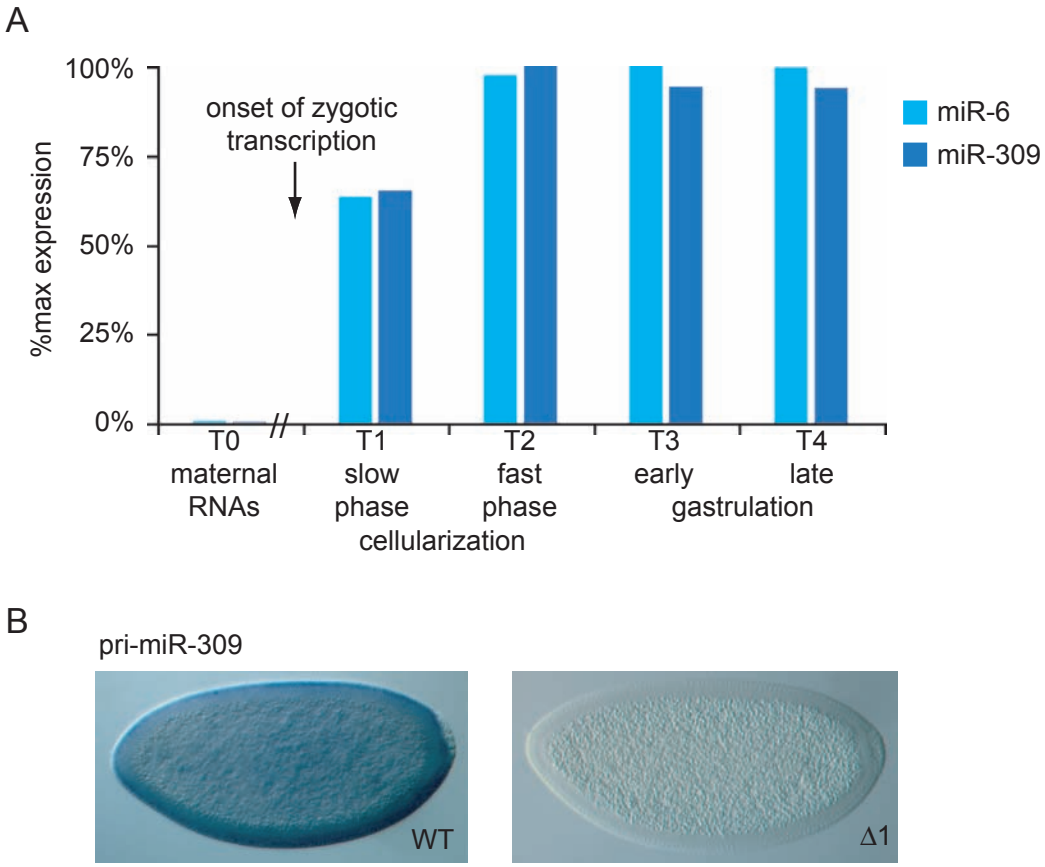

A
100%
75%
50%
25%
0%
%max expression
onset of zygotic
transcription
miR-6
miR-309
T0
maternal
RNAs
T1
slow
phase
T2
fast
phase
T3
early
T4
late
gastrulation
cellularization
B
pri-miR-309
WT
Δ1

Figure 2

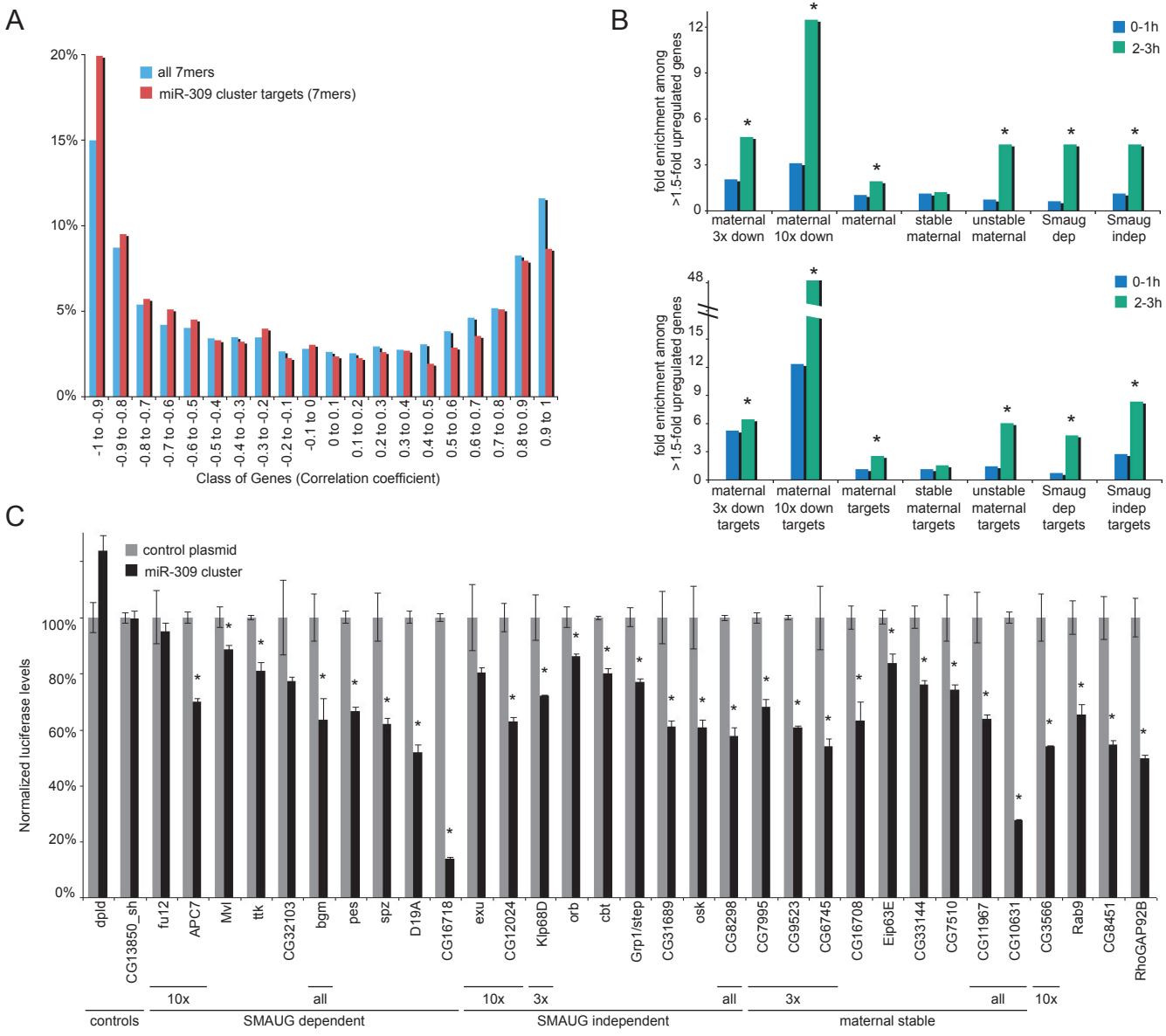
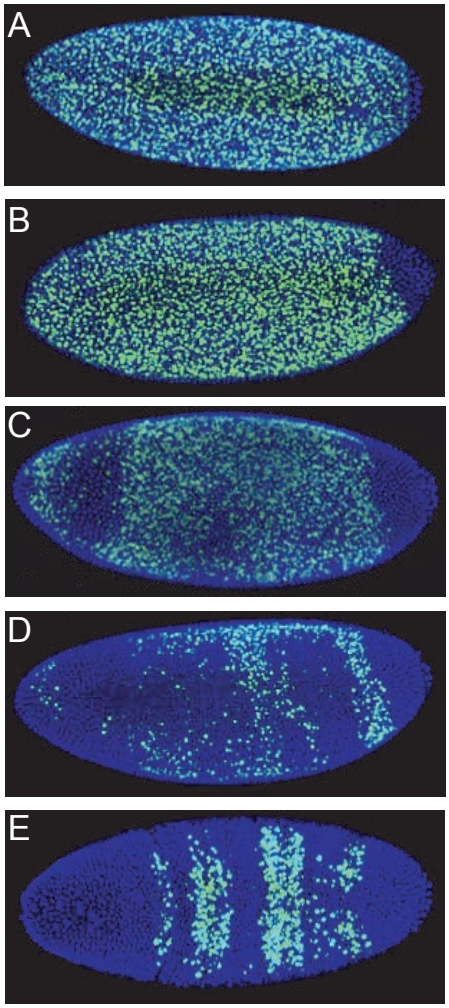
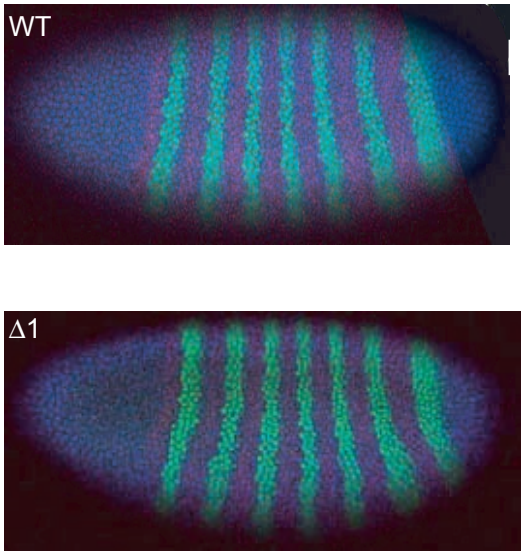

A
20%
15%
10%
5%
0%
all 7mers
miR-309 cluster targets (7mers)
-1 to -0.9
-0.9 to -0.8
-0.8 to -0.7
-0.7 to -0.6
-0.6 to -0.5
-0.5 to -0.4
-0.4 to -0.3
-0.3 to -0.2
-0.2 to -0.1
-0.1 to 0
0 to 0.1
0.1 to 0.2
0.2 to 0.3
0.3 to 0.4
0.4 to 0.5
0.5 to 0.6
0.6 to 0.7
0.7 to 0.8
0.8 to 0.9
0.9 to 1
Class of Genes (Correlation coefficient)
B
fold enrichment among >1.5-fold upregulated genes
0-1h
2-3h
maternal 3x down
maternal 10x down
maternal
stable maternal
unstable maternal
Smaug dep
Smaug indep
maternal 3x down targets
maternal 10x down targets
maternal targets
stable maternal targets
unstable maternal targets
Smaug dep targets
Smaug indep targets
C
Normalized luciferase levels
control plasmid
miR-309 cluster
dpld
CG13850_sh
fu12
APC7
Mvl
ttk
CG32103
bgm
pes
spz
D19A
CG16718
exu
CG12024
Klp68D
orb
cbt
Grp1/step
CG31689
osk
CG8298
CG7995
CG9523
CG6745
CG16708
Eip63E
CG33144
CG7510
CG11967
CG10631
CG3566
Rab9
CG8451
RhoGAP92B
10x
all
10x
3x
all
3x
all
10x
controls
SMAUG dependent
SMAUG independent
maternal stable

Figure 3



F



Supplemental Figure S1

Table S1. Temporal expression of genes containing 7mers complementary to miR-309 cluster miRNA seeds.

Class of Genes (range of correl. coefficients)	# miRNA 7mers in Class	# all 7mers in Class	# miRNA 7mers in all genes	# all 7mers in all genes	P(under)	P(over)
-1 to -0.9	113	323297	979	3121844	8.98E-01	1.22E-01
-0.7 to -0.6	36	96860	979	3121844	8.71E-01	1.72E-01
0.7 to 0.8	41	115909	979	3121844	8.08E-01	2.41E-01
-0.3 to -0.2	27	80022	979	3121844	6.87E-01	3.88E-01
-0.5 to -0.4	24	73351	979	3121844	6.24E-01	4.58E-01
-0.8 to -0.7	35	110958	979	3121844	5.48E-01	5.20E-01
-0.2 to -0.1	17	58674	979	3121844	4.16E-01	6.73E-01
-0.9 to -0.8	54	182430	979	3121844	3.56E-01	6.93E-01
0.6 to 0.7	25	92254	979	3121844	2.59E-01	7.99E-01
0.3 to 0.4	15	57712	979	3121844	2.69E-01	8.03E-01
0.2 to 0.3	17	67129	979	3121844	2.17E-01	8.42E-01
-0.4 to -0.3	21	82590	979	3121844	1.90E-01	8.59E-01
-0.6 to -0.5	24	94059	979	3121844	1.75E-01	8.69E-01
0.8 to 0.9	45	171953	979	3121844	1.19E-01	9.07E-01
0 to 0.1	11	51366	979	3121844	1.23E-01	9.21E-01
0.5 to 0.6	19	86345	979	3121844	6.98E-02	9.53E-01
0.1 to 0.2	8	46967	979	3121844	4.19E-02	9.80E-01
-0.1 to 0	11	69689	979	3121844	1.25E-02	9.93E-01
0.4 to 0.5	10	69090	979	3121844	7.63E-03	9.96E-01
0.9 to 1	46	222714	979	3121844	1.87E-03	9.99E-01

Table S2. Temporal expression of genes containing 6mers complementary to miR-309 cluster miRNA seeds.

Class of Genes (range of correl. coefficients)	# miRNA 6mers in Class	# all 6mers in Class	# miRNA 6mers in all genes	# all 6mers in all genes	P(under)	P(over)
-1 to -0.9	454	324053	3839	3129277	9.99E-01	1.51E-03
-0.8 to -0.7	159	111235	3839	3129277	9.78E-01	2.73E-02
-0.7 to -0.6	130	97077	3839	3129277	8.56E-01	1.66E-01
0.2 to 0.3	84	67264	3839	3129277	5.87E-01	4.57E-01
-0.4 to -0.3	102	82772	3839	3129277	5.38E-01	5.02E-01
-0.9 to -0.8	224	182860	3839	3129277	5.05E-01	5.23E-01
0.3 to 0.4	70	57846	3839	3129277	4.78E-01	5.70E-01
-0.1 to 0	84	69829	3839	3129277	4.49E-01	5.94E-01
0.1 to 0.2	56	47088	3839	3129277	4.33E-01	6.18E-01
-0.3 to -0.2	94	80195	3839	3129277	3.46E-01	6.91E-01
-0.6 to -0.5	108	94267	3839	3129277	2.50E-01	7.79E-01
-0.5 to -0.4	80	73525	3839	3129277	1.51E-01	8.73E-01
-0.2 to -0.1	62	58812	3839	3129277	1.26E-01	8.97E-01
0.6 to 0.7	99	92460	3839	3129277	9.20E-02	9.23E-01
0 to 0.1	52	51489	3839	3129277	8.78E-02	9.31E-01
0.7 to 0.8	123	116133	3839	3129277	5.25E-02	9.56E-01
0.8 to 0.9	184	172327	3839	3129277	2.84E-02	9.76E-01
0.4 to 0.5	61	69233	3839	3129277	5.04E-03	9.96E-01
0.5 to 0.6	77	86524	3839	3129277	2.39E-03	9.98E-01
0.9 to 1	203	223294	3839	3129277	4.96E-06	1.00E+00

Table S3. Up-regulation of gene classes defined by Arbeitman et al. (2002) in the miR-309 cluster mutant.

