Stony Brook University

Novel roles of the conserved Argonaute proteins during mammalian development: from miRNA biogenesis to gene silencing

A Dissertation Presented

by

Sihem Cheloufi

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Genetics

Stony Brook University

May 2010

Stony Brook University
The Graduate School

Sihem Cheloufi

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation

Gregory J. Hannon, Ph.D., Dissertation Advisor
Professor
Watson school of biological sciences
Howard Hughes Medical institute
Cold spring Harbor laboratory

Adrian R. Krainer, Ph.D., Chairperson of Defense
Professor
Watson School of biological sciences
Cold spring harbor laboratory

Scott Lowe, Ph.D.
Professor
Watson school of biological sciences
Howard Hughes Medical institute
Cold spring Harbor laboratory

Dr. Thomas W. White, Ph.D.
Associate Professor
Department of Physiology and Biophysics
Stony Brook University

Dr. Anna-Katerina Hadjantonakis, Ph.D.
Associate member
Developmental Biology
Sloan-Kettering Institute
Associate Professor
Weill Medical College of Cornell University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Novel roles of the conserved Argonaute proteins during mammalian development: from miRNA biogenesis to gene silencing

by

Sihem Cheloufi

Doctor of Philosophy

in

Genetics

Stony Brook University

2010

The Argonaute proteins are highly conserved throughout evolution. They associate with diverse classes of small RNAs to mediate gene silencing in developmental programs and participate in defense responses to viruses, transposons, and cellular stress. Vertebrate genomes encode four argonaute clade proteins. In mice, Argonaute2 (AGO2) is essential for embryogenesis, oogenesis and hematopoiesis. The developmental functions of the remaining Argonaute family members, AGO1, AGO3, and AGO4, remain elusive. My thesis work focuses on exploring the function of all four Argonaute proteins during animal development. Intriguingly, individual deletions of Ago1, Ago3, and Ago4, or deficiency of all three Argonautes had no apparent impact on normal mouse development or fertility. However, we demonstrate that Ago1 and Ago3 may play a role in viral defense. We found that the specialized function of Ago2 in the extraembryonic lineage can, in part, explain its unique requirement during normal development. In addition, AGO2 is the only family member retaining its nucleolytic activity. This enzymatic activity of Argonautes is deeply conserved, despite its having no obvious role in miRNA directed gene silencing. To investigate the evolutionary pressure to conserve Argonaute enzymatic activity, we engineered a mouse with catalytically inactive Ago2 alleles. Homozygous mutants died shortly after birth with an obvious anemia. Examination of microRNAs and their potential targets revealed a loss of miR-451, a small RNA important for erythropoiesis. Though this microRNA is processed by Drosha, its maturation does not require Dicer. Instead, the pre-miRNA becomes loaded into

AGO2 and is cleaved by the AGO2 catalytic center to generate an intermediate 3' end, which is then further trimmed. We demonstrate that this novel AGO2-mediated miRNA biogenesis pathway can be exploited to engineer artificial silencing molecules by mimicking the precursor RNA structure of miR-451. These findings link the conservation of Argonaute catalysis to a conserved mechanism of microRNA biogenesis that is important for vertebrate development. Future studies using the Argonaute genetic models will help us define the biological mechanisms that dictated Argonaute gene family conservation.

To my Father, he will always be living in my mind and in my heart

Algeria, May 2008

A conserved miRNA biogenesis pathway that requires Ago catalysis in vertebrates

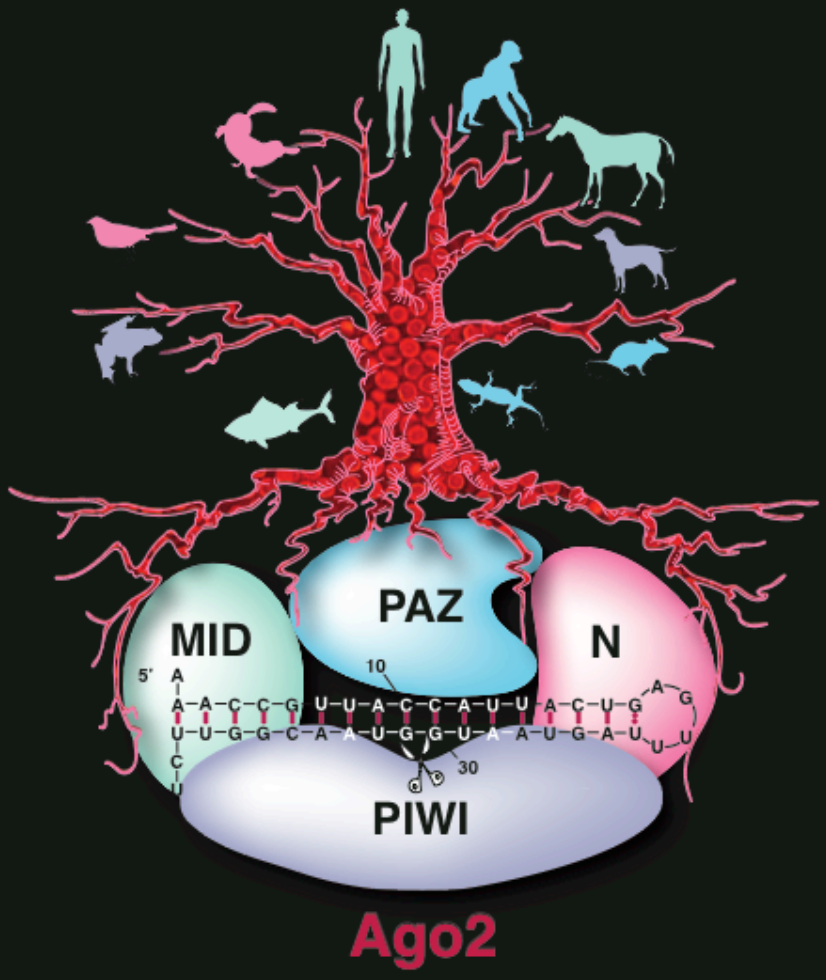

Table of Contents

List of tables

Chapter 2: Mammalian Argonautes: redundant or specialized? Argonautes function in defense against viruses

Chapter3: A dicer independent miRNA biogenesis pathway that requires Argonaute catalysis:

List of figures

Chapter 1: Introduction

Chapter 2: Mammalian Argonautes: redundant or specialized? Argonautes function in defense against viruses

Chapter 3: A dicer independent miRNA biogenesis pathway that requires Argonaute catalysis:

Chapter4: Conclusions and future directions

Preface

Thesis plan

Chapter1: An introduction to the RNAi mechanism in the context of vertebrate development and conservation of Argonaute catalysis. Can be used as a general review for RNAi and development

Chapter2: Mammalian Argonautes: redundant or specialized?
We plan to submit part of this chapter as a paper to the journal of Silence describing the phenotype of the Ago1/3/4 triple knockout mice. It is an important observation to report in the field and many of the reagents generated can be very useful to the scientific community.

Chapter3: A dicer independent miRNA biogenesis pathway that requires Ago catalysis: The discovery of this pathway and the function of Ago2 during development is published as a Nature Article ahead of print:
Nature. 2010 Apr 27. [Epub ahead of print] doi:10.1038/nature09092.

Chapter4: Conclusion and future perspectives
I am currently investigating the potency of mir-451 mimics and their implications in RNAi technology through my postgraduate transition to a post-doctoral fellowship.

Appendix1: Characterization of Dicer-deficient murine embryonic stem cells
This paper was published in the journal of Proc Natl Acad Sci U S A. in the following issue: 2005 Aug 23;102(34):12135-40. Epub 2005 Aug 12.
This project was completed in collaboration with Dr. Elizabeth Murchison a former graduate student in my lab. I contributed in the analysis of centromeric heterochromatin in the dicer null Embryonic Stem cells using histone methyl marks immunoprecipitation assays (fig.1.8C).

Appendix2: Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes.
This paper was published in the journal of Nature 2008 May 22;453(7194):534-8. Epub 2008 Apr 10. This was a collaborative project with Dr. Alexei Aravin and Dr. Oliver Tam in my lab. My contributions were useful comments, discussions and mouse oocyte collections for the small RNA library generation.

Appendix3: Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases

This manuscript was put together in collaboration with Dr. Fedor Karginov, a postdoctoral fellow in the lab. I generated the Ago2 null ES cell lines used in this study to identify Ago2 dependent mRNA cleavage targets. I helped with performing the tissue culture experiments on the Ago2 null and Drosha genetic systems, discussions and comments on the manuscript. This paper is currently under review for the Journal of Molecular Cell.

Appendix4: A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity

This paper is published online in Science Express (6 May 2010/ 10.1126/science.1190809). This study was initiated thanks to our data on the Ago2 catalytically inactive mouse model that provided a road map to investigate the conservation of this pathway in Zebrafish (see Chapter3). My contribution was to guide this study and provide useful discussions for the design of the experiments and helpful comments with writing the manuscript.

Acknowledgments

There are not enough words or space in my thesis to extend my gratitude to all the people and all the living organisms who have been my companions throughout this journey. It has been 10 years of great education and hard work of my life and I cannot say that the road was easy. It is going through those bumpy valleys that one really appreciates how life is not always smooth sailing and one needs to learn how to bypass all the hurdles to find the truth. Life at the bench is always an adventure and our curiosity and hunger for knowledge keep us searching for the secrets of life, even if it is through a faint band on a gel.

I cannot be grateful enough to Greg, who truly believes in me and taught me the most valuable lesson in life, that is to be fearless and to trust and believe in my judgment and really just to go for it and most importantly to always strike while the iron is still hot. I am just extremely lucky to have found a niche here at the McClintock building where I have built a giant network of real friends and colleagues. I have been extremely spoiled by the quality of science that surrounded me for the last few years and I am really lucky to have all the unconditional love and support from my mentor, my colleagues, my friends and my family.

My voyage from Algeria started 13 years ago when I followed my dad's advice to seize the opportunity to learn a new language and explore new intellectual possibilities that are no available in my country. I indeed found my passion, that is to study genetics and development. I cannot thank my family enough for being so supportive and so understanding about my choice, especially my mother Bahia and my two little sisters Esma and Wafaa. I also would like to thank my brothers Amine and Sofiane, who are always supportive despite all the differences and disagreements about my life style. I also feel certain that my dad "May he rest in peace" is very proud of me and he had always believed in me throughout my entire journey and dreamt to live this very moment. I feel so sorry I couldn't make it to the finish line earlier.

I would like to thank the three most influential scientists during my graduate studies: Alexei Aravin, who is to me one of the greatest minds in science, he has truly been a wonderful teacher and a great friend. I also want to thank, Ingrid Ibarra, from whom I learned to be perseverant and to realize that even though science can be a strike of luck sometimes, the real feeling of happiness and satisfaction in finding the truth comes from hard work, no less. I also want to thank another Russian colleague, Vasily Vagin. He is my bench mate who is just so nice to discuss any aspect of my scientific work with, it being molecular biology, development or biochemistry. He is always willing to teach, to listen and to help in the best way he can. He has truly

been my everyday cheerleader for the last 9 months, which have not been so easy.

I have communicated all my gratitude to the beautiful scientific community in my first scientific paper, who helped put such a wonderful story together. Maria Mosquera, Lisa Bianco, Sang young Kim and all the mouse crew at Woodbury who helped with all the animal care and the techniques. I would like to thanks David Frendewey and Aris Economides for all the advice and the help with the transgenic mouse generation. I cannot be thankful enough to all the teachers at the mouse embryology course here at the lab; it has truly been one of the best courses in my life, intense and fun at the same time. I would like to thank everyone in the Hannon lab for helpful discussions and for making the lab such a great work and social environment. I especially would like to thank Oliver Tam who has been so helpful in every difficulty I face in the lab, from fixing my computer to discussing and interpreting my results, to helping me with my writing and to always do this with a smile. I especially want to thank my friend Kata, who has been there for me through bad and good weather, always so generous and understanding. I would like to thank Jo Leonardo and Sabrina Boettcher for all their support and for making my life in the lab always smooth. I want to thanks all the Hannon lab alumni who I had a chance to meet and wish to have the chance to interact them with in the future. Greg's lab is truly a unique environment. Greg really has a gift of picking the right people and making the lab such a great and comfortable environment to do science. I also want to thank all my committee members for all the support and the great advice over the years.

At last, I would like to thank my companion and my dearest friend Jernej Murn, one cannot really describe what love is, it is simply a gift of life and a wonderful thing. He has been so patient with me, even when I bring the mice home at night to watch them give birth in the early hours of the morning. He truly changed my life here at the lab and has been so supportive and so understanding with every little detail. We are both so lucky to have met each other in such a wonderful place.

CHAPTER 1: Introduction

1.1 The RNAi mechanism: a historical overview

The past 12 years have witnessed several fascinating discoveries in the field of RNA interference. Fire and Mello first described the use of long double stranded RNA to effectively suppress in a sequence specific manner gene expression in the nematode *Caenorhabditis elegans (Fire et al., 1998).* They demonstrated that the gene silencing effect can be passed on to the next generation and proposed an amplification step in the process due to a very potent gene silencing effect in response to small amounts of the double stranded RNA trigger. This study instigated the exploration of the molecular components of the newly discovered process of RNA interference (RNAi), its existence in other organisms, and putative applications in molecular biology.

The first clue to how RNAi may be initiated came from a genetic screen for RNAi components in worms (Tabara et al., 1999). This study identified rde-1 gene that is now known to be part of the well-conserved and diverse family of Argonaute proteins (Carmell et al., 2002; Tolia and Joshua-Tor, 2007) (fig.1). This provided the first genetic evidence for the involvement of an Argonaute protein in mediating an RNAi response, and indicated that RNAi may play a role in defense against transposon activation. In addition, this was the first report to address the evolutionary conservation of the RNAi response. The involvement of RNAi in normal development of different organisms then became evident from the different genetic studies of Argonaute homologs (Bohmert et al., 1998; Schmidt et al., 1999; Wilson et al., 1996), reviewed in (Carmell et al., 2002) .

The association between small RNAs and RNAi became apparent from the discovery of *short temporal RNAs* in worms (Lee et al., 1993) (see below) and small antisense RNAs in plants generated in response to viral infection(Hamilton and Baulcombe, 1999). However, the vision of a molecular pathway remained unclear. It wasn't until the beginning of the millennium that dicer was discovered and provided the key to understanding of how these small effector molecules were generated in a cell (Bernstein et al., 2001; Grishok et al., 2001; Hammond et al., 2000). Genetic mutants of dicer and Argonaute proteins in the worm were the first to demonstrate an essential role of the RNAi components during normal development (Grishok et al., 2001). This was followed by elegant studies by the Zamore group and our group in collaboration with the Plasterk group in Holland, that demonstrated a physiological role of dicer in generating the so called "small temporal RNAs", now classified as the miRNA class of endogenous small RNAs (Hutvagner et al., 2001; Ketting et al., 2001). The dissection of the post transcriptional regulation of miRNAs in mammalian cells revealed the role of an RNaseIII-associated microprocessor enzyme, Drosha, in releasing miRNA precursor molecules for export into the cytoplasm and later processing by dicer

(Lee et al., 2003). Although clarification of these small RNA biogenesis mechanisms founded the base for understanding how the small RNA-mediated silencing pathways were initiated, the key effector steps of RNAi in regulating gene expression remained elusive.

Subsequent biochemical studies proved the involvement of yet another third nuclease in the RNAi pathway. This nuclease, ascribed to the RNaseH activity of mammalian Argonaute 2, turned out to be the nuclease responsible for cleavage and destruction of target mRNAs having sequences complementary to small RNAs (Liu et al., 2004; Song et al., 2004).

In parallel to the studies of the RNAi pathway in plants and animals, yeast genetics provided the first evidence for the involvement of RNAi in transcriptional gene silencing. Components of the RNAi machinery in *Schizosaccharomyces pombe* emerged as essential molecules for centromeric repeat silencing and chromosome segregation (Volpe et al., 2002). These findings unraveled a distinct function and site of action of the RNAi machinery that may contribute to our understanding on how RNAi regulation of developmental processes and participate in defense mechanism against viruses and transposon activation. Furthermore, recent analyses of the RNAi components in germ cells uncovered a function for small RNAs in transposon silencing and chromosomal integrity, reviewed in (Malone and Hannon, 2009).

My thesis work concentrates on the developmental roles of Argonautes during mouse embryogenesis with a specific focus on Argonaute's catalytic activity and the reason for its evolutionary conservation. My studies led me to the discovery of a novel small RNA biogenesis pathway that is deeply conserved in vertebrates. This introduction continues with the description of the known biogenesis pathways and effector functions of small RNAs and their mode of action, followed by an overview of how components of these pathways may impact vertebrate development. I conclude by giving an evolutionary perspective on the catalytic role of Argonautes.

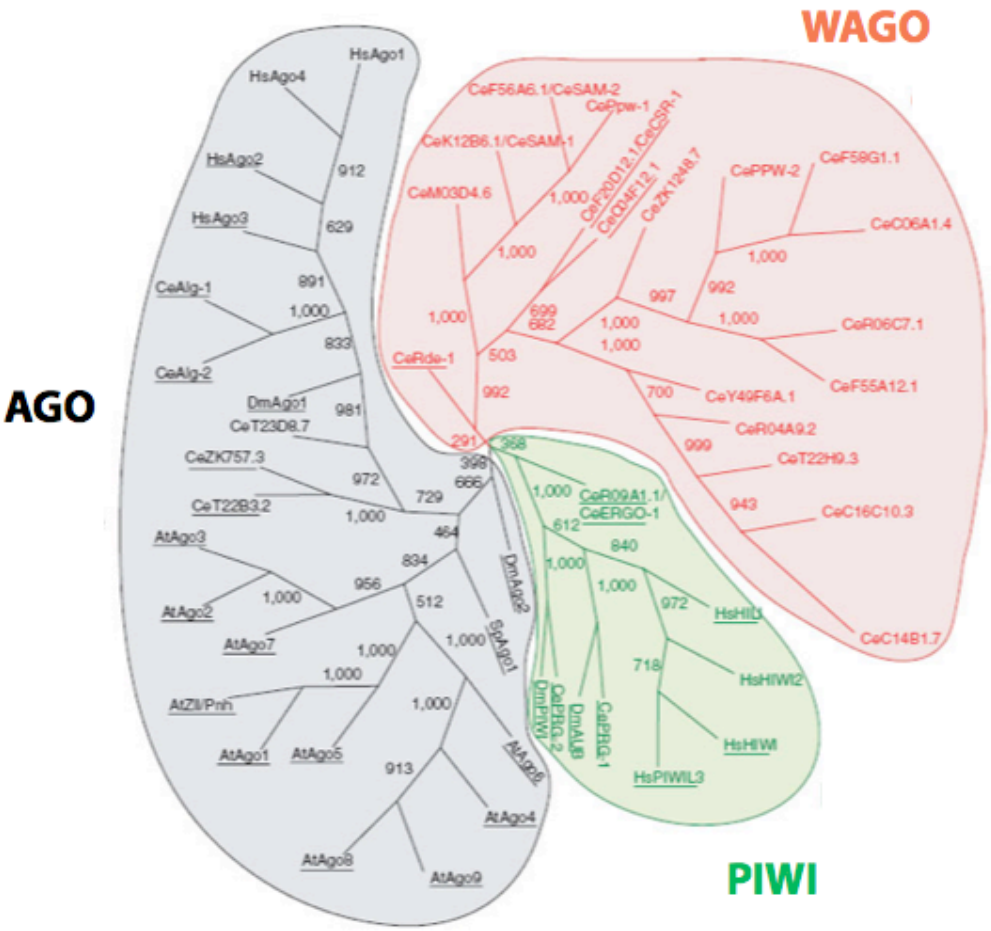

Figure 1: The three clades of the Argonautes

The first two letters of each protein in the tree depicts the organism: Ce: C. elegans, Dm: Drosophila melanogaster, Hs: Homo sapiens, At: Arabidopsis Thaliana and Sp: S. Pombe. Argonautes that contain a complete catalytic motif and/or were shown to have slicer activity are underlined. (Modified from Slicer and the argonautes, Niraj H Tolia & Leemor Joshua-Tor Nature Chemical Biology 3, 36 - 43 (2007)

1.2 Mechanisms of small RNA biogenesis and effector functions

1.2.1 Small RNA biogenesis

Structural and biochemical studies combined with genetic analyses and cell biology of the RNAi machinery components paved the path to understanding how every enzymatic step of the pathway functions. All small RNA biogenesis pathways converge to produce an active RNA molecule that becomes part of a mature RNA silencing complex upon binding of an Argonaute protein (Kim et al., 2009; Sashital and Doudna, 2010). Small RNAs are classified according to the nature of their precursors, their biogenesis process, and association with different Argonaute proteins. The association of different small RNA molecules with different Argonautes can be attributed in part to the size and tissue specificity of Argonaute family members. For example, piRNAs (Piwi-interacting RNAs) (Aravin et al., 2006; Girard et al., 2006), owing their name to the PIWI clade of the Argonaute superfamily, are 26-30nt long small RNA and associate with Argonautes that are specifically expressed in the gonads of animals, while miRNAs and siRNAs are 20-23nt long small RNAs and associate with the more ubiquitously expressed Argonautes (the AGO clade)(Liu et al., 2004; Meister et al., 2004).

1.2.2 Endonucleolytic processes

miRNA biogenesis is the best characterized pathway, since miRNAs were the first discovered naturally occurring class of small RNAs. Sequencing projects have identified close to 1000 unique miRNAs encoded by the human genome (mirBase collection of miRNAs). miRNA genes reside in intergenic regions, in introns, or in exons of coding or non-coding transcriptional units. Their precursor transcripts, also known as primary transcripts, are pol II or pol III transcribed non-coding molecules that can fold into one, two, or multiple hairpin structures, depending on whether they are derived from a mono-, bi-, or poli-cistronic miRNA gene, respectively. The structure of a double stranded RNA hairpin is recognized and captured by an RNAseIII enzyme to produce a mature ~20nt long miRNA duplex. miRNA hairpins undergo two main processing steps (fig 2a-d). The first step is mediated by Drosha/Pasha microprocessor complex in the nucleus. Drosha protein harbors two RNAseIII domains that cleave the hairpin structure of a miRNA precursor about 11nt from the base of the stem of a hairpin(Seitz and Zamore, 2006). This process is dependent on pasha, a double strand RNA binding (DSRB) protein that associates with Drosha (Denli et al., 2004). Recent conformational modeling of the two DSRB domains of pasha provided evidence

for correct recognition of Drosha substrates that have a ~33nt long stem (Sohn et al., 2007). Once Drosha cuts the precursor miRNA hairpin, it is released and exported into the cytoplasm. The export is mediated by exportin 5, a member of the nuclear transport receptor family that binds cooperatively to the Drosha

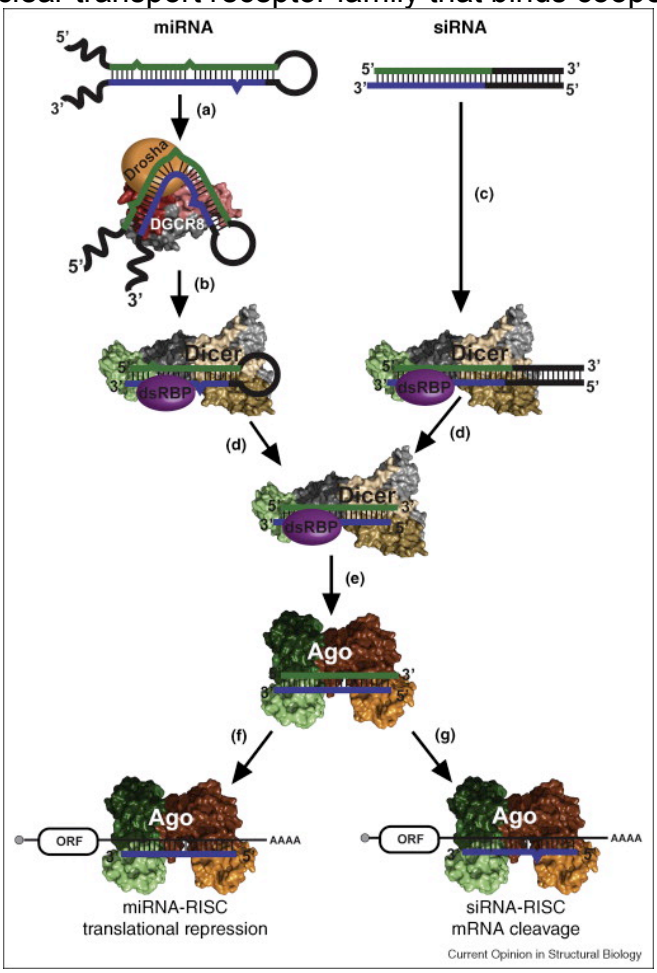

Figure 2: miRNA and siRNA biogenesis and gene silencing

Modified from Sashital DG,Doudna J.A. Current Opinion in structural biology 2010 Feb; 20(1):90-7. Known structures of RNA silencing factors are represented, while unknown structures are shown as 3D shapes (Drosha and TRBP). RNA is depicted with thick lines where blue represents the guide strand sequence, green represents the passenger strand sequence, and black represents sequence that is cleaved off by RNase III enzymes. **(a)** Primary miRNA transcripts form hairpin structures that are recognized by the double-stranded RNA-binding protein (dsRBP) DGCR8 and cleaved 11-bp away from the base of the stem by the RNase III enzyme Drosha. **(b,c)** Pre-miRNA hairpins and double-stranded siRNA precursors are cleaved by Dicer, forming the mature miRNA duplex. Product length is determined by the distance between the 3′-overhang-binding PAZ domain (green) and the RNase III catalytic center (beige, sand). **(d)** The Dicer–dsRBP complex retains the mature product, handing it off to Argonaute (Ago), the effector of RISC. **(e)** After binding the duplex, the passenger strand (green) is removed while the guide strand (blue) remains bound to Ago. **(f,g)** The activated RISC can then go on to target mRNAs, acting through separate mechanisms depending on the type of guide RNA.

product and the GTP-bound form of the cofactor Ran. The intermediate miRNA precursor, also known as the pre-miRNA, is then released in the cytoplasm following GTP hydrolysis (Lund et al., 2004). Once in the cytoplasm, the pre-miRNA is recognized by a second RNAseIII enzyme, dicer, that contains two RNAseIII domains essential for two simultaneous endonucleolytic cuts to cleave off the loop of the hairpin and generate the mature miRNA duplex (Zhang et al., 2004). Despite dicer having its own DSRB domain, it requires an association with another DSRB protein, TRBP, that plays an important role in Argonaute loading (Chendrimada et al., 2005; Haase et al., 2005).

siRNA processing has mainly been studied by introducing long double stranded RNAs or synthetic siRNAs into the cell. Maturation of siRNA duplexes requires only one endonucleolytic step mediated by dicer (fig 2c-d). The endogenous siRNAs have only recently been characterized in murine oocytes, and their biogenesis is difficult to study in their natural environment(Tam et al., 2008; Watanabe et al., 2008). siRNAs originate from pseudogene transcripts that form double stranded RNA either through annealing to homologous spliced transcripts or form inverted repeats.

piRNAs processing is poorly understood. The maturation of piRNA is independent of dicer(Vagin et al., 2006). Drosha requirement for their processing remains to be determined experimentally. piRNAs are generated through a primary and a secondary processing step(Brennecke et al., 2007). Sequence analysis of Ago-piRNA complexes revealed that the secondary processing step requires Argonaute catalysis for maturation of the 5'end of the molecule and an unknown nuclease for maturation of the 3'end. piRNA precursors and their primary processing steps are currently being investigated.

1.2.3 Loading and maturation of an active small RNA molecule

Assembly of a miRNA or an siRNA duplex into the Argonaute protein is dependent on dicer and its partner TRBP. Both Dicer and Argonaute proteins use their signature PAZ domain to anchor the 3' end of the small RNA molecule (Song et al., 2003). In the case of miRNAs, Drosha determines the 3'end of the hairpin that is anchored to dicer. However, after dicer makes the second cut near the terminal loop, a second 3'end is generated and Ago has to choose between the two 3'ends. This is known as the asymmetry problem. Structural modeling of dicer, TRBP, and Ago propose a possible explanation for how dicer releases the small RNA duplex and hands it over to Ago(Sashital and Doudna, 2010). In this model, the strand selection is dependent on location and the flexibility of the TRBP protein within the complex and could orient the correct 3' end of the duplex into the PAZ domain of the Argonaute. TRBP is thought to be responsible for sensing the less thermodynamically stable 5'end of the small RNA duplex and choosing it as the active (guide) strand. Once the duplex is loaded into Ago, the

passenger strand (also referred to as the sense strand) is destroyed by an unknown mechanism. This mechanism is well characterized in the fly where an unwinding mechanism, probably mediated by a helicase, forms the mature guide strand of miRNAs. In the case of siRNAs, the passenger strand is destroyed via a two-step process involving Ago2 catalysis and endonucleolytic cleavage by C3PO (Liu et al., 2009; Matranga et al., 2005). In mammals, miRNA maturation is probably mediated by the unwinding pathway since most mature miRNA duplexes contain bulges and are therefore less likely to be cleaved by Ago. The maturation mode of endogenous siRNAs in the oocytes or synthetic siRNAs in mammalian cells remains to be determined.

1.2.4. Effector functions

The ultimate purpose of generating an Ago-small RNA complex is to regulate gene expression. Different small RNA subclasses have different effector functions. For example, miRNAs mainly function through translational control or mRNA degradation while siRNAs mediates mRNA cleavage. The general rule is for the Argonaute to use a small RNA molecule as a guide to find its complementary nucleic acid sequence. This regulation step can occur either at transcriptional or posttranscriptional level (Kim et al., 2006; Valencia-Sanchez et al., 2006). The transcriptional control by small RNAs in mammals is controversial in the field. Small RNA mediated posttranscriptional regulation has been extensively studied. miRNAs regulate gene expression through mechanisms of inhibiting translation or mRNA degradation through deadenylation and decapping of mRNAs followed by their degradation (Nilsen, 2007). miRNAs recognize their targets through base paring of the miRNA seed sequence (2-8nt stretch at the 5'end of the guide strand) to binding sites within the 3'UTR of the mRNA(Bartel, 2009). Targeted sites within the coding region have also been reported(Tay et al., 2008). Argonaute mediated target cleavage is rare in mammals and has only been reported in the literature for two miRNAs that have perfectly matched sites to their target genes (Davis et al., 2005; Yekta et al., 2004). In collaboration with Fedor Karginov in the lab, we report additional miRNA cleaved targets in the ES cells (see appendix 3). Endogenous siRNAs in the oocytes probably silence their targets through Ago cleavage but this still needs to be experimentally demonstrated. An additional mechanism is to regulate gene expression through sequestering mRNAs in cytoplasmic RNA processing centers (P bodies) (Liu et al., 2005a; Liu et al., 2005b; Nilsen, 2007).

The RNAi pathway components described here were identified either through genetic screens or the study of evolutionary conserved paralogs in different organisms or from small RNA sequencing projects. The roles of several other candidate processing or regulatory molecules that have panned out from different types of screens and proteomic studies remain to be elucidated. While we may understand the core frameworks of these pathways, it will take several

years and decades before we understand their regulation and fine-tuning in different physiological contexts in the organism.

1.3. RNAi and development

The physiological functions of the RNAi machinery are dictated by the regulatory mechanisms and the effector functions of the RNAi pathway. The developmental functions of the RNAi components, namely RNA molecules and their associated proteins, such as Drosha, dicer, Argonautes, and many other regulators, have been studied through forward genetic strategies in vertebrates. Here I describe the consequences of deficiencies in every step of the miRNA pathway.

1.3.1 Drosha microprocessor

To date there have been no reports in the literature to directly link Drosha deficiency and developmental defects, perhaps due to lethality caused by a deficiency in this primary step of miRNA processing. However, individual miRNA biogenesis studies reflect the tight regulation of this processing step in different developmental contexts (Davis and Hata, 2009). For example, miR-21 maturation is important in human vascular smooth muscle cell contraction and SMAD proteins are involved in processing pri-miR-21 by interacting with the Drosha microprocessor complex (Davis et al., 2008). Another example in mouse embryonic stem (ES) cells is LIN28 regulation of let-7 processing by interacting with the loop region of this miRNA (Newman et al., 2008; Viswanathan et al., 2008). Let-7 is a well-conserved miRNA that plays an important role in cellular differentiation (Ibarra et al., 2007; Melton et al., 2010). Perhaps the most direct evidence for the involvement of Drosha during normal development is the phenotype of the DGCR8 deficient mouse embryonic stem cells and the maternal zygotic mutant mouse embryos (Suh et al., 2010; Wang et al., 2007). DGCR8 deficient ES cells manifest severe proliferation and differentiation defects. This phenotype could explain the developmental arrest of DGCR8 maternal zygotic mutant embryos at the early post implantation stage of mouse development.

1.3.2. Dicer

Dicer loss-of-function phenotype allows for testing of the consequences of depletion of all small RNAs that are dependent on dicer's RNaseIII activity. Dicer-deficient phenotypes have been investigated during mouse and Zebrafish development. Maternal zygotic dicer mutant fish undergo axis formation and differentiate into multiple cell types but display abnormal morphogenesis during

gastrulation, brain formation, somitogenesis, and heart development (Giraldez et al., 2005). Interestingly, these defects are rescued by administration of the processed form of miR-430. miR-430 has been demonstrated to be involved in maternal mRNA deadenylation and clearance.

Contrary to the Zebrafish dicer model, it is impossible to generate a maternal zygotic dicer mutant mouse since the mammalian dicer is essential for oogenesis. Deletion of dicer in the growing oocytes results in the arrest during meiosis I with multiple disorganized spindles and severe chromosome segregation defects (Murchison et al., 2007). These findings are similar to the Zebrafish model, and suggest a role for the mammalian dicer in turnover of maternal transcripts that are normally lost during oocyte maturation (Giraldez et al., 2006). However, the analysis of the dicer deficient oocyte transcriptome revealed a possible role of dicer in transposon derived sequence element gene regulations. A recent study from our lab identified a new class of small RNAs specifically expressed in murine oocytes (Tam et al., 2008) (see appendix 2). These are endogenous siRNAs derived from pseudogene transcripts forming a double stranded RNA either through an inverted repeat transcript or through hybridization with their homologous spliced transcripts. We believe that the mouse oocyte represent a unique environment for the formation of a double strand RNA that is probably dependent on dicer for the generation of the mature siRNAs since their targets are upregulated in the dicer null background.

Zygotic loss of dicer in the mouse results in a developmental arrest at the gastrulation stage probably due to a loss of stem cell self-renewal at the early blastocyst stage as judged by aberrant expression of oct-4 at E7.5 (Bernstein et al., 2003). The dicer phenotypes in the mouse appear to be more severe than the Zebrafish phenotypes, perhaps due to a more complex developmental mechanism of mammalian embryogenesis involving the extraembryonic lineage, which is necessary for implantation and growth of the embryos. The miRNAs or maternally deposited siRNAs to regulate the early developmental cell fates remain to be identified. The generation of dicer conditional mouse models has been extremely useful in understanding the developmental requirements of dicer in the mouse. Conditional ablation of dicer in mouse embryonic stem cells results in a dramatic cell proliferation defect confirming the embryonic defect of dicer deficient embryos (Murchison et al., 2005)(see appendix 1). Crosses of the dicer conditional alleles to multiple tissue specific Cre lines almost always result in dramatic phenotypes, suggesting that dicer plays a central role in regulating gene expression(Andl et al., 2006; Chen et al., 2008; Harris et al., 2006; Koralov et al., 2008; Muljo et al., 2005; Schaefer et al., 2007). However, the requirement for specific miRNAs in different tissue types remains unexplored due to technical difficulties in tissue specific rescue experiments in the mouse. The possibility to rescue a miRNA global defect has been addressed in embryonic stem (ES) cells deficient in Dgcr8. In these experiment the cell proliferation phenotype was scrutinized by a miRNA screening strategy to define the mir-290 miRNA cluster as the key player in maintenance of stem cell proliferation (Wang et al., 2008).

Similarly, it has been shown that let-7 counteracts the pluripotent state of ES cells and inhibition of let-7 greatly enhanced the efficiency of iPS cell generation (Melton et al., 2010).

1.3.3. Argonautes

Argonautes are the key effector components in the RNAi pathway. They bind to the processed small RNAs and use them as guides to regulate gene expression to serve different cellular processes. According to sequence similarity they can be divided into three main clades (fig. 1). The AGO clade that is the most conserved family of Argonaute proteins in animals, plants, and fungi. The PIWI clade is restricted to the animal kingdom and the WAGO clade is specific to worms. Here I focus on the AGO clade. The proteins in this clade are ubiquitously expressed and associate with 21-24nt small RNAs. These small RNAs are known as miRNAs and siRNAs, differing mostly in their biogenesis and downstream effects as described above.

Of all the Argonaute family members in vertebrates only Ago2 in the mouse has been knocked out. Ago2 zygotic mutant embryos die at early stages of postimplantation (E7.5-E10.5) (Alisch et al., 2007; Liu et al., 2004; Morita et al., 2007). The phenotypic differences between the different alleles are probably due to the strength of the genetic mutants since all of them are generated by insertional mutagenesis. One study reports an Ago2 phenotype that is somehow similar to the dicer knockout and support the involvement of Ago2 in regulating gene expression at the posttranscriptional level, since Ago2 null embryos show no defects in DNA methylation maintenance in imprinted loci, centromeric repeats and Xist. One interesting study pointed to the role of Ago2 in mediating epithelial to mesenchymal transition during mouse gastrulation and indicated a genetic interaction between Ago2 and brachyury gene, a T box transcription factor that is essential for mesoderm development. Profiling small RNAs from the primitive streak or the gastrulating embryo would provide insights into the candidate miRNAs that could regulate the mesoderm germ layer formation. Finally, analysis of an independent Ago2 insertional allele revealed a role for Ago2 during midgestation where Ago2 mutant embryos had severe heart and neural tube defects. In addition, morphological analysis of the Ago2 mutant embryos proposes possible defect in placental and yolk sac development. In conclusion, the three genetic models of Ago2 in the mouse support a unique role of Ago2 during embryogenesis. A conditional Ago2 allele has been used to bypass the embryonic lethality and demonstrated that Ago2 also plays a unique role during hematopoiesis and oogenesis (Kaneda et al., 2009; Lykke-Andersen et al., 2008; O'Carroll et al., 2007). Ago2 deficient oocytes manifest chromosomal segregation defects during meiosis I similar to the dicer phenotype. Loss of Ago2 during adult hematopoiesis results in erythroid and B cell differentiation defects independently of its catalytic role. However, loss of Ago2 in ES cells and during spermatogenesis shows no obvious defects compared to the dicer

phenotypes (Hayashi et al., 2008). This suggests that Ago2 can have unique and redundant functions with the other Argonaute family members in a tissue specific manner during development. The molecular mechanisms behind its specialized role remain to be investigated.

1.3.4. miRNAs

Although, the dicer conditional allele of the mouse has been extensively used to study the role of miRNAs during development, there are few examples in the literature that point to the essential role of miRNAs in specific tissues. While Argonautes and dicer have ubiquitous expression profiles, there is accumulating evidence from the miRNA sequencing projects supporting tissue specific transcriptional or posttranscriptional regulation of miRNA expression (Landgraf et al., 2007). For example, the two mir-1/mir-133 bicistronic clusters are regulated by critical myocyte differentiation factors in cardiac muscle cells. Deletions and overexpression studies of these clusters highlight the important role of these miRNAs in regulating cellular proliferation in the heart. Additional knockouts of heart specific miRNAs, such as mir-208, highlight a natural role of miRNAs in stress-dependent cardiac growth (Meder et al., 2008). Another example demonstrates the essential role of the imprinted rtl-1 miRNA cluster in normal placental development. Knockout of the maternally expressed rtl-1 miRNA cluster results in placentomegally as a result of a detrimental defect in maternal-fetal nutrient, gas and waste exchange (Sekita et al., 2008). In addition to regulating normal developmental processes, some miRNAs have essential functions in regulating the immune system. Loss of mir-155 results in immunodeficiency as it affects T cells, B cells, and dendritic cell functions (Rodriguez et al., 2007; Thai et al., 2007).

There are almost 1000 human miRNA genes annotated in the miRbase miRNA database, some of which map to clusters in the genomes or belong to miRNA families sharing the same seed sequence. The use of classical miRNA gene knockouts together with antagomir technology and the development of decoy vectors will help us probe the functions of individual miRNAs in normal development, responses to stress, and disease.