Class of genes	# genes in class upregulated	# genes in class	# genes upregulated	# genes on array	enrichment factor	P-value
0-1h						
Maternal mRNAs (3x down-reg)	6	291	138	13615	2.0	0.08
Maternal mRNAs (10x down-reg)	1	32	138	13615	3.1	0.28
Zygotic_early genes	6	470	138	13615	1.3	0.52
Zygotic (10x up)	2	52	138	13615	3.8	0.10
Target genes	21	1774	138	13615	1.2	0.26
Maternal_all target genes	4	159	138	13615	2.5	0.08
Maternal (3x down) target genes	3	57	138	13615	5.2	0.02
Maternal (10x down) target genes	1	8	138	13615	12.3	0.07
2-3h						
Maternal genes (3x down-reg)	42	291	410	13615	4.8	1.8E-17
Maternal genes (10x down-reg)	12	32	410	13615	12.5	6.2E-11
Zygotic_early genes	8	470	410	13615	0.6	0.97
Zygotic (10x up)	1	52	410	13615	0.6	0.80
Target genes	96	1774	410	13615	1.8	3.9E-09
Maternal all target genes	17	159	410	13615	3.6	6.05E-06
Maternal (3x down) target genes	11	57	410	13615	6.4	8.6E-07
Maternal (10x down) target genes	5	8	410	31615	48.2	1.26E-06

Table S4. Up-regulation of maternal mRNA classes defined by Tadros et al. (2007) in the miR-309 cluster mutant.

Class of genes	# genes in class upregulated	# genes in class	# genes upregulated	# genes on array	enrichment factor	P-value
0-1h						
Maternal mRNAs	52	5071	138	13615	1.0	0.49
Stable maternal mRNAs	44	4006	138	13615	1.1	0.29
Unstable maternal mRNAs	8	1065	138	13615	0.7	0.85
Smaug dependent mRNAs	4	710	138	13615	0.6	0.93
Smaug independent mRNAs	4	355	138	13615	1.1	0.49
Target genes	21	1774	138	13615	1.2	0.26
Maternal target genes	11	961	138	13615	1.1	0.38
Stable maternal target genes	8	749	138	13615	1.1	0.49
Unstable maternal target genes	3	212	138	13615	1.4	0.36
Smaug dependent target genes	1	140	138	13615	0.7	0.76
Smaug independent target genes	2	72	138	13615	2.7	0.17
2-3h						
Maternal mRNAs	284	5071	410	13615	1.9	1.2E-40
Stable maternal mRNAs	146	4006	410	13615	1.2	3.5E-03
Unstable maternal mRNAs	138	1065	410	13615	4.3	2.9E-53
Smaug dependent mRNAs	92	710	410	13615	4.3	2.9E-34
Smaug independent mRNAs	46	355	410	13615	4.3	3.6E-17
Target genes	96	1774	410	13615	1.8	3.9E-09
Maternal target genes	72	961	410	13615	2.5	3.4E-13
Stable maternal target genes	34	749	410	13615	1.5	0.01
Unstable maternal target genes	38	212	410	13615	6.0	3.5E-19
Smaug dependent target genes	20	140	410	13615	4.7	7.1E-09
Smaug independent target genes	18	72	410	13615	8.3	2.7E-12

Table S5. Occurrence of 7mer seed matches for individual miRNAs in maternal mRNAs up-regulated in the miR-309 cluster mutant.

			0-1 hour RNAs upregulated >1.5 fold				**2-3 hour** RNAs upregulated >1.5 fold			
"targets of"	**# targets**	**# total genes**	**#in set up-reg**	**# total up-reg**	**enrichment factor**	**P-value**	**#in set up-reg**	**# total up-reg**	**enrichment factor**	**P-value**
cluster	1774	13615	21	138	1.2	2.6E-01	96	410	1.8	3.9E-09
miR-6-1	478	13615	5	138	1.0	5.4E-01	35	410	2.4	1.1E-06
miR-4	739	13615	10	138	1.3	2.2E-01	35	410	1.6	5.3E-03
miR-5	379	13615	4	138	1.0	5.4E-01	21	410	1.8	5.4E-03
miR-309	239	13615	1	138	0.4	9.1E-01	15	410	2.1	5.9E-03
miR-3	239	13615	1	138	0.4	9.1E-01	15	410	2.1	5.9E-03
miR-286	256	13615	2	138	0.8	7.4E-01	10	410	1.3	0.24

microRNA Functions

Natascha Bushati and Stephen M. Cohen

Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany 69117; email: Stephen.Cohen@embl-heidelberg.de

Annu. Rev. Cell Dev. Biol. 2007. 23:175–205

First published online as a Review in Advance on May 21, 2007

The *Annual Review of Cell and Developmental Biology* is online at http://cellbio.annualreviews.org

This article's doi:
10.1146/annurev.cellbio.23.090506.123406

1081-0706/07/1110-0175$20.00

Key Words

posttranscriptional gene regulation, development, disease, genetics

Abstract

microRNAs (miRNAs) are small noncoding RNAs that play important roles in posttranscriptional gene regulation. In animal cells, miRNAs regulate their targets by translational inhibition and mRNA destabilization. Here, we review recent work in animal models that provide insight into the diverse roles of miRNAs in vivo.

Contents

INTRODUCTION

Since the discovery of the founding members of the microRNA (miRNA) family, *lin-4* and *let-7* (Lee et al. 1993, Reinhart et al. 2000, Wightman et al. 1993), hundreds of miRNAs have been identified in plants, animals, and viruses by molecular cloning and bioinformatic approaches (Berezikov et al. 2006, Lagos-Quintana et al. 2001, Lau et al. 2001, Lee & Ambros 2001, Ruby et al. 2006). miRNAs were found to downregulate gene expression by base-pairing with the 3′ untranslated regions (3′UTRs) of target messenger RNAs (mRNAs) (Lee et al. 1993, Reinhart et al. 2000, Slack et al. 2000, Wightman et al. 1993). These discoveries indicated that this class of noncoding RNA molecules may constitute a new layer of regulatory control over gene expression programs in many organisms. Here, we review recent work, principally from animal models, that reveals how miRNAs are generated and act to silence gene expression, how targets of miRNAs can be identified, and how the biological functions of miRNAs can be illuminated by knowledge of gene expression patterns, by mutant phenotypes of miRNAs, and by overexpression of their targets.

microRNA (miRNA): ~22-nt noncoding RNA that serves as a posttranscriptional regulator

miRNA BIOGENESIS AND SILENCING MECHANISM

Most miRNA genes are transcribed by RNA polymerase II (Pol II) to generate a stemloop containing primary miRNA (pri-miRNA), which can range in size from hundreds of nucleotides to tens of kilobases (Cai et al. 2004, Y. Lee et al. 2004) (**Figure 1**). An exception are miRNAs lying within Alu-repetitive elements, which can be transcribed by RNA polymerase III (Borchert et al. 2006). Like mRNAs, Pol II–transcribed pri-miRNAs contain 5′ cap structures, are polyadenylated, and may be spliced (Bracht et al. 2004, Cai et al. 2004). A very recent report shows that most mammalian miRNAs are encoded in introns and that miRNA processing appears to occur before splicing (Kim et al. 2007). The pri-miRNA is processed within the nucleus by a multiprotein complex called the Microprocessor, of which the core components are the RNase III enzyme Drosha and the double-stranded RNA-binding domain (dsRBD) protein DGCR8/Pasha (Denli et al. 2004, Gregory et al. 2004, Han et al. 2004, Landthaler et al. 2004, Lee et al. 2003). This complex cleaves the pri-miRNA

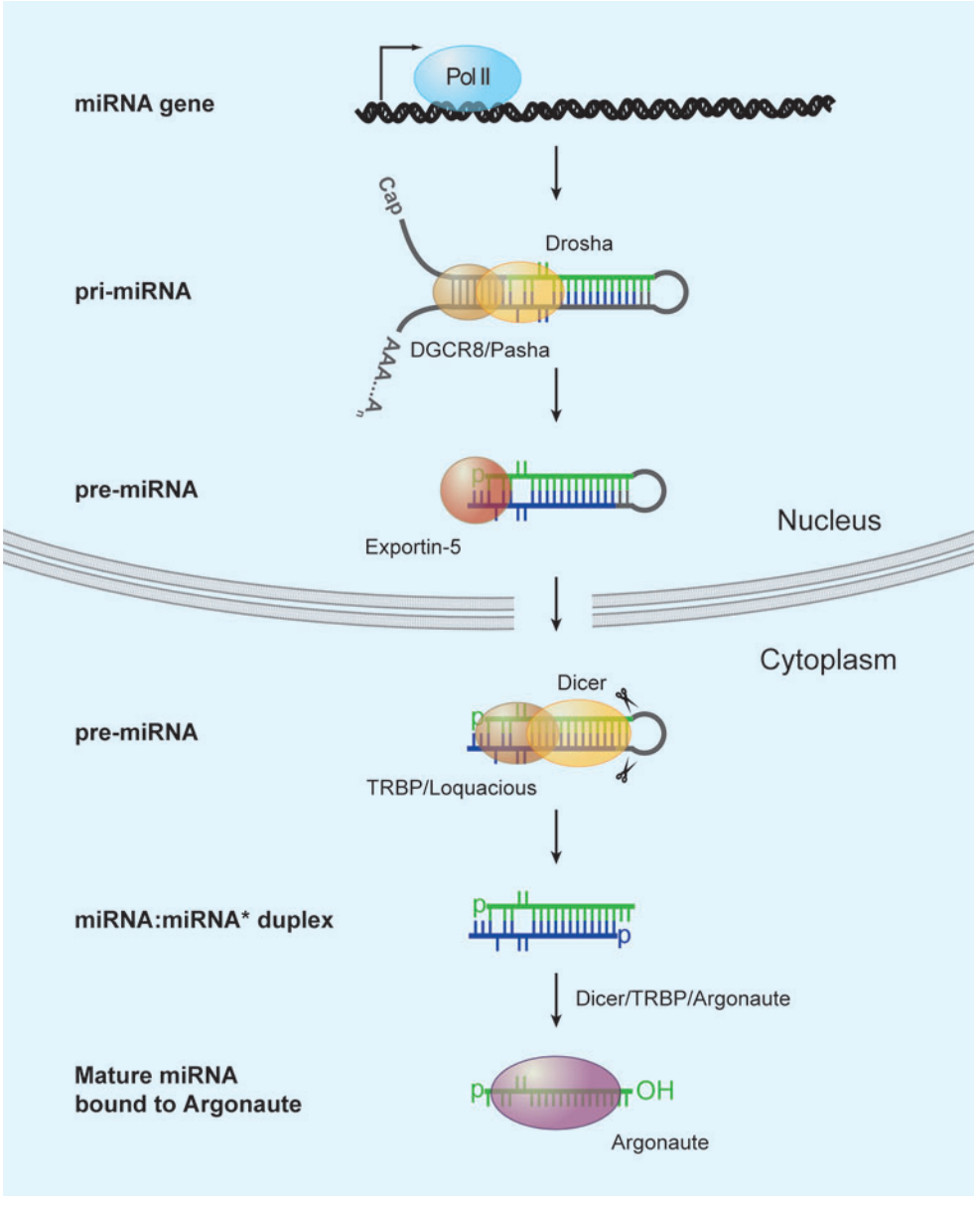


Figure 1

miRNA biogenesis. An miRNA gene is transcribed, generally by RNA polymerase II (Pol II), generating the primary miRNA (pri-miRNA). In the nucleus, the RNase III endonuclease Drosha and the double-stranded RNA-binding domain (dsRBD) protein DGCR8/Pasha cleave the pri-miRNA to produce a 2-nt 3′ overhang containing the ~70-nt precursor miRNA (pre-miRNA). Exportin-5 transports the pre-miRNA into the cytoplasm, where it is cleaved by another RNase III endonuclease, Dicer, together with the dsRBD protein TRBP/Loquacious, releasing the 2-nt 3′ overhang containing a ~21-nt miRNA:miRNA* duplex. The miRNA strand is loaded into an Argonaute-containing RNA-induced silencing complex (RISC), whereas the miRNA* strand is typically degraded.

3′ untranslated region (3′UTR): 3′ untranslated sequences that follow the protein-encoding open reading frame in a messenger RNA

Primary miRNA (pri-miRNA): the miRNA transcript

Precursor miRNA (pre-miRNA): a ~70-nt stemloop structure

Dicer: cytoplasmic RNase III enzyme that processes the pre-miRNA to produce the mature ~22-nt miRNA duplex

Argonaute: catalytic component of the RISC complex

RNA-induced silencing complex (RISC): the ribonucleoprotein complex required for small RNA-mediated gene silencing

Slicing: possible mode of action of the Ago2-containing RISC complex, in which the target mRNA is cleaved and subsequently degraded

stem by measuring the distance from the single-stranded/double-stranded RNA junction (Han et al. 2006), producing a ~70-nt hairpin precursor miRNA (pre-miRNA). The 2-nt 3′ overhang, characteristic of RNase III–mediated cleavage, is recognized by Exportin-5, which transports the pre-miRNA into the cytoplasm via a Ran-GTP-dependent mechanism (Bohnsack et al. 2004, Lund et al. 2004, Yi et al. 2003).

Next, the pre-miRNA is cleaved to produce the mature ~22-nt miRNA:miRNA* duplex by another RNase III enzyme, Dicer, which interacts with the dsRBD proteins TRBP/Loquacious and, in human cells, PACT (**Figure 1**; Chendrimada et al. 2005, Forstemann et al. 2005, Hutvagner et al. 2001, Jiang et al. 2005, Ketting et al. 2001, Lee et al. 2006, Saito et al. 2005). Subsequently, in human cells TRBP recruits the Argonaute protein Ago2 (and perhaps other Ago proteins), and together with Dicer they form a trimeric complex that initiates the assembly of the RNA-induced silencing complex (RISC), a ribonucleoprotein complex (Gregory et al. 2005, Maniataki & Mourelatos 2005). The miRNA strand with relatively lower stability of base-pairing at its 5′ end is incorporated into RISC, whereas the miRNA* strand is typically degraded (Schwarz et al. 2003, Du & Zamore 2005). Recent deep sequencing efforts indicate that the average ratio of miRNA to miRNA* is ~100:1 but can be much lower in cases in which both strands are functional (Ruby et al. 2006).

Once incorporated into RISC, the miRNA guides the complex to its RNA targets by base-pairing interactions. In cases of perfect or near-perfect complementarity to the miRNA, target mRNAs can be cleaved (sliced) and degraded; otherwise, their translation is repressed (Hutvagner & Zamore 2002, Martinez & Tuschl 2004). For slicing to occur, RISC must contain an Argonaute protein capable of endonucleolytic cleavage. Ago2 is the sole enzyme conferring this activity in mammals and is the major enzyme in flies (Liu et al. 2004, Meister et al. 2004, Okamura et al. 2004). Only one endogenous animal miRNA target has been reported to be sliced (Yekta et al. 2004). Most animal miRNAs base-pair imperfectly with their targets and promote translational repression rather than cleavage and degradation. In this mode of repression, target mRNAs are not actively degraded but can be destabilized owing to deadenylation and subsequent decapping (Behm-Ansmant et al. 2006, Giraldez et al. 2006, Jackson & Standart 2007, Wu et al. 2006). If slicing is so rare in animal miRNA-target interactions, it is curious that the mechanism has been so well conserved. Perhaps this mechanism is used more often than we are currently aware, or the conservation has some other basis.

The mechanism of translational repression by miRNAs remains unclear. Indeed, the step at which miRNAs block translation is controversial. There is evidence that miRNAs block translation initiation, whereas other studies suggest a block in elongation (Humphreys et al. 2005, Maroney et al. 2006, Nottrott et al. 2006, Petersen et al. 2006, Pillai et al. 2005). Argonaute proteins bound to miRNAs and their target mRNAs accumulate in processing bodies (P-bodies), cytoplasmic foci that are known sites of mRNA degradation (Liu et al. 2005a, Pillai et al. 2005, Sen & Blau 2005). P-bodies exclude ribosomal components and may therefore serve as sites in which mRNAs can be stored without translation. Several proteins found in P-bodies (GW182, the Dcp1/Dcp2 decapping complex, and the RCK/p54 helicase) can bind to Argonaute proteins, and this interaction mediates translational repression (Behm-Ansmant et al. 2006; Chu & Rana 2006; Liu et al. 2005a,b; Rehwinkel et al. 2005). Yet, doubts remain over the importance of P-bodies in target repression. Disrupting P-bodies does not affect the degree of translational repression, and therefore the P-body localization may be a consequence of repression rather than its cause (Chu & Rana 2006, Jackson & Standart 2007). Moreover, miRNA-mediated repression and P-body localization are reversible (Bhattacharyya

et al. 2006), indicating that the P-bodies may serve as sites of temporary storage of translationally repressed mRNAs.

Many human mRNAs that encode proteins whose levels are under tight control have AU-rich elements in their 3′UTRs. These elements are bound by an AU-rich element–binding protein, ARE. Intriguingly, a specific miRNA, *miR-16*, is required for the rapid turnover of mRNAs containing AU-rich elements in their 3′UTRs, to which *miR-16* binds (Jing et al. 2005). This does not seem to involve site-specific endonucleolytic cleavage because none of the decay intermediates expected from slicer activity were observed. Presumably, *miR-16* collaborates via RISC binding with TPP, which is required for ARE mRNA degradation. A very recent paper also suggests a new role for Argonaute proteins and AU elements in translational regulation (Vasudevan & Steitz 2007).

Although most attention to date has been paid to miRNA action in the cytoplasm, a recent report shows that mature *miR-29b* contains a 6-nt motif at its 3′ terminus, which directs import of the mature miRNA into the nucleus (Hwang et al. 2007), raising intriguing possibilities for other modes of miRNA function.

OTHER SMALL NONCODING RNAs

This review focuses on animal miRNAs and their biological functions. But other types of small noncoding RNA with distinct properties deserve mention. Small interfering RNAs (siRNAs) differ from miRNAs mainly in their origin: They derive from endogenous or exogenous double-stranded RNAs and are processed into siRNAs by Dicer. siRNAs usually induce cleavage of their targets when loaded onto an Ago2-containing RISC. However, siRNAs can also act as miRNAs on targets with imperfect complementarity and induce translational repression (Doench et al. 2003). Depending on the number and position of the mismatches, RISC-mediated cleavage of mismatched targets can also occur, albeit at a lower rate (Martinez & Tuschl 2004, Meister & Tuschl 2004).

Piwi-interacting RNAs (piRNAs) are a third group of small RNAs that has garnered considerable attention recently (Aravin et al. 2006, Girard et al. 2006, Grivna et al. 2006, Lau et al. 2006, Watanabe et al. 2006). They associate with members of the Piwi family, a subtype of Argonaute proteins, including the mouse proteins MILI and MIWI and perhaps MIWI2. piRNAs are 26–31 nt in length. Because piRNAs accumulate at the onset of male meiosis and sperm maturation arrests at different stages in *Mili* and *Miwi* knockout mice, piRNAs are thought to play an essential role during gametogenesis. Recent evidence indicates that the previously identified class of repeat-associated siRNAs (Aravin et al. 2003) is the major class of *Drosophila* piRNAs (Brennecke et al. 2007, Gunawardane et al. 2007, Saito et al. 2006, Vagin et al. 2006). Although the function and targets of mammalian piRNAs are unclear, *Drosophila* piRNAs are involved in silencing transposons in the male and female germ lines. Recently an intricate system for the control of transposons was identified; it involves discrete piRNA-generating loci composed of defective transposon copies and an amplification cycle to boost piRNAs against actively transcribed transposons (Brennecke et al. 2007, Gunawardane et al. 2007). piRNA biogenesis in flies is independent of Dicer and components of the miRNA and siRNA pathways (Vagin et al. 2006). Recent evidence indicates that the Piwi proteins themselves are directly involved in the biogenesis of piRNAs (Brennecke et al. 2007, Gunawardane et al. 2007).

Plant miRNAs differ from animal miRNAs in their biogenesis and in their mode of target regulation: Most plant miRNAs display perfect or near-perfect complementarity to their target mRNAs—sites occur in both coding regions and 3′UTRs—and therefore induce mRNA slicing (Vaucheret 2006). However, the plant miRNA family *miR-854*

miRNA target: a messenger RNA encoding a protein, containing target sites for and regulated by an miRNA

miRNA seed: seven to eight nucleotides at the 5′ end of an miRNA that serve as the primary determinant of target specificity

SNP: single-nucleotide polymorphism

was recently reported to act by translational repression via imperfect binding sites and to be conserved in animals (Arteaga-Vazquez et al. 2006). The observation of potential GU-rich binding sites in the animal ortholog of the plant *miR-854* target *UBP1b* is suggestive, but this target has not yet been validated. These observations are surprising because miRNAs are thought to have evolved independently in the plant and animal kingdoms (Chen & Rajewsky 2007). However, the recent discovery of miRNAs and genes involved in miRNA biogenesis in the unicellular green algae *Chlamydomonas* suggests an older, perhaps common origin for plant and animal miRNAs (D. Baulcombe, personal communication).

TARGET IDENTIFICATION

The first animal miRNA targets were identified by genetics in *Caenorhabditis elegans*. The heterochronic miRNA *lin-4* displays a mutant phenotype that can be suppressed by a second mutation in its target mRNA, *lin-14* (Lee et al. 1993, Wightman et al. 1993). These genetic interactions led to the identification of sequences complementary to *lin-4* in the 3′UTR of *lin-14* (**Figure 2**). Although powerful, this genetic approach can identify only those targets, such as *lin-14*, whose overexpression is directly responsible for the miRNA mutant phenotype, in this case *lin-4*. Few examples of this type have been found, and it is unclear whether this sort of relationship will be more the exception than the rule.

Subsequently, the criteria for miRNA-target interactions were deciphered by mutation of known miRNA-target sites and testing for function in miRNA misexpression assays (Brennecke et al. 2005b, Doench & Sharp 2004, Kiriakidou et al. 2004, Kloosterman et al. 2004, Lewis et al. 2003). These studies focused attention on the importance of pairing to the 5′ end of the miRNA, called the seed region (**Figure 2**). We have grouped target sites into two broad classes: (*a*) 5′ dominant sites, which base-pair precisely to the seed of the miRNA, with or without 3′ pairing support, and (*b*) 3′ compensatory sites, which have insufficient 5′ pairing compensated for by strong pairing to the miRNAs' 3′ region.

In addition to direct experimental tests for site function, further indication of the importance of the seed region in miRNA-target recognition has been inferred from computational studies that showed significant overrepresentation of conservation of matches to miRNA seeds or, in some cases, avoidance of miRNA seed matches (Brennecke et al. 2005b; Farh et al. 2005; Krek et al. 2005; Lewis et al. 2003, 2005; Stark et al. 2005; Xie et al. 2005). A recent analysis of single-nucleotide polymorphism (SNP) genotype data also showed that polymorphism density was significantly lower in conserved target site regions that are complementary to the 5′ portion of the miRNA (Chen & Rajewsky 2006).

miRNA-target prediction has gone through several iterations, with significant improvements to the original efforts (e.g.,

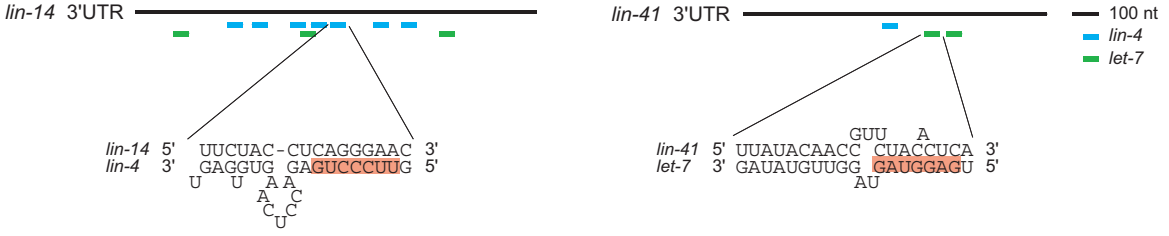


Figure 2

The microRNAs *lin-4* and *let-7* confer regulation to their targets via imprecise base-pairing with the target's 3′UTRs. The miRNA seed region is highlighted in red.

Enright et al. 2003, Lewis et al. 2003, Stark et al. 2003) made on the basis of rules derived from the experimental approaches described above and by the use of evolutionary conservation (Brennecke et al. 2005b, Grun et al. 2005, Krek et al. 2005, Lewis et al. 2005, Rajewsky 2006, Xie et al. 2005). The newer-generation methods have largely converged on a similar approach and have led to estimates that more than 30% of animal genes may be miRNA targets. Recently, a pattern-based algorithm that predicts miRNA-target sites without relying on cross-species conservation or miRNA sequence estimated even larger numbers of miRNA-regulated genes (Miranda et al. 2006). Whether this approach proves to be valid awaits verification.

One approach to assess target predictions uses miRNA misexpression and assays for target downregulation. Overexpression of miRNAs in tissue culture followed by expression profiling provides a global picture of target RNAs that are destabilized by miRNA binding (Lim et al. 2005). The converse approach, bulk depletion of miRNAs followed by mRNA profiling, demonstrates upregulation of many, but not all, predicted target mRNAs (Rehwinkel et al. 2006). Depletion of single miRNAs by antisense methods (Krutzfeldt et al. 2005) or by the use of miRNA mutants (N. Bushati, J. Karres, G. Easow & S.M. Cohen, unpublished results) shows that many RNAs are upregulated; among these are predicted targets but also many RNAs with potential target sites that were not predicted because they are not evolutionarily conserved. However, many potential targets are not affected on the RNA level, suggesting that these approaches will underestimate meaningful miRNA-target relationships. On the protein level, miRNA-target interactions are not amenable to high-throughput approaches and must be tested one by one in reporter assays. Although miRNA misexpression is useful for testing whether regulation is possible, such experiments are not sufficient to draw conclusions about miRNA-target relationships in vivo.

Given the large numbers of predicted miRNA-target genes and the paucity of genetic evidence for miRNA-target relationships, the significance of the predicted targets remains an open question. Computational studies suggest that, on the basis of the occurrence of SNPs within conserved miRNA sites, ~85% of conserved miRNA sites are functionally important (Chen & Rajewsky 2006). The regulation of such sites by miRNAs presumably improves the fitness of the organism in some way, unless the sequences are conserved for reasons unrelated to miRNAs. However, 16 predicted conserved target sites of *lsy-6* were not regulated by endogenous levels of miRNA (Didiano & Hobert 2006). A possible caveat is that target-containing UTRs were expressed at potentially unphysiological levels, perhaps to levels beyond the capacity of the endogenous miRNA to regulate, through the use of multicopy transgenes. Whether this provides a firm basis to doubt the significance of the many evolutionarily conserved seed matches is open to debate.

Accumulating evidence suggests that miRNA-target regulation can be context dependent. For example, (*a*) zebrafish *miR-430* directly contributes to the repression of residual maternal *nanos1* in somatic cells (Mishima et al. 2006). Although both *miR-430* and *nanos1* are also present in primordial germ cells, *nanos1* can overcome the regulation conferred by the miRNA and is robustly expressed. (*b*) In mammalian hepatocarcinoma cells, stress-induced derepression of the *miR-122* target CAT-1 depends on the presence of a binding site for the HuR protein in the UTR, which is distinct from the miRNA sites (Bhattacharyya et al. 2006). Derepression did not occur if the HuR sites were removed in a reporter construct, implying context dependence. (*c*) Didiano & Hobert (2006) have reported that a minimal *lsy-6* target site can be functional when embedded in the 3′UTR of its endogenous target but not when embedded in a heterologous 3′UTR. They took this as evidence that target site function may depend on the specific 3′UTR context.