1.4. The evolutionary conservation of Argonaute's catalytic function

There are four Argonaute clade members in the mouse. Ago2 is found on mouse chromosome 15, while the other three Agos (Ago1, Ago3 and Ago4) reside in the Ago3/1/4 gene cluster on chromosome 4. Ago2 is deeply conserved between species with a homolog found even in various species of

insects. In contrast, the Ago3/1/4 cluster appears to be more recent in origin being conserved only in vertebrates. Previous reports suggest that all the Argonautes have similar expression profiles and associate with identical small RNA partners (Azuma-Mukai et al., 2008; Ender et al., 2008; Liu et al., 2004; Meister et al., 2004). For example, the Argonaute clade members seem to have a redundant role during spermatogenesis, as conditional knockout of Ago2 in testes exhibits no effect on spermatogenesis. In contrast, the loss of *Dicer* leads to dramatic defects in differentiation and proliferation of spermatogonia. This has also been observed in ES cells, where loss of individual Argonaute clade members had no obvious effect on ES cell growth (Su et al., 2009). Given the suspected redundancy between the Argonaute clade members in many systems, it was surprising to discover that Ago2 is essential for mouse embryogenesis, hematopoiesis, and oogenesis. It has remained unclear whether the mutant phenotype results from a functional non-redundancy, or occurs due to differential expression of the Argonaute clade members. In this study, we investigate the biological role of the Ago3/1/4 cluster and determine if the unique requirement for Ago2 is attributable to its expression pattern, preferential loading of particular small RNAs, or its unique role as a catalytic enzyme.

Ago2 has been extensively characterized in various organisms, and has been demonstrated to be the catalytic engine of the RNAi machinery. When associated with a small RNA that has a perfect complementarity to its target RNA sequence, Ago2 can cleave the target through its RNase H domain (with the DDH motif as the critical catalytic triad). We have previously demonstrated that Ago2 is the only Argonaute with an active catalytic domain, while Ago1, Ago4 and Ago3 have all been shown to be catalytically inactive (Liu et al., 2004). It is therefore perhaps not surprising to find frequent silent mutations within the otherwise highly conserved DDH motif of Ago1 and Ago4, as opposed to Ago2 and Ago3. Interestingly, Ago3 retains the conserved catalytic residues, and it remains unknown why a catalytic function cannot be detected for this Argonaute protein. The activity of mammalian Argonautes has been well characterized but the assessment of Ago catalytic competence in other vertebrates is restricted to the analysis of the catalytic triad. It is important to note that this is not always an accurate measure and still remains to be tested experimentally or defined, as we understand better Ago catalysis molecular rules. At least analysis of the catalytic triad phylogeny suggests a split in catalytic vs. non-catalytic Argonautes occurring early in vertebrate lineages (Karginov F and Hannon GJ unpublished).

Although cleavage activity of Argonaute proteins has been shown to be important in other organisms, such as yeast, plants and flies, reviewed in (Tolia and Joshua-Tor, 2007), the biological requirement for the cleavage activity in vertebrates has not been demonstrated conclusively. Perhaps the most compelling evidence for the essential role of Ago cleavage during development comes from the studies done on plants where trans-acting siRNAs produced endogenously had important functions in regulating adaxial-abaxial leaf polarity in Arabidopsis or where plant Ago4 catalysis is essential for transcriptional gene

silencing in a locus dependent manner (Chitwood et al., 2009; Qi et al., 2006). Most small RNAs in differentiated tissues are derived from miRNA precusor hairpins that are processed by dicer and generate imperfect RNA duplexes. Although there are microRNAs that have the ability to cleave target mRNA in various tissues, e.g. in the limb or in the placenta, this cleavage activity *in vivo* has not been correlated with down-regulation of nascent transcripts.

Another potential small RNA partner for Ago2-mediated cleavage are siRNAs. Exogenously introduced siRNAs have been shown to bind Ago2 and induce mRNA cleavage in vertebrates. We and others have also found endogenous siRNAs in the murine growing oocytes, and the observation by Surani and colleagues reveals the requirement for Ago2 during oogenesis in mouse. While this indicates a potential link between endo-siRNAs and Ago2-mediated cleavage in murine oocytes, it does not completely correspond with the phenotype observed in the Ago2 mutant mice where Ago2 seems to be essential for miRNA stability. Despite all the small RNA deep sequencing projects, endogenous siRNA have not been identified in somatic tissues during embryonic or postnatal development, which suggests that the cleavage activity of Ago2 is not required in somatic lineages. While endogenous siRNAs were uncovered in Drosha-deficient ES cells (Babiarz et al., 2008; Chiang et al., 2010), the absence of any phenotype in Ago2 deficient ES cells suggests that they are either non-essential, or do not require Ago2 for their function.

In this study, we investigate the evolutionary pressure to conserve Ago2 catalysis through engineering a catalytic inactive Ago2 allele. Using this approach, we demonstrate that the embryonic lethality of our Ago2 mutant is independent of the cleavage activity. Furthermore, we uncover a new role for Ago2 during placental development and describe the purpose of Ago2 catalysis in postnatal development through the definition of a novel miRNA biogenesis pathway.

CHAPTER 2: Mammalian Argonautes: redundant or specialized? Argonautes function in defense against viruses

2.1 Introduction

Argonaute proteins are key effectors of the small RNA mediated regulatory pathways that are involved in modulating gene expression at the transcriptional, post-transcriptional and translational levels in various sub-cellular compartments (Hutvagner and Simard, 2008). Through associations with processed small RNA molecules, Argonaute proteins play a role in complex developmental processes, from oogenesis and spermatogenesis, through embryogenesis, to a fully developed organism. The Argonaute protein contains three highly conserved functional domains: the PAZ and MID domains, required for anchoring the 3'end and the 5'end of the associated small RNA molecule respectively, and the PIWI domain at the C-terminus containing an RNaseH domain that can function to cleave complementary target RNA

The Argonaute protein family is highly conserved in the animal kingdom and co-evolved with a diverse repertoire of small RNAs, whose biogenesis depends on a series of enzymatic processes (Kim et al., 2009). The RNaseIII enzymes, Drosha and Dicer, are responsible for maturation of two classes of regulatory small RNAs (siRNAs and miRNAs), while a subset of Argonautes produces a germ line specific small RNA class known as piRNAs. Regardless of their biogenesis, all regulatory small RNA pathways achieve their effector functions through the formation of an active Argonaute-small RNA silencing complex. Based on sequence similarities, the Argonaute family is subdivided into three major clades: a germline specific PIWI clade that participates in guarding the genome from transposable elements, a nematode-specific WAGO clade that evolved more diverse functions, and the AGO clade that is specialized in siRNA and miRNA mediated gene silencing and has ubiquitous expression throughout the organism (Tolia and Joshua-Tor, 2007).

There are four members of the AGO clade in the vertebrate genome: *Ago1*, *2*, *3* and *4 (Carmell et al., 2002)*. In the mouse, *Ago2* maps to chromosome 15, while the other homologues (*Ago1/3/4*) are encoded in a highly conserved 190kb genomic cluster. Studies have demonstrated similar expression profiles and miRNA associations for the four mammalian Argonaute members, raising the possibility that they have interchangeable biological roles (Azuma-Mukai et al., 2008; Ender et al., 2008; Liu et al., 2004; Meister et al., 2004). Functional redundancy of Argonaute proteins has been observed in spermatogenesis, as conditional knockout of *Ago2* in testes has no obvious phenotype. Although ablation of all Argonautes causes murine ES cells to undergo uncontrolled apoptosis, overexpression of any mammalian Argonaute was able to rescue the phenotype, further supporting the hypothesis (Su et al., 2009). The Developmental requirements of these homologs remain unclear.

Intriguingly, ablation of murine *Ago2* leads to embryonic lethality at early postimplantation stages (E7.5-E10.5) (Alisch et al., 2007; Liu et al., 2004; Morita et al., 2007), revealing a potential non-redundancy in mammalian Argonaute functions. This observation correlates with the characterization of *Ago2* as the sole catalytic mammalian Argonaute (Liu et al., 2004; Meister et al., 2004), suggesting that *Ago2* catalytic functions might be required for embryonic development. However, we cannot discount the possibility that differential expression of *Ago2* during embryonic development contributes to the mutant phenotype, resulting in asymmetric distribution of Argonaute proteins in the developing conceptus. Furthermore, the developmental requirements of *Ago1*, *3* and *4* remain unclear. In order to address these issues, we generated genetic knockouts of murine Argonaute to investigate their biological roles in mammalian development.

2.2. Phenotypic analysis of mammalian *Argonaute* mutants

In order to generate knockouts of Argonautes family members, we took advantage of the gene knockout strategies offered by the International Gene Trap Consortium (IGTC) and the Mutational Insertion and Chromosome Engineering Resource (MICER) (Adams et al., 2004). The gene trap resource is a repository of ES cell clones with mapped single retroviral insertions across the genome. The retroviral construct harbors a splice acceptor site, a selectable/reporter cassette and a polyA signal allowing hijacking of the transcriptional unit of the host gene when inserted in an exon or intron. The gene trap vector insertion is highly mutagenic and results in a truncated protein when inserted close to the 5'end of the gene and downstream of its untranslated. The MICER resource is also a repository of insertional targeting vectors that have genomic inserts in two different backbones bearing LoxP sites and HPRT mini genes for chromosome engineering purposes. These vectors are highly mutagenic when inserted in the middle of the gene, creating a duplication of the homologous genomic region of the insert and vector insertion, often resulting in a frameshift mutations or truncated proteins.

Ago3 was previously ablated using the MICER technology, with mutants viable, fertile and demonstrating no overt mutant phenotype (Carmell MA and Hannon GJ unpublished). To assess the function of *Ago4*, we utilized an *Ago4*-specific MICER clone to generate an insertional mutant at the locus (Fig. 1a), and demonstrated viability and fertility in *Ago4*-deficient animals (Table 1). Ablation of *Ago1* was achieved using ES cells containing a gene-trap insertion in exon 10 of *Ago1* commercially obtained from BayGenomics/IGTC (Fig. 1a). Similar to *Ago4* mutants, loss of *Ago1* has does not result in lethality or sterility (Table 1).

	Ago3FLJ	Ago1GT	Ago4S1	Ago1gt/3FLJ	A1/3/4df1	A1/3/4df2
wild type	2	7	17	87	28	4
heterozygous	2	10	43	146	33	27
mutant	1	4	13	49	14	12
total	5	21	73	282	75	43
p value		0.64	0.25	0.005	0.04	0.055

Table1: Argonautes genetic set allele segregation analysis of heterozygous intercrosses (p vlaues are derived from chi-square analysis).

The lack of overt mutant phenotypes in single knockouts of *Ago1*, *3* and *4* argues that compensatory mechanisms exist to alleviate loss of individual Argonautes. To address these issues, we generated alleles containing multiple Argonaute knockouts. Utilizing the *Ago1* gene-trap insertion mutant, we generated an Ago1/3 double mutant by targeting *Ago3* with an *Ago3*-specific MICER clone (Fig.1b). Due to the close proximity and co-segregation of the two loci, we ensured that the Ago3 targeting vector is integrated in cis. Loss of *Ago1* and *Ago3* results in viable and fertile offspring, but analysis of over 200 animals reveals lower mutant numbers than expected by Mendelian inheritance (p=0.004) (Table 1).

In parallel, ablation of the Ago1/3/4 genomic cluster (ago1/3/4^{df1}) was achieved by targeting MICER constructs containing loxP sites to *Ago4* and *Ago3* (forming the boundary of the genomic cluster), followed by an *in cis* Cre-mediated chromosomal excision *in vivo* using Cre recombinase driven by the Sox2 promoter (Sox2-Cre) (Fig. 1b-c). Although homozygous mutants are viable and fertile, the proportion of wildtype, heterozygous and homozygous offspring differs from the expected Mendelian ratios (p=0.04) (Table 1). An independent ablation of the *Ago1/3/4* chromosomal cluster (ago1/3/4^{df2}) also hints at non-Mendelian inheritance (p=0.055), though the difference between the expected and observed number of homozygous mutants is not significant (Table 1), suggesting that a possible detrimental effect of removing the chromosomal cluster. Analyses of our Argonaute mutants support the hypothesis that Ago2 is the only Argonaute essential and sufficient for normal mouse development, and despite a non-Mendelian inheritance of the *Ago1/3/4* mutant alleles, we suspect that this is due to chromosomal instability caused by the loss of ~190kb from the chromosome.

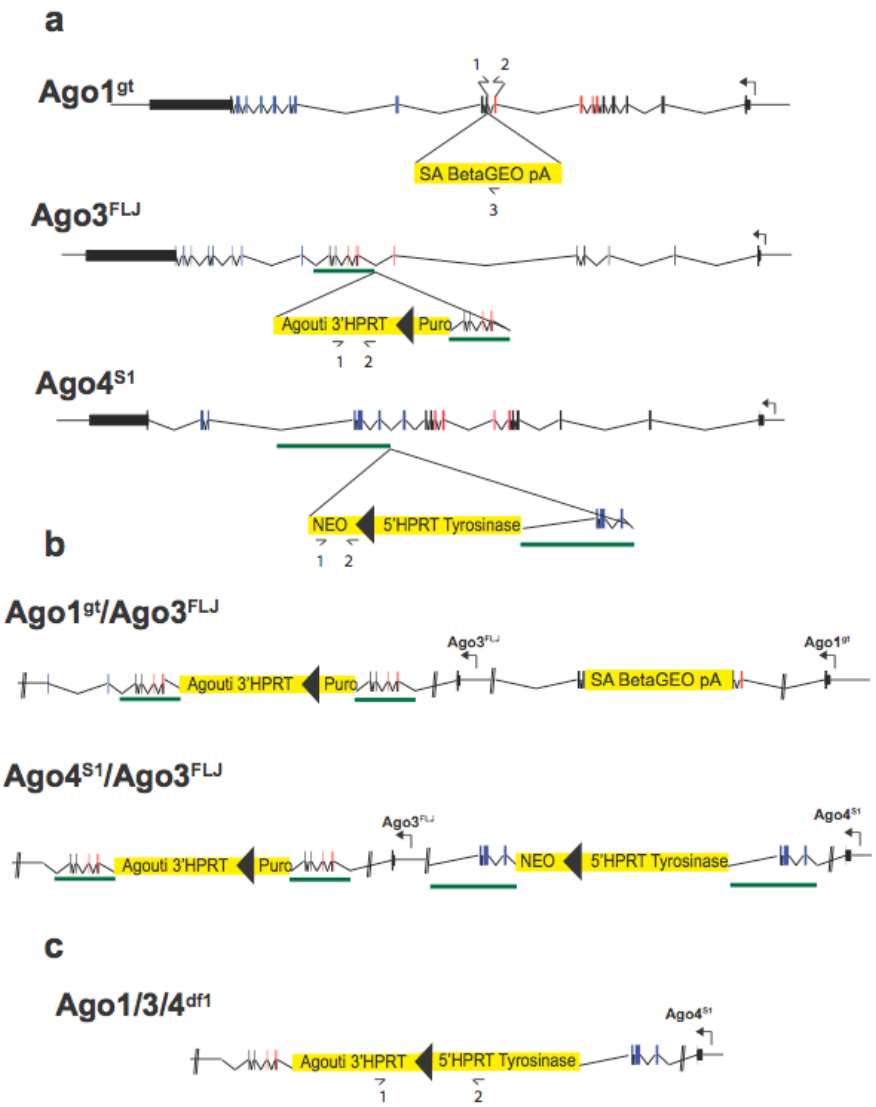

Figure 0.1: Argonaute gene mutants maps and targeting strategies

For all Argonautes genes: ago1, ago3 and ago4, exons, introns surrounding the insertion sites are represented. All transcription start sites are represented with an arrow. 5'UTR and 3'UTR are depicted as black boxes. For every Argonaute protein the exons encoding the PAZ domain are painted in red and the exons encoding the PIWI domain are painted in blue. **a.** Single insertional mutant alleles: Ago1^{gt} :gene trap allele map representing the site of the retroviral integration. SA: splice acceptor. BetaGEO is selection cassette encoding the neomycin resistant gene and a beta-galactosidase enzyme.pA: poly A tail. Ago3^{FLJ}: MICER insertion in the Ago3 gene creating a duplication in the middle of the gene (green) and an insertion of the 3'HPRT vector. Ago4^{S1}: Ago4 MICER insertion allele is similar to Ago3^{FLJ} except that the inserted vector is the 5'HPRT version. **b.** Ago1^{gt}/ Ago3^{FLJ} double knockout allele using the Ago1^{gt} allele targeted with the Ago3^{FLJ} vector. Ago4^{S1}/Ago3^{FLJ} doubly targeted allele with two MICER alleles combining the previous single Ago4 and Ago3 knockouts. **c**. Ago1/3/4^{df1} allele resulting from Cre excision of the Ago4^{S1}/Ago3^{FLJ}. The two HPRT mini genes from each vector are joined to make a functional HPRT gene rendering the cells HAT resistant. In all the schematics the genotyping primers are depicted with numbers: 1-3.

2.3. Characterizing small RNA partners of Argonautes

Although *Ago2* can compensate for the loss of *Ago1/3/4*, the other Argonautes cannot rescue the ablation of *Ago2*. The lack of biological redundancy led us to hypothesize that, in addition to its catalytic activity, *Ago2* associates with unique small RNA partners. Taking advantage of ES cells derived from the Ago2, Ago3/4 and Ago1/3/4 mutant blastocysts, we investigated Argonaute-mediated repression to identify differential activity between the four mammalian Argonautes.

First, we assessed RNAi responses in *Ago2* -/-, *Ago1/3/4* -/- and *Ago3/4* +/- (control) ES cells using siRNA and miRNA luciferase reporter systems. We confirmed that miRNA induces reporter silencing in Ago2 null ES cells, while siRNA-mediated repression is absent, supporting previous observations made in *Ago2*-deficient MEF. In contrast, siRNA and miRNA-mediated repression are present in Ago1/3/4 deficient cells at a level compared to the control (*Ago3/4* +/-) cell line (Fig. 2a-b), indicating that *Ago2* is sufficient to mediate small RNA-mediated regulation in absence of other Argonautes.

We hypothesized that *Ago2* loss results in destabilization of their associated small RNA partners, and therefore allow us to identify them through their depletion in *Ago2*-deficient cells. We generated small RNA libraries from wild type, Ago2 null, Ago3/4 double knockout and Ago1/3/4 deficient ES cells. Analysis of miRNA population revealed invariance in virtually all miRNAs between different genetic backgrounds (fig. 2c), arguing against our hypothesis. To exclude compensatory effects resulting from the loss of *Ago2*, we purified *Ago1* and *Ago2* complexes from wild type ES cells and sequenced their associated small RNA populations. About 90% of the reads in both complexes represent miRNAs, with the remainder mapping to genic regions, repeats or unannotated genomic regions (fig. 2d). Comparing the miRNA proportions between *Ago2* and *Ago1* immunoprecipitated complexes reveals no preferential loading of miRNAs that are highly enriched in ES cells (fig. 2e). Our results are in agreement with other studies, showing no preferential sorting of miRNAs in mammalian Argonaute complexes, though their analyses have been limited to annotated miRNAs (miRBase). By extending our analysis to the unannotated genomic regions, we detect an abundant small RNA read (~ 5000 reads) found in both AGO1 and AGO2 IPs, containing a fixed 5' end and a heterogeneous 3' end (fig. 3a-b). The small RNA maps immediately downstream a t-RNA transcript, and conserved only in rodents (fig. 3b). Interestingly, a 50nt sequence starting from the fixed 5' end of this small RNA molecule can fold into a secondary stem-loop structure (fig. 3c), with the detected small RNA extending from the stem into the loop region of the hairpin. Further characterization is required to determine if this is a novel miRNA, or a different small RNA species.

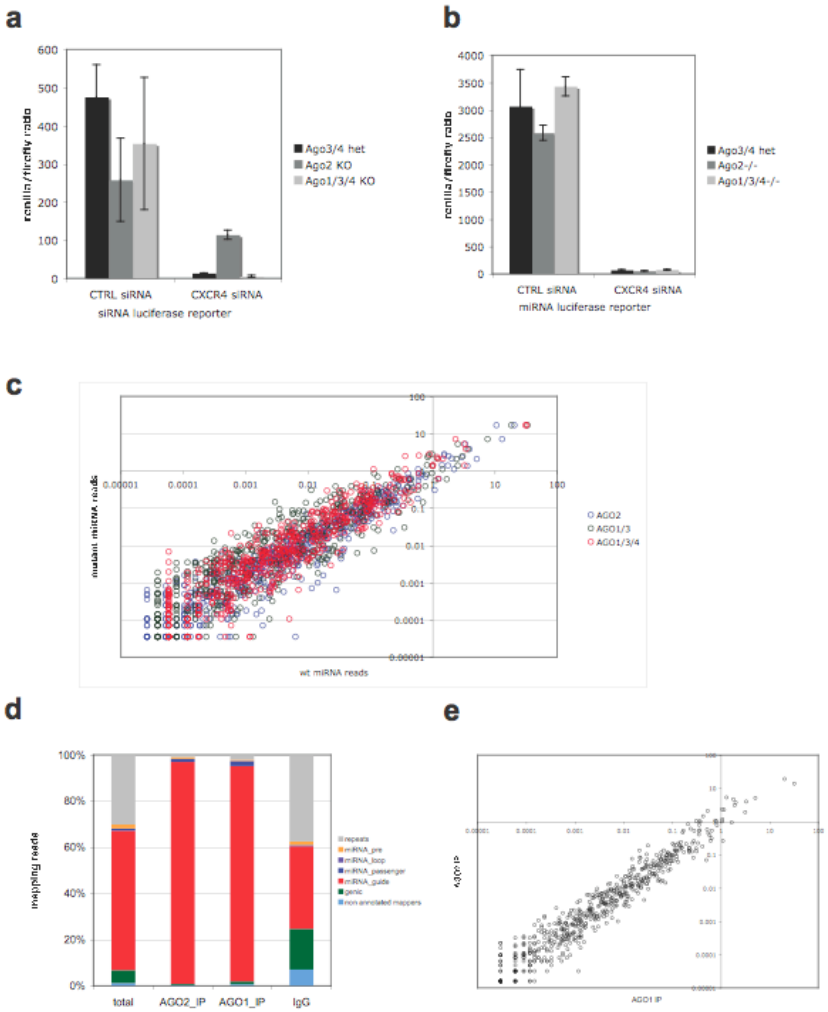

Figure 2: Characterization of Argonautes mutant ES cells and analyzing miRNAs populations in mutants cells and Argonaute complexes.

a. Histogram representing a dual luciferase reporter using a CXCR4 siRNA reporter system testing the Ago2 null cells and Ago1/3/4 KOs. **b.** Histogram representing a dual luciferase reporter using a CXCR4 miRNA reporter system testing the Ago2 null cells and Ago1/3/4 KOs. Error bars represent the standard deviation of three technical replicates **c.** Scatter plot of miRNA percentages plotted in a logarithmic scale from different Ago genetic backgrounds compared to a wild type library. **d.** Histogram showing small RNA library subsets from wild type total RNA and RNA extracted from AGO1 and AGO2 complexes. The IgG library is used as a control. miRNAs reads are subdivided into: miRNA loop, miRNA passenger reads, miRNA pre (reads mapping to precursor region of miRNAs), miRNA guide reads. The remaining categories are reads mapping to repeats, genes or mapping to the genome with no known annotation. **e.** Scatter plot of miRNA percent reads in the AGO1 and AGO2 complexes plotted in a logarithmic scale.

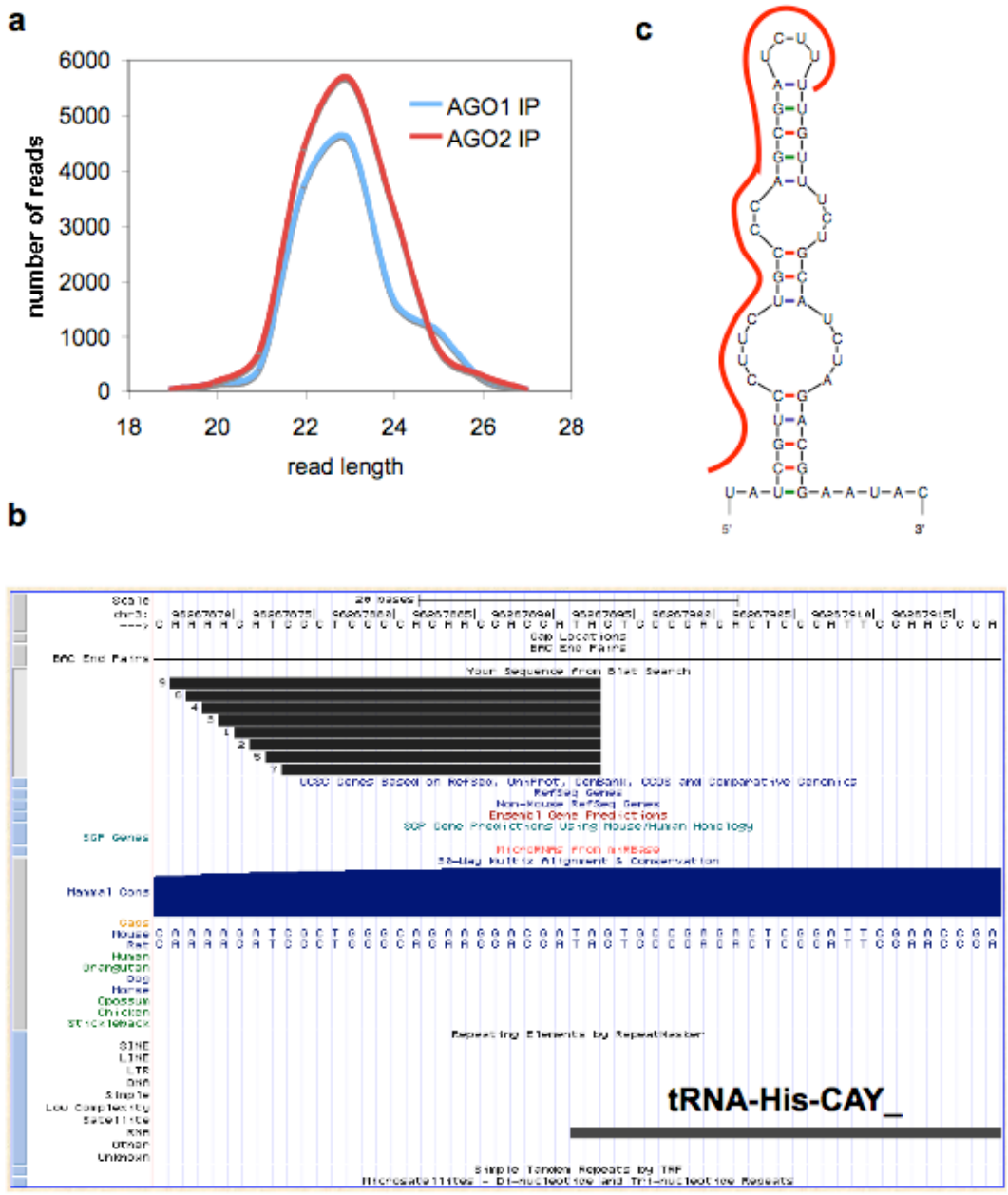

Figure 0.2: Novel small RNA prediction from Argonaute complexes

a. Read length distribution of a small RNA molecule cloned from AGO1 and AGO2 IPs that maps to the genome but is unannotated showing a peak at 23nt. **b**. Genome map showing sequences reads of the non-annotated small RNA from AGO1 and AGO2 IP libraries with a fixed 5'end and a heterogeneous 3'end. Note the tRNA mapping upstream of the small RNA sequence. Both the small RNAs and tRNA map antisense to the genome. **c.** Short hairpin showing secondary structure prediction of ~50nt region encompassing the cloned small RNA. The mature 23nt nucleotide sequence is highlighted in red.

2.4. Roles of Ago1 and Ago3 in viral defense

The absence of overt phenotypes in *Ago1/3/4* mutants does not preclude their requirement under certain stress stimuli or disease states. The variable susceptibility of the mutant mice to these conditions can potentially explain the incomplete penetrance observed in the Ago1/3 and Ago13/4 knockouts. miRNA analyses and Dicer deficiency studies suggest a role of RNAi in viral and stress response (Babar et al., 2008; Weidhaas et al., 2007). For example, Dicer ablation in mice results in increased susceptibility to vesicular stomatitis virus through the activity of miR-24 and miR-93 in targeting viral genes (Otsuka et al., 2007). Our Argonaute mutant mouse models provide a tool to probe for the involvement of Argonaute proteins during viral infection. Dr. Melanie Van Stry, a postdoctoral fellow in Dr. Mark Bix's lab at St. Jude Children's Research Hospital challenged our Ago1/3 double and single knockouts with influenza A virus to assess their immune response. The virus is delivered by aerosols, where it infects and replicates in epithelial cells of the respiratory tract. Ago1/3 DKO mice have increased susceptibility to influenza A virus lab strain with a higher mortality rate of the Ago1/3 DKO. Survival curves reveal a lower survival rate (~30%) of the mutants compared to ~60% in wild type mice (fig. 4.a). Virus titers in the lung of wild type and mutant animals are comparable, indicating *Ago1/3* loss does not enhance viral replication (fig. 4.b). Further analysis of the wild type and mutant host immune systems in response to the viral infection will help determine the role of *Ago1* and *Ago3* in host defense against the flu.

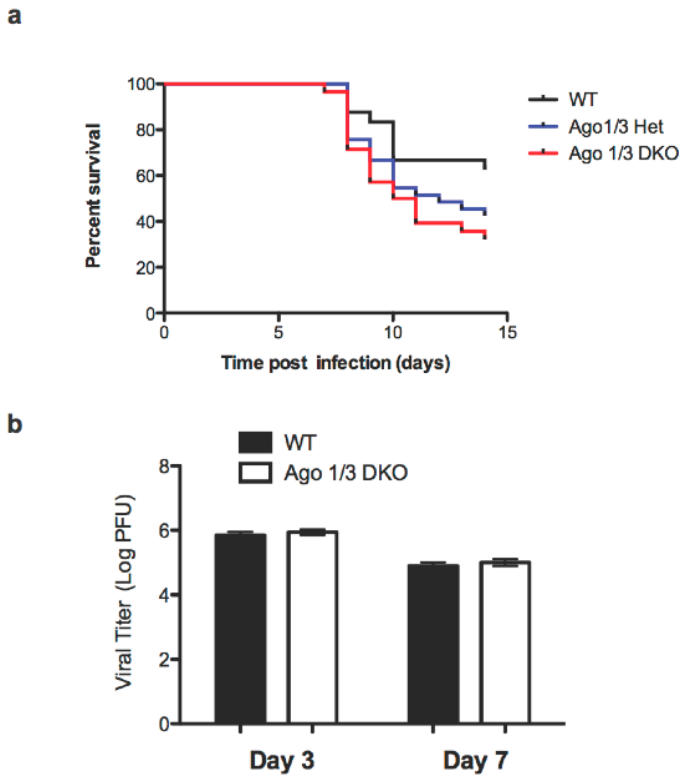

Figure 4: Ago1/3 double KO has increased susceptibility to influenza A (PR8) infection.

a. Mice were monitored for survival. Ago1/3 DKO mice exhibit an increased mortality compared to wt littermates. Data are average of three independent experiments (p=.026, Log-rank Mantel Cox Test). Wild type (N=25), Ago1/3 +/- (N=34) and Ago1/3-/- (N=29). Mice were infected with 2000 EID_{50} of PR8 intranasally. **b.** wild type and Ago DKO mice have similar virus titers in the lungs. Virus titers from lungs of infected animals were determined with standard MDCK plaque assay. Day3 time point: wt (N=15) and Ago1/3 DKO (N=16). Day7 time point: wt (N=18) and Ago1/3 DKO (N=15). Error bar are standard error of the mean of three replicates.

2.5. Discussion

The canonical RNAi pathway proposes Argonautes requires the presence of Dicer-generated small RNAs in order to function, leading to the hypothesis that ablation of mammalian Dicer should phenotypically mimic the loss of all mammalian Argonautes. It is important to compare phenotypic differences of Dicer and Argonaute mutants in order to identify Dicer-independent functions of Argonaute proteins. For example, the loss of *Dicer* leads to dramatic defects in differentiation and proliferation of spermatogonia (Hayashi et al., 2008). This also holds true for Embryonic Stem (ES) cells. We have previously reported that conditional inactivation of dicer in ES cells abrogates miRNA biogenesis and compromises cell proliferation(Murchison et al., 2005). However, the cells survive and we were able to maintain them in culture and use them as a cellular model to study RNAi deficiency. As expected, the dicer null cells fail to mount an RNAi response when transfected with long double stranded RNA. Interestingly, dicer loss affects abundance of transcripts from centromeres without pronounced effects on histone modifications at pericentromeric repeats or DNA methylation in centromeres. A recent study reported the essential role of Argonautes in ES cell survival (Su et al., 2009). The apoptotic effect of Argonautes deficient ES cells can be rescued by introducing activated Akt to target Bim the proapoptotic target. Interestingly, overexpression of any of the four Argonaute members was able to rescue the phenotype. Argonautes are highly conserved throughout evolution, with homologues found as evolutionary distant as archea and eubacteria. The lack of canonical RNAi pathways in these organisms suggests alternative biological roles and functions. Biochemical and structural analysis of the Argonautes protein domains indicate that it has evolved as a small RNA binding protein, with potential endonucleolytic activity dependent on the conservation of catalytic residues within the PIWI domain (Liu et al., 2004). Our extensive analysis of RNA-Argonaute complexes indicates no preferential loading of miRNAs in mammalian Argonaute members, but we cannot exclude the possibility that other small RNA are differentially loaded. We have also identified a novel small RNA associated with AGO1 and AGO2 complexes, whose maturation may be mediated by the processing of a t-RNA precursor transcript encompassing the small RNA. The 3' heterogeneity of this molecule can be explained through exonuclease-mediated trimming of the short hairpin not protected by the Ago1 or Ago2 protein. Similar analysis needs to be extended to small RNAs derived from genic regions and repeat elements to identify additional candidates. Small RNA libraries from different species should also be analyzed in a similar manner. It is important to note that our sequencing studies of these complexes are limited to small RNA fractions and excludes longer non-coding RNAs that associate with Argonautes.

Given the sufficiency of *Ago2* as an effector in small RNA pathways, the conservation of the other Argonautes is intriguing. The early embryonic lethality of *Ago2* mutant mice precludes us from identifying developmental niches where *Ago2* can be compensated by its homologs, though other developmental processes, such as oogenesis, have been identified to be completely dependent on *Ago2* (Kaneda et al., 2009; Lykke-Andersen et al., 2008). Several explanations can account for the unique developmental requirement of *Ago2*: 1) *Ago2* is the only Argonaute to conserve its catalytic activity, 2) The pattern and level of *Ago2* expression during embryonic development are not mimicked or compensated by other Argonautes, and 3) *Ago2* interacts with proteins not associated with the other homologs or has unique protein modifications. We have evidence that embryonic requirement of *Ago2* might be due to a combination of aforementioned reasons, but further investigation is required to dissect the molecular mechanisms. In addition, the postnatal requirement of *Ago2* have not been extensively studied except in red blood cell development, B cell development (O'Carroll et al., 2007), oogenesis and spermatogenesis, and we cannot discount the possibility that *Ago2* is critical for various cellular processes in the adult organism.

The discovery of Ago1/3-dependent susceptibility to influenza A virus raises the possibility that Argonaute proteins have non-redundant roles in response to viral infections or stress. The early lethality of the Ago2 mutants makes it difficult to detect potential synergistic effects of the Argonautes, although conditional ablation of *Ago2* could allow characterization of Argonaute requirements in cellular niches critical for host viral responses. Preliminary, analysis from bronchoalveolar lavage reveals no differences in the leukocytes populations (data not shown), although a more detailed analysis of the lung's residence immune cells and secondary sites of the immune system are underway to test for Ago1 and Ago3 functions in providing a defense mechanism against influenza A virus.

Our collaborators are also investigating Argonautes's role at the primary site during viral infection using primary mouse tracheal epithelial cells (mTECs) as an in-vitro model. We are also profiling small RNAs from infected vs. uninfected cells in the presence or absence of *Ago1/3*, which will reveal small RNAs that might be important for viral response. We are not expecting to see preferential down-regulation of miRNAs in the *Ago1/3* double mutants due to *Ago2* compensation, although there is evidence suggesting that ubiquitous expression of Argonautes previously reported in literature is incorrect, and we may identify differential expression of Argonaute proteins in various population of immune cells isolated by immunophenotyping.

We are considering analyzing the triple KOs (ago1/3/4), which we believe will have a more dramatic susceptibility due to Argonaute dosage, as heterozygous *Ago1/3* DKO have an intermediate response when compared to the wild type and the homozygous KO animals. We cannot exclude the possibility

that other tissues are affected, and our collaborators are performing extensive analysis on the full pathology of the infected animals. This analysis could help identify to the tissue of interest that contributes to increased rate of mortality.

It is important to note that although a subset of *Ago1/3* DKO and *Ago1/3/4* triple knockout animals are viable and allow us to perform different experimental challenges, a proportion of mice are lost by the time of weaning. A detailed statistical analysis of genotypic proportions at different developmental time points is required to uncover the possible mechanism behind non-Mendelian inheritance of the triple knockout mutants.

2.6. Conclusions and future directions

Our Argonaute mouse mutants reveal that Ago2 is essential and potentially sufficient for murine development, although its role in various biological processes remains to be characterized. Our small RNA profiling data indicate that no preferential loading of miRNAs into different Argonautes complexes, confirming previous observations. However, we identified potentially novel small RNAs molecules that are generated through a non-canonical small RNA biogenesis pathway. We provide evidence for the susceptibility of the *Ago1/3* double knockout mice to influenza A virus that remains to be explored at the molecular level, and propose that Argonaute family members may function synergistically to provide a small RNA mediated defense mechanism against cellular stress or disease conditions.

2.7. Methods

ES cell targeting and genotyping strategies:

The Ago3 targeting vector (FLJ12765) was previously available from a Michelle Carmell, a former graduate student in the lab. In order to orient the 3'HPRT mini gene cassette to allow double targeting of the Ago1/3/4 chromosomal cluster for subsequent deletion, the vector backbone had to flipped to the correct orientation using the unique AscI cloning sites of the insert. The Vector was then readily available for targeting in ES cells by generating a GAP with NedI restriction ezyme of about 2.9kb within the 7kb insert. Two ES cell lines were targeted with the AgoFLJ vector for two different purposes: AB2.2 wild type ES cells that are HAT sensitive for chromosome engineering of the Ago1/3/4 locus and Ago1 gene trap ES cells (RRR031) obtained from the gene trap resource to generate the $Ago1^{gt}/Ago3^{FLJ}$ doubly targeted chromosome. For the latter targeting in cis vs trans Ago3 insetions were screened in vivo within the first

generation (fig. 1a-b). All positively targeted ES cell clones were screened using southern blot strategy based the map construction of the inserted vector within the Ago3 locus: HindIII was predicted as diagnostic enzyme that results in a 21kb wild type fragment and 15kb targeted fragment when using a probe mapping to the GAP region. The primers used to amplify the southern blot probe are the following: forward-TCAGAGTTTTGTACTAACCAA & reverse-AAGTGATTTAATAATGTAGAC

The Ago4S1 targeting vector (MHPN30h12) was obtained from the sanger MICER resource (http://www.sanger.ac.uk/PostGenomics/mousegenomics/). The ~11kb insert of Ago4 was linearized with HpaI enzyme and targeted in wild type AB2.2 ES cells to generate Ago4 single knockout mice and Ago3 positively targeted ES cells to generate the conditional Ago1/3/4 chromosomal cluster. The Ago4 targeted map the built to devise a southern blot strategy to screen for positive integrants. Using an external probe and PmeI restriction enzyme, resulting a 13.5kb wild type fragment and 11.7kb targeted fragment. The primers used to amplify the probe are the following: Forward- tttagtataagagtataaacac& reverse- aagggttgggggagaagaggg

All positive ES cell clones were confirmed for correct targeting before blastocyst injection. The chimeric mice were then crossed to B6 mice to test for germline transmission. To screen for positive the heterozygous progeny, mice were genotyped by using primers specific to the corresponding targeting vector. For the Ago4 targeted locus the following primers were used: forward1: atgggatcggccattgaa and reverse2: gaactcgtcaagaaggcg. For the ago3 targeted locus the following primers were used: forward1: gacggtgaagaagcacaggaa and reverse2: ggtccgatgggaaagtgtagc. Homozygous progeny was genotype using the corresponding genotyping strategy decribed above. For the $Ago1^{gt}$ allele, a PCR genotyping strategy was devised to differentiate between wild type, heterozygous and homozygous mice. First the gene trap insertion site was confirmed to map to exon10 of the Ago1 gene. The mice were genotyped using the following primer: forward primerF1 (cccagacagacaggaggaga) is common to the wild type and targeted allele. The reverse primer for the wild type allele R-2 is ccaaattcctggatgtaggg and the reverse primer for the targeted allele R-3 is gttttcccagtcacgacgtt. To generate the $Ago1/3/4^{df1}$, the conditional $Ago4^{S1}/Ago3^{flj}$ were crossed to SOX2 Cre females and the progeny was screened for the deletion using the HPRT primer pair provided by Sanger: F1: ccttcatcacatctcgagcaagacgttcag & R2: aagggtgtttattccccatggactaattatg. The homozygous progeny was screen using the combination of the latter pair and the Ago1 primer pair (see fig. 1 for primer maps).

Luciferase reporter assays

Dual luciferase assays were performed as previously described. Briefly, for testing RNAi responses in ES cells, a renilla plasmids (with no artificial site or 1 perfect (siRNA) CXCR4 site or 6 buldged (miRNA) CXCR4 sites were

cotransfected with or without perfectly matched CXCR4 siRNA duplex (dharmacon).

Small RNA library construction and sequencing

Total RNA from ES cells or immunoprecipitated material was extracted using trizol. Two Small RNA libraries with a size range of 19-24nt were generated using a modified small RNA cloning strategy. Briefly, the small RNA fraction was ligated sequentially at the 3'OH and 5'phosphates with synthetic linkers, reverse transcribed and amplified using solexa sequencing primers. Around 7 million reads were generated for each small RNA library. Sequences were then trimmed from the 3' linker, collapsed and mapped to the mouse genome with no mismatches using multiple annotation tracks, namely: UCSC genes miRNAs and repeats. For this study we used the mirbase database to annotate the cloned miRNAs and analysed the mapping non-annotated reads.

Argonaute immunoprecipitation

Prior to cell collection mouse monoclonal antibodies (anti-ago2 from abnova also known as eif2c2 M01-clone 2E12-1C9, anti-Ago1 clone 6D8.2 and normal IgG both from millipore) were conjugated on protein A Dynabeads magnetic beads (sigma) using Rabbit anti-mouse bridging antibody. ES cells were grown on gelatinized plates using ES cell growth medium as described previously. About 30X 15cm plates were washed in PBS, scraped and span at 1200RPM. The cell pellet was resuspended in and lysis buffer (10mM Hepes pH 7.0, 100mM KCL, 5mM MgCl, 0.5% NP-40, 1% Triton X-100, 10% glycerol, 2mM DTT, mini EDTA free protease inhibitor tablets (ROCHE) 100u/ml RNASIN), dounced and incubated rotating at 4°C for 20min. The lysate were span at maximum speed for 30min at 4°C. The supernatant collected and a fraction was saved for total RNA and protein extractions. The lysates was then diluted 5 times in NT2 buffer (50mM Tris pH 7.4,150mM NaCl, 1mM MgCl2, 0.05% NP40, 100U/ml RNASIN, 1mM DTT), split in three portions and incubate with Ago1, Ago2 and IgG antibodies for 4-6hours, rotating at 4°C. The beads were then collected using a magnetic stand and wash twice with NT2 buffer for 10-15min and two additional times with a higher salt NT2 buffer (300mM). Tubes were changed for the last wash. ¼ of the IP was saved for western and the rest for RNA extraction. The RNA was used to construct the small RNA libraries as described above.

Mouse influenza A viral infections and viral titers

Mouse groups were infected with 2000 EID_{50} of the virulent A/Puerto Rico/8/34 laboratory strain (H1N1; PR8) intranasally and were monitored over the course of two weeks for morbidity (weight loss) and mortality. The amount of virus in the lungs was determined by using standard MDCK plaque assays to measure viral titers in whole lungs at Day3 and Day7 post infection.

2.8. Acknowledgments

I would like to thanks Michelle Carmell for transferring all the reagents smoothly as I started this project. Many thanks to my collaborators Stefan Muljo and Chryssa Kanellopoulou for sharing unpublished reagents of their independent Ago1/3/4 deficiency allele. My sincere acknowledgments to Chris Hsiung for his hard work on the small RNA profiling and for driving the project to completion within a short time during the summer undergraduate research program at the labs. I am most grateful to my Immunologist collaborator Melanie Van Stry for sharing her Ago1/3 viral immunity data and for being so patient explaining to me all the virus experiments and sharing her research proposal with me.

CHAPTER 3: A dicer-independent miRNA biogenesis pathway that requires Ago catalysis

Sihem Cheloufi[1,2], Camila O. Dos Santos [1], Mark M.W. Chong[3,4] and Gregory J. Hannon[1]*

[1] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
Howard Hughes Medical Institute
Cold Spring Harbor,NY 11724

[2] Graduate Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[3] The Kimmel Center for Biology and Medicine at the Skirball Institute
New York University School of Medicine
New York, NY 10016

[4]The Walter and Eliza Hall Institute of Medical Research
Parkville, Victoria 3052
Australia

*To whom correspondence should be addressed

3.1. Abstract

The nucleolytic activity of animal Argonaute proteins is deeply conserved, despite its having no obvious role in microRNA-directed gene regulation. In mice, Ago2 is uniquely required for viability, and only this family member retains catalytic competence. To investigate the evolutionary pressure to conserve Argonaute enzymatic activity, we engineered a mouse with catalytically inactive Ago2 alleles. Homozygous mutants died shortly after birth with an obvious anemia. Examination of microRNAs and their potential targets revealed a loss of miR-451, a small RNA important for erythropoiesis. Though this microRNA is processed by Drosha, its maturation does not require Dicer. Instead, the pre-miRNA becomes loaded into Ago and is cleaved by the Ago catalytic center to generate an intermediate 3' end, which is then further trimmed. Our findings link the conservation of Argonaute catalysis to a conserved mechanism of microRNA biogenesis that is important for vertebrate development.

3.2. Introduction

Argonaute proteins are the key effectors of small RNA-mediated regulatory pathways that modulate gene expression, regulate chromosome structure and function, and provide an innate immune defense against viruses and transposons[1]. The structure of Ago proteins is well conserved, consisting of an amino-terminal domain, the mid domain, and their signature PAZ and Piwi domains. Structure-function relationships in this family are becoming increasingly well understood[2]. The PAZ and Mid domains help to anchor the small RNA guide, with PAZ binding the 3' end using a series of conserved aromatic residues and the Mid domain providing a binding pocket for the 5' end. The Piwi domain contains an RNAse H motif that was cryptic in the primary sequence but easily recognizable in the tertiary structure. Loading of a highly complementary target into an Ago brings the scissile phosphate, opposite nucleotides 10 and 11 of the small RNA guide, into the enzyme active site, allowing cleavage of the RNA to leave 5' P and 3' OH termini[3-7].

Ago proteins can be divided into three clades. The Piwi clade is animal specific, and forms part of an elegant innate immune system that controls the activity of mobile genetic elements[8]. The Wago clade is specific to worms and acts in a variety of different biological processes[9]. The Ago clade is defined by similarity to *Arabidopsis* Ago1[10]. Ago-clade proteins are found in both plants and animals where one unifying thread is their role in gene regulation. In plants, some Ago family members bind to microRNAs and are directed thereby to recognize and cleave complementary target mRNAs[11, 12].

Animal microRNAs function differently from their plant counterparts, with nearly all microRNA-target interactions providing insufficient complementarity to properly orient the scissile phosphate for cleavage. Here, target recognition relies mainly on a "seed" sequence corresponding to miRNA nucleotides 2-8. While complementarity of the target to other segments of the miRNA can contribute to recognition, seed pairing appears to be the dominant factor in determining regulation[13]. A very few extensive microRNA-target interactions can lead to mRNA cleavage in mammals[14, 15]. However, none of these has yet been shown to be critical for target regulation[16-18].

Despite the fact that animal microRNAs regulate targets without Ago-mediated cleavage, the Argonaute catalytic center is deeply conserved. This consists of a catalytic DDH triad that serves as a metal coordinating site[19]. Examining the evolution of mammalian Ago proteins suggests a split into catalytic and non-catalytic family members occurring early in the vertebrate lineage (FV Karginov and GJH, unpublished). Of the four Ago-clade proteins in mammals, only Ago2 has retained both the DDH motif and demonstrable endonuclease activity[20-22]. Ago 1, Ago 3, and Ago4 are linked within a single

~190 KB locus and have lost catalytic competence. An analysis of Ago2 mutant cells has indicated that proteins encoded by the Ago 1/3/4 locus can support miRNA-mediated silencing[20]. This leaves us without a clear explanation for the maintenance of a catalytically competent Ago family member, since miRNAs are the exclusive partners of these proteins in virtually all cell types[20, 23, 24].

In mammals, endogenous siRNAs have been detected in abundance only in oocytes and ES cells[25, 26] (GJH, unpublished). In oocytes, these arise, in part, from dsRNAs formed from the interaction of antisense pseudogene and sense protein-coding transcripts[26, 27]. Loss of key components of small RNA biogenesis or effector pathways, Dicer, DGCR8, or Ago2, from growing oocytes indicates the importance of endo-siRNAs for gene regulation and ultimately for proper chromosome segregation in that cell type[28-31]. Although this phenomenon could provide one explanation for conservation of catalysis, Ago2-null mice are not simply female sterile. Instead, lethality is observed from early to mid gestation[20, 32, 33]. This strongly suggests unique functions for the single catalytic Ago protein outside of the female germline. We envisioned that this could be due to Ago2 being the only family member expressed in a critical cell type or domain in the developing embryo or to a previously unsuspected and critical role for catalysis.

3.3. Ago2 is required for extraembryonic development

A number of studies have indicated that all four mammalian Ago proteins share broadly similar expression patterns and largely identical populations of bound small RNAs[20, 23, 24, 34]. However, most of these were carried out by looking grossly at tissues of adult animals or by using continuous cell lines. We wished to probe the possibility that Ago2 occupies a unique expression domain during embryogenesis that might provide an explanation for the lethal phenotype of mutations in this gene. Toward this end, we took advantage of existing gene trap reporters, which bring lacZ under the control of the Ago1 or Ago2 promoters. Staining of mid-gestation embryos demonstrated broad expression of Ago2 in both the embryonic and extraembryonic compartments. In contrast, Ago1 expression was restricted to the embryo (Fig. 1a). Our previous studies also demonstrated Ago3 expression within the developing embryo[20].

Around embryonic day 8.5, the allantois grows out from the posterior mesoderm of the embryo to invade the chorion. This contact is essential for branching of the allantois structure and formation of a complex network of blood vessels that invade the trophoblast cells to form the labyrinthine layer. This is the site of nutrient, waste, and gas exchange between the maternal blood islands and the embryo[35].

We previously observed growth retardation, cardiac abnormalities, and aberrant neural tube closure in Ago-2-mutant embryos. Based upon

comparisons to other mutants with similar phenotypes, this seemed strongly indicative of a defect in extraembryonic development[36]. An examination of this compartment revealed morphological abnormalities in implanted decidua and hypotrophic placentas in Ago2 mutants (Fig. 1b, c). Histological analysis of placental cross sections

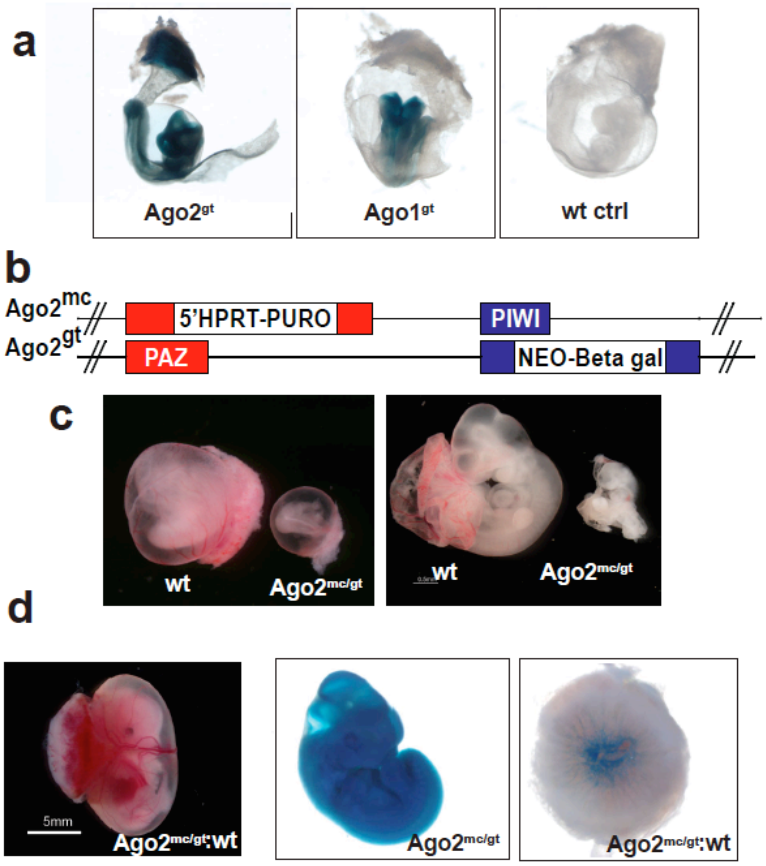

Figure 1: Ago2 is essential for extraembryonic development

a. LacZ whole mount staining of E9.5 embryos bearing the Ago2^{gt} or Ago1 ^{gt} gene trap beta-galactosidase reporter alleles and wild-type controls (wt ctrl). **b.** Allelic combination in Ago2 insertional mutant embryos, with the structure of each allele shown. **c.** Example wild type (wt) and mutant (Ago2m$^{mc/gt}$) embryos from heterozygous intercrosses. Left panel: E10.5 embryos within their embryonic yolk sac and placentas. Right panel: embryos dissected from their extraembryonic components. **d.** Representative E12.5 chimeric embryo developed from Ago2 null ES cells aggregated with wild-type tetraploid embryos. From left to right: whole chimeric embryonic conceptus (Ago2$^{mc/gt}$:wt), Beta-galactosidase staining of the whole embryo showing contribution of Ago2 null ES cells (Ago2 $^{mc/gt}$), Beta-galactosidase staining of the placenta showing contribution of the ES cells to the vasculature invading the placental labyrinth (Ago2$^{mc/gt}$:wt).

revealed a dramatic reduction in the thickness of the labyrinthine layer (Fig. S1a). In contrast, the spongioblast and giant trophoblast layers appeared unaffected. This was consistent with detailed analysis of Ago2 gene trap animals, which verified the presence of the reporter strongly in the affected cell types (Fig. S1b).

The malformation of the labyrinthine layer could be due either to a failure of the embryonic vasculature to invade the placenta, or an inherent defect in the cells derived from the trophoblast (extraembryonic) lineage to support the embryonic vasculature. To distinguish between these two possibilities, we performed tetraploid complementation. Ago2-null embryonic stem (ES) cells were derived from mutant embryos (Fig. S2) and aggregated with tetraploid cells derived from wild-type embryos. The latter can form extraembryonic tissues but cannot contribute to the embryo proper. One Ago2 allele was marked by the gene-trap lacZ reporter, allowing us to visualize the contribution of null cells to the embryo (Fig. 1d). The presence of a wild-type extraembryonic compartment was able to bypass the mid-gestational death of Ago2-null embryos. This verified an essential role for Ago2 in placental development. However, viable mice were not obtained, suggesting additional critical Ago2 functions during later embryonic or perinatal development.

3.4. Ago2 catalytic activity is important for postnatal development

The requirement for Ago2 during embryonic development could be based solely upon its being the only Argonaute protein abundantly expressed in critical extraembryonic cell types or it could depend upon Ago2 being catalytically competent. We, therefore, sought to generate a mouse in which Ago2 was still expressed but had been rendered catalytically inactive. We had previously shown that substitution of alanine for either of the two aspartate residues within the catalytic DDH motif disrupted RNA cleavage without impeding small RNA binding[20]. We therefore replaced the endogenous Ago2 allele with one carrying an ADH rather than a DDH triad (Fig S3).

We intercrossed heterozygous $Ago2^{ADH}$ mice and scored the ratios of each resulting genotype at various developmental time points (Supplementary Table 1). Expected Mendelian ratios were observed up until the point of weaning at which time viable homozygotes were no longer present.

This demonstrates that Ago2 catalysis is not required for extraembryonic development and strongly indicates that it is simply the presence of Ago2 as the dominant microRNA partner in this tissue that leads to its loss of function phenotype. We have identified many abundant placental microRNAs that could underlie the observed effects via their association with Ago2 (Fig. S4). Catalytic competence is, however, required for viability. Mice lacking a catalytic Ago2

became pale after birth and generally died within a few hours (Fig. 2a). Histological examination revealed no gross morphological defects. However, the appearance of these animals was strongly indicative of anemia.

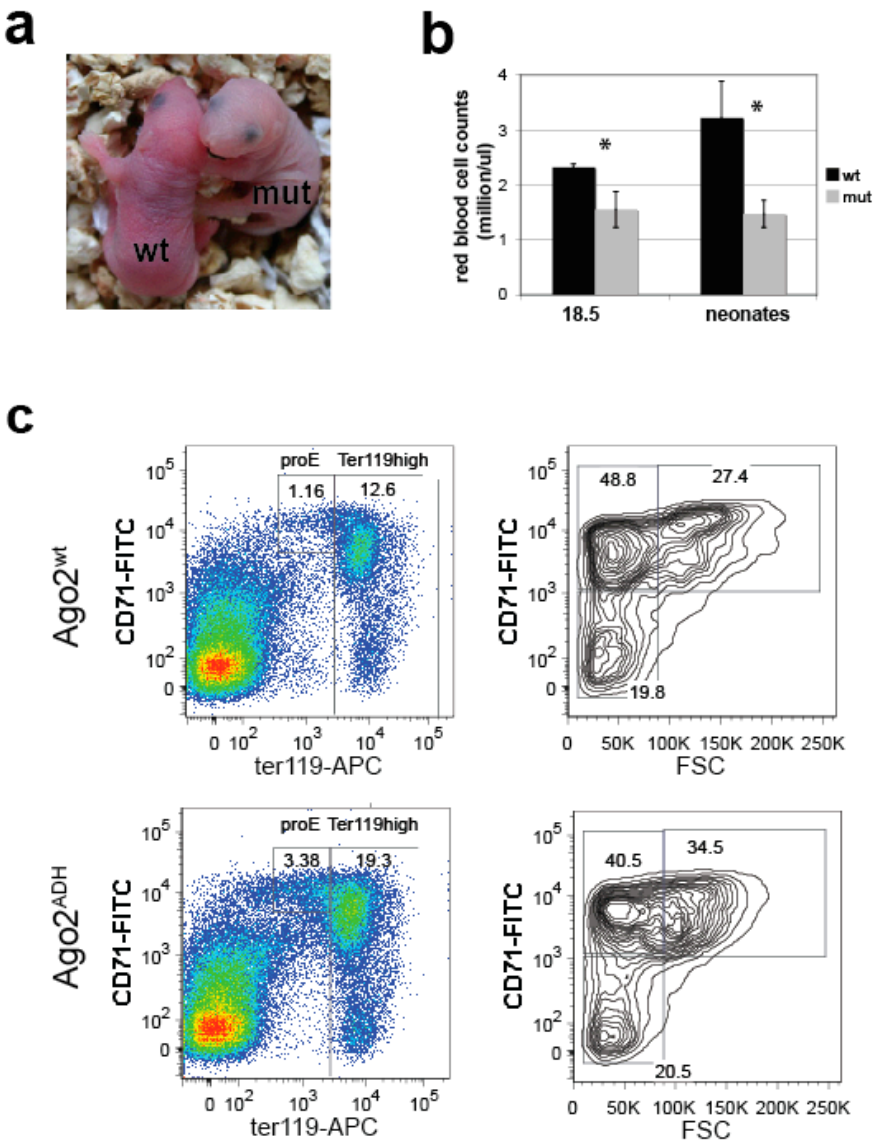

Figure 2: Ago2 catalysis is essential for development

a. Representative neonates from Ago2^{ADH} heterozygous inter-crosses. wt: wild type, mut: Ago2 homozygous mutant. **b.** Peripheral blood count of litter mates from wt and mut. Data are the mean ± SD. * t-test (unequal variance for E18.5 time point p=0.035, equal variance for the birth time point p= 1.95E-07). **c**. Representative FACS analysis using CD71/ter119 erythroid populations marking of individual bone marrow samples of mutant vs wt litter mates. Three independent pairs from three different litters showed virtually identical profiles.

To address this possibility, we performed a complete blood count (CBC). By E18.5, red blood cell (RBC) counts in mutants begin to deviate from those of control animals (Fig. 2b). This was even more dramatic at birth, with mutants showing a roughly 50% reduction peripheral RBCs. This appears specific to erythroid cells, since no alteration in the abundance of other cell types within the hematopoietic compartment was observed.

Loss of circulating RBCs could result from hemolysis, hemorrhage, or a defect in erythropoiesis. We therefore assessed red blood cell production in the hematopoietic sites of neonates using Ter119 and CD71 double marking of differentiating erythroblasts cell populations (Fig. 2c). We detected a 2-3 fold proportional increase in proerythroblasts (Pro-E) in liver, bone marrow, and spleen in the Ago^{ADH} homozygotes. These results suggested that Ago2 catalysis might be important for erythroid maturation during the transition from pro-E to basophilic erythroblasts.

3.5. Non-canonical biogenesis of an erythropoietic miRNA

Our results suggested that miRNA directed target cleavage might prove important for erythrocyte maturation. As a step toward identifying such a target, we profiled miRNAs expressed in the liver, one of the fetal hematopoietic sites. Deep sequencing from wild-type animals and $Ago2^{ADH}$ homozygotes revealed that virtually all microRNAs were present at nearly identical levels. However, one miRNA, miR-451, represented 11% of all miRNA reads in normal fetal liver but was dramatically reduced in the mutants (Fig. 3a).

Previous studies have demonstrated a strong dependency of the development of pro-E into basophilic erythroblasts on the expression of miR-451[37]. Together, miR-451 and miR-144 form a microRNA cluster with robust expression in erythroid cells. This pattern can be explained in part based upon the presence of regulatory sites for the GATA-1 zinc finger transcription factor, which acts as a master regulator of eythroid differentiation[38]. The regulatory circuit seems to be intact in $Ago2^{ADH}$ animals, since we observe no changes in the levels of pri-mir144-451 in homozygous mutants (Fig. 3b). This strongly pointed to an impact of catalysis on miR-451 maturation rather than miR-451 expression.

MicroRNA biogenesis occurs via a two-step processing pathway wherein Drosha initially cleaves the primary microRNA transcript to liberate a hairpin pre-miRNA [39]. This is exported to the cytoplasm and recognized and cleaved by Dicer to yield the mature duplex, which is loaded into Ago. The passenger strand

is removed through unknown mechanisms to yield a complex ready for target recognition.

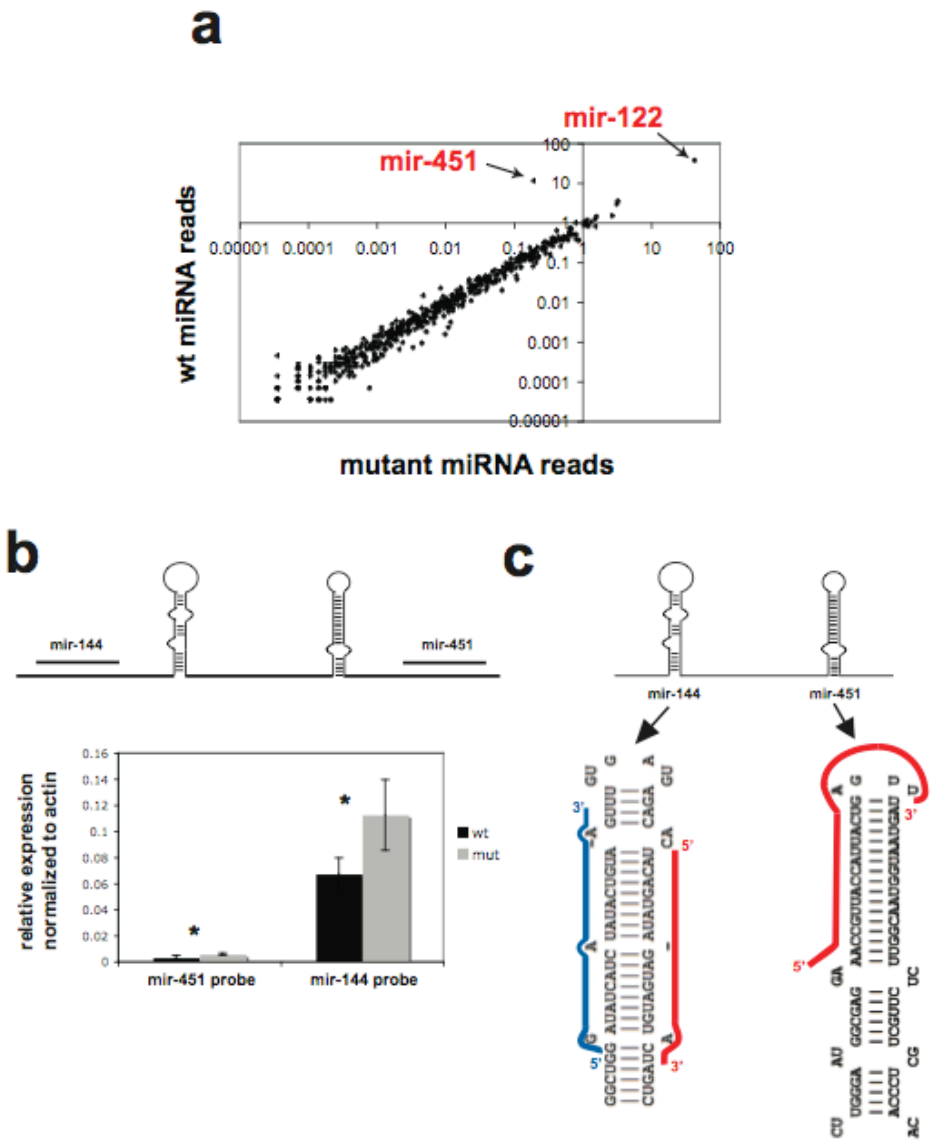

Figure 3: Mature mir-451 expression depends on Ago2 catalysis
a, Scatter plot of miRNA reads in wild-type versus mutant fetal liver. **b.** Quantitative RT-PCR of primary transcript levels of mir-144 and mir-451 in wt and mutant liver samples. Data are the mean of three biological replicates ±SD. * t-test with equal variance $p>0.05$. **c.** The unique structure of the mir-451 hairpin compared to mir-144 with the mirbase annotation of mature mir-451 and mir-144 mapped to the predicted secondary hairpin structure shown. Guide strand in red and passenger strand in blue.

An examination of the miR-451 precursor and its mature strand revealed an unusual feature. As annotated, the 6 terminal nucleotides of the 23nt long mature mir-451 span the loop region and extend into the complementary strand of the hairpin precursor. This arrangement appears incompatible with the well-studied enzymatic activities of Drosha and Dicer, which would normally liberate the mature microRNA mapping to the stem only (Fig. 3c). We therefore explored the possibility that miR-451 might adopt an unusual mode of biogenesis.

We began by assessing the dependency of miR-451 on Drosha. We created a construct, which drives the expression of the miR-144/451 precursor from a strong viral promoter and introduced this into MEF homozygous for a conditional Drosha allele. Following activation of Cre-ER and Drosha loss of function, we noted a 20-fold reduction in levels of mature miR-451. This was even more dramatic than the effect on a miRNA, let-7c, with a well-established dependency on canonical processing factors (Fig. 4a). We also assessed the ability of Drosha to liberate pre-miR-451 *in vitro*. Drosha complexes were affinity purified from human 293T cells and mixed with *in vitro* synthesized fragments of pri-miR-451 or pri-miR-144. In both cases, bands of the appropriate size for the pre-miRNA were observed (Fig. 4b). In the case of pri-mir-451 processing the 5' flank of the transcript folds into an additional hairpin, which may be released by Drosha to give additional fragments. As a result, only one flank was observed. The identities of pre-miRNA bands were confirmed by Northern blotting with oligonucleotide probes corresponding to the predicted species (Fig. 4c, not shown). Considered together, these experiments provide both genetic and biochemical support for Drosha catalyzing the excision of pre-miR-451 from its primary transcript.

Pre-miR-451 has an unusually short, 17 nt stem region. Previous studies indicate that this is too short to be efficiently recognized and processed by Dicer[40]. We therefore examined the role of Dicer in miR-451 maturation. We introduced the pre-miR-451 expression vector into ES cells that are homozygous for Dicer conditional alleles and express Cre-ER. While acute Dicer loss caused a roughly 80-fold reduction in a control ES cell microRNA (miR-294), miR-451 levels did not change (Fig. 4d). A pure population of continuous Dicer-null ES cells showed more than a 500-fold reduction in conventional microRNA, whereas levels of miR-451 were unaffected (Fig. 4e). We also confirmed this results using northern blot analysis of dicer nulls ES cells transiently expressing the mir-451 precursor (Fig. 4f). Finally, we incubated synthetic miR-451 pre-miRNA with recombinant Dicer and observed no mature cleavage products, though pre-let-7 was efficiently processed (not shown). Thus, conversion of pre-miR-451 into a

mature miRNA proceeds independently of Dicer. We therefore strove to identify an alternative maturation pathway.

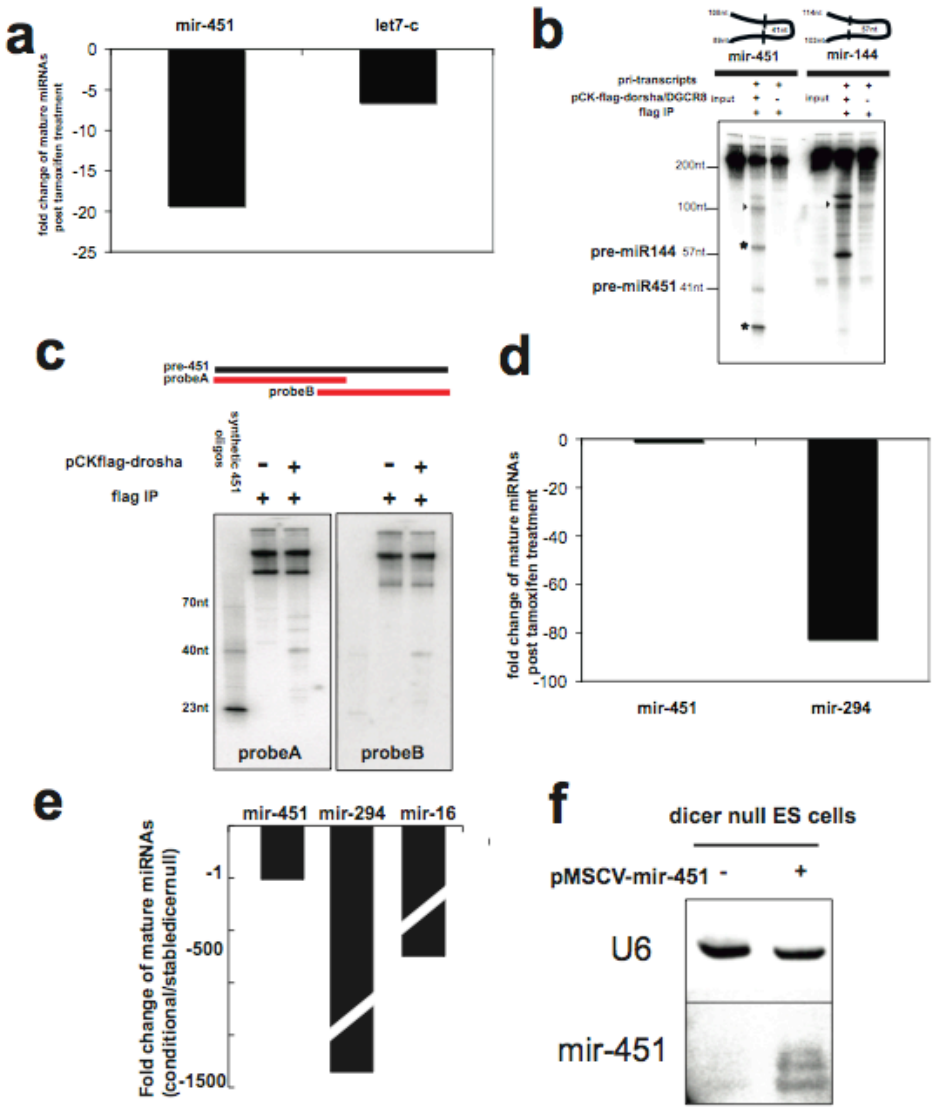

Figure 4: Non-canonical biogenesis of mir-451.
a. Effect on mature miRNA levels of Drosha conditional ablation in Drosha flox/flox Cre-ER MEFs. **b**. In-vitro processing of mir-451 and mir-144 primary transcripts by Drosha immunoprecipitates. pre-miR144 and pre-miR451 are indicated with their corresponding expected sizes. Additional fragments released by possible Drosha processing of the 5' miR-451 flank are indicated with asterisks. Flanks are indicated with arrowheads. **c**. Northern blots for confirmation of in-vitro Drosha processing assays. **d.** Effect on mature miRNA levels of Dicer conditional ablation in Dicer flox/flox Cre-ER ES cells. **e**. effects on mature miRNA levels in Dicer null stable ES cells. **f.** Mature mir-451 expression in dicer-null stable ES cells. U6 is used as a loading control.

By Northern blotting, we examined miR-451 species in wild-type and $Ago2^{ADH}$ mutant livers. This confirmed loss of the mature miRNA in the mutant animals. However, we noted the appearance of an ~40nt band that co-migrated with a synthetic pre-miR-451 and hybridized to probes to its 5' and 3' arms (Fig. 5a). This indicated accumulation of the Drosha cleavage product in mutant animals. Notably, the same bands seen in total RNA were also detected in Ago2 immunoprecipitates (Fig. 5a). This demonstrated the direct loading of the pre-miRNA into Ago2 and raised the possibility that the Ago2 catalytic center might help to catalyze the maturation of this microRNA.

The well-established biochemical properties of Ago2 predict that it would cleave a loaded pre-miR-451 after its 30^{th} base. We searched for evidence of such an intermediate in fetal liver small RNA libraries encompassing an expanded size range. Plotting a size distribution of reads corresponding to a conventional miRNA, miR-144, gave the expected pattern, a sharp peak at ~20 nt. In contrast, miR-451 showed a heterogeneous size distribution, exclusively because of variation at its 3' end. One abundant species corresponded precisely to the predicted Ago cleavage product (Fig 5b,c).

We confirmed the capacity of Ago2 to load and cleave pre-miR-451 using *in vitro* assays (Fig. 5d). Wild-type or catalytically inactive Ago2 complexes, complexes (Fig. 5d) were affinity purified from 293T cells and mixed with 5'-end labeled pre-miR-451. Only wild-type Ago2 produced the expected product, and this depended upon the presence of Mg^{2+} (not shown). No product was produced if we provided a mutant version of the precursor in which a single point mutation disrupted pairing at the cleavage site. Beta elimination and ligation reactions confirmed that cleavage left a free 3' OH terminus as expected of Argonaute proteins. These data strongly support a role for the Ago2 catalytic center in miR-451 maturation. This is perhaps akin to the proposed role of passenger strand cleavage in the maturation of siRISC. Ago1 could load pre-miR-451 but was unable to process it to its mature form (Fig. 5e)

To investigate sequence versus structural requirements for entry into the alternative miRNA biogenesis pathway, we created a structural mimic of miR-451 that might instead produce let-7c. At the concentration tested, this was as potent as the native pre-let-7c in suppressing a GFP or luciferase reporter containing perfect let-7 complementary sites (Fig. 5f, S5a). The miR-451 precursor could also be remodeled to express an shRNA that efficiently represses p53 (Fig. S5b). Given the unique ability of Ago2 to productively process miR-451, these data could have practical implications in the experimental use of shRNAs.

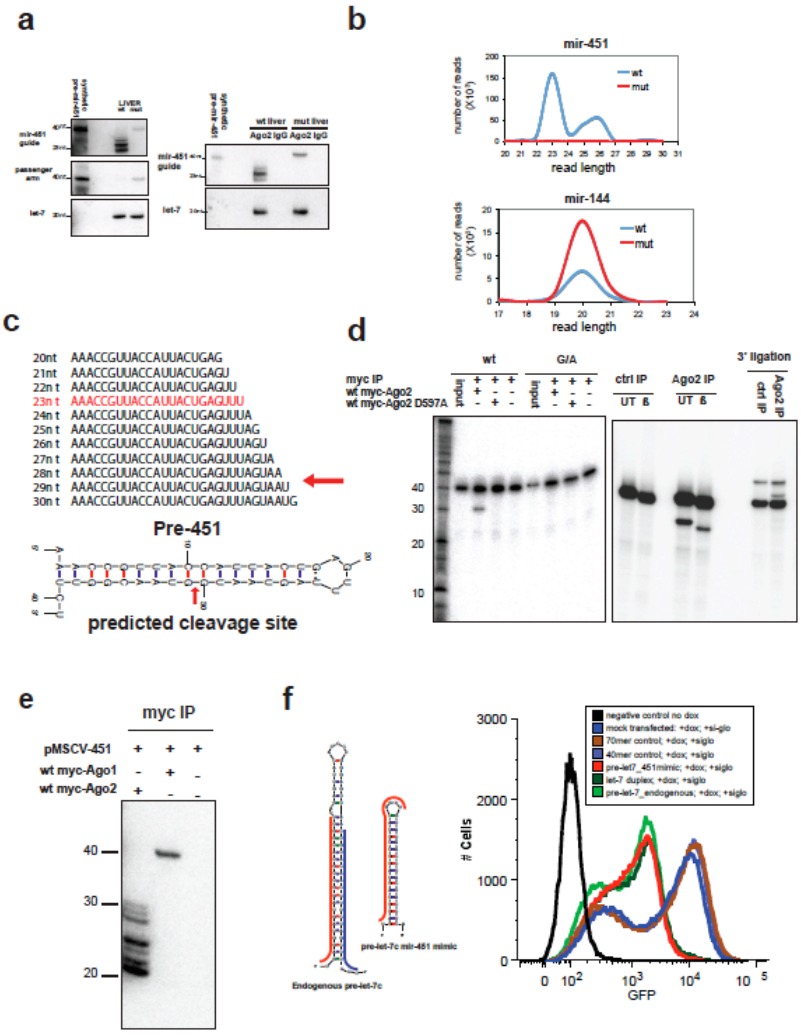

Figure 5: Ago2 catalysis is required for mir-451 biogenesis

a. Left panels: Northern blot on total RNA from wt and mutant livers probing for mir-451 guide strand and passenger arms of the hairpin (indicated). Let-7 is used as a loading control. Right panels: Northern blots of Ago2 and IgG control immunoprecipitates from wt and mutant liver extracts with the indicated probes. **b.** miRNA read length distribution for the indicated miRNA from deep sequencing of WT and mutant livers. **c.** prediction of a mir-451 Ago2 cleavage site. top: mir-451 3'end heterogeneity. Bottom: predicted cleavage site at the 30th phosphodiester bond of pre-mir-451. **d**. left gel: in-vitro cleavage assay of pre-mir-451 by an Ago2 immunoprecipitate. right gel: confirmation of the 3' end character of the Ago2 cleavage product using beta elimination and 3' end ligation reactions. **e.** IP-northern indicating presence of the mature form of mir-451 in Ago2 complexes and loading of pre-miR451 without processing in Ago1 complexes. **f.** Left: schematic depictions of the pre-let-7-mir-451 mimic hairpin compared to th native pre-let-7. Guide strand in red and passenger strand in blue. Right: FACS analysis for GFI in the indicated samples. Cells were co-transfected with PE-siRNA control. 100000 PE po cells were gated and analysed for GFP expression.

=

3.6. Discussion

Considered together, our results suggest a model (Fig S6) in which miR-451 enters RISC through an alternative biogenesis pathway. Though Drosha cleavage proceeds normally, the Dicer step is skipped and the pre-miRNA is loaded directly into Argonaute. This is surprising, considering prior studies indicating a coupling of Dicer cleavage and RISC loading[41, 42]. Such a complex would also lack interactions between the PAZ domain and the 3' end of a conventional Dicer product[43, 44]. A prior report indicated the ability of RISC to accommodate such species and posited a potential for Ago cleavage in the maturation of canonical microRNAs [45]. However, no physiological role for such an activity was demonstrated, and we detect no measurable defects in the processing of canonical miRNAs in $Ago2^{ADH}$ mutants. Mir-451 maturation proceeds with Ago-mediated cleavage producing an intermediate that is further trimmed. While this could occur via either endo- or exo-nucleolytic digestion, the observed distribution of 3' ends, many bearing single non-templated U residues (Fig. S6, S7), seems more consistent with the latter model. Though the precise enzymology of this step remains obscure, preliminary studies fail to support roles for Eri-1 or the exosome complex (not shown).

A previous report noted severe defects in erythropoiesis in Ago2-null cells[46]. In that case, the phenotype could be rescued by transplantation of adult HSC expressing a catalytically inactive Ago2 allele. While the two studies may seem at odds, we feel that the differences simply reflect the quantitative nature of our respective phenotypes and the fact that the strong reduction in miR-451 that we observe results in only a partial loss of RBCs. Anemia in the Ago2 mutants could be a complex phenotype, of which miR-451 loss is only one component. Finally, in the prior study, a different inactive Ago2 allele was expressed in adult HSC from a strong viral promoter, whereas, in the work presented herein, a knock-in allele had its impact on the hematopoietic niche as a whole during fetal development.