However, in this case the conclusion is weakened because the presence of a second site in the endogenous 3′UTR that may contribute to its regulation was not considered.

miRNA EXPRESSION PATTERNS

miRNA-target prediction programs do not take into account whether an miRNA is ever expressed in the same cell as its predicted target. Information about miRNA and target expression patterns can help to assess the likelihood that a predicted miRNA-target relationship is relevant in vivo.

miRNA expression profiles were first generated by small RNA cloning and Northern blotting (Aravin et al. 2003, Berezikov et al. 2006, Chen et al. 2005, Lagos-Quintana et al. 2001, Lau et al. 2001, Lee & Ambros 2001, Ruby et al. 2006). Cloning can identify new miRNAs and provide quantitative information about their expression levels. miRNA microarrays (Baskerville & Bartel 2005, Miska et al. 2004, Nelson et al. 2004, Thomson et al. 2004), quantitative reverse transcriptase–polymerase chain reaction (qRT-PCR) methods, and in situ hybridization have provided further insight into the tissue-specific expression of pri- and mature miRNAs (Aboobaker et al. 2005, Ason et al. 2006, Kloosterman et al. 2006, Wienholds et al. 2005). Some miRNAs with deeply conserved sequences display striking conservation of their spatial expression patterns, providing further evidence for the conservation of their functions. Prominent examples are the muscle-specific *miR-1* and the nervous system–specific *miR-124*, which are conserved in flies, fish, and mammals. Other miRNAs, e.g., *miR-279* and *miR-9a* in *Drosophila*, are expressed in highly dynamic patterns during development of the embryo (Stark et al. 2005). It should be noted that the expression pattern of the pri-miRNA does not necessarily reflect that of the mature miRNA because some miRNAs are regulated post-transcriptionally (Obernosterer et al. 2006, Thomson et al. 2006).

FUNCTIONS IN ANIMAL DEVELOPMENT

Functions Inferred from Bioinformatic Approaches

Transfection of the tissue-specific miRNAs *miR-1* and *miR-124* into HeLa cells shifted the expression profiles toward those of the respective miRNA expressing tissues, affecting 100–200 transcripts (Lim et al. 2005). The interesting insight that emerged from this study was that the transcripts targeted were those that would normally be expressed at a low level in the cell type in which the miRNA is expressed. For example, RNAs targeted by the muscle-specific *miR-1* upon its transfection into nonmuscle cells were RNAs that are normally at low or undetectable levels in muscle. Global analyses of the relations of miRNA expression patterns to those of their conserved targets arrived at a similar conclusion (Farh et al. 2005, Sood et al. 2006, Stark et al. 2005). These studies provided evidence that (*a*) the average miRNA has target sites in hundreds of genes; (*b*) few predicted targets contained multiple conserved binding sites for a single miRNA, suggesting that the large-magnitude regulation conferred by multiple sites for a single miRNA is exceptional; (*c*) abundant mRNAs are under selective pressure to avoid regulation by coexpressed miRNAs; and (*d*) genes involved in basic cellular processes have been selected for loss of sites and for short 3′UTRs to limit the occurrence of potentially detrimental miRNA sites, whereas genes involved in developmental processes show evidence of enrichment for miRNA sites.

An additional observation made primarily in *Drosophila* is that many miRNAs and their targets appear to be expressed in a largely nonoverlapping manner, either temporally or spatially (Stark et al. 2005) (**Figure 3**). In the latter case, targets are typically present in domains adjacent to the miRNA-expressing tissue. Expression profiling of mammalian cells indicated that many conserved targets are present in the tissue expressing the miRNA,

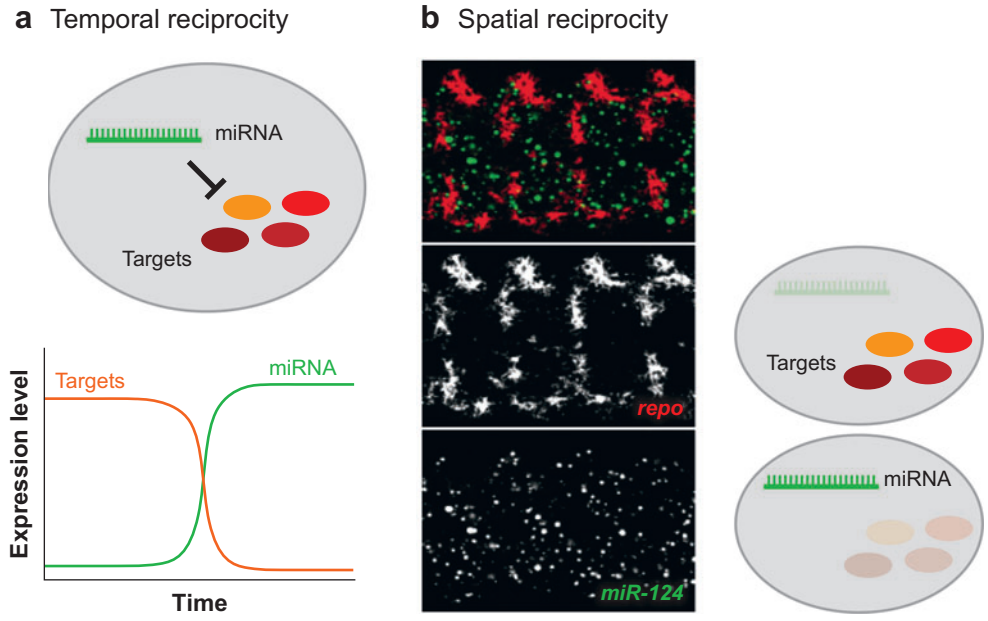


Figure 3

Mutually exclusive expression of miRNAs and their targets. (*a*) miRNAs and their targets can be expressed in a temporally reciprocal manner: Target transcription may be turned off, concurrent with the turning on of miRNA expression. miRNA expression may also actively contribute to target repression, without an underlying change in target transcription. (*b*) In the case of spatial reciprocity, the miRNA and its targets, e.g., *miR-124* (*green*) and its target *repo* (*red*) in the *Drosophila* central nervous system, are often expressed in adjacent domains (from Stark et al. 2005).

albeit at significantly lower levels than in most other tissues (Farh et al. 2005, Sood et al. 2006). In *Drosophila* this possibility is not excluded, although in situ hybridization failed to detect the targets in the miRNA-expressing cell. This probably reflects differences in sensitivity of the methodology because selection for miRNA sites would not emerge if the targets were not present at some level. In some cases, the apparent reciprocity of miRNAs and targets may reflect the action of the miRNA on target-RNA levels. This can be seen when the target-RNA level changes in an miRNA mutant (e.g. Teleman et al. 2006). In other cases, comparison of primary transcript and mature target-RNA patterns has suggested no effect of the miRNA (Stark et al. 2005).

On the basis of these studies, it was suggested that some miRNAs may help maintain and define cell types by dampening the expression of unwanted transcripts, whether from preexisting mRNAs during developmental transitions or simply from leaky transcription. As a cell changes fate, transcription of a target may be turned off while miRNA transcription is being turned on, producing what appear as mutually exclusive patterns (**Figure 3**). Thus, transcriptional regulation is primary, and miRNA regulation is the second level that reinforces the transition as the targets are on their way out (or present only at low levels as leaky transcripts). In this way, miRNAs would provide a fail-safe mechanism to ensure accuracy and confer robustness to underlying gene expression programs.

Although the trend of anticorrelation between miRNA and target expression is common enough to emerge from a global computational study of miRNA and target expression patterns with high statistical

Robustness: resistance of a biological system to perturbation; in genetic terms, invariance of phenotype in response to changing conditions or altered gene dosage

significance, it does not follow that all miRNA-target relationships will be of this type. Further in-depth analysis of miRNAs and their targets as well as of miRNA mutants will be needed to reveal how prevalent this mode of regulation is. Mutants lacking miRNAs for this class of regulation would be expected to suffer subtle changes in the expression of many targets. We consider it likely that the accumulated consequences of many small misregulations may lead to relatively mild and difficult-to-characterize defects in such mutants rather than to obvious phenotypes.

DEPLETION OF ALL miRNAs

An approach to examine the spectrum of possible miRNA functions in development is to eliminate the production of all miRNAs. Dicer mutants have proven informative in the study of the miRNAs expressed earliest during development. Moreover, conditional inactivation of Dicer in specific mouse tissues has shed some light on the functions of some differentially expressed miRNAs. This approach is useful in asking if any miRNA might be involved in a particular biological process.

Animals that are unable to produce mature miRNAs do not survive or reproduce (Bernstein et al. 2003, Forstemann et al. 2005, Ketting et al. 2001, Wienholds et al. 2003). However, the degree to which development is impaired varies. In *C. elegans*, *dicer-1* mutants display defects in germ-line development and a burst vulva phenotype reminiscent of the *let-7* mutant (Ketting et al. 2001, Knight & Bass 2001). Simultaneous removal of maternal and zygotic *dicer-1* is embryonic lethal, suggesting an essential role for miRNAs during embryogenesis (Grishok et al. 2001). In *Drosophila*, depletion of Loquacious, the partner of Dicer-1, causes female sterility (Forstemann et al. 2005), and *dicer-1*-mutant germ-line stem cells display cell division defects (Hatfield et al. 2005). Unfortunately, the full phenotype of *dicer-1*-mutant flies has not been described, but mutant eyes are small, with disorganized ommatidial arrays and missing interommatidial bristles (Y.S. Lee et al. 2004). Unlike their single mutants, *ago1*, *dicer-1* double mutants exhibit strong segmentation defects (Meyer et al. 2006), suggesting that miRNAs may regulate patterning of the *Drosophila* embryo.

miRNAs have also been implicated in memory formation in *Drosophila*. The putative RNA helicase Armitage, which is involved in the maturation of RISC (Tomari et al. 2004), is colocalized at synapses with calcium/calmodulin-dependent protein kinase II (CaMKII), a kinase required for memory (Ashraf et al. 2006). Upon neural stimulation in olfactory-avoidance learning, Armitage is degraded, and CaMKII translation increases, leading to the establishment of a stable memory. The CaMKII 3′UTR contains some miRNA-binding sites and is required for this regulation. In *armitage* and *dicer-2* mutants, CaMKII expression is significantly higher. However, because *Drosophila dicer-2* has been implicated in the biogenesis of siRNAs rather than of miRNAs (Y.S. Lee et al. 2004), it remains unclear if the canonical miRNA pathway will have a role in the establishment of long-term memory.

dicer-1-mutant zebrafish die by 2–3 weeks of age, without obvious defects except for a general growth arrest (Wienholds et al. 2003). miRNAs do not have a function in the zebrafish germ line, so the germ line can be transferred from zygotic mutant fish into a normal host. When mated to mutant males, the resulting females can produce eggs lacking maternal and zygotic *dicer-1* (Giraldez et al. 2005). Although these embryos cannot produce any mature miRNAs, their early development is not severely perturbed. Abnormalities have been seen in morphogenetic processes during gastrulation, somitogenesis, and heart and brain development, ultimately leading to late-embryonic lethality. These comparatively mild defects contrast sharply with the case of *Dicer*-mutant mice, which arrest development and die during gastrulation, before axis formation (Bernstein et al. 2003).

In studies using conditional mouse *Dicer*, embryonic stem cells were impaired in their ability to proliferate (Murchison et al. 2005), and those selected for survival failed to differentiate (Kanellopoulou et al. 2005). In contrast to the zebrafish situation, recent reports indicate an essential role for dicer-dependent miRNA biogenesis in the oocyte to support normal development of the mouse embryo (Murchison et al. 2007, Tang et al. 2007). Other *Dicer*-depleted tissues were able to differentiate and pattern properly, yet the morphogenesis of hair follicles, the lung epithelium, and limbs was perturbed (Andl et al. 2006, Harfe et al. 2005, Harris et al. 2006, Yi et al. 2006), and apoptosis was induced. Specific deletion of *Dicer* in the mouse thymus early in T cell development reduced the proliferation and survival of αβ T cells, but the CD4/CD8 lineage choice was not affected (Cobb et al. 2005). These studies suggest that some miRNAs play important roles in these tissues, but leave open the question of what aspect of their biology is being regulated. The identification of which miRNAs act and on which targets will be needed for a deeper understanding of their roles in development.

These approaches suggest that miRNAs are likely required in a broad range of biological processes during animal development. Surprisingly, at least in vertebrates, basic patterning events and differentiation do not seem to rely heavily on miRNA-mediated regulation, whereas physiological processes may do so.

SPECIFIC miRNA FUNCTIONS

If one considers the abundance of miRNA genes, it may seem surprising that so few miRNA mutants have been recovered in the multitude of genetic screens done in *Drosophila* or *C. elegans*. Why have all these miRNA genes been missed?

An obvious reason might be the small size of miRNAs. In most cases, the seven nucleotides that constitute the seed must be affected to eliminate miRNA function, making miRNAs difficult-to-hit targets in chemical mutagenesis. However, this explanation does not hold true for screens based on transposon insertion for loss- or gain-of-function. Aside from the trivial possibility that miRNA genes may have been overlooked historically because they lacked protein-coding regions, redundancy between miRNAs that share the same seed sequence and that are coexpressed, may mask the effects of loss of single miRNAs (Abbott et al. 2005). Another possibility is that many miRNA mutants show subtle defects or low-penetrance defects that may be difficult to identify in high-throughput genetic screens.

To address these issues, a large-scale mutagenesis project was undertaken to identify mutants for all miRNAs in *C. elegans* (E.A. Miska, E. Alvarez-Saavedra, A.L. Abbott, N.P. Lau, A. Helmann, S. McGonagle, D.P. Bartel, V. Ambros & H.R. Horvitz, personal communication). Relatively few miRNA genes individually generate robust abnormal phenotypes. In some cases, this can be attributed to overlapping functions; multiple mutants that remove related miRNAs reveal stronger defects than do the single mutants. However, some miRNA multiple mutants still produce few or not discernible defects. This group may reflect those miRNAs that confer precision to developmental processes rather than making developmental decisions.

Misexpression and Overexpression

When an miRNA is misexpressed, it has the potential to regulate many targets that it might never encounter in its endogenous expression domain (Farh et al. 2005, Lim et al. 2005, Sood et al. 2006, Stark et al. 2005). Shutting down such targets can be expected to cause strong phenotypes. But because the miRNA and target may not normally be significantly coexpressed, it is difficult to extrapolate from the misexpression effect to the normal function of the miRNA. miRNA misexpression can produce intriguing defects that have limited relevance to what was learned from mutants removing the miRNA (Teleman

et al. 2006), although both defects may result from misregulation of the same target in different contexts. Our unpublished observations suggest that this may often be the case (N. Bushati, J. Karres, J. Varghese & S.M. Cohen). We suggest that reciprocity of outcome should be considered an essential criterion in assessing miRNA function.

In *Drosophila*, two families of Notch target genes, the Enhancer of split-Complex and the Bearded-Complex genes, contain conserved motifs in their 3′UTRs complementary to the seed sequences of a family of related miRNAs (Lai 2002, Stark et al. 2003). Some of these miRNAs, when misexpressed, can induce phenotypes reminiscent of loss of Notch pathway function (Stark et al. 2003, Lai et al. 2005). Similarly, misexpression of *miR-iab-5p* can repress Ubx and induce a homeotic phenotype (Ronshaugen et al. 2005). But it remains to be determined if mutants removing these miRNAs impact Notch signaling or Ubx function in vivo. However tantalizing the misexpression results, we consider studies based solely on miRNA misexpression to have limited predictive value. In some cases, they may help to identify the correct target gene (e.g., Teleman et al. 2006), but this cannot be taken for granted. Below we outline several examples in which misexpression and loss-of-function approaches have provided consistent results, giving us confidence in their insights into miRNA functions.

When overexpressed, the pancreatic islet–specific *miR-375* inhibits glucose-induced insulin secretion. This can be mimicked by knockdown of its target, *myotrophin* (Poy et al. 2004). Importantly, depletion of *miR-375* increases *myotrophin* levels and enhances glucose-stimulated insulin secretion, indicating that *miR-375* is an inhibitor of glucose-stimulated insulin secretion. Using a similar approach in cultured hippocampal neurons, Schratt et al. (2006) found that *miR-134* regulates dendritic spine size by inhibiting translation of *Limk1*.

In hematopoietic stem cells, enforced expression of the B cell–specific miRNA *miR-181* stimulates their differentiation to B-lineage cells (Chen et al. 2004), but it remains to be determined if loss of *miR-181* impedes differentiation. A very recent report (Li et al. 2007) provides evidence that *miR-181* overexpression and depletion produce opposing effects on antigen sensitivity in T cells.

miR-1 and *miR-133* are absent from undifferentiated myoblasts and strongly upregulated upon differentiation into myotubes (Boutz et al. 2007). Tissue-specific overexpression of *miR-1* in the developing mouse heart induces premature differentiation of myocytes (Zhao et al. 2005). *miR-1* misexpression can accelerate myoblast differentiation by the targeting of histone deacetylase 4, a repressor of muscle differentiation, whereas depletion of *miR-1* impedes differentiation, indicated by a decrease in myogenic markers (Chen et al. 2006). *miR-1* and *miR-133* form one genomic cluster and are coexpressed in heart and skeletal muscle. *miR-133* promotes myoblast proliferation by targeting serum response factor (Chen et al. 2006), but in a different experimental setup myoblast differentiation was not observed (Boutz et al. 2007). *miR-133* was required during differentiation to downregulate nPTB protein, a repressor of alternative splicing, resulting in splicing of a group of silenced exons in mature myotubes. *miR-1* and *miR-133* are seemingly required to define and maintain the properties of differentiated muscle cells, in agreement with the conclusions drawn from *miR-1* misexpression (Lim et al. 2005). In a very recent report, a mouse lacking one of the *miR-1* genes had defects in heart development and function, providing robust genetic evidence for the importance of this miRNA family in cardiogenesis (Zhao et al. 2007).

In *C. elegans*, *let-60/RAS* contains several putative *miR-84/let-7*-binding sites and can be downregulated by *miR-84* and *let-7*. *let-7* mutants display a burst vulva phenotype that can be suppressed by RNAi of *let-60/RAS* (Johnson et al. 2005), suggesting that excess *let-60/RAS* activity contributes to the defect. Overexpression of *miR-84*, a member of the

let-7 family, can rescue the multivulva phenotype caused by *let-60/RAS* gain-of-function alleles (Johnson et al. 2005). These results implicate *let-7* as an important in vivo regulator of *let-60/RAS*, but final conclusions about the relevance of *miR-84* await mutant analysis.

MODES OF miRNA FUNCTION: miRNA MUTANTS

How essential are the contributions of specific miRNAs to developmental programs? Because many miRNAs are highly conserved, their functions must be advantageous, even if they appear vanishingly subtle to us in the lab. As for other regulatory molecules, individual miRNAs can be expected to influence development to different degrees. Whether a general mode of action is attributable to most miRNAs is still a subject of discussion (Bartel & Chen 2004, Hornstein & Shomron 2006, Stark et al. 2005). Most likely one size won't fit all. Comparison of the insights gained from experimental analysis of miRNAs and computational studies suggests that there may be several conceptually distinct modes in which miRNAs act. Here we present evidence for several such modes of action and discuss what sort of evidence would be required to support the case for others.

miRNAs Acting as Developmental Switches

When deleted, miRNAs required for developmental decision-making can be expected to cause strong phenotypes. The first miRNAs identified, *C. elegans lin-4* and *let-7*, are of this kind. Indeed, it could hardly be otherwise, given that they were found in forward genetic screens for mutant phenotypes (Chalfie et al. 1981, Lee et al. 1993, Reinhart et al. 2000, Wightman et al. 1993). Without *lin-4*, the animal is unable to make the transition from the first to the second larval stage owing to a differentiation defect. This developmentally retarded phenotype is caused by a failure to posttranscriptionally downregulate the *lin-14* gene, which harbors seven *lin-4* target sites in its 3′UTR (**Figure 2**) (Lee et al. 1993, Wightman et al. 1993). In the adult worm, *lin-4*-mediated downregulation of *lin-14* has been implicated in life-span regulation (Boehm & Slack 2005).

Depletion of *let-7*, a highly conserved miRNA, also leads to a heterochronic defect, failure of larval-to-adult transition (Reinhart et al. 2000). Several genes are direct targets of *let-7* during this transition: *lin-41*, *hbl-1*, and *daf-12* and the forkhead transcription factor *pha-4* (Abrahante et al. 2003, Grosshans et al. 2005, Lin et al. 2003, Slack et al. 2000). *let-7* is a member of a family of miRNAs including *miR-48*, *miR-84*, and *miR-241*, which function in a cooperative manner to control the earlier L2-to-L3 transition, probably by targeting *hbl-1* (Abbott et al. 2005, Li et al. 2005). Additionally, in adult worms *miR-48* and *miR-84* are coordinately required to bring the larval molting cycle to an end (Abbott et al. 2005). These *C. elegans* miRNAs act as key regulators of developmental timing. The temporal regulation of *let-7* during development and the presence of a target site in *lin-41* (**Figure 2**) are conserved (Pasquinelli et al. 2000), raising the possibility that this regulatory relationship controls some aspect of developmental timing in other organisms.

Loss of *lsy-6* in *C. elegans* induces a cell-fate switch: Instead of having two morphologically distinct gustatory neurons on left and right sides, known as ASEL and ASER, respectively, two ASER neurons are specified (Johnston 2003). Asymmetry is lost owing to the failure of *lsy-6* to repress cog-1 expression in ASEL. die-1, present only in ASEL, is required for *lsy-6* expression (Chang et al. 2004). die-1 in turn is downregulated in ASER by *miR-273*, which is expressed predominantly in the ASER because it is activated there by the *lsy-6* target cog-1 (Johnston et al. 2005). Therefore, the cell-fate decision and stabilization of the gustatory ASE neurons are controlled via a double negative-feedback loop in which two transcription factors use miRNAs to repress each other (**Figure 4**).

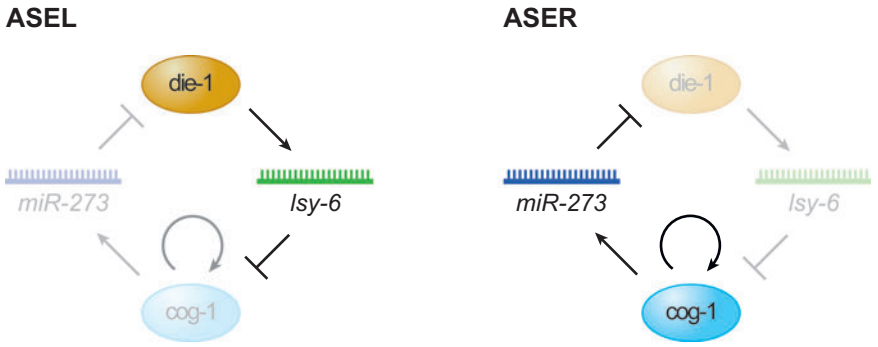


Figure 4

miRNAs acting in a double negative-feedback loop control neuronal asymmetry in *C. elegans*. In *C. elegans*, *lsy-6*, induced by the transcription factor die-1, determines adoption of the ASEL fate by repressing cog-1 expression. In the ASER neuron, cog-1 positively regulates its own expression and induces *miR-273*, which suppresses die-1 and, consequently, *lsy-6*.

miR-181 expression is upregulated during terminal differentiation of myoblasts, and the depletion of *miR-181* impedes their differentiation, partly owing to the upregulation of one of its targets, Hox-A11, an inhibitor of differentiation (Naguibneva et al. 2006). However, *miR-181* overexpression is not sufficient to induce differentiation. In adult muscle, *miR-181* is expressed at very low levels, suggesting that it is involved in muscle establishment, not maintenance. Whether or not this miRNA acts as a developmental switch in vivo awaits analysis of a knockout mouse.

Fine-Tuning Developmental Programs: Robustness

On the basis of the computational studies described above, many miRNAs are expected to regulate a large set of targets that may be expressed at low levels in the miRNA-expressing cells. This may provide a second layer of regulation to reinforce transcriptional controls. miRNA mutants of this class may be expected to have subtle phenotypes and, perhaps, phenotypes that are difficult to study, owing to the heterogeneous nature of their targets.

Members of the *miR-430* multigene family are expressed at high levels at the onset of zygotic transcription in zebrafish embryos. Supplying miRNA-depleted embryos with *miR-430* by injection rescues the brain morphogenesis defects seen in these embryos (Giraldez et al. 2005). Comparing expression profiles of miRNA-depleted embryos with those of embryos resupplied with *miR-430* revealed that several hundred transcripts, likely direct *miR-430* targets based on miRNA seed matches, are misregulated in the absence of *miR-430* (Giraldez et al. 2006). Approximately 40% are maternally deposited RNAs, suggesting that *miR-430* is required to accelerate the clearance of preexisting maternal mRNAs in the embryo when zygotic transcription starts (**Figure 5**). This study provides experimental evidence that an endogenous miRNA targets hundreds of transcripts in vivo and that their regulation by the miRNA serves to ensure robustness of the developmental program. A similar situation exists in *Drosophila*, in which the six miRNAs encoded of the *miR-309* miRNA gene cluster contribute to the clearance of maternally encoded mRNAs at the onset of zygotic transcription (N. Bushati, A. Stark, J. Brennecke & S.M. Cohen, unpublished observations).

Surprisingly, most *Drosophila* larvae lacking the highly conserved, muscle-specific miRNA *miR-1* develop apparently normal, functional muscles (Sokol & Ambros 2005). Twenty percent of the mutant embryos do not hatch, but the lethality is not due to obvious

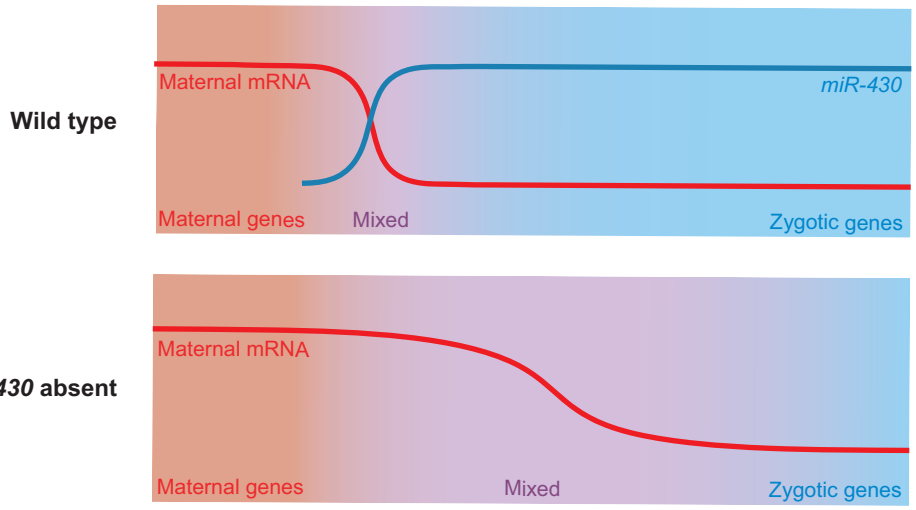


Figure 5

Fine-tuning embryonic development. In early zebrafish embryogenesis, the miRNA *miR-430* regulates the transition from maternal to zygotic mRNA transcription by targeting maternal mRNAs for degradation. In the absence of *miR-430*, maternal mRNA expression overlaps with zygotic gene expression and interferes with morphogenesis.

muscle defects (Brennecke et al. 2005a). The remaining 80% die with massively disrupted muscles when the mutants begin the rapid phase of larval growth upon feeding. However, if the mutant larvae are fed on sugar, conditions under which they live but do not grow, muscle function is normal, and the animals survive. It is unclear which target(s) is misregulated in the *miR-1* mutant or how growth causes muscle degeneration. Perhaps this defect is the consequence of many subtle errors earlier during development, when *miR-1* may be required to confer robustness to the identity and/or physiology of muscle cells. But a specific role during muscle growth is also possible.

Zebrafish *miR-214* is expressed in mesoderm during early somitogenesis and throughout embryogenesis. Antisense oligonucleotide–mediated depletion of *miR-214* reduced Hedgehog signaling in the presomitic mesodermal cells adjacent to the notochord and therefore reduced the number of slow-muscle cells (Flynt et al. 2007). During normal development, these cells experience high Hedgehog levels, whereas their more lateral neighbors are exposed to less Hedgehog. Su(fu), a negative regulator of Hedgehog signaling, is upregulated in *miR-214*-depleted embryos, probably owing to the lack of regulation by *miR-214*. Su(fu) participates in Hedgehog signaling by retaining both the activator and repressor forms of the Gli transcription factors in the cytoplasm. miRNA-mediated downregulation of Su(fu) may maximize the response to different levels of Hedgehog signaling.