While the anemia of the $Ago2^{ADH}$ animals is profound, it is unclear whether this alone is sufficient to cause the perinatal death of mutant animals. Though no other well-defined pathologies were observed, we did find hemorrhages in the lungs and intra-abdominal bleeding in some animals. Moreover, miR-451 is expressed in the gastric epithelium, where its loss could have impacts that we have as yet failed to detect[47]. We leave open the possibility that additional microRNAs might rely on the alternative biogenesis pathway for maturation, though none presently annotated in miR-base share the usual structure or 3' heterogeneity of miR-451. Additional defects could also arise from a role of Ago2-mediated target mRNA cleavage. Thus far, miR-196 and microRNAs from the imprinted rtl1 cluster have been ascribed this property; however, none of the phenotypes that we observe can be explained by defects in these particular species[16-18].

The structure of the miR-451 precursor and the extension of its mature sequence around the loop and into the complementary strand of the precursor are present throughout vertebrate evolution (Supplementary Fig. S8). Thus, a conserved pathway of miR-451 maturation may provide at least some of the evolutionary pressure to maintain a catalytic Argonaute protein in animals.

3.7. Acknowledgements

We thank Maria Mosquera, Sang Young Kim, David Frendewey, and Aris Economides for help with generating and caring for animals. Andras Nagy, Michael Shen, Jenej Murn, Vasily Vagin, and Alexei Aravin provided helpful discussion. Narry Kim, Dan Littman, Ingrid Ibarra, and Elvin Wagenblast provided critical reagents, and Oliver Tam, Ravi Sachidanandam, Zhenyu Xuan, Dick McCombie, and Michelle Rooks provided support for generation and analysis of deep sequencing data. This work was supported by grants from the NIH and by a kind gift from Kathryn W. Davis.

3.8. Methods Summary

Mouse strains were generated as described in the Methods. Embryos were harvested at E3.5 for ES cells derivation, at midgestation for embryo phenotyping and beta-galactosidase reporter analysis and at E12.5 for chimeric analysis. Neonates were dissected to collect peripheral blood and hematopoietic organs (liver, bone marrow and spleen) to perform a complete blood count and erythroid FACS analysis, respectively. Small RNA fractions were isolated from wild type and mutant livers. These were cloned and deeply sequenced as previously described. For gene expression analysis quantitative RT-PCR was performed using custom designed primers or TaqMan specific probes. Immunoprecipitations were performed as described in the methods. Small RNA northerns were performed using miRNA-specific DNA probes. In vitro Drosha, Dicer and Ago2 processing experiments were performed as previously described. Substrates for enzymatic reactions were prepared either by in-vitro UTP (α^{32}P) labeled transcripts or ATP (γ^{32}P) end labeled synthetic RNAs. Conditional cell lines were treated with tamoxifen for inducible gene deletion and with doxycycline for inducible gene expression. All cell lines were transfected as described in the methods and harvested for gene expression analysis or reporter assays.

3.9. Methods

Mouse strains

Ago2 insertional mutant mouse strains, generated previously[20], were used for mutant analysis, ES cell derivation and reporter analysis. Ago1 gene trap strain was generated through germline transmission of Ago1 gene trap ES cells from Bay Genomics (RRR031). Ago2 catalytically inactive mutant knock-in mice were generated through germline transmission of positive ES cell clones targeted with Bacterial artificial chromosome (RP23-56M12) that has been modified with a point mutation D598A in the PIWI domain of Ago2.

Beta-galactosidase staining

For whole mount staining, embryos from different stages were dissected together with their extra-embryonic compartments in PBS. Beta galactosidase staining was performed using millipore's staining reagents. X-gal staining was performed overnight at room temperature. For placental sections, whole placentas were first stained for B-gal, sectioned and counterstained with Haematoxylin and Eosin.

Ago2 mutant crosses and Embryonic Stem (ES) cell derivation

Ago2 mutant phenotype was re-examined combining two insertional alleles for ease of genotyping the homozygous progeny and to take advantage of the Ago2 beta gal reporter allele. Ago2 null ES cells were derived as previously described[48]. Null cells were genotyped using primers specific to both insertional alleles. Null cells were genotyped using primers specific to both insertional alleles. $Ago2^{mc}$ allele: forward (GACGGTGAAGAAGCACAGGAA), reverse (GGTCCGATGGGAAAGTGTAGC). $Ago2^{gt}$ allele: forward (ATGGGATCGGCCATTGAA), reverse (GAACTCGTCAAGAAGGCG).

RT-PCR, western blot and Immunoprecipitation

Ago2 RT-PCR primers were designed downstream of both insertional alleles: Ago2F: TGTTCCAGCAACCTGTCATC, Ago2R:GATGATCTCCTGTCGGTGCT Actin primers were used as a normalization control. ActinF: ATGCTCCCCGGGCTGTAT, ActinR: CATAGGAGTCCTTCTGACCCATTC. QRT-PCR was performed using invitrogen superscript III and Applied biosystem cyber green PCR reagent. miRNA levels were measured using Applied Biosystems pri or mature miRNA assays. Ago2 western blot and immunoprecipitation analysis were performed using abnova eif2c2 antibody (M01). P53 western was performed using santa cruz mouse monoclonal antibody (Pab240).

ES-tetraploid aggregation

Ago2 null ES cells were injected into tetraploid blastocyst as previously described[49]. Embryos were transferred to foster mothers and dissected at E12.5. Beta gal stained was performed as described above.

Peripheral Blood collection and FACS analysis

Blood was collected from decapitated fetuses (pre-mortem) using heparanized microcapillaries and the CBC count was performed using the hemavet. For FACS analysis, single cells were isolated from neonatal liver, spleen and bone marrow and co-stained with Ter119 and CD71 antibodies (BD) and analyzed on LSRII flow cytometer (BD) as previously described[50]. The Same number of events of each sample were collected according to doublet descrimination gating and analyzed as follows: the ProE cell population was defined by CD71high/ter119 medium positive events. The ter119 high population was further subdivided into basophilic, late basophilic/polychromatic and orthochromatic/reticulocyte cell populations according to CD71 and FSC parameters to define the subsequent differentiating erythroblasts[51].

Small RNA cloning and bioinformatics annotation

Total RNA was extracted from E18.5 livers using trizol. Two Small RNA libraries with a size range of 19-30nt and 30-40nt were generated using a modified small RNA cloning strategy[52, 53]. Briefly, the small RNA fraction was ligated sequentially at the 3'OH and 5'phosphates with synthetic linkers, reverse transcribed and amplified using solexa sequencing primers. Around 7 million reads were generated for each small RNA library. Sequences were then trimmed from the 3' linker, collapsed and mapped to the mouse genome with no mismatches using multiple annotation tracks, namely: UCSC genes, miRNAs and repeats. For this study we used the mirbase database to annotate the cloned miRNAs.

Cell culture, plasmids, transfections and sensor assays

Mir-144-451 expression vector was constructed by cloning the genomic cluster into pMSCV retroviral vector. Cre-ER MEFs and ES cells were cultured as previously described[48]. Excision of dicer and drosha allele was mediated through tamoxifen treatment (100nM) for 5days followed by transient transfection of mir-451 expressing plasmid using lipofectamine (Invitrogen). For in vitro procesessing assays and northern blots 293T cells were cultured in DMEM + 10% FBS and cotransfected using LT-1 Mirus reagent with flag tagged drosha and DGCR8 constructs, myc tagged Ago2 or Ago1 with MSCV-miR144-451 expression vector or myc tagged Ago2 alone. Dual luciferase assays were perfomed as previously described. For validation of the Ago2 null ES cells, a luciferase plasmid with no artificial site was cotransfected with a perfectly matched siRNA duplex (dharmacon). Testing the functionality of mir-451 mimics was performed using three strategies: 1) cotransfection of let-7-mir-451 mimics, pre-let-7 or let-7 duplex or CTRL RNAs (dharmacon) at a 100nM concentration with let-7c luciferase reporter construct containing two perfect matching sites in the 3'UTR in HEK293 cells 2) Similarly, tetracycline inducible Let-7 GFP sensor ES cells containing two perfectly matched sites cotransfected with PE-labeled siRNA and let-7-mir-451 mimics (50nM) followed by GFP analysis of PE positive cell population using LSRII flow cytometer (BD). GFP sensor was induced using dox (1ug/ml). 3) For p53 knockdowns, ES cells were transfected with p53 shRNA and p53-mir-451 mimics followed by p53 induction using Adriamycin (0.5ug/ml) within the last 8hours before harvest. All cells were harvested 48hours post-transfection.

Drosha and dicer in-vitro processing assays

PCR fragment mapping to mir-451 and mir-144 were amplified out of the human genome with T7 promoter sequence. Pri-451 and Pri-144 RNA transcripts were generated using the genomic PCR product and Ambion's T7 in-vitro transcription kit. Transcripts were gel purified and used in a drosha in-vitro processing assay as previously described[54, 55]. For the dicer in vitro processing assay, ^{32}P end labeled synthetic pre-miR451 and pre-let-7 was incubated with 1

unit of human recombinant dicer (genlantis) in 30mM tris-HCL ph6.7, 50mM NaCl and 3mM MgCl2 buffer at 37°C over a 4 hours time course.

RNA northern blot analysis

Small RNA gel:

RNA was isolated using trizol (glycogen was used as a carrier in the case of RNA isolated from immunoprecipitated material). Total RNA was quantified using the nanodrop, and 5-10 µg of total RNA was resolved by 20% denaturing polyacrylamide/urea gel electrophoresis (sequaGel Urea system, National diagnostics) in 0.5X TBE buffer. 5´-32P-radiolabeled small RNA marker was used as size markers (USB). All samples were resuspended in 2X RNA loading buffer from ambion (95% Formamide, 18 mM EDTA and 0.025% each of SDS, Xylene Cyanol, and Bromophenol Blue). After electrophoresis, the polyacrylamide gel was transferred to Hybond N+ (GE healthcare) in 0.5x TBE by semi-dry transfer (Transblot SD, Bio-Rad) at 20 V for 1–2 h. The RNA was crosslinked to the membrane by UV irradiation (1200 µjoules/cm; Stratalinker, Stratagene, and pre-hybridized in either Church buffer (0.5M NaPO4 ph7, 1mM EDTA, 7% SDS) or *ULTRAhyb*® oligo hybridization solution (ambion) for 1 h at 37 (DNA probes) or 65°C (RNA probes).

Probe preparation

10 pmol of RNA (Dharmacon) or DNA (IDT) probe was 5´-32P-radiolabeled with polynucleotide kinase (NEB) and γ-32P-ATP (7,000 µCi/mmol; Elmer) and purified using a Sephadex G-25 spin column (Roche).

A typical labelling reaction is as follows:
50ul total reaction:
10pmole of RNA/DNA probe
5ul 10X buffer
1ul kinase
2-5ul (γ-32P-ATP)
bring reaction up to 50ul with water

Hybridization

To detect U6 snRNA, 1/50th of the 32P-radiolabeled probe was diluted with unlabeled probe. The 32P-radiolabeled probes were hybridized in Church buffer or the ambion commercial sln for 4–12 h. For RNA probes, hybridization was at 65°C; for DNA probes, hybridization was at 37°C. After hybridization, membranes were washed twice with 2x SSC/0.1% (w/v) sodium dodecyl sulfate (SDS) and once with 1x SSC/0.1% (w/v) SDS for 15-20 min. Membranes were analyzed by phosophorimagery (Fuji, Tokyo, Japan). To strip probes, membranes were microwaved in 0.2XSSC/0.05% (w/v) SDS for 1 min (twice),

then re-exposed to confirm probe removal. Probe sequences are designed as perfect and complementary sequences to 21-23 small RNA molecules.

Ago2 cleavage assays and beta elimination

Ago2 myc tagged constructs (wt and D797A) were transfected in 293T cells. Lysates were collected after 48hours, immunoprecipitated using myc agarose beads. The catalysis reaction was carried out on beads using 5' P32 end labeled synthetic pre-miR-451 (dharmacon) as previously described[20]. Beta elimination was performed through treating the purified RNA from the Ago2 beads with Sodium periodate for 30min at room temperature followed by ethanol precipitation. The RNA was resuspended in loading buffer containing TBE and run on a 20% acrylamide gel where the beta elimination reaction occurs.

3.10. References

1. Hutvagner, G. & Simard, M. J. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol 9, 22-32 (2008).
2. Joshua-Tor, L. The Argonautes. Cold Spring Harb Symp Quant Biol 71, 67-72 (2006).
3. Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev 15, 188-200 (2001).
4. Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J 20, 6877-88 (2001).
5. Yuan, Y. R. et al. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell 19, 405-19 (2005).
6. Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev 18, 975-80 (2004).
7. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr Biol 14, 787-91 (2004).
8. Malone, C. D. & Hannon, G. J. Small RNAs as guardians of the genome. Cell 136, 656-68 (2009).
9. Yigit, E. et al. Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell 127, 747-57 (2006).
10. Bohmert, K. et al. AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J 17, 170-80 (1998).
11. Baumberger, N. & Baulcombe, D. C. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci U S A 102, 11928-33 (2005).
12. Qi, Y., Denli, A. M. & Hannon, G. J. Biochemical specialization within Arabidopsis RNA silencing pathways. Mol Cell 19, 421-8 (2005).
13. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. Cell 136, 215-33 (2009).
14. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. Science 304, 594-6 (2004).
15. Davis, E. et al. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol 15, 743-9 (2005).
16. Harfe, B. D., McManus, M. T., Mansfield, J. H., Hornstein, E. & Tabin, C. J. 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 (2005).
17. Sekita, Y. et al. Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. Nat Genet 40, 243-8 (2008).
18. Hornstein, E. et al. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. Nature 438, 671-4 (2005).

19. Tolia, N. H. & Joshua-Tor, L. Slicer and the argonautes. Nat Chem Biol 3, 36-43 (2007).
20. Liu, J. et al. Argonaute2 is the catalytic engine of mammalian RNAi. Science 305, 1437-41 (2004).
21. Rivas, F. V. et al. Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 12, 340-9 (2005).
22. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305, 1434-7 (2004).
23. Azuma-Mukai, A. et al. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc Natl Acad Sci U S A 105, 7964-9 (2008).
24. Ender, C. et al. A human snoRNA with microRNA-like functions. Mol Cell 32, 519-28 (2008).
25. Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev 22, 2773-85 (2008).
26. Tam, O. H. et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 453, 534-8 (2008).
27. Watanabe, T. et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453, 539-43 (2008).
28. Kaneda, M., Tang, F., O'Carroll, D., Lao, K. & Surani, M. A. Essential role for Argonaute2 protein in mouse oogenesis. Epigenetics Chromatin 2, 9 (2009).
29. Ma, J. et al. MicroRNA Activity Is Suppressed in Mouse Oocytes. Curr Biol 20, 265-270.
30. Murchison, E. P. et al. Critical roles for Dicer in the female germline. Genes Dev 21, 682-93 (2007).
31. Suh, N. et al. MicroRNA Function Is Globally Suppressed in Mouse Oocytes and Early Embryos. Curr Biol 20, 271-277.
32. Alisch, R. S., Jin, P., Epstein, M., Caspary, T. & Warren, S. T. Argonaute2 is essential for mammalian gastrulation and proper mesoderm formation. PLoS Genet 3, e227 (2007).
33. Morita, S. et al. One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. Genomics 89, 687-96 (2007).
34. Sasaki, T., Shiohama, A., Minoshima, S. & Shimizu, N. Identification of eight members of the Argonaute family in the human genome small star, filled. Genomics 82, 323-30 (2003).
35. Rossant, J. & Cross, J. C. in Mouse development Pattering, Morphogenesis and Organogenesis (eds. Rossant, J. & Tam, P.) 155-180 (2002).
36. Rossant, J. & Cross, J. C. Placental development: lessons from mouse mutants. Nat Rev Genet 2, 538-48 (2001).

37. Papapetrou, E. P., Korkola, J. E. & Sadelain, M. A Genetic Strategy for Single and Combinatorial Analysis of miRNA Function in Mammalian Hematopoietic Stem Cells. Stem Cells (2009).
38. Dore, L. C. et al. A GATA-1-regulated microRNA locus essential for erythropoiesis. Proc Natl Acad Sci U S A 105, 3333-8 (2008).
39. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10, 126-39 (2009).
40. Siolas, D. et al. Synthetic shRNAs as potent RNAi triggers. Nat Biotechnol 23, 227-31 (2005).
41. Chendrimada, T. P. et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 436, 740-4 (2005).
42. Wang, H. W. et al. Structural insights into RNA processing by the human RISC-loading complex. Nat Struct Mol Biol 16, 1148-53 (2009).
43. Song, J. J. et al. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol 10, 1026-32 (2003).
44. Wang, Y., Sheng, G., Juranek, S., Tuschl, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. Nature 456, 209-13 (2008).
45. Diederichs, S. & Haber, D. A. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell 131, 1097-108 (2007).
46. O'Carroll, D. et al. A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. Genes Dev 21, 1999-2004 (2007).
47. Bandres, E. et al. microRNA-451 regulates macrophage migration inhibitory factor production and proliferation of gastrointestinal cancer cells. Clin Cancer Res 15, 2281-90 (2009).
48. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. Manipulating the Mouse Embryo: A Laboratory Manual (CSHL press, 2003).
49. Nagy, A. & Rossant, J. in Gene Targeting A Practical Approach (ed. Joiner, A. L.) 189-192 (Oxford University Press, 2000).
50. Socolovsky, M. et al. Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. Blood 98, 3261-73 (2001).
51. Liu, Y. et al. Suppression of Fas-FasL coexpression by erythropoietin mediates erythroblast expansion during the erythropoietic stress response in vivo. Blood 108, 123-33 (2006).
52. Aravin, A. & Tuschl, T. Identification and characterization of small RNAs involved in RNA silencing. FEBS Lett 579, 5830-40 (2005).
53. Pfeffer, S. et al. Identification of microRNAs of the herpesvirus family. Nat Methods 2, 269-76 (2005).
54. Lee, Y. et al. The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415-9 (2003).

55. Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. Nature 432, 231-5 (2004).

3.11. Supplementary material

Supplementary Tables

	development stages	E14.5	E15.5	E16.5	E18.5	new born	weaning
observed ratios	wt	5	6	6	3	16	13
	het	6	7	3	17	39	27
	mut	8	6	4	9	10	0
	total	19	19	13	29	65	40
	p values	0.172	0.518	0.112	0.188	0.156634	**0.0013**

Table1: Ago2 allele segregation analysis from Ago2 catalytically inactive heterozygous intercrosses

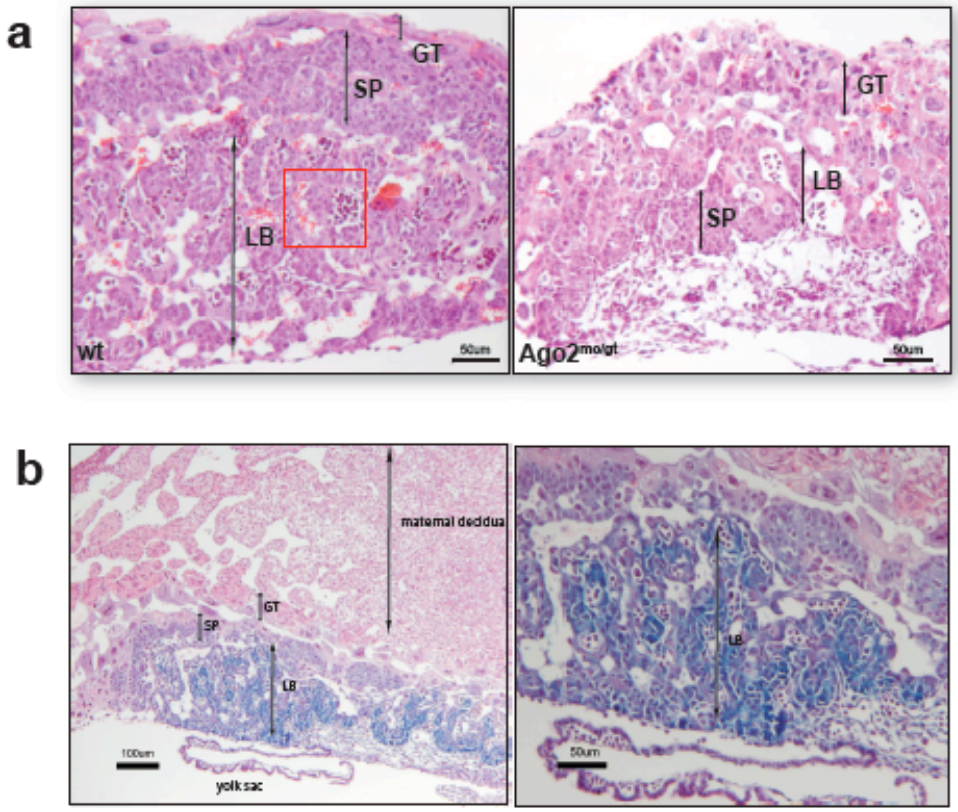

Figure S1: Labyrinthine layer defect in the Ago2 mutant mice

a, Histological analysis of wild type placenta left and mutant placenta right. Ago2 mutant embryos have a reduced labyrinthine layer. **(LB:** Labyrinth, **SP:** spongioblast**, GT:** giant trophoblast**).** Red rectangle depicting the site of nutrient exchange between the maternal blood islands and the fetal blood (nucleated red blood cells) **b,** Cross section of lacZ stained E9.5 placenta from embryos heterozygous for the Ago2^{gt} allele. Left panel: low magnification capturing the three layers of the placenta and the maternal decidua. Right panel: higher magnification capturing the labyrinth layer where Ago2 is highly expressed**.**

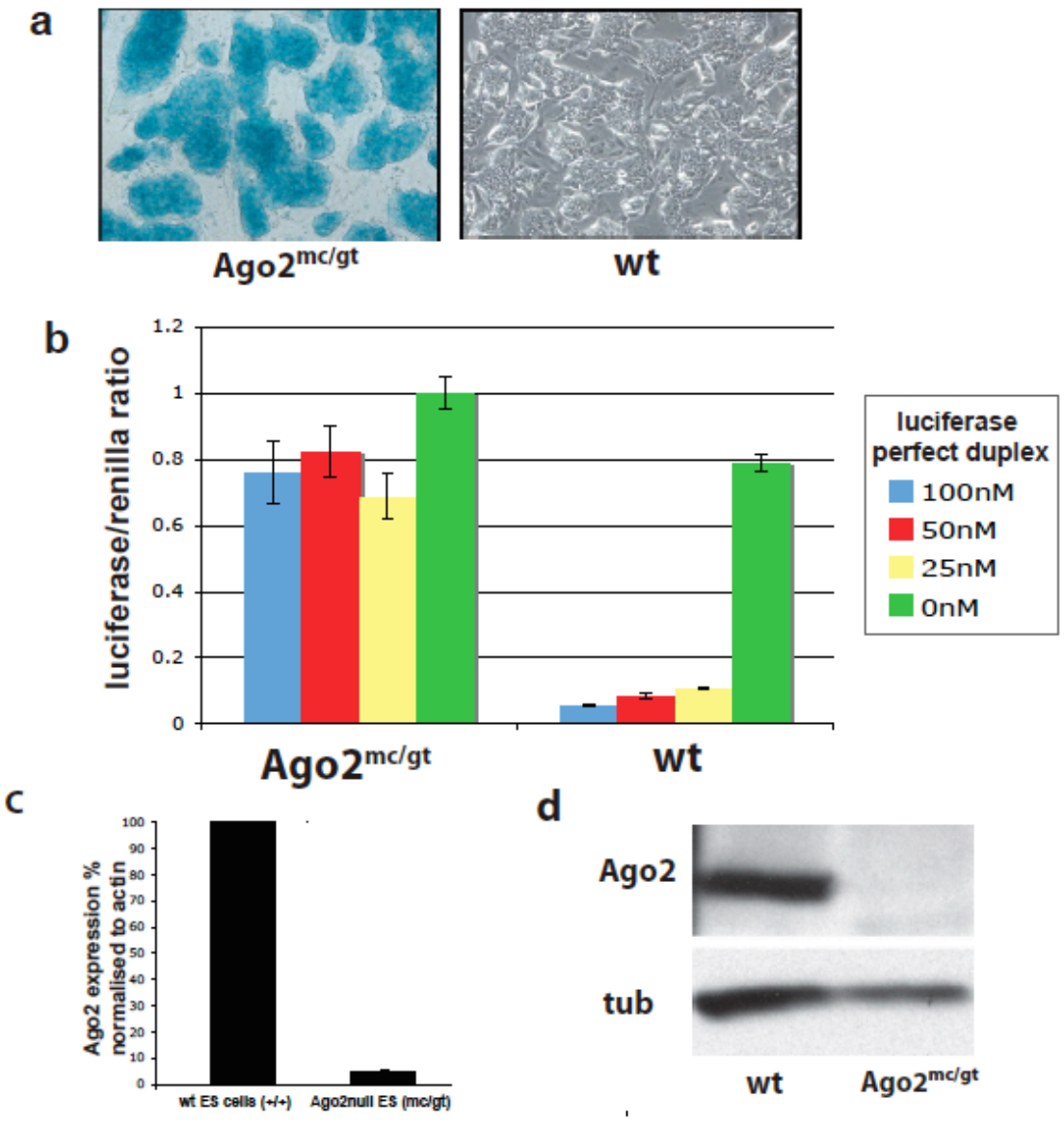

Figure S2: Characterization of Ago2 null ES cells

a, Ago2 null ES cells derived from blastocyts (left), wt ES cells as a control for the reporter staining. **b,** Dual luciferase siRNA silencing assay: Ago2 null cells fail to mount an siRNA mediated silencing response. Here, a perfectly matched luciferase siRNA duplex (Dharmacon) is titrated. Data are the mean of three technical replicates ±SD. **c,** Quantitative RT-PCR showing depletion of Ago2 message in the Ago2 null ES cell line. **d,** Western blot analysis: Ago2 protein is not detected in the Ago2 null ES cells.

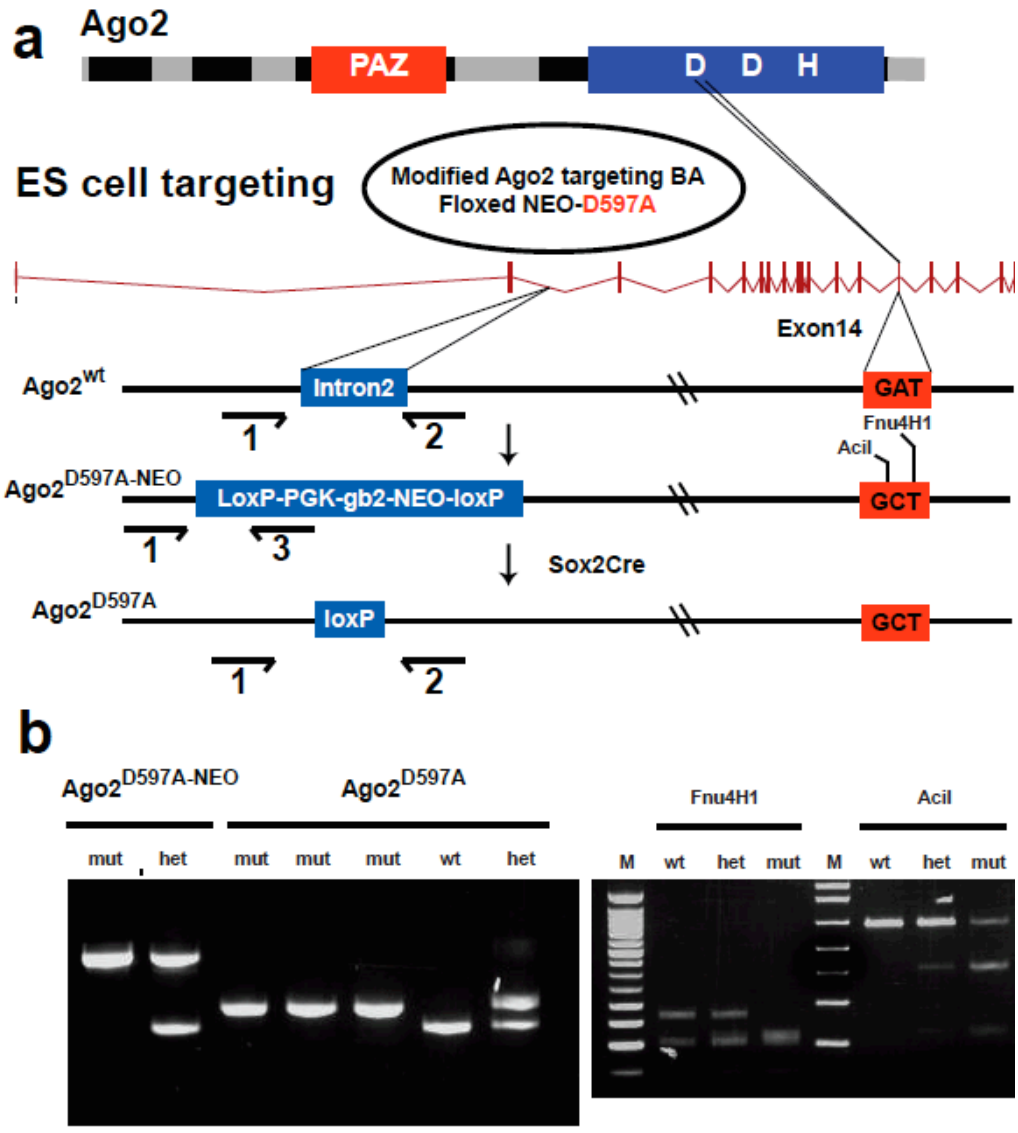

Figure S3: Schematic of the Ago2 catalytically inactive knock-in allele

a, Ago2 catalytically inactive knock-in allele targeting strategy. **b**, Genotyping strategy using either the selection cassette (left gel) or restriction fragment length polymorphism created by the point mutation (right gel).

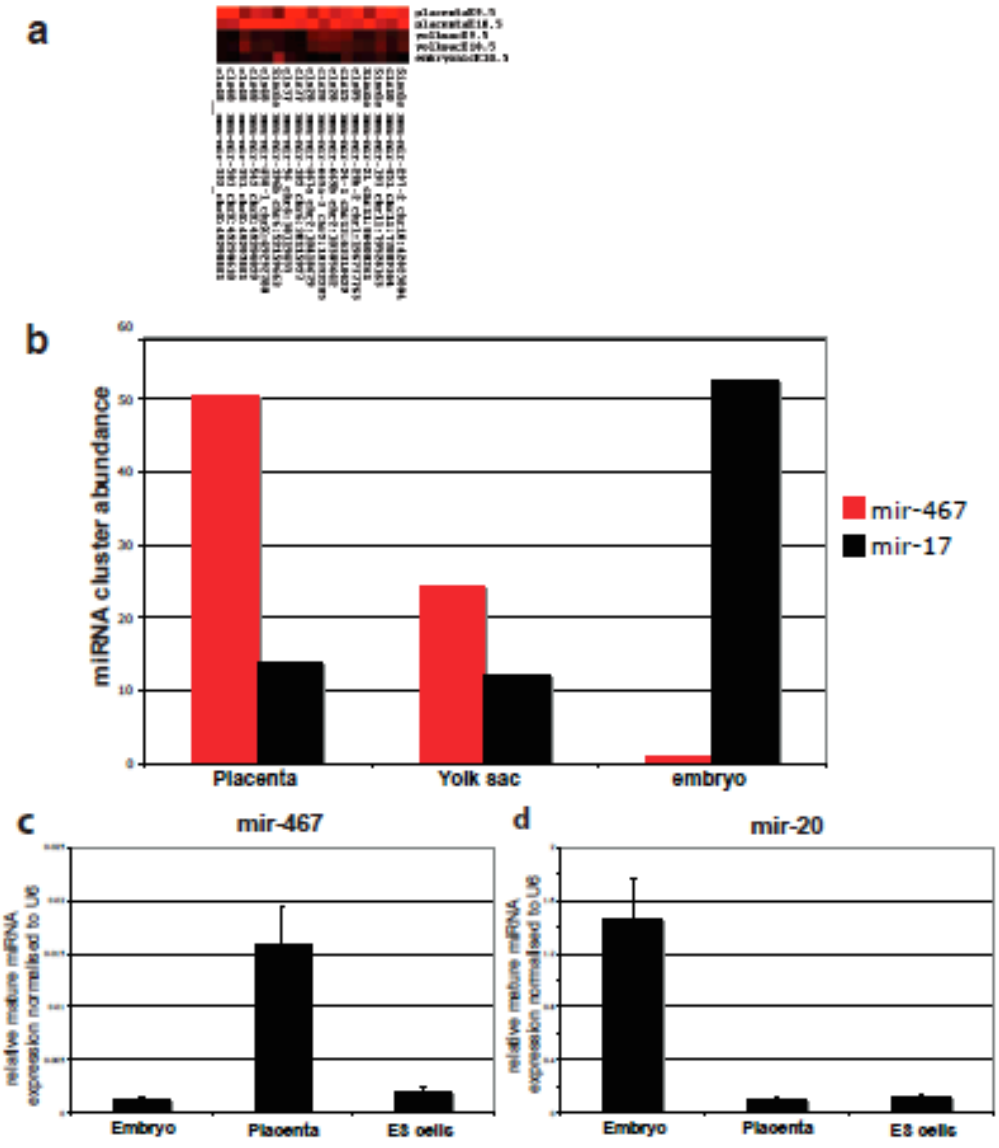

Figure S4: Extraembryonic-specific miRNAs are strong candidates for Ago2 function in the placenta

a, miRNA expression profile in placenta, yolk sac, embryo from left to right. (extraembryonic profiles are taken at two time points during midgestation (E9.5 and E10.5) each miRNA is classified as singleton or as a member of a cluster, mirbase name annotation and chromosomal location. **b,** mir-467 is a placental specific miRNA cluster representing half of the miRNA population in the placenta as represented by total reads of miRNA from sequencing miRNA libraries.mir-467 abundance in the placenta is compared to mir-17 (an embryonic specific miRNA cluster).**c,** Q-PCR validation of mir-467a and mir-20 members of the mir-467 and mir-17 clusters respectively.

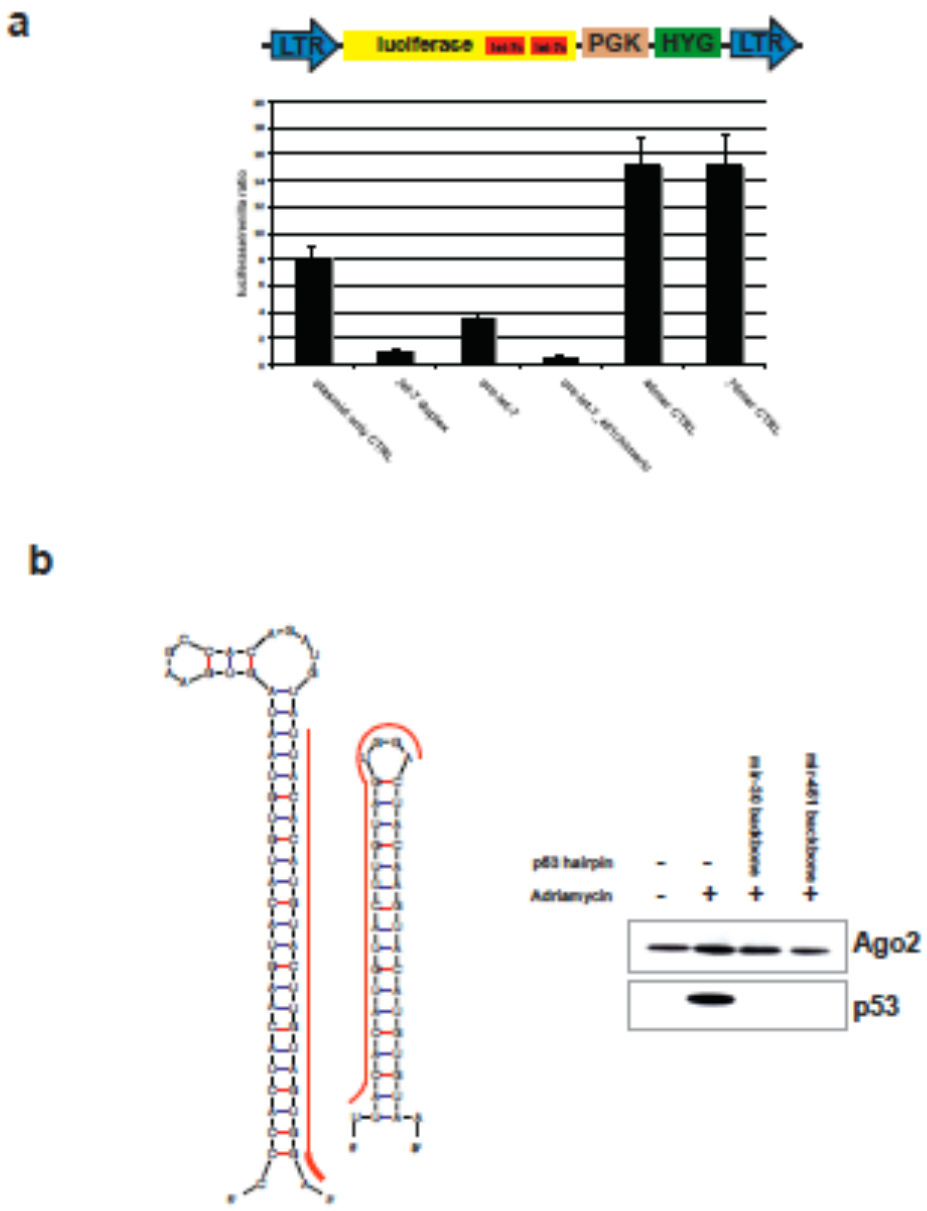

Figure S5: mir-451 mimics generation

a, Dual luciferase assay reporting mature let-7 activity. Top: Schematic of let-7 MSCV-luciferase reporter construct containing two perfectly matching let-7c sites. Bottom: Histogram showing luminescence values of luciferase/renilla ratios. Data are the mean of three technical replicates ± SD. **b,** left panel: schematic of p53 hairpin design in the mir-30 backbone or following the mir-451 fold. Right panel: Western blot analysis showing p53 knowdown in ES cells upon transfection of p53 hairpins and induction of p53 with adriamycin.

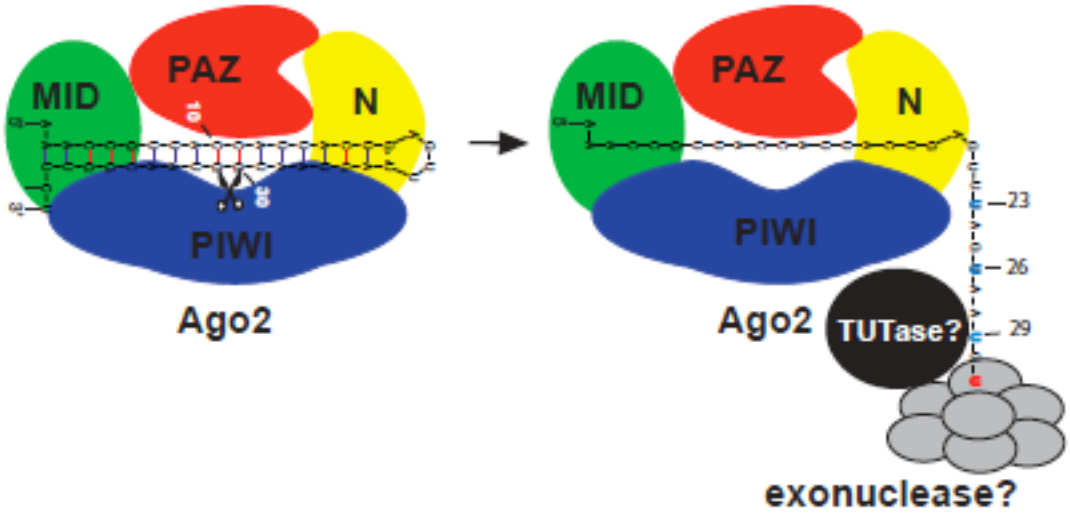

Figure S6: Hypothetical Model for mir-451 loading and 3'end trimming
Trimming of mir-451 post Ago2 catalysis through cycles of uridine transfer mediated by TUTases and exosome trimming. The domains of the Ago2 protein are color coded: Amino terminal domain (yellow), PAZ domain (Red), MID domain (green) and PIWI domain (blue). mir-451 molecule is drawn here according to its predicted secondary structure. Scissors denote Ago2 mediated catalysis. mir-451 cleavage products are drawn post Ago2 catalysis. The 3'end product can be degraded through Xrn1 exonuclease while the 5'end product (the precursor to the mature mir-451 forms) is protected by Ago2 while its tailed may be modified by TUTases and trimmed by the exoxsome.

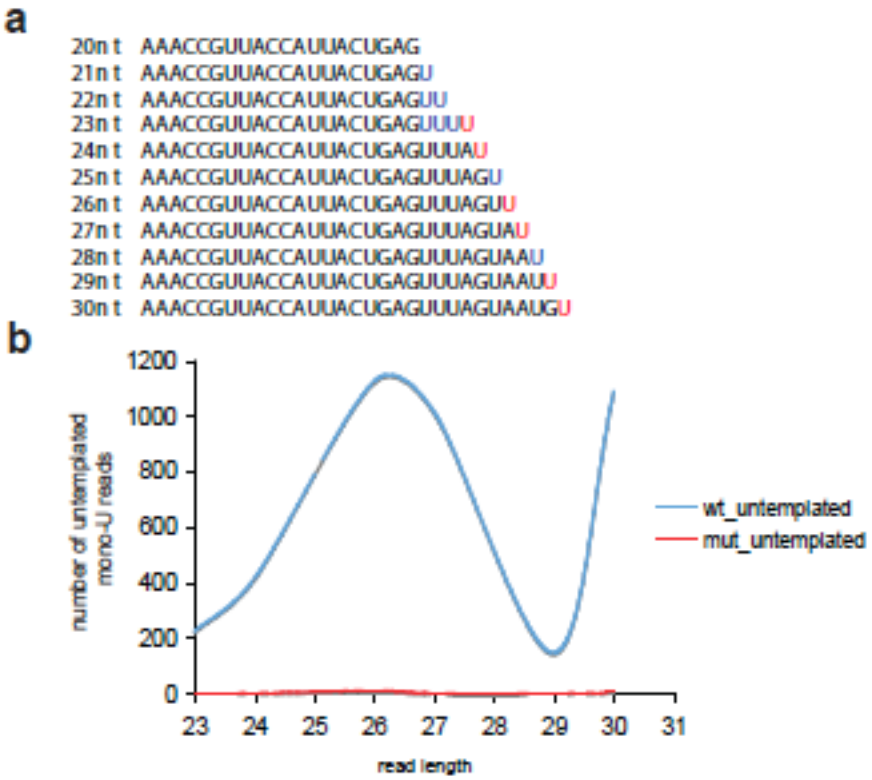

Figure S7: Sequence analysis of untemplated mono-U transferred to the endogenous 3'ends of mir-451

a, untemplated mono-Us at the 3'end of mir-451. Endogenous Us in blue and untemplated Us in red. Untemplated mono-Us are only detected post cleavage. b, endogenous mir-451 read length distribution representing abundance of untemplated mono-U non mapping reads only. See (fig. 5c) for non-templated endogenous reads.

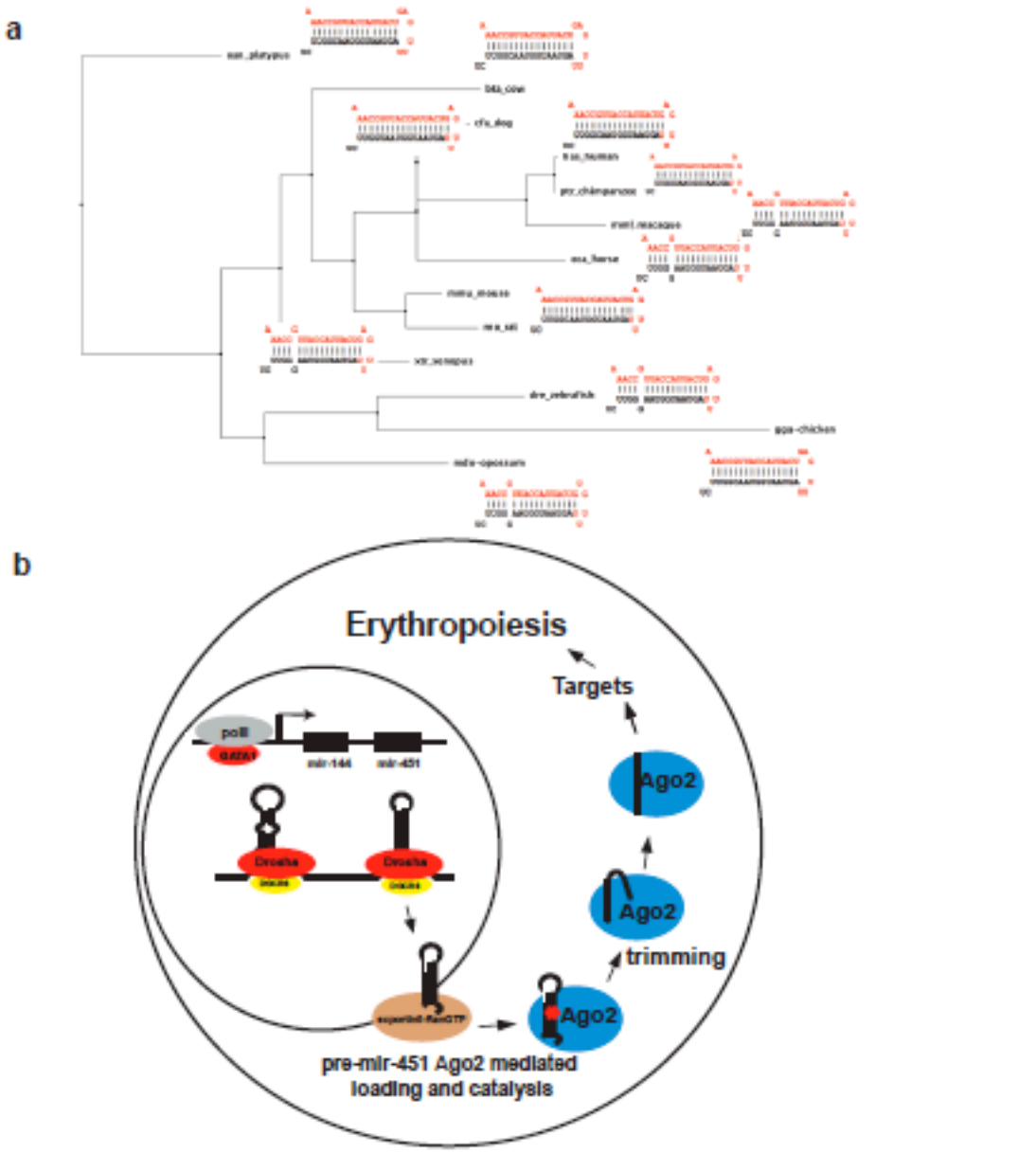

Figure S8: A conserved miRNA biogenesis pathway that depends on Ago2 catalysis to serve normal erythropoiesis

a, A neighbor-joining tree constructed using Jalview program depicting mir-451 hairpin conservation in vertebrates. Mouse and rat hairpins are identical. The guide strand is shown in red. Sequences of mir-451 were extracted from mirbase. b, mir-451 bicitronic cluster under the control of GATA-1 transcription factor. mir-451 primary transcript is processed by Drosha/DGCR8 complex in the nucleus then possibly exported to the cytoplasm where the pre-mir451 is loaded onto Argonaute2 which then cleaves its passenger arm to generate the mature species of mir-451 that target genes to regulate erythropoiesis.

CHAPTER 4: Conclusions and future directions

RNAi was established first as a technical tool in the laboratory. Using double stranded RNA, scientists were able to silence complementary sequences in *C. Elegans* in a simple and efficient manner. Subsequent genetic screens identified Argonaute genes to play an important role in mediating the RNAi response, and it became apparent that Argonautes are essential for the development of a variety of organisms (ranging from plants to flies). Although Argonautes are required in all model organisms, we do not fully comprehend how they orchestrate different cellular processes to ensure proper growth and development. I chose to focus on mammalian Argonautes, using the mouse as a model system, to study their impact on prenatal and postnatal development. My thesis research contributes to a better understanding of the spatio-temporal requirements of Argonautes during mouse development, and defines a novel mechanism for the biogenesis of small RNAs during erythropoiesis. Characterization of mammalian Argonautes would also improve the efficacy and specificity of RNAi technology used in mammalian systems.

Our ability to introduce genetic modifications in mouse embryonic stem cells and generate mutant mouse models allows us to directly probe gene function *in vivo*. At the commencement of my PhD, *Ago2* was the only mammalian Argonaute mutated *in vivo*, and found to be essential during normal mouse embryogenesis. Through subsequent genetic analysis of other Argonautes, Ago2 was able to compensate for the loss of all other Argonautes (*Ago1, Ago3, Ago4*) during embryonic and postnatal development. Although a subset of mutant animals lacking *Ago1*, *Ago3* and *Ago4* perished prior to weaning, we suspect that this could be the result of chromosomal instability caused by the deletion of a ~200kb region where *Ago1/3/4* resides. This chromosomal deletion could act as a risk factor, making animals more susceptible to stress, disease onset or environmental influences. Interestingly, the loss of Ago1 and Ago3 increases susceptibility to influenza A virus, with a decreased survival rate compared to wild type littermates. The mechanisms of Argonautes in viral response remain to be determined, although candidate miRNAs in viral host cells or immune system are suspected to confer resistance through their association with Argonautes.

Genetic analysis of murine Argonautes raised a very important question: Are Argonaute proteins redundant or specialized in their functions during development? The arguments for redundant functions are mainly attributed to their ubiquitous and overlapping patterns of expression and association with identical small RNA partners. The counterarguments revolve around the unique catalytic function and developmental requirement of Ago2. Observations from several studies raises the possibility that Argonautes can be either redundant or specialized dependent on developmental context. Ago2 is essential during erythroid and B cell differentiation independent of its catalytic activity, and cannot be compensated by its family members. The protein is also essential for oocyte maturation, and although its catalytic requirement has not been directly

investigated, identification of endogenous small RNAs in oocytes argues that Ago2 catalysis is required in biogenesis and/or effector functions. In contrast, Ago2 is redundant with other Argonautes in embryonic stem cells and during spermatogenesis. Since no preferential loading of miRNAs has been demonstrated in different Argonaute complexes, my research focused on their expression patterns and catalytic requirement to explain the differential requirement of Argonautes during development.

The mutant phenotype of Ago2 null embryos suggests defects in the extraembryonic compartment. Using gene trap mouse lines targeting Argonaute proteins, we perform comparative expression analysis during midgestation, coincident with the embryonic lethality of Ago2 mutants. Although we found interesting differences in embryonic *Ago1* and *Ago2* expressions, placental tissues exclusively express *Ago2*. Chimeric analysis in the mouse demonstrate that *Ago2* is required for providing the maternal-fetal interphase, with a possible role in the development of the labyrinthe layer, the structural network for nutrient, gas and waste exchange between the fetal and maternal blood. The exclusive expression of Ago2 in the placenta could explain the embryonic lethality exhibited by Ago2-deficient embryos. Placental specific miRNA candidates were identified through small RNA analysis, and gene expression studies or proteomic analysis will be instrumental in identifying biological processes affected in *Ago2* deficient placentas.

The catalytic activity of Argonautes is well conserved throughout evolution form plant to fungi to animals. Using a knock-in mouse model that expresses a catalytically inactive *Ago2*, we demonstrate unperturbed development past midgestation, although mutant offspring die from severe anemia shortly after birth. Small RNA profiling of hematopoietic regions uncovered a non-canonical miRNA, miR-451, that requires *Ago2* catalysis for its maturation. Although Drosha/Rnasen processes its primary transcript, the short hairpin miR-451 precursor is directly loaded into *Ago2*, bypassing the need of *Dicer*. The unique structural features of miR-451 enables Ago2 to cleave the hairpin structure to an intermediate 30nt long small RNA. The 30nt small RNA is further processed by an unknown mechanism that may require uridyl transfer and exo or endo nuclease activity to form the mature microRNA. The latter step of the pathway remains poorly characterized, and requires further biochemical and genetic studies both to elucidate the mechanisms of biogenesis and identify the biological relevance of the 3'end heterogeneity. MicroRNA-451 is highly conserved among vertebrate species. In collaboration with Antonio Giraldez group at Yale University, we tested the conservation of this pathway in the Zebrafish model organism. miR-451 is also processed independently of *Dicer* in the fish, with *Ago2* mutant manifesting a blood phenotype as well (see appendix 4).

Our studies have unveiled a conserved miRNA biogenesis pathway that require the conservation of catalysis in Argonaute proteins and a small RNA

precursor molecule that fits into the catalytic Ago2 to generate an active molecule important for erythropoiesis. With catalytic Argonautes found in all organisms involved in small-RNA mediated repression, it is no surprise that Ago catalysis proves to be important in diverse developmental processes. The pressure to retain catalysis appears to be dictated by evolutionary events that preferentially load perfectly paired RNA strands into the Ago catalytic site (miRNA/target pairs, siRNA duplexes or piRNA/precursor transcripts). The discovery of mir-451 biogenesis has revealed Argonaute's flexibility in binding physiologically important RNA molecule where the cleavage reaction may occur intra-molecularly.

The postnatal lethality and severe anemic phenotype of Ago2 catalytically inactive mutants is unlikely to be caused purely by defective biogenesis of mir-451. Two independent miR-451 knockout mouse models manifest mild anemia, enhanced in response to oxidative stress, but fail to phenocopy the Ago2 catalytically inactive mutants (Camilla dos Santos and Eric Olson, personal communications). In addition, reduction of miR-451 with molecular "sponges" induces a severe block during early erythropoiesis. These findings indicate that mir-451 might belong to a yet undiscovered group of small RNAs that are processed through Ago2-mediated catalysis, and required for late prenatal or perinatal development, although they remain to be identified. Ago2 complexes purified from wild type and catalytic inactive animals, and sequenced for an expanded range of RNA sizes (beyond 19 to 24 nucleotides) would reveal accumulation of RNA precursors and downregulation of their corresponding mature forms, providing candidates for subsequent functional analyses.

Our understanding of *Ago2* catalysis revolves around fundamental rules based on structural studies and genetic manipulation of protein residues or artificial RNA/RNA pairs. Argonaute mediated catalysis of precursor miRNAs raises very interesting questions in the field. First, how and where at the sub-cellular level are these structures recognized by Argonautes? Second, does Argonaute cleave these substrates autonomously or does it depend on protein modifications or additional binding partners? The latter is very likely, since both RNaseIII processing enzymes of the canonical miRNA pathway (*Dicer* and *Drosha*) require the assistance of double stranded RNA binding proteins. Third, are there structure or sequence features of mir-451 that directs the microRNA into the Ago2-dependent pathway? To address the latter question, we engineered synthetic shRNAs containing either an endogenous mature miRNA (let-7) and an siRNA targeting the p53 coding sequence, with the hairpin structure predicted to mimic miR-451. These artificial shRNAs are as efficient as the endogenous pre-miRNA in silencing let-7 sensors, and reduces p53 expression comparable to commercially available miR-30 based hairpins in at least two different cell lines tested (ES cells and HEK293T cells). The ability to utilize miR-451 biogenesis mechanisms is an indication that Ago2-catalysis-dependent silencing pathway is active *in vivo* in diverse cell types, and has huge

implications in both biological and technological applications of small RNA-mediated silencing.

Current RNAi technology in mammals utilizes the canonical miRNA pathway, where primary transcripts are processed by DROSHA/RNASEN-DGCR8, and then processed by DICER into small RNA duplexes (21-22nt). These are then loaded into the four Argonaute proteins (AGO1, AGO2, AGO3 and AGO4) to form an active RNA Induced Silencing Complex (RISC). The physiological mechanisms of small RNA sorting into different Argonaute proteins, and strand selection from the small RNA duplexes are not very well understood. By utilizing the Argonaute-dependent biogenesis pathways, we can design novel shRNA applications where the canonical biogenesis pathway is absent or disrupted, with several key advantages. First, only a catalytic Ago2 is required for small RNA-mediated repression, both in biogenesis and effector function. Second, thermodynamic asymmetry of the duplex (to ensure the correct small RNA is generated) is no longer a design concern, as the Ago2-depedent biogenesis mechanism produces one active (mature) small RNA species. This eliminates the possibility of off target effects resulting from the incorporation of the non-functional strand into an active RISC. The miR-451 shRNA design is highly flexible, requiring only high sequence complementarity in the precursor hairpin to mimic the structural features of the microRNA.

An example of a miR-451 mimic shRNA targeting p53 is shown in figure 1 (fig.1). A reverse complement of the 22nt target site of p53 coding region is generated, which will serve as the “active” small RNA species in RISC. Starting from the 5’ end, we designate the first 17 to 18 nucleotides to form one half of the stem region, then appending perfect complementary sequences to the 3’ end of the RNA molecule. When properly folded, this will generate a perfectly paired stem region of 17 to 18 nucleotides in length, with the remaining 4-5 nucleotides of the 22nt “active” small RNA sequence forming a small loop region. The resulting ~40nt hairpin can then be cleaved by *Ago2* (and not *Dicer*), and further processed to generate the mature 22nt small RNA.

In addition to their use as synthetic shRNAs in gene knockdown studies, miR-451 structural mimics can aid in understanding small RNA-mediated repression *in vivo*. The non-canonical biogenesis of mir-451 structural mimics allows us to bypass defects in the canonical (*Dicer*-dependent) biogenesis pathway, and provide stable expression of silencing molecules in a variety of previous untenable systems. Conditional ablation of *Dicer* results in dramatic phenotypes in virtually all cell types or tissues tested, but further characterization is impaired by the inability to utilize conventional (and *Dicer*-dependent) shRNA technologies. The *Ago2*-dependent (but *Dicer*-independent) biogenesis pathway would allow the development of an inducible mir-451 based expression system to probe for miRNA candidates or their targets, and potentially rescue the *Dicer*-deficient phenotypes. For example, cell proliferation defects in *Dicer*-deficient ES

cells can be dissected by introducing mir-451 mimics corresponding to miRNAs normally expressed in ES cells. Similarly, genes that are up-regulated can be repressed with miR-451 mimics specific for the gene(s) of interest to elucidate the critical components that are misregulated in the absence of *Dicer*. This approach can be extended to any cell type or tissue culture system where conditional *Dicer* ablation occurs.

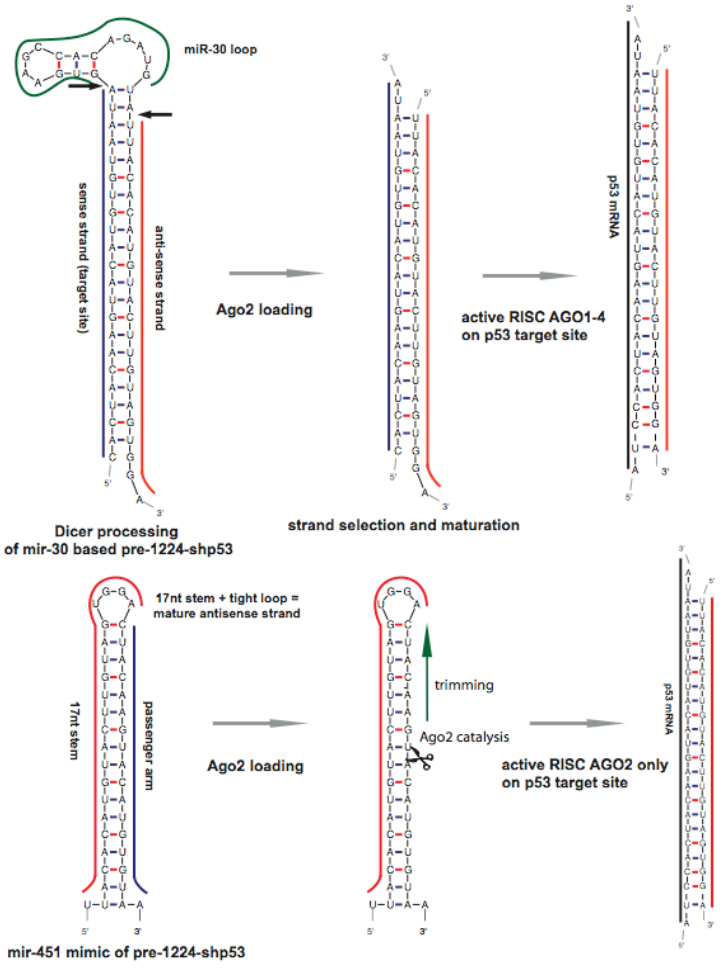

Figure 1: mir-451 mimic design

Schematic example for generating a mir-451 mimic molecule. Here p53-shRNA-1224 is shown as an example targeting the following site in the p53 message: UCCACUACAAGUACAUGUGUAA. Top panel: depicting the canonical miRNA processing pathway using the mir-30 backbone of the p53 hairpin. The mir-30 loop sequence in green. The strands are color coded: antisense strand in red and sense strand corresponding to the target site in blue. DICER RNAseIII cut sites are depicted using arrows. Bottom panel: showing the generation of the p53-1224 shRNA mimicking miR-451 structure that can be channeled through the Ago2 mediated miRNA biogenesis pathway. The antisense strand (red) spanning the stem is designed to extend into the loop. The passenger arm is highlighted in blue. Ago2 catalysis of the predicted phosphodiester bond is shown using scissors. This pathway generates only Ago2 active RISC.

Over the last 5 years, we have addressed the basic question of Argonaute's requirement during normal mammalian development using a combination of mouse genetics, molecular biology and biochemical approaches. This has led us to the discovery of Argonaute's function during placental development and viral defense, and also enabled us to explain the evolutionary pressure to conserve the catalytic activity of *Ago2* that is required for neonatal development. We have also identified a novel biogenesis pathway for microRNAs that is dependent on the catalytic activity of *Ago2*, and have demonstrated the basic mechanism of mir-451 maturation, a small RNA important for hematopoietic development. We believe that the characterization of a novel biological function for Argonaute proteins would open new avenues of research into small RNA-mediated regulation, while enabling the development of novel shRNA technologies (miR-451 mimics) that could potentially enhance the flexibility and specificity of gene knockdowns in mammalian systems. The greater understanding of the role of Argonaute proteins has firmly placed them back at the core of the small-RNA regulatory pathway as both an initiator and effector of RNAi. Advances in human genetic studies will allow us to predict whether the RNAi machinery is involved in complex developmental genetic disorders that are found to be associated with chromosomal regions harboring these components.

References

2008. *MicroRNAs: From basic science to disease biology*. cambridge university press, New York.

Adams DJ, Biggs PJ, Cox T, Davies R, van der Weyden L, Jonkers J, Smith J, Plumb B, Taylor R, Nishijima I et al. 2004. Mutagenic insertion and chromosome engineering resource (MICER). *Nat Genet* **36**: 867-871.

Alisch RS, Jin P, Epstein M, Caspary T, Warren ST. 2007. Argonaute2 is essential for mammalian gastrulation and proper mesoderm formation. *PLoS Genet* **3**: e227.

Andl T, Murchison EP, Liu F, Zhang Y, Yunta-Gonzalez M, Tobias JW, Andl CD, Seykora JT, Hannon GJ, Millar SE. 2006. The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles. *Curr Biol* **16**: 1041-1049.

Aravin A, Tuschl T. 2005. Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett* **579**: 5830-5840.

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744-747.

Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC. 2008. Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. *Proc Natl Acad Sci U S A* **105**: 7964-7969.

Babar IA, Slack FJ, Weidhaas JB. 2008. miRNA modulation of the cellular stress response. *Future Oncol* **4**: 289-298.

Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev* **22**: 2773-2785.

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

Baumberger N, Baulcombe DC. 2005. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* **102**: 11928-11933.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-366.

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ. 2003. Dicer is essential for mouse development. *Nat Genet* **35**: 215-217.

Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C. 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J* **17**: 170-180.

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

Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16**: 2733-2742.
Cerutti H, Casas-Mollano JA. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet* **50**: 81-99.
Chen JF, Murchison EP, Tang R, Callis TE, Tatsuguchi M, Deng Z, Rojas M, Hammond SM, Schneider MD, Selzman CH et al. 2008. Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure. *Proc Natl Acad Sci U S A* **105**: 2111-2116.
Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* **436**: 740-744.
Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE et al. 2010. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev*.
Chitwood DH, Nogueira FT, Howell MD, Montgomery TA, Carrington JC, Timmermans MC. 2009. Pattern formation via small RNA mobility. *Genes Dev* **23**: 549-554.
Davis BN, Hata A. 2009. Regulation of MicroRNA Biogenesis: A miRiad of mechanisms. *Cell Commun Signal* **7**: 18.
Davis BN, Hilyard AC, Lagna G, Hata A. 2008. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature* **454**: 56-61.
Davis E, Caiment F, Tordoir X, Cavaille J, Ferguson-Smith A, Cockett N, Georges M, Charlier C. 2005. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol* **15**: 743-749.
Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231-235.
Dore LC, Amigo JD, Dos Santos CO, Zhang Z, Gai X, Tobias JW, Yu D, Klein AM, Dorman C, Wu W et al. 2008. A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc Natl Acad Sci U S A* **105**: 3333-3338.
Elbashir SM, Lendeckel W, Tuschl T. 2001a. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**: 188-200.
Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. 2001b. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J* **20**: 6877-6888.
Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. 2008. A human snoRNA with microRNA-like functions. *Mol Cell* **32**: 519-528.
Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**: 806-811.
Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF. 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**: 833-838.

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

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**: 631-640.

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. 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.

Haase AD, Jaskiewicz L, Zhang H, Laine S, Sack R, Gatignol A, Filipowicz W. 2005. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* **6**: 961-967.

Hamilton AJ, Baulcombe DC. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**: 950-952.

Hammond SM, Bernstein E, Beach D, Hannon GJ. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**: 293-296.

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**: 10898-10903.

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**: 2208-2213.

Hayashi K, Chuva de Sousa Lopes SM, Kaneda M, Tang F, Hajkova P, Lao K, O'Carroll D, Das PP, Tarakhovsky A, Miska EA et al. 2008. MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS One* **3**: e1738.

Hornstein E, Mansfield JH, Yekta S, Hu JK, Harfe BD, McManus MT, Baskerville S, Bartel DP, Tabin CJ. 2005. The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development. *Nature* **438**: 671-674.

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**: 834-838.

Hutvagner G, Simard MJ. 2008. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* **9**: 22-32.

Ibarra I, Erlich Y, Muthuswamy SK, Sachidanandam R, Hannon GJ. 2007. A role for microRNAs in maintenance of mouse mammary epithelial progenitor cells. *Genes Dev* **21**: 3238-3243.

Joshua-Tor L. 2006. The Argonautes. *Cold Spring Harb Symp Quant Biol* **71**: 67-72.

Kaneda M, Tang F, O'Carroll D, Lao K, Surani MA. 2009. Essential role for Argonaute2 protein in mouse oogenesis. *Epigenetics Chromatin* **2**: 9.

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**: 2654-2659.

Kim DH, Villeneuve LM, Morris KV, Rossi JJ. 2006. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**: 793-797.

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**: 126-139.

Koralov SB, Muljo SA, Galler GR, Krek A, Chakraborty T, Kanellopoulou C, Jensen K, Cobb BS, Merkenschlager M, Rajewsky N et al. 2008. Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. *Cell* **132**: 860-874.

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

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

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**: 415-419.

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**: 1437-1441.

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**: 1261-1266.

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R. 2005b. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nat Cell Biol* **7**: 719-723.

Liu Y, Pop R, Sadegh C, Brugnara C, Haase VH, Socolovsky M. 2006. Suppression of Fas-FasL coexpression by erythropoietin mediates erythroblast expansion during the erythropoietic stress response in vivo. *Blood* **108**: 123-133.

Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, Kinch LN, Grishin NV, Liu Q. 2009. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science* **325**: 750-753.

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

Lykke-Andersen K, Gilchrist MJ, Grabarek JB, Das P, Miska E, Zernicka-Goetz M. 2008. Maternal Argonaute 2 is essential for early mouse development at the maternal-zygotic transition. *Mol Biol Cell* **19**: 4383-4392.

Ma J, Flemr M, Stein P, Berninger P, Malik R, Zavolan M, Svoboda P, Schultz RM. 2010. MicroRNA activity is suppressed in mouse oocytes. *Curr Biol* **20**: 265-270.
MacRae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. 2008. In vitro reconstitution of the human RISC-loading complex. *Proc Natl Acad Sci U S A* **105**: 512-517.
Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136**: 656-668.
Maniataki E, Mourelatos Z. 2005. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev* **19**: 2979-2990.
Martinez J, Tuschl T. 2004. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev* **18**: 975-980.
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**: 607-620.
Meder B, Katus HA, Rottbauer W. 2008. Right into the heart of microRNA-133a. *Genes Dev* **22**: 3227-3231.
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**: 185-197.
Melton C, Judson RL, Blelloch R. 2010. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* **463**: 621-626.
Morita S, Horii T, Kimura M, Goto Y, Ochiya T, Hatada I. 2007. One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. *Genomics* **89**: 687-696.
Muljo SA, Ansel KM, Kanellopoulou C, Livingston DM, Rao A, Rajewsky K. 2005. Aberrant T cell differentiation in the absence of Dicer. *J Exp Med* **202**: 261-269.
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**: 12135-12140.
Murchison EP, Stein P, Xuan Z, Pan H, Zhang MQ, Schultz RM, Hannon GJ. 2007. Critical roles for Dicer in the female germline. *Genes Dev* **21**: 682-693.
Nagy A, Gertsenstein M, Vintersten K, Behringer R. 2003. *Manipulating the Mouse Embryo: A Laboratory Manual*. CSHL press.
Nagy A, Rossant J. 2000. Production and analysis of ES cell aggregation chimeras. in *Gene Targeting A Practical Approach* (ed. AL Joiner), pp. 189-192. Oxford University Press.
Newman MA, Thomson JM, Hammond SM. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA* **14**: 1539-1549.
Nilsen TW. 2007. Mechanisms of microRNA-mediated gene regulation in animal cells. *Trends Genet* **23**: 243-249.

O'Carroll D, Mecklenbrauker I, Das PP, Santana A, Koenig U, Enright AJ, Miska EA, Tarakhovsky A. 2007. A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. *Genes Dev* **21**: 1999-2004.
Orban TI, Izaurralde E. 2005. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**: 459-469.
Otsuka M, Jing Q, Georgel P, New L, Chen J, Mols J, Kang YJ, Jiang Z, Du X, Cook R et al. 2007. Hypersusceptibility to vesicular stomatitis virus infection in Dicer1-deficient mice is due to impaired miR24 and miR93 expression. *Immunity* **27**: 123-134.
Papapetrou EP, Korkola JE, Sadelain M. 2009. A Genetic Strategy for Single and Combinatorial Analysis of miRNA Function in Mammalian Hematopoietic Stem Cells. *Stem Cells*.
Pase L, Layton JE, Kloosterman WP, Carradice D, Waterhouse PM, Lieschke GJ. 2009. miR-451 regulates zebrafish erythroid maturation in vivo via its target gata2. *Blood* **113**: 1794-1804.
Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grasser FA, van Dyk LF, Ho CK, Shuman S, Chien M et al. 2005. Identification of microRNAs of the herpesvirus family. *Nat Methods* **2**: 269-276.
Qi Y, Denli AM, Hannon GJ. 2005. Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell* **19**: 421-428.
Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008-1012.
Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**: 901-906.
Rivas FV, Tolia NH, Song JJ, Aragon JP, Liu J, Hannon GJ, Joshua-Tor L. 2005. Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* **12**: 340-349.
Rodriguez A, Vigorito E, Clare S, Warren MV, Couttet P, Soond DR, van Dongen S, Grocock RJ, Das PP, Miska EA et al. 2007. Requirement of bic/microRNA-155 for normal immune function. *Science* **316**: 608-611.
Rossant J, Cross JC. 2001. Placental development: lessons from mouse mutants. *Nat Rev Genet* **2**: 538-548.
-. 2002. extraembryonic lineages. in *Mouse development Pattering, Morphogenesis and Organogenesis* (eds. J Rossant, P Tam), pp. 155-180.
Sasaki T, Shiohama A, Minoshima S, Shimizu N. 2003. Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics* **82**: 323-330.
Sashital DG, Doudna JA. 2010. Structural insights into RNA interference. *Curr Opin Struct Biol* **20**: 90-97.
Schaefer A, O'Carroll D, Tan CL, Hillman D, Sugimori M, Llinas R, Greengard P. 2007. Cerebellar neurodegeneration in the absence of microRNAs. *J Exp Med* **204**: 1553-1558.

Schmidt A, Palumbo G, Bozzetti MP, Tritto P, Pimpinelli S, Schafer U. 1999. Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. *Genetics* **151**: 749-760.

Schwarz DS, Tomari Y, Zamore PD. 2004. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. *Curr Biol* **14**: 787-791.

Seitz H, Zamore PD. 2006. Rethinking the microprocessor. *Cell* **125**: 827-829.

Sekita Y, Wagatsuma H, Nakamura K, Ono R, Kagami M, Wakisaka N, Hino T, Suzuki-Migishima R, Kohda T, Ogura A et al. 2008. Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nat Genet* **40**: 243-248.

Shen B, Goodman HM. 2004. Uridine addition after microRNA-directed cleavage. *Science* **306**: 997.

Siolas D, Lerner C, Burchard J, Ge W, Linsley PS, Paddison PJ, Hannon GJ, Cleary MA. 2005. Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* **23**: 227-231.

Socolovsky M, Nam H, Fleming MD, Haase VH, Brugnara C, Lodish HF. 2001. Ineffective erythropoiesis in Stat5a(-/-)5b(-/-) mice due to decreased survival of early erythroblasts. *Blood* **98**: 3261-3273.

Sohn SY, Bae WJ, Kim JJ, Yeom KH, Kim VN, Cho Y. 2007. Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* **14**: 847-853.

Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**: 1026-1032.

Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**: 1434-1437.

Su H, Trombly MI, Chen J, Wang X. 2009a. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes Dev* **23**: 304-317.

Su J, Zhu Z, Wang Y, Jang S. 2009b. Isolation and characterization of Argonaute 2: a key gene of the RNA interference pathway in the rare minnow, Gobiocypris rarus. *Fish Shellfish Immunol* **26**: 164-170.

Suh N, Baehner L, Moltzahn F, Melton C, Shenoy A, Chen J, Blelloch R. 2010. MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Curr Biol* **20**: 271-277.

Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC. 1999. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**: 123-132.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453**: 534-538.

Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. 2008. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**: 1124-1128.
Thai TH, Calado DP, Casola S, Ansel KM, Xiao C, Xue Y, Murphy A, Frendewey D, Valenzuela D, Kutok JL et al. 2007. Regulation of the germinal center response by microRNA-155. *Science* **316**: 604-608.
Tolia NH, Joshua-Tor L. 2007. Slicer and the argonautes. *Nat Chem Biol* **3**: 36-43.
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**: 320-324.
Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev* **20**: 515-524.
Valenzuela DM, Murphy AJ, Frendewey D, Gale NW, Economides AN, Auerbach W, Poueymirou WT, Adams NC, Rojas J, Yasenchak J et al. 2003. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* **21**: 652-659.
Viswanathan SR, Daley GQ, Gregory RI. 2008. Selective blockade of microRNA processing by Lin28. *Science* **320**: 97-100.
Wang HW, Noland C, Siridechadilok B, Taylor DW, Ma E, Felderer K, Doudna JA, Nogales E. 2009. Structural insights into RNA processing by the human RISC-loading complex. *Nat Struct Mol Biol* **16**: 1148-1153.
Wang Y, Baskerville S, Shenoy A, Babiarz JE, Baehner L, Blelloch R. 2008a. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* **40**: 1478-1483.
Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R. 2007. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* **39**: 380-385.
Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. 2008b. Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**: 209-213.
Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539-543.
Weidhaas JB, Babar I, Nallur SM, Trang P, Roush S, Boehm M, Gillespie E, Slack FJ. 2007. MicroRNAs as potential agents to alter resistance to cytotoxic anticancer therapy. *Cancer Res* **67**: 11111-11116.
Whyatt D, Lindeboom F, Karis A, Ferreira R, Milot E, Hendriks R, de Bruijn M, Langeveld A, Gribnau J, Grosveld F et al. 2000. An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature* **406**: 519-524.
Wilson JE, Connell JE, Macdonald PM. 1996. aubergine enhances oskar translation in the Drosophila ovary. *Development* **122**: 1631-1639.
Yekta S, Shih IH, Bartel DP. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**: 594-596.

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC. 2006. Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* **127**: 747-757.

Yuan YR, Pei Y, Ma JB, Kuryavyi V, Zhadina M, Meister G, Chen HY, Dauter Z, Tuschl T, Patel DJ. 2005. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* **19**: 405-419.

Zamore PD, Tuschl T, Sharp PA, Bartel DP. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25-33.

Zhang H, Kolb FA, Brondani V, Billy E, Filipowicz W. 2002. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J* **21**: 5875-5885.

Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. 2004. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**: 57-68.

Appendix1: Characterization of Dicer-deficient murine embryonic stem cells

Elizabeth P. Murchison[1], Janet F. Partridge[2], Oliver H. Tam, **Sihem Cheloufi** and Gregory J. Hannon[1*]

[1] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, NY, 11724

[2] Department of Biochemistry
MS340, Room D4058
St. Jude Children's Research Hospital
332 N. Lauderdale
Memphis, TN 38105-2794 USA

to whom correspondence should be addressed

Abstract

Dicer is an RNase III-family nuclease that initiates RNA interference (RNAi) and related phenomena by generation of the small RNAs that determine the specificity of these gene silencing pathways. We have previously shown that Dicer is essential for mammalian development, with Dicer-deficient mice dying at embryonic day 7.5 with a lack of detectable multipotent stem cells. To permit a more detailed investigation of the biological roles of Dicer, we have generated embryonic stem cell lines in which their single Dicer gene can be conditionally inactivated. As expected, Dicer loss compromises maturation of microRNAs and leads to a defect in gene silencing triggered by long double-stranded RNAs (dsRNAs). However, the absence of Dicer does not affect the ability of small interfering RNAs to repress gene expression. Of interest, Dicer loss does compromise the proliferation of ES cells, possibly rationalizing the phenotype previously observed in Dicer-null animals. Dicer loss also affects the abundance of transcripts from mammalian centromeres but does so without a pronounced affect on histone modification status at pericentric repeats or methylation of centromeric DNA. These studies provide a conditional model of RNAi deficiency in mammals that will permit the dissection of the biological roles of the RNAi machinery in cultured mammalian cells.

Introduction

Dicer is a multidomain ribonuclease with specificity for dsRNA. Dicer's catalytic role in the production of small RNAs is central to dsRNA-mediated gene silencing or RNA interference (RNAi) (Bernstein et al 2001; Bernstein et al 2003; Hannon 2002). Dicer is essential for the response of many organisms to dsRNAs encountered from exogenous sources, including those generated by viral infection or those that have been experimentally delivered. Additionally, Dicer must process endogenous dsRNAs that trigger silencing responses directed at repetitive elements such as transposons and centromeric sequences. Finally, Dicer must promote the maturation of endogenous noncoding RNAs, the microRNAs (miRNAs) that enter the RNAi pathway to regulate the expression of protein coding genes. In all of these cases, Dicer is directly responsible for producing from a longer precursor the ~22-nt dsRNAs bearing signature 2-nt 3' overhangs that are a hallmark of RNAi and related pathways.

The Dicer enzyme is organized as a modular structure, with a canonical Dicer containing an N-terminal DEAD-box helicase domain, a domain of unknown function (DUF283), a PAZ domain, a pair of catalytic RNase III domains, and a C-terminal dsRNA-binding domain. Structural models of Dicer predict that the RNase III domains combine as an intramolecular dimer to produce a single compound catalytic center that measures the position of scissile bonds precisely from an end of its dsRNA substrate (Carmell & Hannon 2004; Siolas et al 2005; Zhang et al 2002a; Zhang et al 2004).

Mounting evidence also suggests additional roles for Dicer enzymes. In Drosophila, Dcr-2 acts, along with its partner R2D2, in a loading complex responsible for sensing small interfering RNA (siRNA) asymmetry and placing the correct strand of the siRNA into RISC (Khvorova et al 2003; Lee et al 2004b; Liu et al 2003a; Pham et al 2004; Schwarz et al 2003; Tomari et al 2004). Similar activities must also exist for miRNAs because one strand of the precursor miRNA predominantly contributes to the production of a functional RISC (Khvorova et al 2003; Schwarz et al 2003). In mice and humans, Dicer is encoded by a single locus, whose protein product must account for all proposed Dicer activities.