Antisense-mediated silencing of the abundant liver-specific *miR-122* in the mouse leads to significant upregulation of >100 mRNAs containing *miR-122* seed matches in their 3′UTRs. These mice are healthy and show substantially reduced levels of circulating cholesterol and triglycerides. *miR-122* therefore has a function in hepatic lipid metabolism and may be required to fine-tune this process (Esau et al. 2006, Krutzfeldt et al. 2005). Studies such as this suggest the therapeutic potential of miRNA depletion.

In mouse embryos, an *miR-196* gene is located in all four mammalian HOX clusters. *miR-196* is expressed in the hindlimb but not

Antisense oligonucleotide–mediated depletion: an approach to eliminate miRNA function by the injection/transfection of complementary modified RNA, which binds and sequesters the miRNA and so prevents it from binding its targets

in the forelimb. Its targets, *Hoxa7* and *Hoxb8*, are induced by retinoic acid in the vertebrate forelimb but not in the hindlimb. *miR-196* cleaves its target HOXB8 (Yekta et al. 2004) and may block translation of *Hoxa7* when these RNAs are present in the hindlimb (Hornstein et al. 2005). Thus, *miR-196* suppresses unwanted transcripts in a domain in which they are normally transcriptionally repressed, supporting a primary transcriptional level of regulation. This provides an in vivo example of how apparently spatially reciprocal expression domains of miRNAs and targets can reflect an important regulatory relationship.

A similar relationship exists between *miR-278* and its target, *expanded*, in adipose tissue in *Drosophila*. *miR-278*-mutant flies are lean owing to insulin insensitivity in the adipose tissue, in which *miR-278* is most strongly expressed (Teleman et al. 2006). A similar defect results in mice with adipose-tissue-specific knockout of insulin responsiveness (Teleman et al. 2006). *expanded* mRNA and protein levels are normally very low in adipose tissue but increase considerably in the *miR-278* mutant. Genetic analysis has shown that *expanded* overexpression causes the lean phenotype. In this case, the miRNA contributes strongly to downregulation of its target's transcript levels, thereby helping to shape its expression domain.

Proliferation and Apoptosis

The *Drosophila* miRNA *bantam* was identified in a screen for regulators of tissue growth in which its overexpression induced overgrowth. *bantam* mutants are small and die as early pupae (Brennecke et al. 2003). To allow net tissue growth, apoptosis must be overcome. The 3′UTR of the proapoptotic gene *bid* contains five *bantam* target sites, and part of the ability of *bantam* to promote growth can be explained by this regulatory relationship. But this is not sufficient to explain the mutant phenotype because blocking apoptosis is not enough to cause growth. The targets involved in the generation of a positive-growth-regulatory output by *bantam*, e.g., by inhibition of a negative growth regulator, have not yet been identified. Recently, the Hippo signaling pathway, which coordinately controls cell proliferation and apoptosis, has been shown to regulate *bantam* expression (Nolo et al. 2006, Thompson & Cohen 2006).

Feedback Loops

Several miRNAs are part of regulatory feedback loops. The roles they play can be switch-like, as described above for *C. elegans lsy-6* and *miR-273*, which are thought to act in a double negative-feedback loop to specify left-right asymmetry of gustatory neurons (Chang et al. 2004, Johnston & Hobert 2003) (**Figure 4**).

In other cases, the role of the miRNA in the feedback loop can be quite subtle, reinforcing or stabilizing decisions made by other factors. In the *Drosophila* eye, *miR-7* reinforces a developmental decision via a reciprocal negative-feedback loop (Li & Carthew 2005) (**Figure 6**). During photoreceptor differentiation, a transient EGF signal leads to degradation of the protein Yan and, by the release of Yan-mediated repression of *miR-7* transcription, to elevated *miR-7* expression. Subsequently, *miR-7* represses Yan. A stable change in cell fate from progenitor to photoreceptor is thereby achieved, and the differentiated fate is maintained by the presence of the miRNA. The overall eye development of *miR-7*-mutant flies appears normal. Only when the system was sensitized by the introduction of a *Yan* allele, which does not respond normally to epidermal growth factor receptor (EGFR), were defects observed, demonstrating that the miRNA does not control the feedback loop but reinforces a decision made in response to an extracellular signal. This is another example of mutually exclusive expression of an miRNA and its target because *Yan* is downregulated when *miR-7* expression is induced.

A similar negative-feedback loop operates during granulocytic differentiation (Fazi et al.

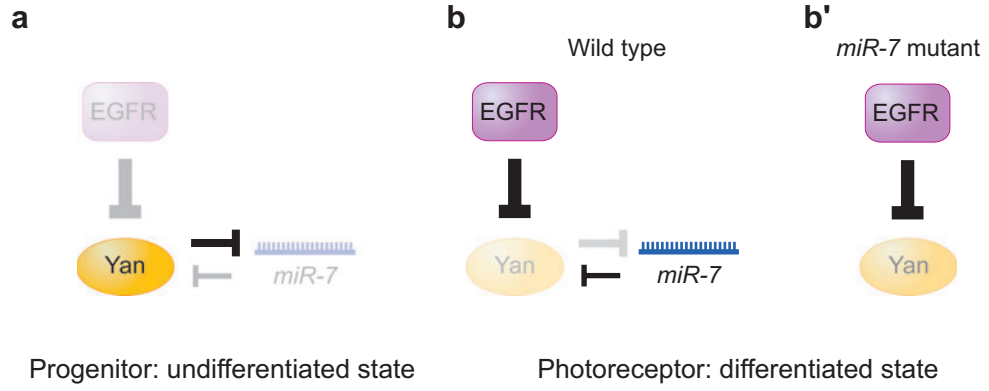


Figure 6

miR-7 reinforces photoreceptor differentiation. (*a*) Yan represses *mir-7* transcription in the progenitor. (*b*,*b*′) During photoreceptor differentiation, epidermal growth factor receptor (EGFR) signaling induces Yan degradation, which leads to the release of *miR-7* repression. (*b*) *miR-7* then reinforces Yan downregulation. (*b*′) In the absence of *mir-7*, Yan is still repressed by EGFR, and therefore photoreceptor differentiation occurs normally.

2005): In undifferentiated myeloid precursors, the transcription factor nuclear factor I-A (NFI-A) keeps levels of *miR-223* expression low. When ectopically expressed, *miR-223* enhances the differentiation of myeloid precursors into granulocytes. Upon stimulation with retinoid acid, which induces differentiation, CCAAT/enhancer binding protein α (C/EBPα) replaces NFI-A and induces high levels of *miR-223* transcription. *miR-223* in turn represses NFI-A translation via a binding site in its 3′UTR, thus ensuring its own expression. Depletion of *miR-223* followed by stimulation with retinoic acid leads to reduced expression of a granulocytic differentiation marker (Fazi et al. 2005).

In *C. elegans*, *miR-61* is directly transcriptionally activated in secondary vulval precursor cells by LIN-12/Notch (Yoo & Greenwald 2005). *miR-61* represses Vav-1, which would otherwise repress LIN-12, thereby inducing a positive-feedback loop, reinforcing LIN-12 activation and specification of secondary vulval cell fate. In contrast to *miR-7*, which ensures robustness of a decision made by EGFR signaling, *miR-61* may play a major role in cell-fate specification.

Thresholding

Flies lacking *miR-9a*, a conserved nervous system–specific miRNA, produce extra sense organs (Li et al. 2006). This defect is due to failure to repress the *miR-9a* target *senseless*. Indeed, the *miR-9a*-mutant phenotype is suppressed by the removal of one copy of the *senseless* gene, which limits the extent to which *senseless* can be overexpressed. An interesting aspect of this mutant is that loss of *senseless* regulation causes a sporadic defect, with up to 40% of animals affected. Why is that so? During selection of the sensory organ precursor (SOP) from cells of a proneural cluster, *senseless* expression is turned on by proneural proteins and feedbacks positively to reinforce proneural gene expression. The resulting high level of proneural activity induces a signal that the SOP uses to repress SOP fate in surrounding cells. Early in the process, *miR-9a* is broadly expressed in the neuroectoderm and limits proneural gene expression by keeping *senseless* expression low. During this phase, *miR-9a* sets a threshold that *senseless* expression has to overcome to induce SOP fate, thereby selecting only the cell with the highest transcriptional peak. In animals lacking *miR-9a*, *senseless* levels sporadically exceed

Sensory organ precursor (SOP): the cell that gives rise to the sensory organ

Proneural cluster: a group of cells that are competent to become neuronal precursors

the threshold, leading to ectopic SOP formation. It should be noted that the miRNA is not part of a feedback loop but limits the level at which a transcriptional feedback system can be activated. Whether other miRNAs have comparable roles in thresholding (Cohen et al. 2006) remains to be seen.

An attempt to investigate *miR-9a* and other miRNA functions has been made by the injection of complementary 2′-*O*-methyl oligoribonucleotides into fly embryos (Leaman et al. 2005). However, where comparisons have been possible the results obtained by antisense-mediated depletion have differed significantly from those observed in the corresponding miRNA-null mutants. Depletion of *miR-1*, *miR-9a*, *miR-279*, *miR-6*, *miR-286*, and several miRNAs of the *miR-310* family was reported to cause a variety of severe embryonic defects. Comparable defects were not seen in embryos mutant for these miRNAs (Li et al. 2006, Sokol & Ambros 2005; N. Bushati & S.M. Cohen, unpublished data). It is unclear how to explain this discrepancy.

TARGET TYPES: SWITCHING OFF VERSUS TUNING TARGET EXPRESSION LEVELS

A few years ago, Bartel & Chen (2004) proposed the terms switch target and tuning target to describe two classes of relationships between miRNAs and targets (**Figure 7**). These terms were coined to describe what the miRNA does to its target and should not be confused with what regulation of the target does to the cell or organism.

A switch target is one for which the miRNA reduces target expression to a level below which it has any meaningful activity in the cell, effectively switching it off. Most of the targets of *miR-1* or *miR-124*, which are normally expressed at low levels in the miRNA-expressing cells, would likely qualify as switch targets because the job of the miRNA appears to be to reduce their potential leaky expression to inconsequential levels. In these and other examples discussed above, such as that of *miR-196*, the miRNAs do not appear to act as developmental switches. However, important miRNA targets, including those of the heterochronic miRNAs *lin-4* and *let-7*, may act as developmental switches (Lee et al. 1993, Wightman et al. 1993). Although these are also likely to be switch targets, it remains to be demonstrated that the residual expression after miRNA-mediated repression is without function.

The notion of a tuning target reflects the role of the miRNA in setting a defined level of target expression while being coexpressed with the miRNA. Tuning targets can make use of miRNAs to smooth out fluctuations in their expression or, by altering miRNA level, to ensure that target levels are suitable for the prevailing conditions. The critical distinction between the switch and tuning modes lies in whether the residual level of target expression is required, i.e., has a specific function, in that cell or if the job of the miRNA is merely to reduce target expression to an inconsequential level. One could make this distinction genetically, for example, by selectively eliminating the target gene in the miRNA cells, using a conditional knockout approach. To date no such analysis has been presented, but we suggest this as a rigorous standard.

miRNAs IN HUMAN DISEASE

One of the human diseases in which miRNAs have been implicated is the neuropsychiatric disorder Tourette's syndrome (TS) (Abelson et al. 2005). The 3′UTR of *SLITRK1* contains a *miR-189*-binding site, which is mutated in some TS patients. This polymorphism replaces a GU base pair with AU pairing, leading to stronger regulation by the miRNA. In situ hybridization of *SLITRK1* mRNA and *miR-189* revealed coexpression in neuroanatomical circuits most commonly implicated in TS. Obviously this mutation is only one out of many rare mutations leading to this complex disease, but it demonstrates how an miRNA can be involved in the establishment of a disease phenotype.

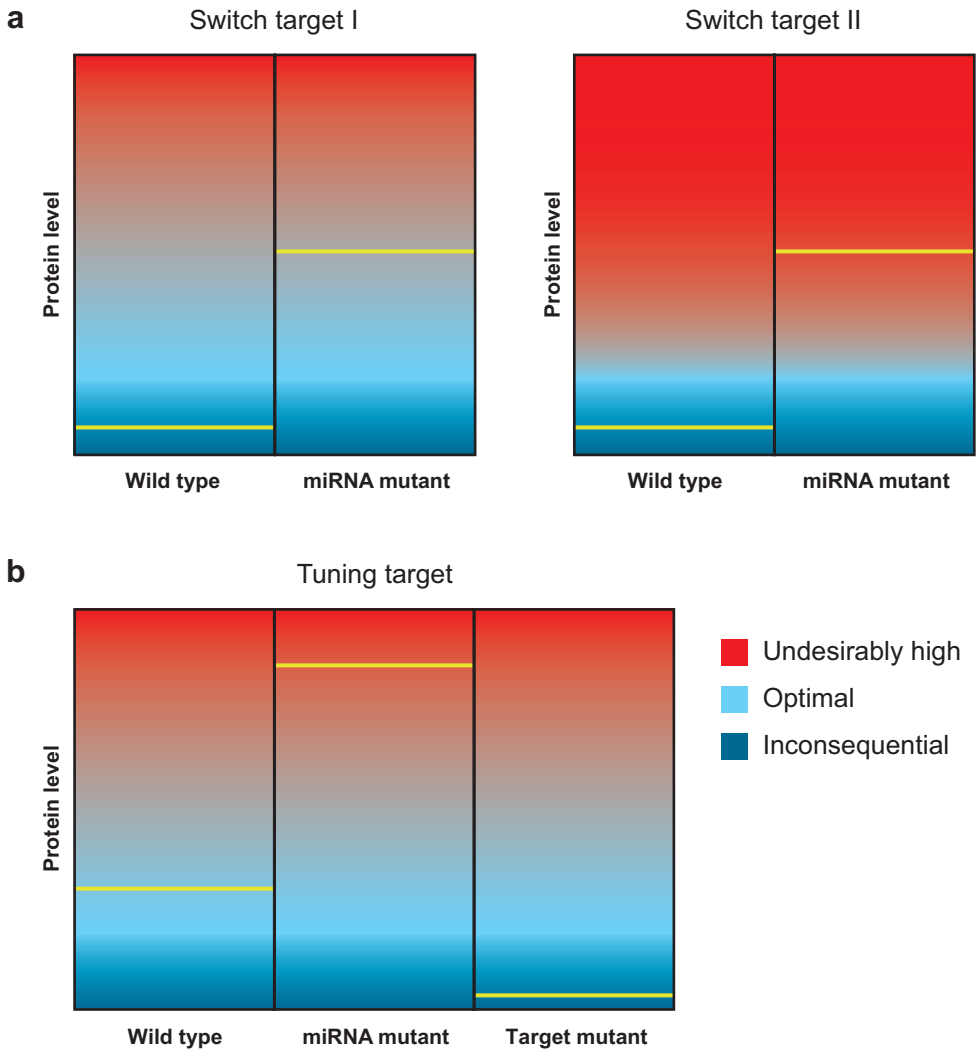


Figure 7

Switch and tuning targets revisited. Red, light blue, and dark blue indicate areas in which protein levels are undesirably high, optimal, and inconsequential, respectively. (*a*) Switch targets are downregulated to inconsequential levels. In the absence of the miRNA, these targets are expressed at levels that may be harmless (switch target I) or at detrimental levels (switch target II). The consequences can be subtle or severe, depending on the type of target gene and its relative level of misexpression. (*b*) miRNAs adjust the expression of tuning targets to optimal levels. In the absence of the miRNA, the tuning target is expressed at a detrimental level. In contrast to switch targets, tuning targets have a function in the miRNA-expressing cell, so reducing them further has a negative effect.