Figure 1.1

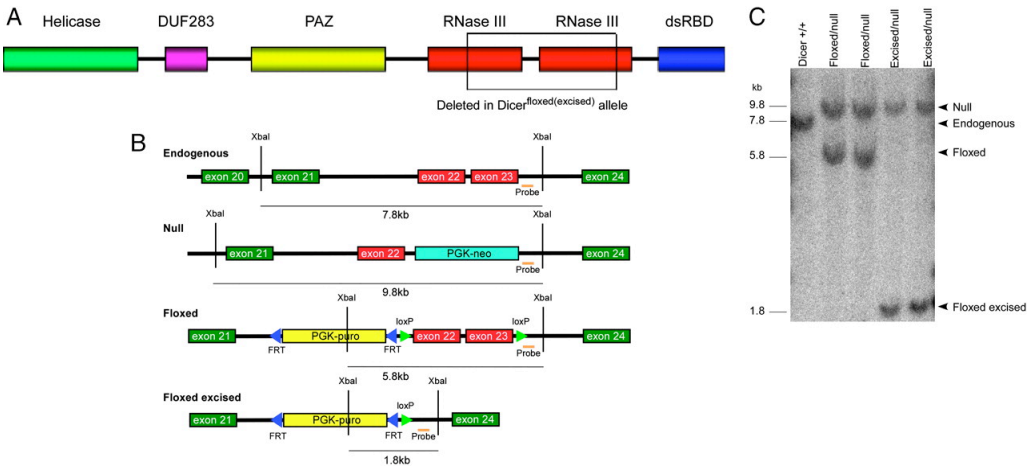

Figure 1.1. Conditional Dicer ES cells. (A) Dicer domain structure. Dicer floxed allele was created by knocking in a floxed cassette, flanking the RNaseIII domain-encoding exons 22 and 23 with loxP sites. **(B)** ES cells that are heterozygous for a Dicer-null allele (created by deletion of an RNaseIII domain) have previously been described (Bernstein et al 2003). The introduction of the Dicer floxed allele into these ES cells created Dicer floxed ES cells. Treatment with Cre leads to excision of the majority of both RNaseIII domains, creating a second nonfunctional allele. The four different Dicer alleles (wild-type, null, conditional but not excised, and conditional excised) can be distinguished by Southern blotting. **(C)** A Southern blot used to confirm the genotypes of the cell lines used in this study is shown.

Based on genetic studies in other systems, mouse Dicer has a predicted involvement in numerous biological processes. Thus, it is not surprising that Dicer-deficient mice die very early in development, around embryonic day 7.5, with essentially a complete loss of pluripotent stem cells (Bernstein et al 2003). In addition, mouse embryos hypomorphic for Dicer die mid-gestation (Yang et al 2005). Dicer-deficient zebrafish progress much further in development, probably owing to the presence of maternally deposited Dicer transcripts (Wienholds et al 2003). Depletion of maternal Dicer achieved by germ-line transplantation in zebrafish revealed that Dicer is required for morphogenesis but not cell fate specification during zebrafish embryogenesis and that the absence of miRNAs is responsible, at least in part, for this phenotype (Giraldez et al 2005).

Recently it has become clear that Dicer is essential for viability and proliferation of some cell types but is dispensable for others. Chicken DT40 cells in which Dicer has been inactivated by homologous recombination become aneuploid and undergo growth arrest upon loss of Dicer (Fukagawa et al 2004). Specific deletion of Dicer during T cell development in the mouse thymus showed that survival is compromised in the αβ- but not γδ-lineage, but that transcriptional gene silencing during CD4/8 differentiation is not perturbed (Cobb et al 2005). In addition, Dicer-deficient mouse ES cells selected for survival *in vitro* are defective in differentiation (Kanellopoulou et al 2005). These phenotypes are consistent with an essential role for Dicer, at least in some cell types, in cellular metabolism and proliferation. However, the precise underlying cause of the requirement for Dicer in these cells remains unclear.

A series of recent studies has implicated the RNAi machinery in the establishment and maintenance of heterochromatin at centromeres (Fukagawa et al 2004; Hall et al 2002; Kanellopoulou et al 2005; Mochizuki & Gorovsky 2005; Pal-Bhadra et al 2004; Provost et al 2002; Verdel et al 2004; Volpe et al 2003; Volpe et al 2002). In *Schizosaccharomyces pombe*, mutants in the RNAi pathway have a number of mitotic and meiotic defects and display a loss of heterochromatin at the centromeres and a derepression of transcription of centromeric repeats (Hall et al 2003; Hall et al 2002; Provost et al 2002; Volpe et al 2003; Volpe et al 2002). Roles for RNAi at heterochromatic mating-type loci are more restricted, with the machinery mainly acting during the initiation phase rather than in the maintenance of heterochromatin (Hall et al 2002; Jia et al 2004). Similar roles for RNAi in the husbandry of repetitive elements and formation of heterochromatin have been observed in *Tetrahymena* (Mochizuki et al 2002; Mochizuki & Gorovsky 2005), chicken DT40 cells (Fukagawa et al 2004), mouse ES cells (Kanellopoulou et al 2005), and plants (Gendrel & Colot 2005) but notably not in *Neurospora* (Chicas et al 2004; Chicas et al 2005; Freitag et al 2004). We set out to examine models of Dicer function in mammalian cells by expanding on our earlier constitutive mutant animals through the creation of conditional Dicer-null ES cells.

ES cells are a transient pluripotent cell population found in the mammalian blastocyst that can be isolated and cultured *in vitro* without a loss of their ability to contribute to all mouse tissues. Mouse ES cells contain Dicer and express a substantial number of miRNAs, including some that are unique to ES cells

(Houbaviy et al 2003). Here we describe the creation and characterization of Dicer conditional ES cell lines that can be induced to inactivate Dicer upon exposure to Cre recombinase. We find that, upon loss of Dicer activity, ES cells have a significant proliferation defect. However, this defect can be overcome with time, probably because of the accumulation of additional mutations. As expected, Dicer-deficient ES cells are unable to process pre-miRNAs or dsRNAs. Interestingly, Dicer-null cells are still able to mount an siRNA-mediated gene silencing response. In addition, ES cells lacking Dicer accumulate transcripts derived from the centromeric major satellite but retain the integrity of centromeric heterochromatin as indicated by the presence of cytosine methylation and histone H3 lysine 9 trimethylation.

Materials and methods

Gene targeting

A conditional targeting construct was linearized with Pvu-I. The 4E4 Dicer heterozygous ES cells (Bernstein et al 2003) were resuspended at a density of 11 x 10^6 cells/ml and electroporated with the linearized targeting vector at 230V, 500μF using a Biorad genepulser. Electroporated cells were plated at

Figure 1.2

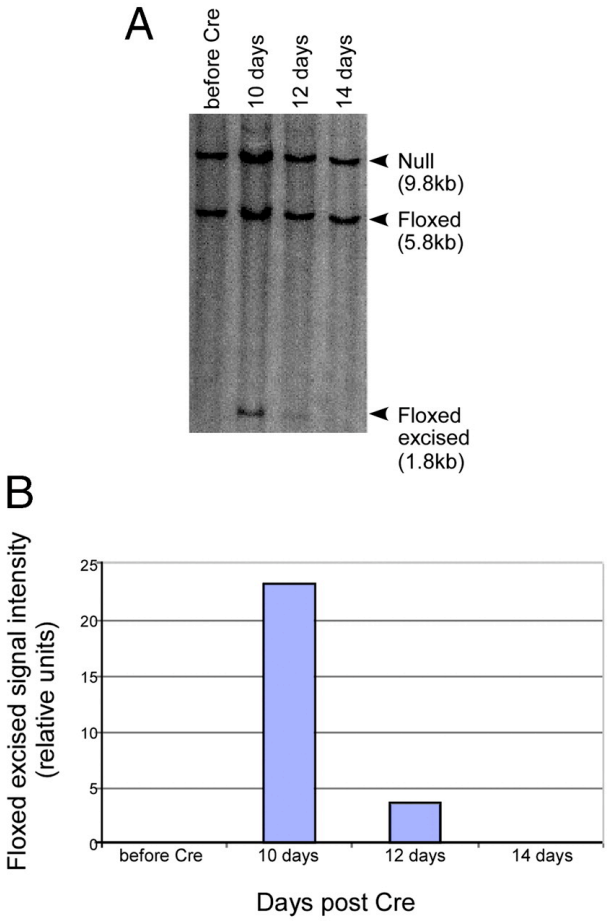

Figure 1.2. Dicer-null cells show proliferation defects. (A) Dicer floxed ES cells were transiently transfected with a Cre-GFP expression plasmid and sorted for GFP-positive cells, and the population was analyzed at different times after Cre expression by Southern blotting. The $Dicer^{flox(excised)/null}$ population was lost from the mixed population at 10-15 days after excision. **(B)** The relative representation of the $Dicer^{flox(excised)/null}$ versus the $Dicer^{flox/null}$ cells in the population was quantified by image analysis of the gel shown in **(A)**.

a density of 2 x 10^7cells/ml and, 24 hours after electroporation, were selected with puromycin (5ug/ml). Puromycin resistant clones were picked after five days of selection, and transferred to 96-well plates. Three days later these were split 1:4, and one plate of cells was harvested for genotyping.

Cell culture

During gene-targeting steps, ES cells were grown on an irradiated STO cell feeder layer. During Cre-transfection steps, ES cells were either grown on an irradiated STO feeder layer, or were grown on plates treated with 0.1% gelatin. ES cells were grown in Knockout DMEM (Gibco) supplemented with 2mM L-glutamine, 49.5U/ml penicillin, 38.8µg/ml streptomycin and 0.1mM β-mercaptoethanol. If grown on gelatin, media was supplemented with 1000U/ml LIF (Chemicon).

Cre Transfection and FACS

ES cells were transfected with pCMV-Cre-EGFP (gift from D. Lee and D.W. Threadgill, UNC) using Lipofectamine 2000 at 5µg/ml and pCMV-Cre-EGFP at 1.5ug/ml, according to the manufacturer's recommendations (Invitrogen). GFP-positive cells were enriched by sorting using a FACSVantage DiVa cell sorter.

Southern analysis

Genomic DNA was prepared by phase-lock gel columns (Eppendorf). Genomic DNA was digested overnight, separated on a 1% TAE agarose gel, transferred onto a hybond N+ nylon membrane overnight, UV crosslinked and hybridized with a random primed α^{32}P-dCTP labeled probe. Probes were amplified by PCR from the genome. In the case of the genotyping probe primers were as follows: floxF, TTGGAGCTGTCTAGTTAGTTATGC; floxR, GTTGCAAGATAAACATGGTCACAA. For the minor satellite probe primer sequences were: JPO106, AGTGTATATCAATGAGTTACAATG, and JPO107, CATCTAATATGTTCTACAGTGTGG. The major satellite probe and mitochondrial probes were a gift from T. Bestor, Columbia (Walsh & Bestor 1999). Southern blots were imaged and quantified using a Fuji phosphoimager and software.

Antibodies and Western blots

Whole cell extract was prepared, run on a 6% polyacylamide gel, and transferred onto a nitrocellulose membrane. α-Dicer antibody (DICER8) was obtained from the Filipowicz lab (Billy et al 2001) and used at 1:1000. The α-CRM1/exportin-1 (BD Biosciences) was used at 1:2000.

Northern analysis

1µg of total RNA extracted by Trizol reagent (Intvitrogen) was run on a 15% polyacrylamide gel and transferred onto a Hybond N+ nylon membrane. DNA oligo probes were labeled with γ-^{32}P (miR292-as ACACTCAAAACCTGGCGGCACTT; miR293 ACACTACAAACTCTGCGGCACT (Houbaviy et al 2003).

Figure 1.3.

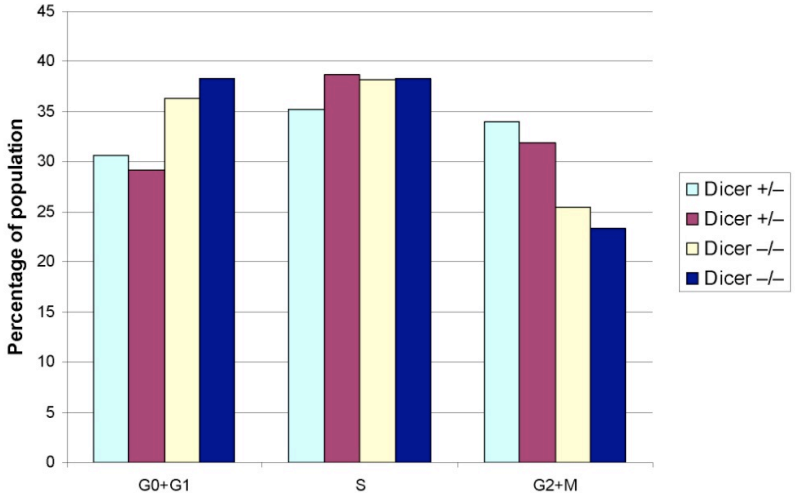

Figure 1.3. Cell-cycle profiles of Dicer null cells. Cell-cycle profiles of the indicated cell lines were determined by staining with propidium iodide.

DsRNA silencing assays

DsRNA was prepared by bidirectional T7 transcription using a Megascript kit (Ambion) according to the manufacturer's directions. T7-tagged firefly luciferase dsRNA was amplified from pGL3 template (Promega) with primer sequences : ffluc(fwd), TAATACGACTCACTATAGGGATAAAGAAAGGCCCGGCGCC, and ffluc(rev), TAATACGACTCACTATAGGGACGAACGTGTACATCGACTGAAAT; pEGFP-N1 (Clontech) was used as a PCR template for GFP dsRNA, using primers T7 GFP F, TAATACGACTCACTATAGGGATCCTGGTCGAGCTGGAC, and T7 GFP R, TAATACGACTCACTATAGGGTGCTCAGGTAGTGGTTGT. DsRNA was transfected into ES cells using Lipofectamine 2000 (Invitrogen) at 10µg/ml with dsRNA at a concentration of 2µg/ml, along with reporter plasmids pSV40renilla and pGL3 (Promega) at 0.1µg/ml and 1µg/ml concentrations respectively. Luciferase activity was measured 48 hrs post-transfection using a Dual Luciferase Detection Kit (Promega) and an ASI luminometer.

SiRNA silencing assays

SiRNAs used in the study were sod40/21.251 and 21.233/sod40 (Schwarz et al 2003) and a control siRNA corresponding to the target sequence AAACCCTAGCGCCATCGTGCC. Reporter plasmids pGL2-p10wt-sense and pGL2-p10wt-antisense were gifts from P. Zamore, UMass. SiRNAs were transfected into ES cells at a total concentration of 100nM, with the total concentration being maintained by mixing variable ratios of target siRNA and

control siRNA. Reporters were cotransfected at 1µg/ml (pGL2-p10wt-sense and pGL2-p10wt-antisense) along with pSV40renilla (Promega), at 0.1µg/ml. Lipofectamine 2000 (Invitrogen) was used at 10µg/ml. Dual luciferase assays were performed at 48hrs post-transfection as described above.

RNA preparation

Cells grown in 6 well tissue culture dishes were washed with PBS. 1ml Trizol (Invitrogen) was added to each well, and total cellular RNA was prepared according to manufacturers instructions. Contaminating DNA was removed as follows. RNA was diluted with DEPC H_20, and buffer RLT was added (Qiagen RNeasy kit), prior to shredding using a QIAshredder. RNA was then purified on an RNeasy mini column, with on column DNAse1 digestion. Eluted RNA was re-DNAsed in solution for 15 minutes at 30^0C using Invitrogen DNAse I, prior to 2 further rounds of RNA clean up with on column DNAse1 digestion using Qiagen RNeasy kit. RNA concentration was measured by UV absorbance, and the integrity of the RNA was checked by capillary electrophoresis (Agilent technologies).

cDNA preparation and quantitation

A total of 0.1 ug of purified RNA was used in cDNA synthesis with 0.1 ug random nonomer primers in a 20 ul reaction volume. 1 ul of this cDNA was used as template for quantitative PCR analysis of actin transcript levels. 1.25 ul 5uM actin F and actin R primers (actin F:ACCCACACTGTGCCCATCTAC**,** actin R: AGCCAAGTCCAGACGCAGG**)** were used with 1ul cDNA, 12.5 ul Quantitect SYBR mix (Qiagen) in a 25ul reaction volume. PCR conditions were 50^0C for 2 min, 95^0C for 10 min, followed by 40 cycles of 95^0C for 15 seconds alternating with 60^0C for 1 minute, on an ABI PRISM 7700 sequence detector. Relative amplification of duplicate samples for each cDNA were compared with 5 fold serial dilutions of wild type AB2.2 cDNA, with C_t values plotted against Log dilutions.

For centromeric transcript analyses, 1 ug RNA was used for cDNA synthesis with 90 pmoles of oligo dT_{12}, 150 pmoles of random nonamers, 150 pmoles of MajF1 or 150 pmoles of MajR1 (MajF1: GACGACTTGAAAAATGACGAAATC, MajR1: CATATTCCAGGTCCTTCAGTGTGC (Lehnertz et al 2003) in a 20 ul reaction, and 1ul of this cDNA was used for PCR analysis of satellite transcripts. PCR conditions used were 94^0C for 4 min, followed by 25 or 30 cycles of 94^0C for 30 sec, 67^0C for 30 sec and 72^0C for 1 min. PCR products (30 cycles) were resolved by electrophoresis on 1.2% agarose gels, and stained with ethidium bromide prior to photography. PCRs performed with 25 cycles were 'spiked' by addition of α^{32}P-dCTP, and products were fractionated on 4% PAG prior to drying gels, and quantification of bands by phosphoimaging (Storm, Molecular Dynamics) and use of Imagequant software.

Chromatin Immunoprecipitations

Chromatin immunoprecipitations were done by using anti-trimethyl H3K9 antibody (1:100, Upstate Biotechnology, Lake Placid, NY) according to Upstate Biotechnology protocol. Immunoprecipitated DNA was used in PCRs with primers specific for the major satellite (Lehnertz et al 2003).

Results

Generation of Dicer deficient ES cells

In order to generate ES cells that conditionally express Dicer, we replaced a wild-type Dicer allele with a floxed allele in ES cells that already contained a null Dicer allele (Bernstein et al 2003) (Fig. 1.1). We screened puromycin resistant clones for those in which the intact allele of Dicer had been replaced by the floxed allele. In these clones, we expected that functional Dicer would be expressed from the floxed locus, in which exons 22 and 23, encoding both catalytic RNase III domains, were flanked by loxP sites (Fig. 1.1). Two independent Dicer conditional clones were recovered and named 19F9 and 2D5.

To examine the phenotypic effects of Dicer loss, we used the $Dicer^{flox/null}$ clones to generate Dicer-null cells. Expression of Cre recombinase in $Dicer^{flox/null}$ ES cells was expected to lead to recombination between the loxP sites and excision of exons 22 and 23, resulting in Dicer-deficient ES cells. We transiently expressed Cre in $Dicer^{flox/null}$ clone 19F9 by transfection of a Cre-expression plasmid, enriching for transfected cells by FACS for those that express the co-expressed GFP marker. This population was plated onto gelatin-coated plates and analyzed by Southern blotting for the presence of Dicer-deficient cells. Initially highly enriched, Dicer-deficient cells represented ~30% of the population at 10 days after Cre treatment. However, the Dicer-deficient sub-population was rapidly depleted, and by two weeks post Cre transfection, was no longer detectable by southern (Fig. 1.2). These results demonstrated that Dicer-deficient ES cells are at a disadvantage as compared to the $Dicer^{flox/null}$ cells, indicating the absence of Dicer causes either a significant alteration in the growth of ES cells or compromises their survival.

In order to distinguish between these possibilities, we plated 19F9 populations enriched for Cre-GFP transfection at clonogenic densities. Single clones were picked six days after Cre transfection and analyzed by PCR for recombination between loxP sites. Among different experiments, between fifty and ninety percent of clones analyzed had undergone Cre-mediated excision events. Clones were then transferred to individual wells of a 96-well plate, and cultured in isolation either on gelatin or on an irradiated feeder layer. Strikingly, within a week of culture in individual wells, all clones that had not undergone Cre-mediated excision events had confluence, whereas no Dicer deficient clones had reached a similar density. In fact, Dicer deficient clones required between one and three months to become confluent on a single well of a 96-well plate,

with clones grown on feeders proliferating slightly faster than clones grown on gelatin. In addition, more than 50 percent of clones that were grown on gelatin completely ceased proliferating and were eventually discarded. Ten percent of clones grown on feeders had a similar outcome. After several weeks, Dicer-deficient clones that eventually achieved confluency in a well of a 96-well plate were passaged, and a small number of these had proliferation rates sufficient to establish continuous Dicer deficient cell line

Figure 1.4

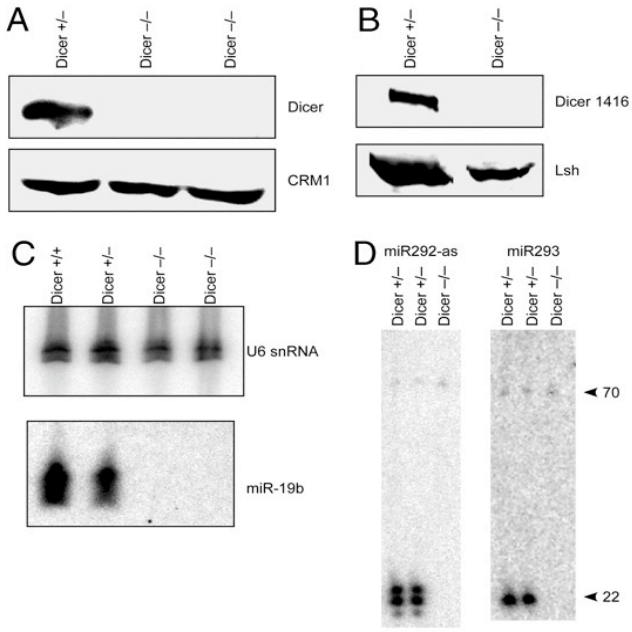

Figure 1.4. No full-length Dicer protein can be detected in whole-cell lysates of Dicer-deficient cells. (A) Western blotting using an antibody recognizing the RNase III region of Dicer. **(B)** Western blotting using an antibody recognizing the PAZ domain of Dicer. Dicer-null ES cells fail to process pre-miRNAs into mature miRNAs. Total RNA extracted from floxed/null and floxed excised/null ES cells was probed with ^{32}P-labeled oligonucleotides complementary to miR-19b (with equal loading indicated by blotting for U6 snRNA) **(C)** or miR-293 and miR-292-as **(D)**.

In this study we primarily analyzed two representative Dicer deficient cell lines 27H10 and 27G5 isolated from populations derived on gelatin, along with their 'sibling' conditional cell line, 27D4. For comparison, we also examined two cell

lines that retain their original genotype (Dicer$^{flox/null}$) despite having been exposed to Cre recombinase and isolated alongside the Dicer-null cell lines from single parental cells.

The significant proliferation lag seen in Dicer-deficient cells is consistent with the possibility that loss of Dicer leads primarily to an acute loss of proliferative potential that can eventually be rescued by the accumulation of compensatory events, either mutations or stable changes in gene expression. These 'escapers' eventually achieve a growth rate approaching that of Dicer wild-type or heterozygous cells. Interestingly, the Dicer-deficient 'escapers' have an altered cell cycle profile, as compared to Dicer$^{flox/null}$ cells, with an slight increase in G1 and G0 cells in and a corresponding decrease of cells in G2 and M (Fig. 1.3). The presence of a feeder layer both enhanced the survival rate and reduced the proliferation delay in Dicer deficient ES cells, although these growth conditions did not affect either survival or growth rate of Dicer heterozygous or wild-type ES cells. It seems likely that feeder cells either supply factors in trans to Dicer deficient ES cells, or that growth of Dicer deficient ES cells is sensitive to cell density.

Properties of Dicer deficient ES cells

The genotypes of Dicer-deficient cell clones 27H10 and 27G5 were confirmed by Southern blotting, along with heterozygous floxed/null clones 19F9 and 27D4, and parental 129-derived wild-type ES cell clone AB2.2 (Fig. 1.1C). As expected, no full-length Dicer protein can be detected in floxed excised/null ES cell whole cell extracts by Western blotting using antibodies recognizing either the N or C termini of Dicer (Fig. 1.4A and B).

Since Dicer is encoded by a single locus in mouse, it was predicted that the single mouse Dicer enzyme is responsible for processing of pre-miRNAs into mature miRNAs and for processing of long double-stranded RNAs (dsRNAs) into siRNAs to initiate post-transcriptional gene silencing (PTGS). To test whether mouse Dicer is responsible for pre-miRNA processing we analyzed mature miRNA accumulation in Dicer-deficient cells. Northern blotting using several miRNA probes revealed, as expected, a complete absence of mature miRNAs in Dicer deficient cells, and in some cases a slight accumulation of Dicer substrates, the pre-miRNAs (Fig. 1.4C and D).

ES cells have previously been shown to lack prominent non-specific (e.g., PKR) responses to long double-stranded RNAs and instead to mount a sequence-specific silencing response to these triggers (Billy et al 2001; Paddison et al 2002; Yang et al 2001). To confirm that Dicer is essential for these sequence directed responses, we transfected 500nt dsRNAs corresponding in sequence either to firefly luciferase or GFP into our Dicer-deficient cells and their Dicer-expressing siblings. In each case, reporter plasmids encoding firefly luciferase and renilla luciferase were cotransfected with the dsRNA. At forty-eight hours post transfection, cells were analyzed for both luciferase activities. Wild-type and heterozygous ES cell lines (AB2.2, 19F9, 27D4) greatly reduced in firefly luciferase activity when transfected with firefly dsRNA, as compared to

cells transfected with GFP dsRNA (Fig. 1.5). However, Dicer-deficient cells (27H10 and 27G5) failed to silence firefly luciferase to a significant extent following firefly dsRNA treatment (Fig. 1.5). Considered together, these studies confirm that a single mouse Dicer enzyme is required for both miRNA and dsRNA processing pathways.

Figure 1.5

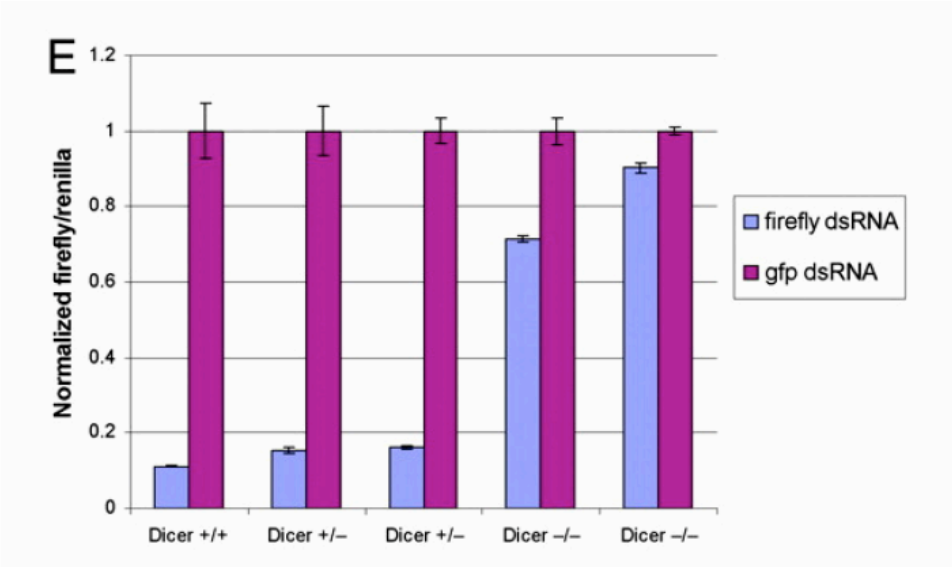

Figure 1.5. Dicer-null ES cells fail to initiate gene silencing in response to dsRNA. dsRNAs, 500 nt in length, corresponding to either firefly luciferase or GFP were introduced into ES cells along with firefly and *Renilla* luciferase reporters.

Responses to small interfering RNAs (siRNAs)

Synthetic siRNAs mimic Dicer cleavage products and presumably act downstream of Dicer. Therefore, there is not an a priori expectation that these RNAi triggers would require Dicer for their action. However, previous studies suggested that Dicer is required for siRNAs to trigger silencing in mammalian cells (Doi et al 2003). A potential mechanistic explanation for this observation came from studies of a Drosophila Dicer, Dcr-2. These showed that Dcr-2 is required along with R2D2 as part of the RISC loading complex (RLC) that assembles siRNAs with Argonaute proteins into an active RISC (Liu et al 2003a; Tomari et al 2004). The RLC in Drosophila is central to the recognition of asymmetry in siRNAs with differential binding of Dicer and R2D2 to specific termini of the siRNA, depending upon which end has greater single-stranded character, determining the strand that becomes the guide strand in RISC (Tomari et al 2004). Given the availability of genetic Dicer mutants in mouse cells, we asked whether siRNAs are functional in the absence of Dicer in mouse ES cells and whether siRNA asymmetry is accurately interpreted.

A highly asymmetric siRNA was previously shown in Drosophila lysates to target the antisense strand but not to efficiently target the sense strand of sod-1 (P. Zamore, personal communication). We transfected a mixture of this siRNA and a second siRNA that does not target our reporter into cells, varying the concentration of the relevant siRNA but keeping the overall siRNA concentration

constant. Along with the siRNAs, we co-transfected two reporters. One was a Renilla luciferase plasmid as an internal control. The second was plasmid encoding a Firefly luciferase, which had been fused to the antisense strand of sod-1. When the dual luciferase activity was analyzed 48 hours post-transfection, there was a clear concentration-dependent reduction in luciferase activity from the antisense-sod-1 reporter in response to the sod-1 siRNA (Fig. 1.6a). Interestingly, the effect was comparable in wild-type ES cells (AB2.2), Dicer$^{flox/null}$ cells (19F9, 27D4) or cell that are null for Dicer (27H10, 27G5) (Fig. 1.5A). These studies suggested that siRNAs can indeed enter RISC and function in the absence of an active Dicer protein in mammals. Although it as been reported elsewhere that Dicer-deficient mammalian

Figure 1.6

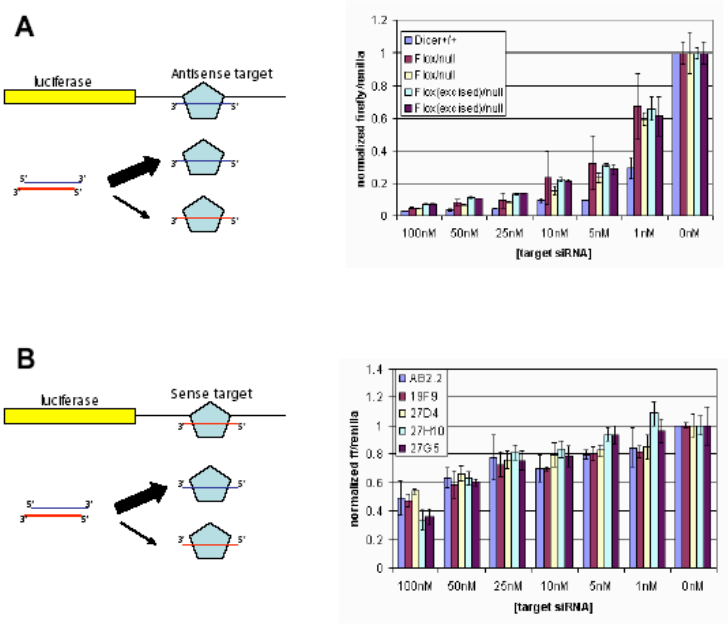

Figure 1.6. Dicer is dispensable for siRNA function and asymmetry detection. ES cells of various Dicer genotypes (AB2.2, Dicer+/+; 19F9, Dicer+/-; 27D4, Dicer+/-; 27H10, Dicer-/-; 27G5, Dicer-/-) were transfected with an identical asymmetric siRNA duplex at different concentrations. Two reporter constructs, encoding Firefly or Renilla luciferase were cotransfected, together with an siRNA targeting GFP that served to maintain a constant concentration of total siRNA. The Firefly constructs contained a perfectly complementary site to either the siRNA (blue) **(A)** or siRNA* (red) **(B)** strand of the duplex. SiRNA efficiency was quantified by dual luciferase assays 48 hours post transfection.

cells are competent for siRNA mediated silencing (Kanellopoulou et al 2005), our data extend these findings by showing that lack of Dicer does not change the IC_{50} for inhibition by a small RNA.

To test whether asymmetric assembly of siRNAs into RISC can proceed in the absence of Dicer we examined the activity of the asymmetric siRNA against a reporter that contained the opposite (sense) strand of sod-1. If asymmetry is lost, the prediction is that this luciferase reporter would be targeted as efficiently as the sod-1 antisense reporter. If asymmetry is maintained the relative efficiency with which the asymmetric siRNA targets the sense and antisense sod-1 reporters would be predicted to substantially differ. We observed that the asymmetric siRNA was able to reduce expression of the sense strand reporter by only about 50 percent at the highest concentration used, a concentration roughly 40-50 fold greater than the amount of siRNA required to reduce the antisense sod-1 reporter to similar levels. The potency of silencing was not significantly different between cells that were wild-type, heterozygous or null for Dicer (Fig. 1.6b).

These observations indicate that Dicer deficient ES cells retain their ability to incorporate siRNAs into RISC and to detect siRNA asymmetry. Thus, RISC loading in mammals must differ in some way from the process that has been studied in Drosophila. In mice, recognition of asymmetry may depend more heavily on as yet unidentified ortholog of Drosophila R2D2 with the role played by Dicer in this process being either diminished or completely eliminated. Of course we cannot eliminate the possibility that small amounts of a truncated, catalytically inert, Dicer protein are produced in our Dicer-null cells that might be selectively able to participate in RISC assembly.

Accumulation of centromeric transcripts in Dicer deficient ES cells

In a number of organisms, RNAi has been implicated in the regulation of heterochromatin, especially at centromeres (Fukagawa et al 2004; Mochizuki et al 2002; Mochizuki & Gorovsky 2005; Pal-Bhadra et al 2004; Provost et al 2002; Volpe et al 2003; Volpe et al 2002). This was first shown in *S. pombe*, where substantial genetic and biochemical evidence has demonstrated a role for the RNAi machinery in maintaining the heterochromatic nature and function of centromeres. It has been suggested that transcripts derived from heterochromatic repeats are recognized and processed by Dicer, producing siRNAs that enter complexes involved in nucleating or maintaining centromeric heterochromatin (Reinhart & Bartel 2002; Verdel et al 2004; Volpe et al 2002). Mouse centromeres are composed of highly repetitive heterochromatic loci, divided into two classes of repeats known as the major satellite and the minor satellite. In accord with the situation previously characterized in *S. pombe*, transcripts derived from both strands of the satellite repeats have been detected in mouse cells and in mouse embryos (Cobb et al 2005; Kanellopoulou et al 2005; Lehnertz et al 2003; Martens et al 2005; Rudert et al 1995). Therefore, these studies are consistent with a possible role for Dicer in maintaining the integrity of mammalian centromeres

Figure 1.7

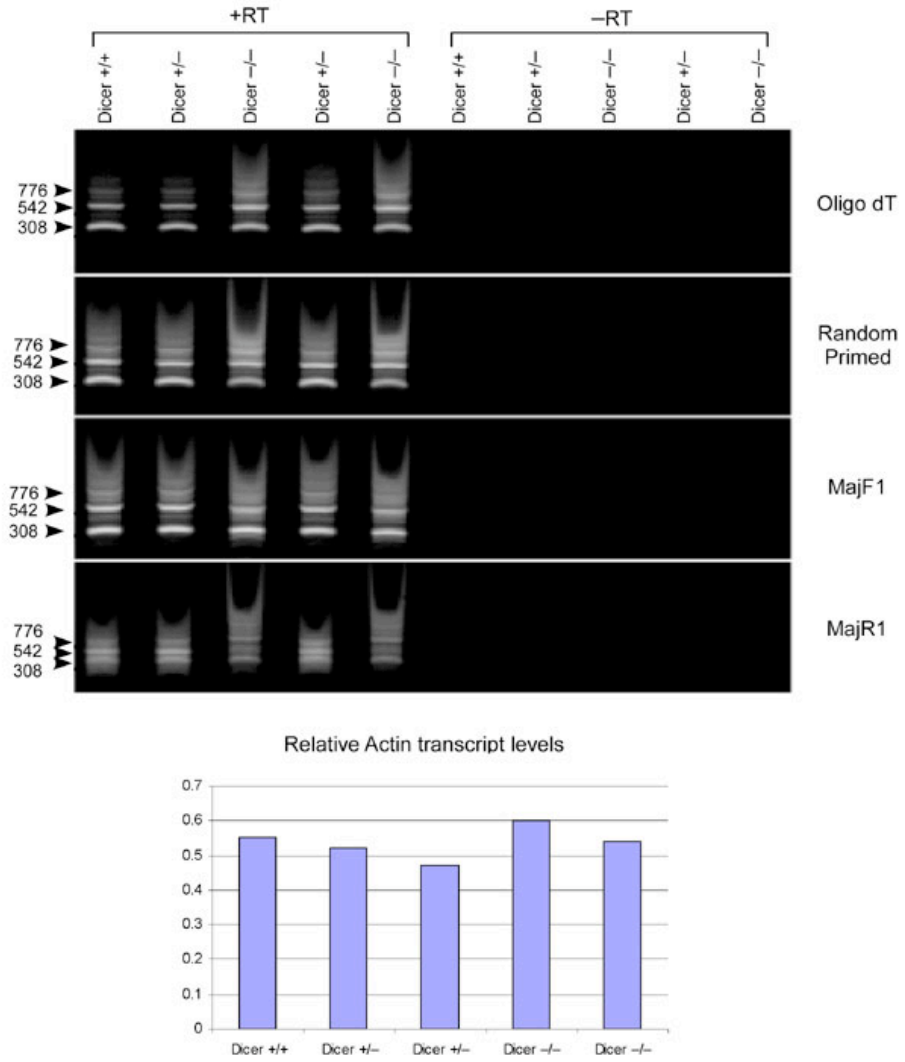

Figure 1.7. Accumulation of major satellite transcripts in Dicer-/- ES cells. RT-PCR analyses were performed on oligo(dT)-primed, random primed, or forward or reverse major satellite primed cDNAs (MajF1 and MajR1) prepared from ES cell total RNA of the indicated genotypes. –RT samples were treated identically, except that no reverse transcriptase was added for cDNA syntheses. The relative amount of total cDNA was quantified by measuring actin transcript levels in random primed cDNA samples by real-time PCR.

To investigate the role of Dicer at mammalian centromeres, we used RT-PCR to quantify transcript accumulation from the major satellite repeats. Using actin as a standardization control, we found an accumulation of major satellite transcripts in cDNA primed with random primers, oligo-dT, and major satellite specific primers (Fig. 1.7). This effect was quantified, and Dicer-deficient cells (27H10, 27G5) showed a consistently greater than two-fold increase in major satellite transcript levels, as compared to heterozygous (27D4) and wild-type cells (AB2.2) (Fig. 1.8a).

There are several possible interpretations of these findings. In one model, loss of Dicer would lead to deconstruction of centromeric heterochromatin, similar to the phenotype seen in *S. pombe*. The status of heterochromatin at centromeres can be monitored in part by examining patterns of cytosine methylation, one important marker of facultative heterochromatin in mammals. We used methylation sensitive restriction enzymes to assay the methylation status of the major and minor satellite repeats by Southern blotting in cells that contain or lack an active Dicer enzyme. Genomic DNA from cells of various genotypes was cut with either MspI (methylation insensitive) or HpaII (methylation sensitive) and hybridized with major satellite or minor satellite probes. A mitochondrial DNA probe served as a loading control since mitochondrial DNA is unmethylated. We did not observe any significant change in the methylation status of cells irrespective of their Dicer genotype (Fig. 1.8b). To determine whether histone methylation is altered at the centromeres in Dicer-null cells, we used chromatin immunoprecipitation assays with antibodies specific to histone H3 trimethyl lysine 9, a histone modification characteristic of pericentric heterochromatin (Martens et al 2005). We found no change in the enrichment of this modification at the major satellite repeats in Dicer-null cells, although there was a significant decrease in the enrichment of this modification at the pericentric region in Suv39h1 and h2 double null mouse embryonic fibroblasts (Fig. 1.8c) (Garcia-Cao et al 2004; Peters et al 2001). These results raise another possibility, namely that heterochromatin remains intact at centromeres, at least as reflected by DNA methylation status and that Dicer and the RNAi machinery are normally involved in degrading transcripts that are generated from these repeated loci. These studies do not rule out the possibility that Dicer might be involved in initiating DNA methylation at centromeric regions, for example during neocentromere formation, or that changes in Dicer might affect chromatin structure by altering histone methylation patterns.