Ninety percent of patients suffering from DiGeorge syndrome lack one copy of the chromosomal region 22q11 (Lindsay 2001). Interestingly, the Drosha partner DGCR8 maps to this region (Denli et al. 2004, Gregory et al. 2004, Han et al. 2004, Landthaler et al. 2004, Shiohama et al. 2003), but it remains to be determined if reduced miRNA levels are an underlying cause of DiGeorge syndrome.

Expansion of the polyglutamine repeats in Ataxin-3 has been linked to neurodegeneration. Depletion of *dicer* in human cells led to a significant enhancement of Ataxin-3-induced toxicity (Bilen et al. 2006). Which miRNAs are involved in this human neurodegenerative disorder remains to be determined.

miRNAs as Oncogenes and Tumor Suppressors

miRNA expression levels are altered in primary human tumors (Calin et al. 2004, Lu et al. 2005). One study reported that global miRNA expression was lower in cancer tissues than in normal tissues (Lu et al. 2005), but another did not find this trend (Volinia et al. 2006). Because many miRNAs are expressed in differentiated cell types, global miRNA levels may reflect the differentiation state of the tissue and perhaps may be attributable to a failure of Drosha processing (Thomson et al. 2006). Intriguingly, significantly differing miRNA profiles can be assigned to various types of tumors, suggesting that miRNA profiling has diagnostic and perhaps prognostic potential (Calin & Croce 2006, Lu et al. 2005). However, for most miRNAs it is unknown whether they actually play an active role in tumor formation.

Loss of miRNAs in cancer tissue may suggest a role of miRNAs as tumor suppressors. The *miR-15a* and *miR-16-1* genes are located in chromosome region 13q14, which is deleted in most cases of chronic lymphocytic leukemia (Calin et al. 2002). These miRNAs target *B cell lymphoma 2* (*Bcl2*), an antiapoptotic gene, suggesting that loss of *miR-15a* and *miR-16-1* in B cells may lead to the inhibition of apoptosis, giving rise to malignancies (Cimmino et al. 2005). However, causality has not been established, and the correlation is not universal. The same miRNAs are overexpressed in a type of pancreatic tumor (Volinia et al. 2006).

miRNAs with oncogenic potential are expressed from the *miR17-92* locus 13q31, which is amplified in some tumors, e.g., tenfold in B cell lymphoma samples (He et al. 2005). Overexpression of this cluster in a mouse model of human B cell lymphoma accelerated *c*-Myc-induced tumorigenesis, and the apoptosis normally seen in *c*-Myc-induced tumors was suppressed (He et al. 2005). *c*-Myc can directly transcriptionally activate the *miR17-92* cluster and the proapoptotic E2F1, which in turn is targeted by two miRNAs of the cluster (O'Donnell et al. 2005). These miRNAs would therefore support a shift from apoptosis toward proliferation by the downregulation of E2F1. Moreover, *c*-Myc-induced activation of the *miR17-92* cluster leads to enhanced tumor angiogenesis in mouse colonocytes, probably via direct, miRNA-mediated downregulation of antiangiogenic proteins (Dews et al. 2006).

Primary human fibroblasts expressing the clustered miRNAs *miR-372* and *miR-373* were able to overcome oncogenic Ras-mediated arrest and, therefore, induced tumorigenesis (Voorhoeve et al. 2006). In part, this effect is mediated by targeting of the tumor suppressor LATS2. *miR-372* and *miR-373* are expressed specifically in testicular germ cell tumors (Voorhoeve et al. 2006).

Interestingly, both miRNA clusters are highly expressed in embryonic stem cells (Suh et al. 2004, Thomson et al. 2004), which suggests that they contribute to tumorigenesis by exerting their normal function at the wrong time and place in the organism. Another miRNA with oncogenic potential is *miR-155*, which is overexpressed in several kinds of B cell lymphomas and can induce preleukemic pre–B cell proliferation in mice when it is overexpressed specifically in B cells (Costinean et al. 2006). For a more in-depth review of miRNAs in cancer, see Calin (2006).

miRNAs and Viruses

Viruses use miRNAs in their effort to control their host cell; reciprocally, host cells use miRNAs to target essential viral functions.

miRNAs have been found in nuclear DNA viruses like the herpesvirus, but to date none have been found in RNA viruses (Cullen 2006; Pfeffer et al. 2004, 2005). Most RNA viruses are restricted to the cytoplasm and are therefore not expected to encode miRNAs because miRNA transcripts need to be processed in the nucleus by Drosha.

The SV40-encoded miRNA *miR-S1* helps keep the infected cell hidden from the immune system. It is expressed late in the viral replication cycle, when it acts to degrade early viral mRNAs encoding T antigen, limiting exposure of the infected cell to cytotoxic T lymphocytes (Sullivan et al. 2005). *miR-LAT* of herpes simplex virus-1 inhibits apoptosis of latently infected neurons by targeting the proapoptotic transforming growth factor-β (TGF-β) and one of its mediators, SMAD3 (Gupta et al. 2006). The miRNA thereby keeps its host cell alive.

Endogenous, cellular miRNAs that target viral RNAs have been reported as well. In one scenario, the cell uses the miRNA to impede viral replication. *miR-32* restricts the replication of the retrovirus PFC-1 in cell culture (Lecellier et al. 2005). In another scenario, the virus takes advantage of an endogenous cellular miRNA. Replication of hepatitis C viral RNA is facilitated by binding of the liver-specific miRNA *miR-122* to the 5′ noncoding region of the viral genome (Jopling et al. 2005).

CONCLUSIONS

miRNAs have captured the attention of the biological sciences community in part because of their novelty and in part because of the excitement of exploring the regulatory potential of a new type of molecule. This review aims to illustrate that the analysis of miRNA functions in vivo has begun to shed light on the types of biological processes that miRNAs regulate. As well, we have begun to understand the variety of ways miRNAs can act in the context of other regulatory mechanisms. It appears that their biological roles in development and disease as well as their modes of action will be diverse.

SUMMARY POINTS

1. A developmental switch is a change in the program of a cell in response to the input of new information. This can be accomplished by a change in transcription or an miRNA-mediated change in posttranscriptional gene expression.
2. Resistance to perturbation is a common property of developmental systems. miRNA-mediated posttranscriptional gene regulation is thought to contribute to robustness, in part through noise reduction.
3. miRNA targets are often expressed at very low levels, possibly at that of noise, in miRNA-expressing cells. In such cases, the job of the miRNA is to keep low-level expression to inconsequential levels.
4. Some miRNAs act as components of regulatory feedback loops. In some cases they are key components. In other cases, they are modulators that ensure robustness or set a threshold for switch activation.
5. Some miRNAs switch off target gene expression, reducing them to inconsequential levels. Failure to regulate targets in this way may have severe consequences or subtle effects, depending on the nature of the targets.

6. Some miRNAs may maintain the level of target gene expression within an optimal window, neither too high nor too low. Here the critical features are that the target has a specific function in the cell at the level set by the miRNA and that expression above or below the set level may be detrimental.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Barry Thompson and Julius Brennecke for their helpful suggestions. Work in the authors' lab is supported by the EU-FP6 project Sirocco LSHG-CT-2006-037900 and by the European Molecular Biology Laboratory.

LITERATURE CITED

Abbott AL, Alvarez-Saavedra E, Miska EA, Lau NC, Bartel DP, et al. 2005. The *let-7* MicroRNA family members *mir-48*, *mir-84*, and *mir-241* function together to regulate developmental timing in *Caenorhabditis elegans*. *Dev. Cell* 9:403–14

Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, et al. 2005. Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* 310:317–20

Aboobaker AA, Tomancak P, Patel N, Rubin GM, Lai EC. 2005. *Drosophila* microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proc. Natl. Acad. Sci. USA* 102:18017–22

Abrahante JE, Daul AL, Li M, Volk ML, Tennessen JM, et al. 2003. The *Caenorhabditis elegans* hunchback-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. *Dev. Cell* 4:625–37

Andl T, Murchison EP, Liu F, Zhang Y, Yunta-Gonzalez M, et al. 2006. The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles. *Curr. Biol.* 16:1041–49

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442:203–7

Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, et al. 2003. The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* 5:337–50

Arteaga-Vazquez M, Caballero-Perez J, Vielle-Calzada JP. 2006. A family of microRNAs present in plants and animals. *Plant Cell* 18:3355–69

Ashraf SI, McLoon AL, Sclarsic SM, Kunes S. 2006. Synaptic protein synthesis associated with memory is regulated by the RISC pathway in *Drosophila*. *Cell* 124:191–205

Ason B, Darnell DK, Wittbrodt B, Berezikov E, Kloosterman WP, et al. 2006. Differences in vertebrate microRNA expression. *Proc. Natl. Acad. Sci. USA* 103:14385–89

Bartel DP, Chen CZ. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.* 5:396–400

Baskerville S, Bartel DP. 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 11:241–47

Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev.* 20:1885–98

Berezikov E, Cuppen E, Plasterk RH. 2006. Approaches to microRNA discovery. *Nat. Genet.* 38(Suppl.):S2–7

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, et al. 2003. Dicer is essential for mouse development. *Nat. Genet.* 35:215–17

Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. 2006. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125:1111–24

Bilen J, Liu N, Burnett BG, Pittman RN, Bonini NM. 2006. MicroRNA pathways modulate polyglutamine-induced neurodegeneration. *Mol. Cell* 24:157–63

Boehm M, Slack F. 2005. A developmental timing microRNA and its target regulate life span in *C. elegans*. *Science* 310:1954–57

Bohnsack MT, Czaplinski K, Gorlich D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10:185–91

Borchert GM, Lanier W, Davidson BL. 2006. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* 13:1097–101

Boutz PL, Chawla G, Stoilov P, Black DL. 2007. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev.* 21:71–84

Bracht J, Hunter S, Eachus R, Weeks P, Pasquinelli AE. 2004. Trans-splicing and polyadenylation of *let-7* microRNA primary transcripts. *RNA* 10:1586–94

Brennecke J, Aravin AA, Stark A, Dus M, Kelis M, et al. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128:1089–103

Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. 2003. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* 113:25–36

Brennecke J, Stark A, Cohen SM. 2005a. Not miR-ly muscular: microRNAs and muscle development. *Genes Dev.* 19:2261–64

Brennecke J, Stark A, Russell RB, Cohen SM. 2005b. Principles of microRNA-target recognition. *PLoS Biol.* 3:e85

Cai X, Hagedorn CH, Cullen BR. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10:1957–66

Calin GA, Croce CM. 2006. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* 6:857–66

Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, et al. 2002. Frequent deletions and down-regulation of micro-RNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA* 99:15524–29

Calin GA, Liu CG, Sevignani C, Ferracin M, Felli N, et al. 2004. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. *Proc. Natl. Acad. Sci. USA* 101:11755–60

Chalfie M, Horvitz HR, Sulston JE. 1981. Mutations that lead to reiterations in the cell lineages of *C. elegans*. *Cell* 24:59–69

Chang S, Johnston RJJ, Frokjaer-Jensen C, Lockery S, Hobert O. 2004. MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. *Nature* 430:785–89

Chen CZ, Li L, Lodish HF, Bartel DP. 2004. MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303:83–86

Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, et al. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.* 38:228–33

Chen K, Rajewsky N. 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.* 38:1452–56

Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8:93–103

Chen PY, Manninga H, Slanchev K, Chien M, Russo JJ, et al. 2005. The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev.* 19:1288–93

Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, et al. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436:740–44

Chu CY, Rana TM. 2006. Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *PLoS Biol.* 4:e210

Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, et al. 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl. Acad. Sci. USA* 102:13944–49

Cobb BS, Nesterova TB, Thompson E, Hertweck A, O'Connor E, et al. 2005. T cell lineage choice and differentiation in the absence of the RNase III enzyme Dicer. *J. Exp. Med.* 201:1367–73

Cohen SM, Brennecke J, Stark A. 2006. Denoising feedback loops by thresholding: a new role for microRNAs. *Genes Dev.* 20:2769–72

Costinean S, Zanesi N, Pekarsky Y, Tili E, Volinia S, et al. 2006. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in Eμ-miR155 transgenic mice. *Proc. Natl. Acad. Sci. USA* 103:7024–29

Cullen BR. 2006. Viruses and microRNAs. *Nat. Genet.* 38(Suppl.):S25–30

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* 432:231–35

Dews M, Homayouni A, Yu D, Murphy D, Sevignani C, et al. 2006. Augmentation of tumor angiogenesis by a Myc-activated microRNA cluster. *Nat. Genet.* 38:1060–65

Didiano D, Hobert O. 2006. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.* 13:849–51

Doench JG, Petersen CP, Sharp PA. 2003. siRNAs can function as miRNAs. *Genes Dev.* 17:438–42

Doench JG, Sharp PA. 2004. Specificity of microRNA target selection in translational repression. *Genes Dev.* 18:504–11

Du T, Zamore PD. 2005. microPrimer: the biogenesis and function of microRNA. *Development* 132:4645–52