Figure 1.8

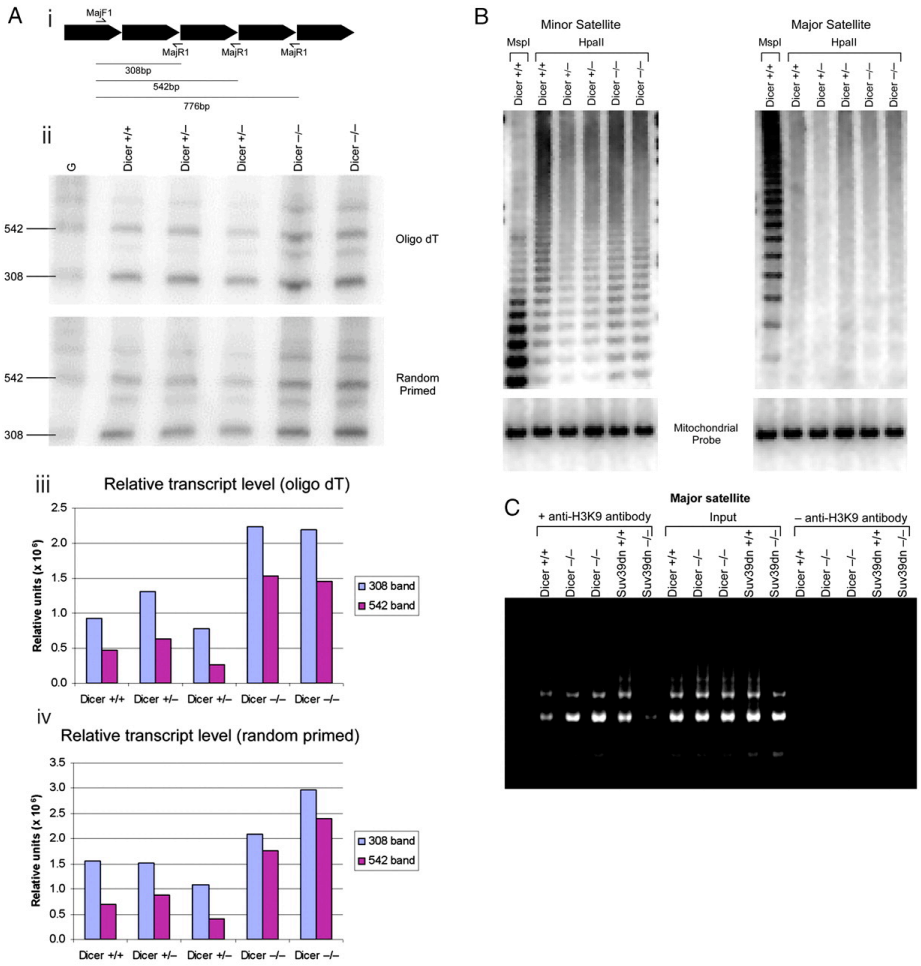

Figure 1.8. Changes in centromeric RNA levels in Dicer-null cells but no change in histone modification status or cytosine methylation in Dicer-null cells. (A) RT-PCR was used to quantify centromere-derived transcripts. **(B)** Southern blotting was used to determine the methylation status of the major and minor satellite repeats. **(C)** Antibodies directed against trimethyl histone H3 lysine 9 were used for chromatin immunoprecipitations at the major satellite repeats. Immunoprecipitate (with or without antibody) and input were amplified with primers specific to major satellite repeats with 20 cycles. Along with Dicer wild-type, heterozygous, and null are shown Suv39h1 and h2 double null mouse embryonic fibroblasts together with wild-type mouse embryonic fibroblast controls.

Discussion

We have constructed conditional Dicer mouse ES cell lines that can be induced by Cre recombinase to excise the catalytic RNase III-encoding domains of Dicer. This system has allowed us to examine the acute effects of Dicer loss on ES cells in culture, as well as to establish Dicer-deficient cell lines.

Studies of Dicer-deficient mice and zebrafish embryos have revealed that Dicer plays critical roles during development (Bernstein et al 2003; Giraldez et al 2005; Wienholds et al 2003; Yang et al 2005). However in part owing to the extremely early lethality in Dicer-deficient organisms, the causes underlying the phenotypes have remained unclear. Using our floxed Dicer ES cell line, we have shown that loss of Dicer in ES cells leads to a pronounced proliferation defect, an effect that can be eventually overcome, perhaps by accumulation of secondary mutations. This phenotype could result of an absence of one or more mature miRNAs, perhaps one required to repress a cell cycle inhibitor that would otherwise be expressed in ES cells. Alternatively, a more global effect on genome organization and structure could trigger checkpoint responses that arrest proliferation.

We show here that removal of Dicer from ES cells results in accumulation of transcripts derived from the major satellite of the mouse centromere. Interestingly, the mouse centromeric satellite repeats are transcriptionally active, with transcripts derived from both strands detectable in normal ES cells. Additionally the transcripts seem to be dynamically regulated during mouse development (Cobb et al 2005; Rudert et al 1995). Transcription across the satellite repeats seems to be controlled, at least in part, by histone modification status, as mouse ES cells deficient in the histone methyltransferase Su(var)3-9 have a slight increase in the accumulation of transcripts derived from the major and minor satellites (Lehnertz et al 2003). Dicer deficient ES cells have a clearly increased level of major satellite transcripts but we cannot distinguish between the possibilities that this is due to increased transcriptional linked to alterations in heterochromatin structure, to transcript stabilization, or perhaps even to a secondary effect of loss of Dicer on the expression of a gene that would then effect expression from centromeres. Interestingly, we detect no effect of Dicer loss on cytosine methylation or histone H3 lysine 9 trimethylation status at the centromeres. Thus, our results suggest that Dicer is dispensable for the maintenance of pericentric heterochromatin in mouse.

A recent study has shown that Dicer is required by mouse ES cells for differentiation *in vitro* and that loss of Dicer can lead to transcriptional derepression of major and minor satellite repeats accompanied by loss of cytosine methylation at these regions (Kanellopoulou et al 2005). In our current study, however, we were unable to detect DNA methylation defects in ES cells deficient in Dicer. There are several explanations for these apparently contradictory results. It is possible that small amounts of catalytically inert, truncated Dicer protein are produced from either or both the alleles described here or that described by Kanellopoulou et al and that this is sufficient to retain or perhaps inhibit cytosine methylation at the centromeres. Alternatively, perhaps

selective pressure accompanied by prolonged cell culture has led to two distinct outcomes with respect to DNA methylation at the major and minor satellite repeats. Interestingly, loss of Dicer in developing mouse thymocytes leads to no apparent loss of constitutive heterochromatin, suggesting that there may be cell-type-specific regulation of heterochromatin.

Dicer has previously been implicated in *Drosophila* as an integral component of the RLC, a multiprotein machine required for discrimination of the guide strand of the siRNA and its entry into RISC (Liu et al 2003a; Pham et al 2004; Tomari et al 2004). The RLC has not been characterized in mammalian cells, and it is unclear whether mouse Dicer plays an analogous role to Dcr-2 in Drosophila. Our finding that Dicer deficient mouse cells are able to load siRNAs into RISC and to detect siRNA asymmetry, suggests that catalytically competent Dicer is not required in the mouse RLC. In mice, loading of siRNAs may depend more heavily on an as yet unidentified ortholog of *Drosophila* R2D2, with the role played by Dicer in this process being either diminished or completely eliminated. Of course, we cannot rule out the possibility that small amounts of a truncated, catalytically inert Dicer protein are produced in our Dicer-null cells that might be selectively able to participate in RISC assembly.

The conditional Dicer flox/null ES cell that we have generated will be useful in dissecting in a mammalian setting the many roles of Dicer, and by implication the RNAi machinery, in gene regulation, genomic organization and genome defense.

Acknowledgements

We thank members of the G.J.H. laboratory for helpful discussions. C. Marsden, P. Moody, and J. Morris provided critical technical assistance. We are grateful to W. Filipowicz, C. Kanellopoulou (Dana Farber Cancer Institute), D. Livingston, D. Lee, D. Threadgill, T. Jenuwein (Research Institute of Molecular Pathology, Vienna), and T. Bestor for their generous gifts of materials. E.P.M. is an Elisabeth Sloan Livingston Fellow and is supported by a fellowship from the Department of Defense Breast Cancer Research Program (X81XWH-05-1-0256). This work was supported by National Cancer Institute Cancer Center Support Grant Developmental Funds 2 P30 CA 21765-25, -26 (to J.F.P.) and the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital. G.J.H. is supported by funds from the Department of Defense Breast Cancer Research program and the National Institutes of Health.

Appendix2: Pseudogene-derived siRNAs regulate gene expression in mouse oocytes

Oliver H. Tam[1]*, Alexei A. Aravin[1]*, Paula Stein[2], Angelique Girard[1], Elizabeth P. Murchison[1], **Sihem Cheloufi[1]**, Emily Hodges[1], Martin Anger[2%], Ravi Sachidanandam[1], Richard M. Schultz[2], and Gregory J. Hannon[1#]

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

[2] Department of Biology
University of Pennsylvania
433 S. University Avenue
205 Lynch Laboratories
Philadelphia, PA 19104-6018

[%]Present Address:
University of Oxford
Dept. of Biochemistry
South Parks Road
Oxford, OX1 3QU, UK

*These authors contributed equally to this work

To whom correspondence should be addressed: hannon@cshl.edu

Pseudogenes populate the mammalian genome as remnants of artifactual incorporation of coding mRNAs into transposon pathways [1]. Here, we show that a subset of pseudogenes generates endogenous small interfering RNAs (endo-siRNAs) in mouse oocytes. These endo-siRNAs are often processed from double-stranded RNAs formed by hybridization of spliced transcripts from protein coding genes to antisense transcripts from homologous pseudogenes. An inverted repeat pseudogene can also generate abundant small RNAs directly. A second class of endo-siRNAs may enforce repression of mobile genetic elements, acting in concert with piwi-interacting RNAs (piRNAs). Loss of Dicer increases expression of endo-siRNA targets, demonstrating their regulatory activity. Our findings provide a function for pseudogenes in regulating gene expression via the RNAi pathway and may, in part, explain the evolutionary pressure to conserve Argonaute-mediated catalysis in mammals.

Small RNA-directed gene silencing pathways have been adapted to accept numerous inputs and to impact many types of downstream targets. In few places is this more apparent than in animal germ lines where two classes of small RNAs, miRNAs and piRNAs, with distinct biogenesis mechanisms and biological functions have been reported. Although miRNAs, as a group, are ubiquitously expressed, piRNAs have thus far been found only in germ cells and in a few gonadal somatic cells types[2]. piRNAs repress the activity of mobile genetic elements, forming a small RNA-based, innate immune system with both genetically encoded and adaptive components [2-9].

In mice, a homozygous mutation in any single Piwi family member causes male sterility accompanied by gonadal hypotrophy [5, 10, 11]. In Mili and Miwi2 mutants, meiosis is not completed and germ cells are progressively lost [5]. This correlates with an activation of transposons, particularly the non-LTR retrotransposon, L1 [5, 12]. DNA methylation of L1 elements is correspondingly lost. In contrast, females bearing homozygous mutations in individual Piwi genes are apparently normal and fertile [5, 10, 11]. Since female germ cells must also control transposons, we sought to characterize their small RNA profiles to determine whether a piRNA system, similar to that operating in spermatocytes, also exists in oocytes.

Approximately 6,000 fully grown oocytes, arrested in prophase of meiosis I, were collected. Small RNA fractions from 19-24 nucleotides (lower fraction) and 24-30 nucleotides (upper fraction) were gel purified and used to prepare small RNA libraries. These were deeply sequenced [3, 4]. A total of 1,037,355 sequences were obtained that could be mapped to the mouse genome (753,981 from the lower fraction and 283,374 from the upper fraction; Table S1) In the lower fraction, 126515 non-redundant sequences were obtained, falling into 24271 non-overlapping clusters. In the upper fraction, 97807 non-redundant sequences fell into 15032 non-overlapping clusters.

An examination of the small RNAs in the upper fraction of the oocyte library revealed a piRNA population that resembled those found in early stage spermatocytes [3] (Fig. 1a, left). Roughly 62% of small RNAs correspond to annotated repeats (Table S2), with 3% matching genic sequences and 3% matching un-annotated, intergenic sites. The function of the latter species remains mysterious. Roughly 30% of the library corresponded to presumed breakdown products of abundant, non-coding RNAs, such as rRNAs, tRNAs and snoRNAs.

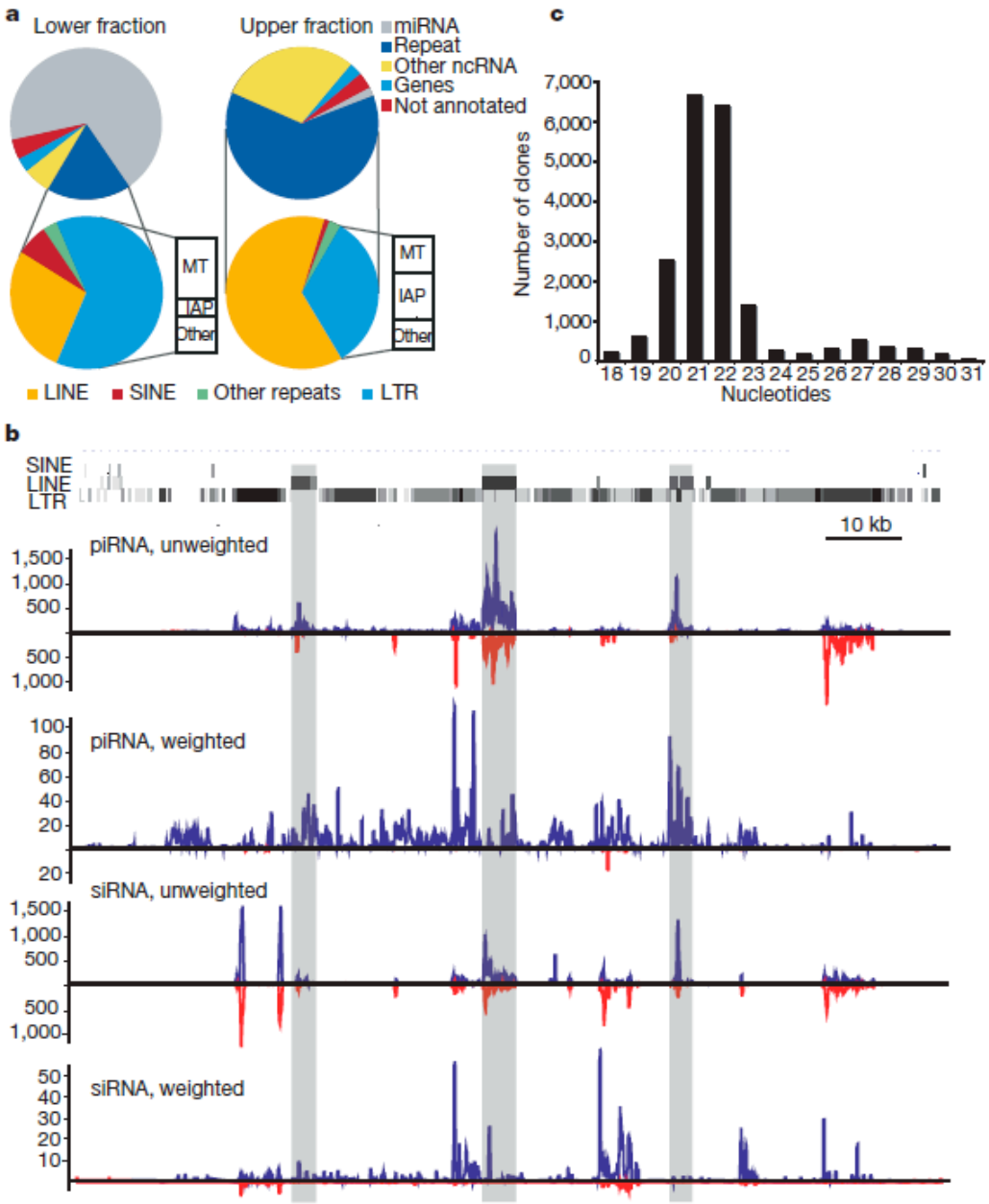

Figure 1. Both piRNA and siRNA systems control transposons in mouse oocytes. a. Small RNA libraries from 19-24 nt. (lower fraction, left) and 24-30 nt. RNAs (upper fraction, right) were deeply sequenced. Reads were assigned an annotation as previously described [12]. The fraction of reads in each category is depicted. The repeat-annotated small RNAs were designated as LINE, SINE, LTR and other, with the LTR category further divided between MT, IAP and other. **b.** A representative piRNA locus on chromosome 10 is shown with the content of LINE, SINE and LTR fragments. Shading indicates the degree of match to the consensus element. Frequency plots for piRNAs and siRNAs from this locus are shown below. "Unweighted" plots each match to the cluster. "Weighted" normalizes each match, dividing by its genomic frequency [4]. Blue and red lines indicate small RNAs mapping to the upper and lower genomic strand respectively. **c.** RNAs from 19-30 nucleotides were deeply sequenced. Reads matching the piRNA cluster shown in **b** were used to construct a frequency plot by length.

As expected, oocyte piRNAs arise from discrete genomic loci in a strand asymmetric fashion (Table S3). A number of these loci share structural similarities to *Drosophila* piRNA loci, which act as master controllers of mobile elements [3]. One example (Fig. 1b) spans ~120 kB of chromosome 10 and contains an abundance of LINE and LTR elements. These have an orientation bias that results in the generation of predominantly antisense piRNAs (Fig. 1b, piRNA, weighted, Fig. S1).

piRNAs have been proposed to act with transcripts from active transposons in a feed-forward amplification loop that confers signature features upon a piRNA population that is mounting an ongoing transposon defense [2-4, 6, 7]. Primary piRNA-directed cleavage of transposon mRNAs creates the 5' ends of secondary piRNAs [4, 6]. This produces primary and secondary piRNAs pairs that overlap by 10 nucleotides at their 5' ends. The 5' U bias of primary piRNAs thus leads to an enrichment of an A at position 10 of secondary piRNAs. These characteristics are prevalent in piRNA populations from mouse oocytes, particularly those that can be mapped to the L1 and IAP elements (Fig. S2).

As expected, annotated microRNAs comprised the majority (69%) of 19-24 nt. RNAs (Fig. 1a, right, Table S4). Amongst the highly abundant species are members of the let-7 family (let-7a/c/f), generally abundant miRNAs (miR-22, miR-16, miR-21, miR-26, miR-93, and miR29a/b), and microRNAs abundant in ovary and placenta (miR-322, miR-503, miR-451). Finally, we detect miRNAs specifically expressed in male and female gonad (miR-103) [13].

A substantial fraction 19-24 nt. RNAs matched annotated transposons (Table S2). Many that mapped uniquely to the genome could be traced to oocyte piRNA loci (Fig. 1b, siRNA, weighted). These species could represent piRNA degradation products, or oocyte piRNA clusters might generate both siRNAs and piRNAs.

Therefore, we independently mapped piRNAs and candidate siRNAs to consensus L1 and IAP sequences (Fig. S3). Each gave characteristic profiles. Moreover, piRNAs and candidate siRNAs show distinctly different nucleotide biases, with piRNAs displaying their characteristic enrichment for a 5' uridine residue and an A at position 10 (Fig. S2). Candidate siRNAs lack a 10A bias and show enrichment for both A and U residues at their 5' ends (Fig. S2). Finally, we gel purified 19-30 nucleotide RNAs from mouse oocytes as a single fraction and deeply sequenced this population. A length distribution of small RNAs that match the piRNA cluster shown in Figure 1b yields two distinct peaks (Fig. 1c). 21-22 nucleotide siRNAs apparently predominate over the piRNA population, which averages ~27 nucleotides. We conclude that transposon-rich loci in oocytes give rise to both siRNAs and piRNAs. Though siRNAs are apparently more abundant, piRNA cloning frequencies could be reduced by the 2'-O-methyl modification that occurs on their 3' termini [2]. Our results raise the possibility that piRNA and siRNA systems may act redundantly to repress transposons in mouse

oocytes, perhaps explaining the lack of substantial phenotypic consequences of individual Piwi mutations in females [5, 10, 11].

Although many transposons were targeted by both piRNAs and siRNAs, some relied more heavily on a particular pathway. For example, MTB and MTC matched almost exclusively to siRNAs. Moreover, the most prominent cluster that produces MTB/MTC small RNAs contains an inverted repeat with strong potential to produce a Dicer substrate. Interestingly, this transposon class showed increased expression in Dicer-null oocytes, consistent with its being regulated predominantly if not exclusively by the siRNA system [14].

Small RNA libraries often contain genic sequences. In other tissues, these correspond exclusively to sense sequences that likely represent contaminating degradation products. However, in oocytes, numerous sense and antisense siRNAs corresponding to protein coding genes could be identified (Table S5). As mammals lack any identifiable RNA-dependent RNA polymerase, this raised the question of how antisense siRNAs might be generated.

Based upon polymorphisms, uniquely mapping sense siRNAs could often be assigned to the functional protein-coding copy of a gene, while antisense siRNAs mapped to a homologous pseudogene (Table S5, Fig. 2a). Thus, oocyte endogenous (endo-) siRNAs might be processed from double-stranded RNAs that form by hybridization of transcripts derived from two unlinked loci. A similar process in which transcripts from active transposons hybridize to antisense transposon fragments transcribed from piRNA clusters could explain the genesis of transposon siRNAs.

siRNAs from gene-pseudogene pairs arise exclusively from regions of complementarity between the partners. Because many sense-oriented siRNAs cross exon-exon junctions, we hypothesize that mature, spliced mRNAs from genes interact with antisense pseudogene transcripts to form Dicer substrates (Fig. S4). In one case (Fig. 2b), both sense and antisense siRNAs to the GTPase activating protein for Ran (Ran-GAP) were produced from a pseudogene locus containing an ~300 bp inverted repeat with an intervening ~800 base loop. siRNAs were derived only from the potentially double-stranded portion of this locus.

In some cases, Dicer processing of dsRNA substrates proceeds in an apparently processive fashion from a discrete initiation site, producing “phased” small RNAs with a 22 nt periodicity[15]. Transposon-derived and genic siRNAs showed this property only very weakly (Fig. S2). Interestingly, piRNAs also show a similar, very weak phasing signal, though with a period of ~27 nt. rather than 21 nt.

Pseudogenes have often diverged substantially from their functional ancestors. Thus, we wished to examine the possibility that pseudogene-derived antisense siRNAs could regulate corresponding protein coding genes. We mapped antisense siRNAs to potentially relevant regulatory targets. Many small RNAs aligned to their targets either with no mismatches or with mismatches lying outside regions essential for Slicer cleavage [16, 17] (Fig. 2a,c). Thus, antisense, pseudogene-derived siRNAs might be capable of regulating homologous protein coding genes through a conventional RNAi mechanism.

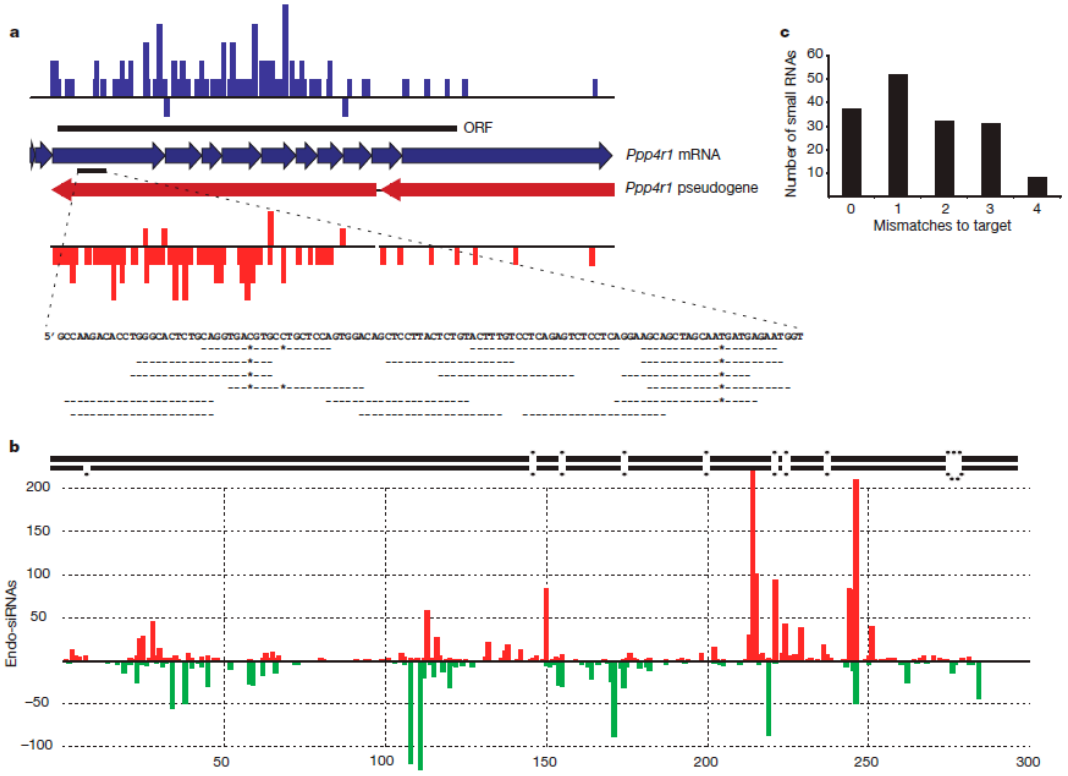

Figure 2. Gene-pseudogene interactions produce endogenous siRNAs. **a.** Endo-siRNAs unambiguously mapped to the functional Ppp4r1 mRNA are plotted in blue, above the mRNA (individual exons indicated as thick arrows). The extent of the ORF is indicated. Endo-siRNAs from the Ppp4r1 pseudogene are shown in red below the pseudogene. Arrows indicate two segments of Ppp4r1 homology. siRNAs plotted above each line are sense oriented, with respect to the functional mRNA, and those below are antisense. Shown below is an enlargement of one section of the mRNA with individual antisense siRNAs aligned. - = match, * = mismatch. **b.** Endogenous siRNAs homologous to Rangap are shown below a schematic of the genomic inverted repeat structure from which they arise (the ~800 base loop is not depicted). Those shown above and below the X-axis come from the upper or lower arm of the hairpin, respectively. No siRNAs were sequenced from the ~800 base loop that separates the two stem arms. **c.** siRNAs antisense to Ppp4r1 counted and plotted by the number of mismatches to the functional mRNA.

To test the regulatory potential of pseudogene-derived siRNAs, we assessed the effects of Dicer loss on their putative targets. We had previously shown that deletion of *Dicer* in growing oocytes caused the production of non-functional gametes with defects in spindle organization and chromosome segregation [14, 18]. We compared the expression of candidate endo-siRNA targets in wild-type and Dicer-null cells [14]. Many genes with abundant, pseudogene-derived siRNAs showed significant increases in expression upon Dicer loss (Fig. 3a). We verified candidates derived from the array data by semi-quantitative RT-PCR (Fig. 3b).

Collectively our data indicate that in mammalian oocytes, protein-coding mRNAs interact with pseudogene transcripts to form dsRNAs that are processed into endo-siRNAs. Examination of *Dicer* knockouts indicates a function for endo-siRNAs in gene regulation. At present, we cannot distinguish whether these small RNAs direct target cleavage or whether the act of siRNA production *per se*, which consumes the coding mRNA, is sufficient for repression. However, the specific case of HDAC1 may point to a RISC-based mechanism. Few uniquely mapping siRNAs are generated from the HDAC1 gene itself, suggesting that it is not used prominently as a Dicer substrate. Instead, most uniquely mapping sense and antisense siRNAs can be assigned to a series of HDAC1 pseudogenes. Based upon its increased expression in Dicer-null oocytes, we propose that pseudogene-derived, antisense siRNAs direct RISC to cleave HDAC1 mRNAs.

The catalytic potential of at least one Argonaute protein has been conserved through mammalian evolution from platypus to humans [19] (Murchison et al., in press). However, mammalian microRNAs, with one known exception, act through translational mechanisms without the need for mRNA cleavage [20]. The discovery of endogenous siRNAs in mammalian oocytes not only expands the realm of mammalian small RNA classes but also provides one possible explanation for the evolutionary pressure to conserve Argonaute enzymatic activity.

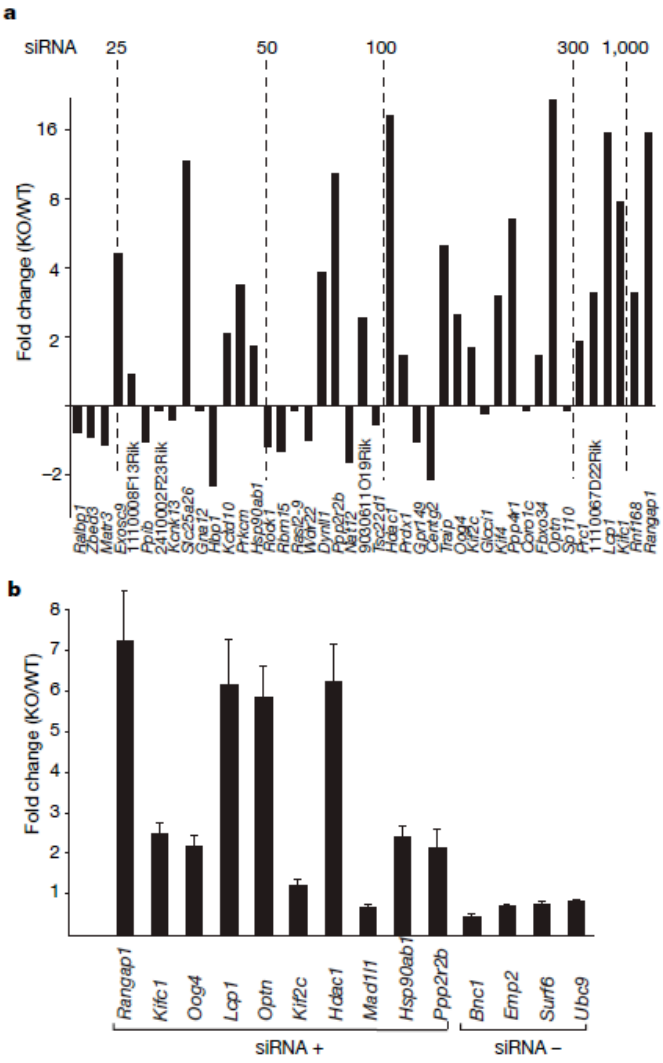

Figure 3. Endo-siRNAs play a role in gene regulation. a. We had previously compared expression levels in wild-type and Dicer-null oocytes by microarray [14]. Genes with large number of siRNAs were examined. For those with significant changes in expression ($p<0.1$), the fold change in Dicer-null versus wild-type was plotted. The graph was arranged according to the number of antisense siRNAs per gene in our dataset, increasing from left to right (with benchmarks shown). The identity of the gene represented by each bar is given below. **b.** Fourteen genes (with and without siRNAs in our dataset, indicated by siRNA+ and siRNA-), comprising a set in which some showed significant changes and some did not, were tested for changes in mRNA levels in Dicer-null versus wild-type oocytes by Q-PCR. Gene names are indicated below. Error bars representing s.e.m (2 replicates).

Pseudogenes have long been considered non-functional artifacts of transposition pathways acting on protein coding mRNAs. In a few cases, regulatory roles have been posited for pseudogenes, largely through antisense mechanisms [21-23]. Our findings provide a role for a subset of mammalian pseudogenes in the production of functional siRNAs. The production of dsRNAs by interaction between sense and antisense transcripts from distinct loci has not

been observed in other tissues and may require the unique environment of oocytes, which both substantially lack a PKR pathway and are geared for mRNA stabilization and storage [24 25, 26]. The fact that many targets of this pathway are related to microtubule dynamics (including microtubule-based processes, p=0; kinesin complex, p=0; motor activity, $p<1x10^{-254}$; spindle, $p<8x10^{-239}$; and microtubule associated complex, $p<3x10^{-60}$; Fig. S5) suggests that the regulatory circuits which we describe may have important biological roles, as the consequences of *Dicer* loss in growing oocytes is disruption of proper spindle formation and defects in chromosome segregation [14, 18].

Acknowledgements

We thank members of the Hannon laboratory for helpful discussions. OHT is a Bristol-Meyers Squibb fellow and AG is a Florence Gould Foundation Scholar of the Watson School of Biological Sciences. EPM is supported by a fellowship from the Australian-American Association. This work was supported in part by grants from the NIH (RMS and GJH) and kind gifts from Kathryn W. Davis and the Stanley family (GJH and EH). GJH is an Investigator of the Howard Hughes Medical Institute. Small RNA datasets can be accessed in GEO via the following accession numbers GSM261957, GSM261958, and GSM261959.

Methods Summary

Mouse oocytes were harvested from superovulated mice and used to prepare small RNA fractions. These were cloned and deeply sequenced as previously described [2, 4]. Bioinformatic analysis of the sequences was performed as described in the detailed online methods. For semi-quantitative RT-PCR, RNA was extracted from fully-grown oocytes from $Dicer^{flox/flox}$ and $Dicer^{flox/flox}$, Zp3-Cre mice. QPCR was performed using TaqMan probes.

Literature Cited

1. D'Errico, I., Gadaleta, G. & Saccone, C. Pseudogenes in metazoa: origin and features. Brief Funct Genomic Proteomic 3, 157-67 (2004).
2. Aravin, A., Hannon, G. & J, B. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. Science in press (2007).
3. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316, 744-7 (2007).
4. Brennecke, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128, 1089-103 (2007).
5. Carmell, M. A. et al. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell 12, 503-14 (2007).

6. Gunawardane, L. S. et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science 315, 1587-90 (2007).
7. Houwing, S. et al. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell 129, 69-82 (2007).
8. Saito, K. et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev 20, 2214-22 (2006).
9. Vagin, V. V. et al. A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313, 320-4 (2006).
10. Kuramochi-Miyagawa, S. et al. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development 131, 839-49 (2004).
11. Deng, W. & Lin, H. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Dev Cell 2, 819-30 (2002).
12. Aravin, A. et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442, 203-7 (2006).
13. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401-14 (2007).
14. Murchison, E. P. et al. Critical roles for Dicer in the female germline. Genes Dev 21, 682-93 (2007).
15. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121, 207-21 (2005).
16. Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev 18, 975-80 (2004).
17. Schwarz, D. S., Tomari, Y. & Zamore, P. D. The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr Biol 14, 787-91 (2004).
18. Tang, F. et al. Maternal microRNAs are essential for mouse zygotic development. Genes Dev 21, 644-8 (2007).
19. Joshua-Tor, L. The Argonautes. Cold Spring Harb Symp Quant Biol 71, 67-72 (2006).
20. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-97 (2004).
21. Korneev, S. A., Park, J. H. & O'Shea, M. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene. J Neurosci 19, 7711-20 (1999).
22. Weil, D., Power, M. A., Webb, G. C. & Li, C. L. Antisense transcription of a murine FGFR-3 psuedogene during fetal developement. Gene 187, 115-22 (1997).
23. Zhou, B. S., Beidler, D. R. & Cheng, Y. C. Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. Cancer Res 52, 4280-5 (1992).
24. Stein, P., Zeng, F., Pan, H. & Schultz, R. M. Absence of non-specific effects of RNA interference triggered by long double-stranded RNA in mouse oocytes. Dev Biol 286, 464-71 (2005).

25. Bettegowda, A. & Smith, G. W. Mechanisms of maternal mRNA regulation: implications for mammalian early embryonic development. Front Biosci 12, 3713-26 (2007).
26. Stitzel, M. L. & Seydoux, G. Regulation of the oocyte-to-zygote transition. Science 316, 407-8 (2007).
27. Schultz, R. M., Montgomery, R. R. & Belanoff, J. R. Regulation of mouse oocyte meiotic maturation: implication of a decrease in oocyte cAMP and protein dephosphorylation in commitment to resume meiosis. Dev Biol 97, 264-73 (1983).
28. Lee, J. S., Katari, G. & Sachidanandam, R. GObar: a gene ontology based analysis and visualization tool for gene sets. BMC Bioinformatics 6, 189 (2005).

Online Methods

Mouse strains

Either CF-1 or CD-1 wild-type mice of 4-6 weeks were purchased from Harlan or Charles River Lab, respectively, and used to obtain oocytes for small RNA isolation. The Dicer$^{flox/flox}$ and Dicer$^{flox/flox}$ Zp3-Cre mice, as previously reported [14], were used to obtain Dicer-deficient oocytes.

Generation of oocyte small RNA libraries

Wild-type mice were primed with 5 IU PMSG 48 hours prior to sacrifice, and fully-grown GV oocytes were collected as previously described [27]. Total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA) according the manufacturer's protocol, and small RNA cloning was performed as described [4].

Quantitative real-time PCR

Total RNA was extracted from fully-grown oocytes from Dicer$^{flox/flox}$ and Dicer $^{flox/flox}$, Zp3-Cre mice using the Absolutely RNA Microprep Kit (Stratagene, La Jolla, CA). cDNA was prepared by reverse transcription of total RNA with Superscript II and random hexamer primers. One oocyte equivalent of the resulting cDNA was amplified using TaqMan probes and the ABI Prism Sequence Detection System 7000 (Applied Biosystems, Foster City, CA). Three replicates of 45 oocytes each were used for RNA isolation and two replicates were run for each real-time PCR reaction; a minus template served as control. Quantification was normalized to the endogenous upstream binding factor (*Ubf*) within the log-linear phase of the amplification curve obtained for each probe/primer using the comparative C_T method (ABI PRISM 7700 Sequence Detection System, User Bulletin #2, Applied Biosystems, 1997). The TaqMan gene expression assays used were: Mm00441071_m1 (Rangap1), Mm00835842_g1 (Kifc1), Mm00620601_m1 (Oog4), Mm00786153_s1 (Lcp1), Mm00728630_s1 (Kif2c), Mm02391771_g1 (Hdac1), Mm00487521_m1 (Mad1l1), Mm00725286_m1 (Optn), Mm00833431_g1 (Hsp90ab1), Mm00511698_m1 (Ppp2r2b), Mm00801709_m1 (Emp2), Mm00486494_m1 (Surf6), Mm00456972_m1 (*Ubf*). For Bcn1 and Ubc9, custom TaqMan Gene Expression Assays were used that had the following primers and probes: Bcn1 forward primer 5′ ACTGGACGCTTCAGGATTACATC3′, Bcn1 reverse primer 5′ GTCATGATGCTCCAGTGATCCA3′, Bcn1 probe 5′FAM TTCCCAGAGGCATCCTG3′; Ubc9 forward primer 5′ CAGGTGAGAGCCAAGGACAAA, Ubc9 reverse primer 5' GGCCCACTGTACAGCTAACA, Ubc9 probe 5'FAM CTGGCCTGCATTGATC.

Bioinformatic analysis

Small RNAs were sequenced using the Illumina 1G platform. Sequencing of the upper and lower fraction libraries produced 2785080 reads, of which 1037355 (37%) could be mapped to the mouse genome (release mm9, Jul 2007) with no mismatches. The small RNAs are matched to a suffix array generated from the

mouse genome, keeping track of exact matches to genome. Repeat masking was not used, but small RNA sequences with more that 10 identical nucleotides in a row were removed from consideration. Annotation categories were assigned based on the annotation of corresponding genomic sequences extracted from the UCSC genome browser. The genome was annotated with mRNAs, ncRNAs and repeats. The annotations at the mapping positions (up to 5 mappings per small RNA) were used, along with a majority rule, to assign an annotation to each small RNA. The sequences were also re-analyzed to allow 1-2 mismatches to the genome. While the number of mapped sequences increased (from 37% to 54%), the genomic origin of repeat-associated small RNAs became ambiguous (Figure S1), and therefore non-informative.

To extract small RNA clusters (both piRNA and siRNA), the genome was scanned to look for regions that had more than 10 uniquely mapping small RNAs, and the boundaries were defined as the location of the first/last small RNA in the cluster. To identify sequences that match the consensuses for transposable elements, the small RNAs were aligned to consensus sequences from release 11.08 of Repbase (www.girinst.org). The following consensuses were used: L1_MM LINE L1 and IAPLTR1a_I_MM for the IAP retrotransposon. Matches to consensus sequences with up to three mismatches were recovered and included in the analysis. Nucleotide biases were calculated for small RNAs matching L1 and IAP consensuses as described [4]. To identify gene/pseudogene pairs, the genomic sequences of the siRNA clusters were extracted from the UCSC genome browser, and re-matched to the genome using BLAT (genome.ucsc.edu). Genomic regions with greater than 95% identity were identified, and small RNAs (both sense and antisense) mapped to these locations were extracted.

Gene ontogeny analysis of endo-siRNA targets was carried out as previously described[28] using GOBAR, which uses a hypergeometric statistic to identify nodes that are significantly enriched. A bootstrapping technique, involving repeated sampling from the reference set, is used to assign significance values to the results.

31. Schultz, R. M., Montgomery, R. R. & Belanoff, J. R. Regulation of mouse oocyte meiotic maturation: implication of a decrease in oocyte cAMP and protein dephosphorylation in commitment to resume meiosis. Dev Biol 97, 264-73 (1983).
32. Lee, J. S., Katari, G. & Sachidanandam, R. GObar: a gene ontology based analysis and visualization tool for gene sets. BMC Bioinformatics 6, 189 (2005).

Appendix3: Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases

Fedor V. Karginov[1], **Sihem Cheloufi[1,2]**, Alexander Stark[3], Andrew D. Smith[4], and Gregory J. Hannon[1]*

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

[2]Graduate Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[3]Research Institute of Molecular Pathology (IMP) Dr. Bohr-Gasse 7, A-1030 Vienna, Austria.