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. 2003. MicroRNA targets in *Drosophila*. *Genome Biol.* 5:R1

Esau C, Davis S, Murray SF, Yu XX, Pandey SK, et al. 2006. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab.* 3:87–98

Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, et al. 2005. The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* 310:1817–21

Fazi F, Rosa A, Fatica A, Gelmetti V, De Marchis ML, et al. 2005. A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPα regulates human granulopoiesis. *Cell* 123:819–31

Flynt AS, Li N, Thatcher EJ, Solnica-Krezel L, Patton JG. 2007. Zebrafish *miR-214* modulates Hedgehog signaling to specify muscle cell fate. *Nat. Genet.* 39:259–63

Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, et al. 2005. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.* 3:e236

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, et al. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* 308:833–38

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, et al. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312:75–79

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442:199–202

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 123:631–40

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, et al. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432:235–40

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, et al. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 106:23–34

Grivna ST, Beyret E, Wang Z, Lin H. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* 20:1709–14

Grosshans H, Johnson T, Reinert KL, Gerstein M, Slack FJ. 2005. The temporal patterning microRNA *let-7* regulates several transcription factors at the larval to adult transition in *C. elegans*. *Dev. Cell* 8:321–30

Grun D, Wang YL, Langenberger D, Gunsalus KC, Rajewsky N. 2005. microRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput. Biol.* 1:e13

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, et al. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* 315:1587–90

Gupta A, Gartner JJ, Sethupathy P, Hatzigeorgiou AG, Fraser NW. 2006. Anti-apoptotic function of a microRNA encoded by the HSV-1 latency-associated transcript. *Nature* 442:82–85

Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* 18:3016–27

Han J, Lee Y, Yeom KH, Nam JW, Heo I, et al. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125:887–901

Harfe BD, McManus MT, Mansfield JH, Hornstein E, Tabin CJ. 2005. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proc. Natl. Acad. Sci. USA* 102:10898–903

Harris KS, Zhang Z, McManus MT, Harfe BD, Sun X. 2006. Dicer function is essential for lung epithelium morphogenesis. *Proc. Natl. Acad. Sci. USA* 103:2208–13

Hatfield SD, Shcherbata HR, Fischer KA, Nakahara K, Carthew RW, Ruohola-Baker H. 2005. Stem cell division is regulated by the microRNA pathway. *Nature* 435:974–78

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, et al. 2005. A microRNA polycistron as a potential human oncogene. *Nature* 435:828–33

Hornstein E, Mansfield JH, Yekta S, Hu JK, Harfe BD, et al. 2005. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature* 438:671–74

Hornstein E, Shomron N. 2006. Canalization of development by microRNAs. *Nat. Genet.* 38(Suppl.):S20–24

Humphreys DT, Westman BJ, Martin DI, Preiss T. 2005. MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc. Natl. Acad. Sci. USA* 102:16961–66

Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* 293:834–38

Hutvagner G, Zamore PD. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297:2056–60

Hwang HW, Wentzel EA, Mendell JT. 2007. A hexanucleotide element directs microRNA nuclear import. *Science* 315:97–100

Jackson RJ, Standart N. 2007. How do microRNAs regulate gene expression? *Sci. STKE* 2007:re1

Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. 2005. Dicer-1 and R3D1-L catalyze microRNA maturation in *Drosophila*. *Genes Dev.* 19:1674–79

Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, et al. 2005. Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* 120:623–34

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, et al. 2005. RAS is regulated by the *let-7* microRNA family. *Cell* 120:635–47

Johnston RJ, Hobert O. 2003. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* 426:845–49

Johnston RJJ, Chang S, Etchberger JF, Ortiz CO, Hobert O. 2005. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl. Acad. Sci. USA* 102:12449–54

Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* 309:1577–81

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, et al. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* 19:489–501

Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15:2654–59

Kim YK, Kim VN. 2007. Processing of intronic microRNAs. *EMBO J.* 26:775–83

Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, et al. 2004. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.* 18:1165–78

Kloosterman WP, Steiner FA, Berezikov E, de Bruijn E, van de Belt J, et al. 2006. Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res.* 34:2558–69

Kloosterman WP, Wienholds E, Ketting RF, Plasterk RH. 2004. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res.* 32:6284–91

Knight SW, Bass BL. 2001. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 293:2269–71

Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, et al. 2005. Combinatorial microRNA target predictions. *Nat. Genet.* 37:495–500

Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, et al. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 438:685–89

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* 294:853–58

Lai EC. 2002. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* 30:363–64

Lai EC, Tam B, Rubin GM. 2005. Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev.* 19:1067–80
Landthaler M, Yalcin A, Tuschl T. 2004. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.* 14:2162–67
Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294:858–62
Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, et al. 2006. Characterization of the piRNA complex from rat testes. *Science* 313:363–67
Leaman D, Chen PY, Fak J, Yalcin A, Pearce M, et al. 2005. Antisense-mediated depletion reveals essential and specific functions of microRNAs in *Drosophila* development. *Cell* 121:1097–108
Lecellier CH, Dunoyer P, Arar K, Lehmann-Che J, Eyquem S, et al. 2005. A cellular microRNA mediates antiviral defense in human cells. *Science* 308:557–60
Lee RC, Ambros V. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294:862–64
Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–54
Lee Y, Ahn C, Han J, Choi H, Kim J, et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425:415–19
Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN. 2006. The role of PACT in the RNA silencing pathway. *EMBO J.* 25:522–32
Lee Y, Kim M, Han J, Yeom KH, Lee S, et al. 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23:4051–60
Lee YS, Nakahara K, Pham JW, Kim K, He Z, et al. 2004. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* 117:69–81
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
Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. Prediction of mammalian microRNA targets. *Cell* 115:787–98
Li M, Jones-Rhoades MW, Lau NC, Bartel DP, Rougvie AE. 2005. Regulatory mutations of *mir-48*, a *C. elegans let-7* family MicroRNA, cause developmental timing defects. *Dev. Cell* 9:415–22
Li X, Carthew RW. 2005. A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the *Drosophila* eye. *Cell* 123:1267–77
Li Y, Wang F, Lee JA, Gao FB. 2006. MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes Dev.* 20:2793–805
Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, et al. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433:769–73
Lin SY, Johnson SM, Abraham M, Vella MC, Pasquinelli A, et al. 2003. The *C. elegans* hunchback homolog, *hbl-1*, controls temporal patterning and is a probable microRNA target. *Dev. Cell* 4:639–50
Lindsay EA. 2001. Chromosomal microdeletions: dissecting del22q11 syndrome. *Nat. Rev. Genet.* 2:858–68
Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, et al. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305:1437–41
Liu J, Rivas FV, Wohlschlegel J, Yates JR, Parker R, Hannon GJ. 2005a. A role for the P-body component GW182 in microRNA function. *Nat. Cell Biol.* 7:1261–66

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R. 2005b. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat. Cell Biol.* 7:719–23

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* 435:834–38

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. 2004. Nuclear export of microRNA precursors. *Science* 303:95–98

Maniataki E, Mourelatos Z. 2005. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.* 19:2979–90

Maroney PA, Yu Y, Fisher J, Nilsen TW. 2006. Evidence that microRNAs are associated with translating messenger RNAs in human cells. *Nat. Struct. Mol. Biol.* 13:1102–7

Martinez J, Tuschl T. 2004. RISC is a 5′ phosphomonoester-producing RNA endonuclease. *Genes Dev.* 18:975–80

Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschl T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* 15:185–97

Meister G, Tuschl T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature* 431:343–49

Meyer WJ, Schreiber S, Guo Y, Volkmann T, Welte MA, Muller HA. 2006. Overlapping functions of argonaute proteins in patterning and morphogenesis of *Drosophila* embryos. *PLoS Genet.* 2:e134

Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, et al. 2006. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126:1203–17

Mishima Y, Giraldez AJ, Takeda Y, Fujiwara T, Sakamoto H, et al. 2006. Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. *Curr. Biol.* 16:2135–42

Miska EA, Alvarez-Saavedra E, Townsend M, Yoshii A, Sestan N, et al. 2004. Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.* 5:R68

Murchison EP, Partridge JF, Tam OH, Cheloufi S, Hannon GJ. 2005. Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl. Acad. Sci. USA* 102:12135–40

Murchison EP, Stein P, Xuan Z, Pan H, Zhang MQ, et al. 2007. Critical roles for Dicer in the female germline. *Genes Dev.* 21:682–93

Naguibneva I, Ameyar-Zazoua M, Polesskaya A, Ait-Si-Ali S, Groisman R, et al. 2006. The microRNA *miR-181* targets the homeobox protein Hox-A11 during mammalian myoblast differentiation. *Nat. Cell Biol.* 8:278–84

Nelson PT, Baldwin DA, Scearce LM, Oberholtzer JC, Tobias JW, Mourelatos Z. 2004. Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat. Methods* 1:155–61

Nolo R, Morrison CM, Tao C, Zhang X, Halder G. 2006. The *bantam* microRNA is a target of the hippo tumor-suppressor pathway. *Curr. Biol.* 16:1895–904

Nottrott S, Simard MJ, Richter JD. 2006. Human *let-7a* miRNA blocks protein production on actively translating polyribosomes. *Nat. Struct. Mol. Biol.* 13:1108–14

O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. 2005. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435:839–43

Obernosterer G, Leuschner PJ, Alenius M, Martinez J. 2006. Post-transcriptional regulation of microRNA expression. *RNA* 12:1161–67

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* 18:1655–66

Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, et al. 2000. Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* 408:86–89

Petersen CP, Bordeleau ME, Pelletier J, Sharp PA. 2006. Short RNAs repress translation after initiation in mammalian cells. *Mol. Cell* 21:533–42
Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, et al. 2005. Identification of microRNAs of the herpesvirus family. *Nat. Methods* 2:269–76
Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, et al. 2004. Identification of virus-encoded microRNAs. *Science* 304:734–36
Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, et al. 2005. Inhibition of translational initiation by let-7 MicroRNA in human cells. *Science* 309:1573–76
Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, et al. 2004. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432:226–30
Rajewsky N. 2006. microRNA target predictions in animals. *Nat. Genet.* 38(Suppl.):S8–13
Rehwinkel J, Behm-Ansmant I, Gatfield D, Izaurralde E. 2005. A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. *RNA* 11:1640–47
Rehwinkel J, Natalin P, Stark A, Brennecke J, Cohen SM, Izaurralde E. 2006. Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in *Drosophila melanogaster*. *Mol. Cell. Biol.* 26:2965–75
Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403:901–6
Ronshaugen M, Biemar F, Piel J, Levine M, Lai EC. 2005. The *Drosophila* microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes Dev.* 19:2947–52
Ruby JG, Jan C, Player C, Axtell MJ, Lee W, et al. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127:1193–207
Saito K, Ishizuka A, Siomi H, Siomi MC. 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in *Drosophila* cells. *PLoS Biol.* 3:e235
Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, et al. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.* 20:2214–22
Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, et al. 2006. A brain-specific microRNA regulates dendritic spine development. *Nature* 439:283–89
Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115:199–208
Sen GL, Blau HM. 2005. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat. Cell Biol.* 7:633–36
Shiohama A, Sasaki T, Noda S, Minoshima S, Shimizu N. 2003. Molecular cloning and expression analysis of a novel gene *DGCR8* located in the DiGeorge syndrome chromosomal region. *Biochem. Biophys. Res. Commun.* 304:184–90
Slack FJ, Basson M, Liu Z, Ambros V, Horvitz HR, Ruvkun G. 2000. The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* 5:659–69
Sokol NS, Ambros V. 2005. Mesodermally expressed *Drosophila* microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes Dev.* 19:2343–54
Sood P, Krek A, Zavolan M, Macino G, Rajewsky N. 2006. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl. Acad. Sci. USA* 103:2746–51
Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. 2005. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* 123:1133–46
Stark A, Brennecke J, Russell RB, Cohen SM. 2003. Identification of *Drosophila* microRNA targets. *PLoS Biol.* 1:E60

Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, et al. 2004. Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* 270:488–98

Sullivan CS, Grundhoff AT, Tevethia S, Pipas JM, Ganem D. 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* 435:682–86

Tang F, Kaneda M, O'Carroll D, Hajkova P, Barton SC, et al. 2007. Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* 21:644–48

Teleman AA, Maitra S, Cohen SM. 2006. *Drosophila* lacking microRNA miR-278 are defective in energy homeostasis. *Genes Dev.* 20:417–22

Thompson BJ, Cohen SM. 2006. The Hippo pathway regulates the *bantam* microRNA to control cell proliferation and apoptosis in *Drosophila*. *Cell* 126:767–74

Thomson JM, Newman M, Parker JS, Morin-Kensicki EM, Wright T, Hammond SM. 2006. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* 20:2202–7

Thomson JM, Parker J, Perou CM, Hammond SM. 2004. A custom microarray platform for analysis of microRNA gene expression. *Nat. Methods* 1:47–53

Tomari Y, Du T, Haley B, Schwarz DS, Bennett R, et al. 2004. RISC assembly defects in the *Drosophila* RNAi mutant armitage. *Cell* 116:831–41

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313:320–24

Vasudevan S, Steitz JA. 2007. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2. *Cell* 128:1105–18

Vaucheret H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.* 20:759–71

Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci. USA* 103:2257–61

Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, et al. 2006. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell* 124:1169–81

Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, et al. 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* 20:1732–43

Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, et al. 2005. MicroRNA expression in zebrafish embryonic development. *Science* 309:310–11

Wienholds E, Koudijs MJ, van Eeden FJ, Cuppen E, Plasterk RH. 2003. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat. Genet.* 35:217–18

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* 75:855–62

Wu L, Fan J, Belasco JG. 2006. MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl. Acad. Sci. USA* 103:4034–39

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, et al. 2005. Systematic discovery of regulatory motifs in human promoters and 3′UTRs by comparison of several mammals. *Nature* 434:338–45

Yekta S, Shih IH, Bartel DP. 2004. MicroRNA-directed cleavage of *HOXB8* mRNA. *Science* 304:594–96

Yi R, O'Carroll D, Pasolli HA, Zhang Z, Dietrich FS, et al. 2006. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat. Genet.* 38:356–62

Yi R, Qin Y, Macara IG, Cullen BR. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 17:3011–16

Yoo AS, Greenwald I. 2005. LIN-12/Notch activation leads to microRNA-mediated down-regulation of Vav in *C. elegans*. *Science* 310:1330–33

Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, et al. 2007. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129:303–17

Zhao Y, Samal E, Srivastava D. 2005. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436:214–20