[4]Molecular and Computational Biology
University of Southern California
Los Angeles, CA, USA

*To whom correspondence should be addressed. Email: hannon@cshl.edu

Abstract

The functional lifespan of a mammalian miRNA is determined by a variety of mechanisms. Among these, animal microRNAs in complex with Argonaute proteins bind to many mRNA targets, generally with imperfect complementarity, leading to destruction of the mRNA through the normal cellular decay machinery. The ancestral “slicer” endonuclease activity of Argonaute2, which requires more extensive complementarity with the target RNA, is not used in this pathway. Only two examples of microRNA- guided slicing of mRNA targets have been reported, neither of which can fully account for the deep conservation of Ago2 catalytic activity or the requirement for an intact Ago2 nuclease domain for mouse viability. Here, we assess the endonucleolytic function of Ago2 and other nucleases by identifying cleavage products retaining 5`-phosphate groups in mouse ES cells on a transcriptome-wide scale. We detect a significant signature of Ago2-dependent cleavage events and validate several targets. Unexpectedly, a broader class of Ago2-independent cleavage sites is also observed, indicating participation of additional nucleases in site-specific mRNA cleavage. Within this class, we identify a cohort of Drosha- dependent mRNA cleavage events that functionally regulate mRNA levels in mES cells, including one in the Dgcr8 mRNA. Together, these results highlight the underappreciated role of endonucleolytic cleavagein controlling mRNA fates in mammals.

Introduction

The last two decades have provided a deep appreciation for the numerous roles of small RNAs in eukaryotic cells. Perhaps the best characterized species are the subset that form components of RNAi-related pathways.These 21-30 nucleotide RNAs join Argonaute proteins and guide them to their targets via complementary base pairing.Once bound, effector complexes can elicit a variety of outcomes, with one of the most conserved being target RNA cleavage catalyzed by the Argonaute RNAse H-related nuclease domain.

Among the classes of Argonaute-associatedRNAs,microRNAs form a conserved regulatory axis, which exerts post-transcriptional control over a broad set of cellular mRNAs (Carthew and Sontheimer, 2009). microRNAs are derived from partly duplexed precursors, called pri-miRNAs, via processing by RNAseIII family enzymes, Drosha and Dicer in animals or DCL1 in plants (Voinnet, 2009). Though superficially similar, plant and animal microRNAs differ not only in their mechanisms of biogenesis but also in the modes by which they regulate gene expression.In plants, extensive miRNA-mRNA pairing most often leads to cleavage of target mRNAs (Voinnet, 2009).In animals, target recognition is dominated by a short sequence comprising bases ~2-8 of the small RNA, termed the seed.The remainder of the small RNA pairs imperfectly, if at all, with its regulatory target. In these cases, a miRNA-target interaction commonly results in degradation of the mRNA by general cellular mRNA decay pathways rather than by Argonaute cleavage. Nevertheless, at least one mRNA, HOXB8, can be demonstrably cleaved in response to interaction with a highly complementary microRNA, miR-196 (Yekta et al., 2004).Additionally, cleavage products corresponding to miRNA target sites in the imprinted Rtl1/Peg11 locus have been observed (Davis et al., 2005).

The four mammalian Ago homologs evolved from a common ancestor early in the vertebrate lineage (unpublished observations).A significant expansion of novelmiRNAfamiliescoincidedwiththisAgo diversification,potentially explaining the emergent tissue complexity in vertebrates (Heimberg et al., 2008). Given the predominantly non-catalytic mode of miRNA action in mammals and the largely indiscriminate association of miRNAs with the Ago homologs (Azuma- Mukai et al., 2008; Ender et al., 2008), it is surprising that one of the Argonautes, Ago2, conserved its catalytic function throughout this expansion (Liu et al., 2004). Ago2 is the only mammalian Ago protein required for viability (Liu et al., 2004) (S. Cheloufi and GJH, unpublished), and the catalytic activity of Ago2 contributes to its essential role (Cheloufi et al., submitted). Ago2 catalysis is required for the unusual biogenesis of an erythroid microRNA, miR-451, but the degree of anemia caused by lack of this regulator probably does not fully explain the perinatal death of animals harboring only catalytically inactive Ago2 alleles

(Cheloufi et al., submitted). Thus, there is a strong likelihood that cleavage by Ago2 has additional roles both during late embryogenesis and in adult animals.
The ends of cellular mRNAs are protected by a cap structure, and the major non-specific ribonucleases generally leave 5' OH groups.Thus, 5' phosphorylated RNA species are thought to be enriched in products of specific cellular RNA processing events.This property has been exploited to identify the mRNAs subject to microRNA-directed cleavage in plants by transcriptome-wide analysis of mRNA fragments bearing 5' monophosphate termini (Addo-Quaye et al., 2008; German et al., 2008). Those studies confirmed many known microRNA targets and led to the identification of novel cleavage products of microRNA- and tasiRNA-primed RISC. The data also implied the existence of cleavage sites that could not be explained by known small RNAs.

Endonucleolytic cleavage likely plays a broader role in mRNA regulation than is currently appreciated, and evidence to this effect is beginning to emerge. In addition to its 3`-5` exonuclease function, the exosome component Dis3/Rrp44 possesses a functionally relevant endonuclease activity via its PIN domain (Lebreton et al., 2008). Similarly, the PIN domain of SMG6 can cleave mRNAs near premature termination codons during NMD (Eberle et al., 2009; Huntzinger et al., 2008). Site-specific cleavage events in particular mRNAs have also been observed.Forexample,Droshacanregulatetheexpressionofits DGCR8/Pasha cofactor by cleaving a site within its 5' UTR that resembles a pri- miRNA, the canonical Drosha substrate (Han et al., 2009b). The stress-induced IRE1α enzyme promotes decay of a number of ER-localized mRNAs via endonucleolytic cleavage (Han et al., 2009a).Additional examples include cleavage of c-myc by G3BP and APE1 (Barnes et al., 2009; Tourriere et al.,2001) and cleavage of α-globin mRNA by an erythroid-enriched endonuclease (Wang and Kiledjian, 2000).

In this study, we examine transcriptome-wide mRNA cleavage patterns resulting in 5`-phosphorylated fragments in mammals.We identify a class of miRNA-guided, Ago2-dependent mRNA cleavage events that may contribute to the conservation of Ago2 catalytic potential. Surprisingly, we also discover a cohort of Drosha-dependent cleavage sites that a broader extent of mRNA processing by this enzyme than was previously appreciated. We also noted a large class of evolutionarily conserved mRNA cleavage sites that depended neither on Ago2 nor Drosha. These highlight the participation of additional, yet to be described, nucleases in this mode of mRNA metabolism.

Results

Combined bioinformatic and experimental approaches reveal a set of miRNA cleavage targets

To define a set of potential targets of miRNA-guided cleavage, we performed a computational search of the transcriptome for extensive complementarity

between miRNAs and mRNA transcripts. Based on known criteria for cleavage-competent pairing (Elbashir et al., 2001; Haley and Zamore, 2004; Martinez and Tuschl, 2004) we considered sites with perfect pairing at miRNA positions 9, 10, and 11, and ranked the sites by the number of non-GU mismatches and total mismatches. The resulting lists for human and mouse datasets showed a large number of targets with perfect or near-perfect complementarity (Supplementary File 1). Cleavage of two predicted mir-151-5p targets, ATPAF1 and LYPD3, was confirmed by gene-specific 5` RACE (Yekta et al., 2004) of 293S cellular RNA. Both reactions produced correctly sized amplicons, with 5/8 and 8/8 clones reflecting the predicted cleavage site upon sequencing (Figures 1A and 1B). The predicted mir-151-5p target sites in the 3` UTRs of these genes showed significant conservation, suggesting a functionally relevant interaction (Figures 1C and 1D).Interestingly, mir-151-5p is derived from a LINE2 element (Smalheiser and Torvik, 2005) and complementarity to this microRNA can be found in a number of genes harboring LINE2 insertions, including the above transcripts. In these cases, however, the region of high conservation is confined to the miR-151-5p complementarity and does not extend throughout the LINE2 fragment.

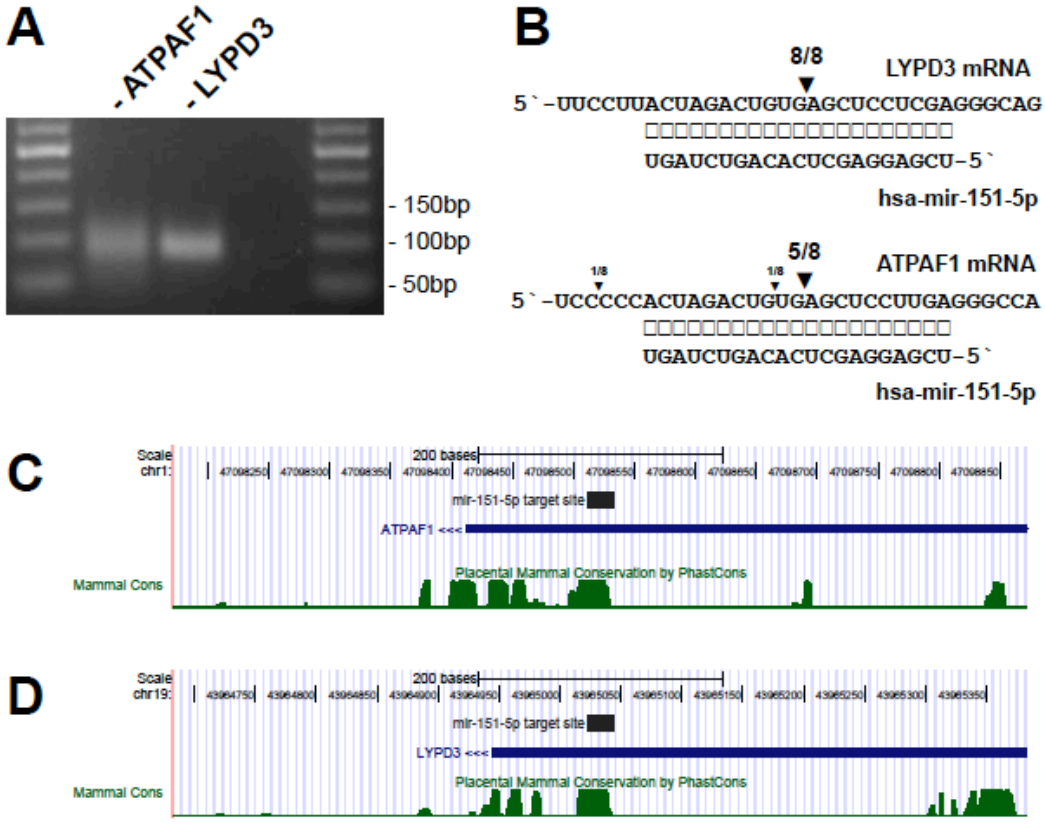

Figure 1. Validation of bioinformatically predicted miRNA cleavage targets by 5' RACE in 293S cells. (A) Nested PCR of the 5` RACE produces amplicons of the expected size for ATPAF1 and LYPD3, but not a negative control (lane 3). (B) Cloning and sequencing of the PCR amplicons indicates products corresponding to the expected miRNA cleavage site. (C) and (D) Conservation of the mir-151-5p target site in the 3` UTRs of ATPAF1 and LYPD3.

As an orthogonal approach, we undertook a purely experimental search for mRNA cleavage sites similar to that previously applied for the identification of Arabidopsis miRNA targets (Addo-Quaye et al., 2008; German et al., 2008). This procedure (Figure 2A) begins with the isolation of polyA+ RNA.Linkers are then ligated specifically to 5' ends that bear monophosphate termini. After randomly primed reverse transcription and PCR amplification, libraries are analyzed by high throughput sequencing, generating a global set of RACE tags that reveal, based upon the presence of the linker, the precise 5' ends of the 3' fragments resulting from mRNA cleavage.

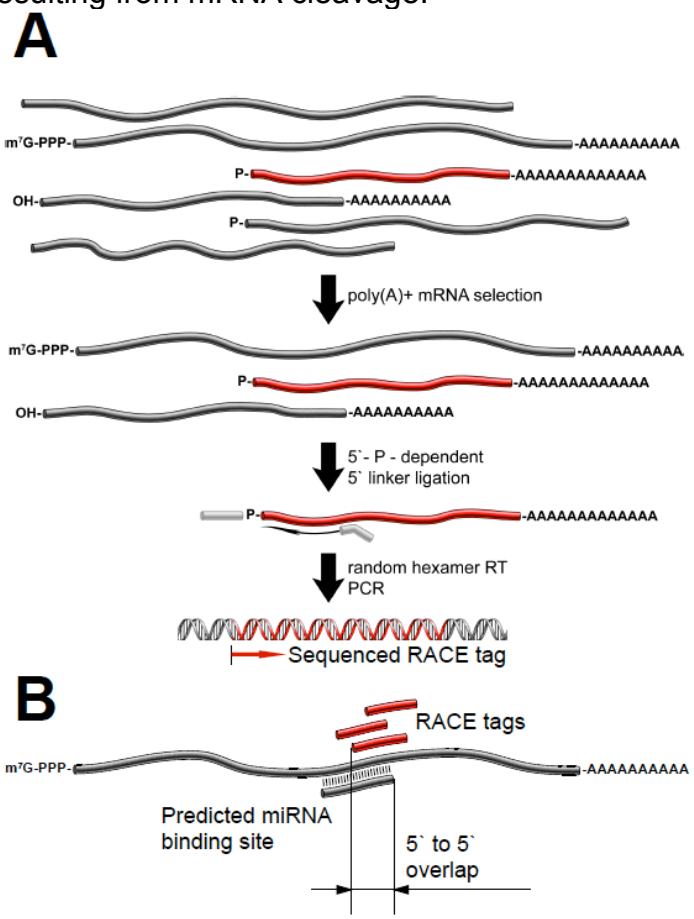

Figure 2.(A)Experimental scheme for transcriptome-wide 5'-phosphate- dependent RACE. PolyA+ RNAs are selected from total RNA and ligated to a 5' adapter.The resulting RNAs are reverse-trancribed using random hexamer primers, PCR-amplified, and sequenced. (B) The experimental RACE tag data was combined with bioinformatic miRNA target site predictions by computing the 5' to 5' overlap between the tag and the miRNA at the binding site.

As a means to discriminate definitively microRNA-directed cleavage events from those produced by other cellular nucleases, we took advantage of the fact that Ago2 is the only catalytically active Argonaute family member. We established a series of ES cell lines using blastocysts collected from an intercross of animals heterozygous for Ago2 insertional mutations within the PIWI and the PAZ domains (Liu et al., 2004). From these, we identified Ago2-/- ES lines and used wild-type ES cells for comparison.

Global RACE tags from both cell lines were intersected with the computational predictions of miRNA cleavage sites.Each time an overlap was detected, we calculated the position of the 5' end of the RACE tag with respect to the 5' end of the miRNA predicted to pair at that site (Figure 2B).When considering all positions with up to 2 non-GU mismatches, libraries from wild-type cells showed a peak of tag abundance centered 10 nucleotides away from the miRNA 5' end (Figure 3A, Supplementary File 2). This is consistent with the known biochemical properties of Argonaute proteins, which cleave the phosphodiester bond opposite nucleotides 10-11 of the small RNA guide (Schwarz et al., 2004).The significance of this signal was indicated by three additional observations. First, libraries from Ago2-/- ES cells did not show a similar relationship between RACE Tags and predicted sites of miRNA-mRNApairing.Second,thepeak disappeared if we performed similar analyses after randomizing the miRNA sequences (Figure 3B). Finally, if we included in our analysis sites with up to 5 non-GU mismatches, the enrichment at position 10 disappeared (Supplementary Figures 1A,B). These observations strongly support the ability of global RACE to identify targets of miRNA-directed Ago2 cleavage in mammalian cells.

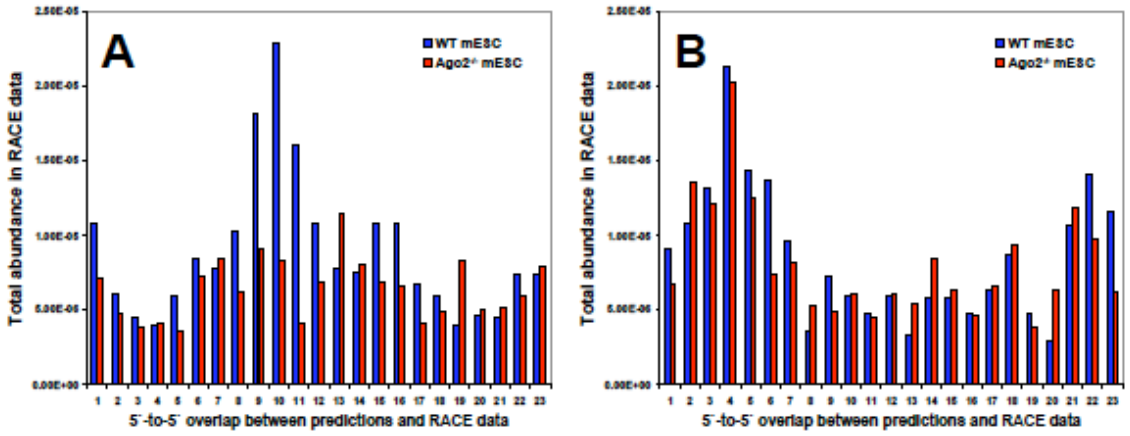

Figure 3. Combined experimental and bioinformatic approaches identify miRNA- guided mRNA cleavage sites in mouse embryonic stem cells. (A) Cumulative RACE tag abundances in WT (blue) and Ago2-/- (red) libraries as a function of overlap with predicted miRNA target sites with 2 or fewer non-GU mismatches. (B) Same as in (A), but for randomized miRNAs.

We chose two potential miRNA cleavage targets for further confirmation. Plekhm1 showed a strong Ago2-dependent RACE signal at a site with complementarity to miR-106b (Figure 4A).Consistent with this observation, conventional RACE amplified a product in wild-type, but not Ago2-/-- mESCs (Figure 4B), and the cleavage site was confirmed by clone sequencing (Figure 4C). The Plekhm1 target site was conserved throughout the length of the miRNA pairing (Figure 4D), and we could detect cleavage at the same site in the orthologous mRNA in human cells (Supplementary Figure 2). Pfkfb1 showed a weaker Ago2-dependent cleavage site in RACE libraries at a position with complementarity to the let-7 miRNA family. By conventional RACE, cleavage at this site was not detectable in ES cells; however, let-7 expression is

quite weak in this cell type.We could readily detect cleavage at the let-7 complementary site in mouse embryonic fibroblasts (Figures 4B, C), where let-7 expression is more robust.Again, the cleavage event occurred in wild-type, but not Ago2-/- MEFs.

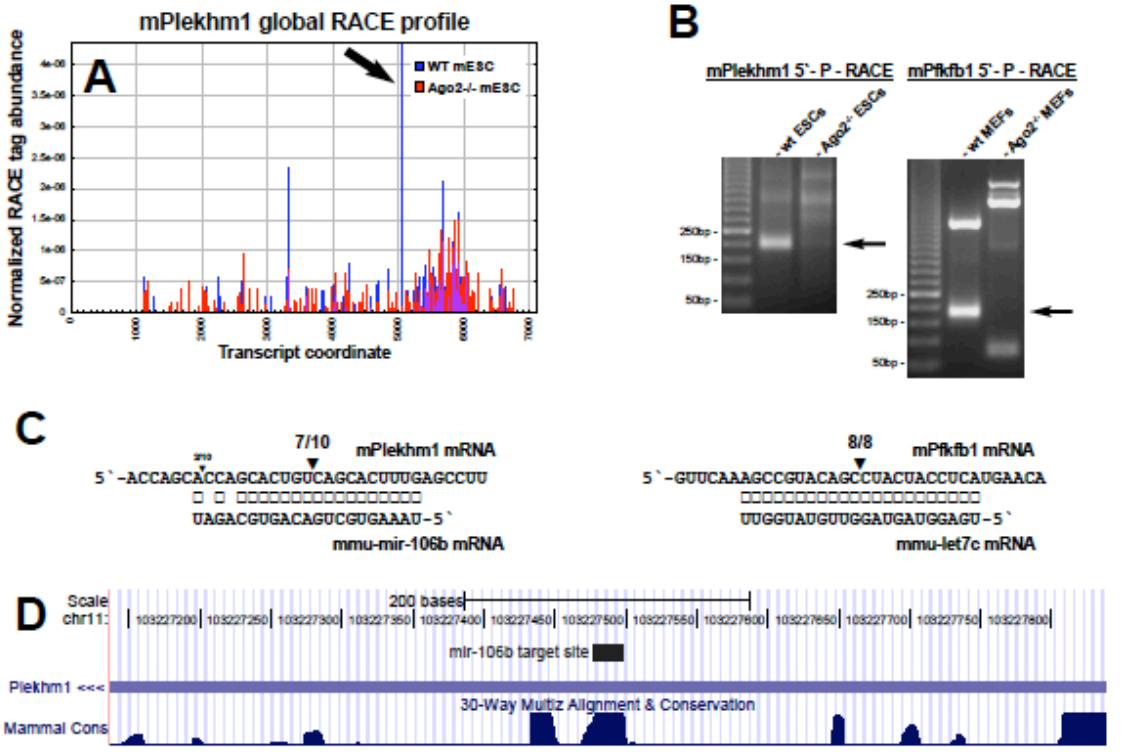

Figure 4. Validation of miRNA-guided cleavage sites. (A) Profile of RACE tag abundances mapping to the Plekhm1 transcript in WT (blue) and Ago2-/- (red) libraries. The identified Ago2-dependent cleavage site is shown by an arrow. (B) Conventional RACE validation of Plekhm1 and Pfkfb1 cleavages in WT, but not Ago2-/- cells.The expected amplicon sizes are indicated by arrows.(C) Complementarity between mir-106b, let-7c and their respective targets, Plekhm1 and Pfkfb1. Triangles indicate the number of RACE clones at the corresponding position. (D) Conservation of the mir-106b target site in the Plekhm1 3` UTR.

Drosha-dependent mRNA cleavage

While a number of the sites detected in our global RACE were clearly Ago2-dependent, there were also numerous presumptive endonucleolytic cleavage sites that persisted in the absence of Ago2. Among this class was a strong site near the 5' end of the DGCR8 mRNA (Figure 5A).Kim and colleagues previously demonstrated that Drosha, which partners with DGCR8 in the Microprocessor complex, directly regulates DGCR8 production by recognizing a site within its 5' UTR that mimics a microRNA precursor (Han et al., 2009b). The so-called A2 site, cleaved by Drosha in vitro, falls within one nucleotide of the site detected in our global RACE analysis (Supplementary Figure 3).

To determine the extent of the transcript population that might be recognized and cleaved by Drosha, we took advantage of a conditional Drosha-null ES cell line. Following cre-mediated inactivation of Drosha in ES cells, levels of Drosha mRNA decrease by 5-12 fold (data not shown).In accord with these observations, a comparison of global RACE libraries from these cells to those from wild-type (cre-uninduced) ESCs revealed loss of the miR-106b-directed, Ago2-dependent Plekhm1 site (Figure 5B).We also saw loss of the Ago2-independent site in DGCR8, consistent with its being Drosha mediated (Figure 5A).

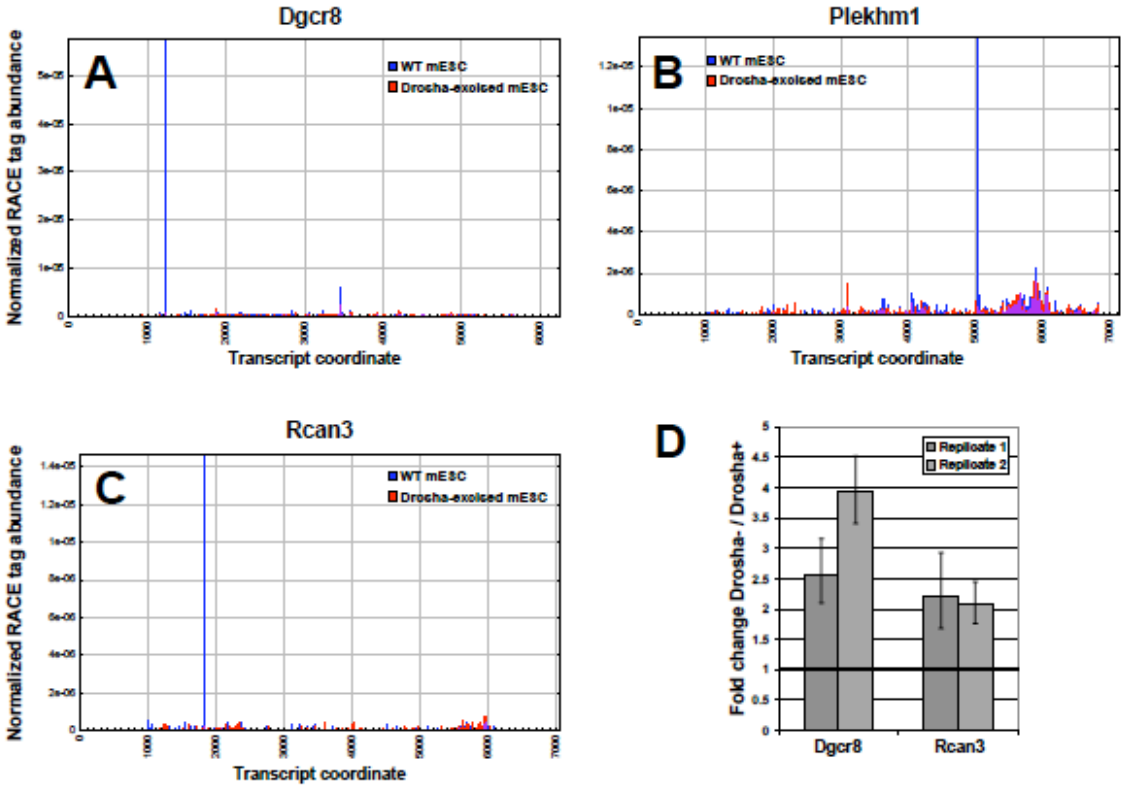

Figure 5. Drosha-dependent cleavage sites in mammalian mRNAs. (A) Profile of RACE tag abundances for the Dgcr8 transcript in WT (blue) and Drosha- excised (red) libraries. (B), (C) Same as in (A), but for Plekhm1 and Rcan3. (D) Transcript levels of Dgcr8 and Rcan3 in Drosha-excised mES cells relative to wild-type controls, as measured by QPCR.

Using their behavior in genetic mutants as a guide, we searched for additional sites of Drosha-dependent cleavage within mRNAs.We noted a substantial number of sites (Supplementary File 3), which varied in the strength of their cleavage signature within wild-type libraries and in the degree to which they responded to Drosha loss. The novel genes in this category are exemplified by RCAN3 (Figure 5C). Furthermore, we observed an increase in total DGCR8 and RCAN3 levels in Drosha-excised cells by QPCR (Figure 5D), strongly suggesting a regulatory nature for these cleavage events.

These sites are likely not responding to Drosha loss indirectly via microRNA depletion because they are not strongly changed in Ago2-null cells nor do they overlapwithpredictedsitesofmicroRNA-mRNAinteraction.Wecannot definitively state that the sites are cleaved directly by Drosha, though they show both the appropriate genetic dependency and phosphate polarity characteristic of this enzyme class. Moreover, many sites can be folded into local or more long-range secondary structures that could provide the double-stranded substrates preferred by this enzyme class.Thus, direct Drosha processing remains the mostparsimoniousexplanation forourobservations;however, further experiments will be required to provide definitive support for a direct enzyme- substrate interaction.

Additional endonucleolytic mRNA cleavage events

If we eliminate sites conforming to our expectations for Ago2- and Drosha-mediated cleavage, the global RACE data still reveals many robust cleavage events that might be catalyzed by other endonucleases. In many cases, these signals were strikingly strong and arose from relatively abundant transcripts (Figure 6 and Supplementary File 4).

Because the nucleases that catalyze such events are unknown, it is difficult to assess the biological significance of these observations.One way to prioritize potentially significant sites might be to examine conservation, both of the cleavage event and of sequence contexts that might ultimately lead to the identification of instructive motifs.We therefore compared global RACE data from human 293S cells to signatures seen in mouse ES cells. Although these cell types are not particularly similar, we were able to find many instances where orthologous genes were expressed in both species and where we observed similaroridenticalsitesof presumptiveendonucleolyticprocessing (Supplementary File 4). Two examples are shown in Figure 6. We confirmed the veracity of the genome-wide data with individual, gene-specific RACE for TRA2A in 293S cells (Supplementary Figure 4).

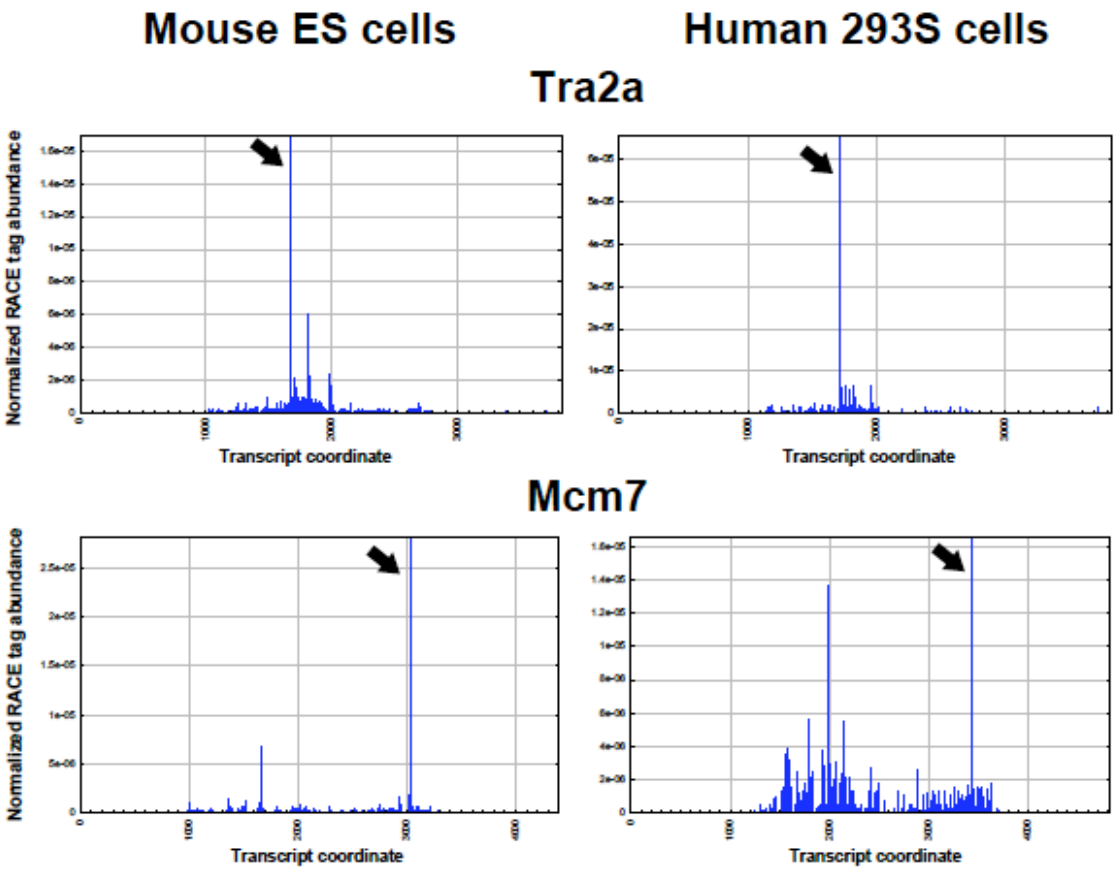

Figure 6. Examples of conserved Ago2- and Drosha-independent cleavage sites in mammalian mRNAs. Profiles of RACE tag abundances for TRA2A and MCM7 in mouse ES (left) and human 293S (right) libraries.

Discussion

We began the present work in part to gain insight into the evolutionary pressure to conserve Argonaute catalysis in vertebrates. Catalytic potential is maintained despite overwhelming evidence that microRNAs can operate largely without the need for an intact Ago catalytic site. Toward this end, we sought to determine the repertoire of microRNA-directed Ago cleavage targets through the use of a genome-wide RACE method that captures RNAs bearing a 5' phosphate termini.

A strong indication that we were identifying relevant sites came from our observed enrichment for a 10 nucleotide overlap between the 5' end of microRNAs and the 5' ends of RACE products. This depended on the integrity of Ago2 and on the comparison of cleavage sites being made to bona fide rather than scrambled microRNA sequences. The Ago2 target set was further enriched through bioinformatic predictions, and these combined approaches have now added substantially to the number of known microRNA-directed cleavage sites in

mammals. Previously, a site within a coding mRNA, HOXB8, had been validated (Yekta et al., 2004).A non-coding RNA from the Rtl1 locus has also been proposed to be regulated by miRNA-directed cleavage (Davis et al., 2005; Seitz et al., 2003).

Thus far, we have examined only one cell type, ES cells, exhaustively, and these are relatively microRNA poor.This likely causes us to minimally estimate the true extent of Ago cleavage targets that might be present were we to analyze many separate cell types. For example, we could only weakly detect cleavage of Pfkfb1 in ES cells, while this was readily apparent in MEFs. Similarly, we saw no signature in the global data of the HOXB8-miR-196 interaction because neither the microRNA nor its partner are abundant in ES cells (Yekta et al., 2004). Thus, we propose that a wide variety of miRNA cleavage targets may provide one factor which contributes to the evolutionary pressure to maintain Ago2 catalysis.

The majority of 5' monophosphorylated mRNA fragments that we detected were not dependent upon the presence of Ago2, but a number of Ago2-independent sites did depend upon another RNAi pathway component, Drosha.Previous studies had demonstrated the ability of Drosha to regulate gene expression by cleaving a microRNA-mimetic structure in the 5' end of the DGCR8 gene. We also uncovered this site as a Drosha-dependent event in a global RACE comparison of wild-type and Drosha-mutant ES cells. However, we also noted a variety of other Drosha cleavage sites, which appeared with varying strengths in our dataset.

Computational folding predictions often place the cleavage sites in regions of extensive secondary structure; however most do not present canonical pri-miRNA-like hairpins (data not shown). Our current understanding suggests that Drosha complexes recognize and cleave helical regions adjacent to single-stranded segments. This is likely determined in part by the co-factors with which it associates, namely DGCR8/Pasha for the canonical RNAi pathway.It is presently unclear whether the sites that we observe are dependent upon this cofactor or whether alternative cofactors might confer upon Drosha a preference for different structural motifs, which might be common to the cleavage sites that we observe.

Over the past several years, a remarkable variety of new long and small RNA products have come to light.Overall, it seems as if the vast majority of the mammalian genome is transcribed, and many of its transcriptional products become further processed after their production. As one example, we recently reported the identification of new small RNA species that appear to be derived from mature mRNAs by endonucleolytic cleavage and cytoplasmic capping (Fejes-Toth et al., 2009), a hypothesis that has gained support through the identification and isolation of a cytoplasmic capping complex capable of acting on

5’ monophosphorylated substrates (Otsuka et al., 2009).This suggested that there must exist a wide variety of endonucleolytic processing events that goes far beyond what is currently appreciated.

In accord with this notion, we observe a large collection of endonucleolytic cleavage sites that are independent of both Ago2 and Drosha.Both their abundance and their evolutionary conservation are suggestive of important functions. Though we have not yet identified the nucleases responsible for such events, a number of possibilities present themselves.Several endonucleases are known to leave 5’ phosphate termini, including Dicer and RNase P. Several more are likely to leave 5’ phosphates as judged from their similarity to known nuclease families. These include the stress-induced endonuclease IRE1, RNase L, and APE1. The PIN domains of Dis3/Rrp44 and SMG6 nucleases adopt an RNase H-like fold that could generate 5’-phosphorylated products (Glavan et al., 2006).As most of these enzymes lack inherent substrate specificity, their targeted cleavage sites are likely to be selected by partner binding proteins.

Considered as a whole, our findings not only reveal a previously unappreciated breadth in the roles of RNAi family enzymes in mRNA cleavage but also suggest that mRNAs are subject to a surprising diversity of endonucleolytic, post-transcriptional processing events.The path toward understanding the precise biological impact of these processing events will likely require our linking individual cleavage sites with the nucleolytic complexes that generate them.

Acknowledgements

The authors would like to thank Assaf Gordon and Oliver Tam for bioinformatic support. FVK is supported by a postdoctoral fellowship from the American Cancer Society, PF-07–058-01-GMC. This work was supported by grants from the NIH and by a kind gift from Kathryn W. Davis. GJH is a professor of the Howard Hughes Medical Institute.

Experimental Procedures

Bioinformatic Predictions of near-perfect mRNA targets

For human and mouse analyses, the reverse complementary sequences of all miRBase Release 14 miRNAs were aligned against Ensembl Release 54 mRNA transcripts, allowing for up to 5 mismatches and no insertions/deletions.The resulting predictions were dominated by a multitude of hits generated from miRNAs with degenerate or simple repetitive sequences.To eliminate such questionable hits, the analysis was limited to miRNAs numbering below 400 as a proxy of their abundance. The predictions were filtered for perfect matches at miRNA positions 9-11, and sorted by the number of non-GU mismatches and total mismatches. As a control, we predicted targets for control miRNAs, which

were created by shuffling the real miRNAs while preserving the dinucleotide composition (Workman and Krogh, 1999).

Individual 5'-phosphate dependent 5' RACE

Individual cleavage products were detected using the Invitrogen GeneRacer kit from 5 µg of total cellular RNA. The GeneRacer 5' oligo was ligated directly to total RNA samples, requiring the presence of 5' phosphates on ligated molecules. PCR reactions were carried out in a final 50 µl reaction using KOD Hot Start DNA polymerase (Novagen) containing 5 µl 10X buffer, 3 µl 25mM MgSO4, 5 µl 2mM dNTPs, 1.5 µl 10µM primers, 1 µl template DNA, and 1 µl enzyme. PCR conditions: 94 ºC for 2 minutes, followed by 35 cycles of 94 ºC for 30 seconds, 58-63 ºC (gene-specific) for 30 seconds, and 70 ºC for 1 minute, with a final extension at 70 ºC for 10 minutes.

Global 5' RACE library preparation

Poly(A)+ mRNAs were isolated from 200-500 µg of total RNA using the Invitrogen Dynabeads mRNA Direct kit. Briefly, 500 µl of oligo(dT) beads were washed once in binding buffer, resuspended in 500 µl binding buffer plus 500 µl of total RNA, and rotated at room temperature for 5 minutes.Beads were washed twice with 500 µl buffer B and eluted with 200 µl 10mM Tris-HCl, pH 7.5 by heating to 65 ºC for 2 minutes and rapidly removing the eluate. The eluate was re-purified over the same beads, eluted in 100 µl, and ethanol-precipitated. The RNA was ligated to a 5` linker, requiring the presence of a 5` phosphate on the ligated RNA, in a 20 µl volume (2 µl 10X T4 RNA ligase buffer, 2 µl DMSO, 2 µl 50 µM SBS3 linker, 2 µl T4 RNA ligase (Ambion), 12 µl reconstituted RNA pellet) at 37 ºC for 1.5 hours.The reaction was phenol-chloroform-extracted, chloroform extracted, purified over 50 µl of oligo(dT) beads as above, and eluted in 25 µl. Reverse transcription reactions were carried out by handom hexamer priming: 11 µl of ligated RNA, 1 µl of 100 µM SBS8-N6 primer, and 1 µl of 10mM

dNTPs were annealed at 65 ºC for 5 minutes and mixed with 4 µl 5X 1st strand buffer, 1 µl 100mM DTT, 1 µl RNaseIN, 1 µl SuperScript III (Invitrogen), and incubated at 50 ºC for 1 hour, followed by 70 ºC for 15 minutes. Template RNA was degraded by addition of 1 µl RNase H (Invitrogen) and incubating at 37 ºC for 20 minutes. PCR reactions on the resulting cDNAs were carried out in a final 50 µl reaction using KOD Hot Start DNA polymerase (Novagen) containing 5 µl 10X buffer, 3 µl 25mM MgSO4, 5 µl 2mM dNTPs, 3 µl 10µM PE-P5-SBS3 and PE-P7-SBS8 primers, 1 µl template DNA, and 1 µl enzyme. PCR conditions: 95 ºC for 2 minutes, followed by 30 cycles of 95 ºC for 15 seconds, 60 ºC for 30 seconds, and 72 ºC for 1 minute, with a final extension at 72 ºC for 7 minutes. Products were run on a 2% low-melt agarose gel, and the 150-500 bp (sometimes 150-1000bp) range was excised, purified, and sequenced by Illumina high-throughput sequencing. For Ago2-dependent studies, a total of 6 independent wild-type samples and 10 Ago2-/- samples were prepared and

sequenced on one Illumina lane each.For 293S samples, a total of 7 independent samples were sequenced.For Drosha-dependent studies, two conditional knockout replicates were performed. Drosha flox/flox, CreER mouse embryonic stem cells were treated with 100nM 4-OH tamoxifen for 4-9 days, along with untreated controls.Each replicate sample was sequenced on 4-5 Illumina lanes.

Bioinformatic analysis of sequenced libraries

We used the ENSEMBL transcripts as our transcriptomic reference (Birney et al., 2004). Reads were mapped with the RMAP mapping tool (Smith et al., 2009), allowing up to 2 mismatches.All reads were 36nt. Mapping was done to a reference that included a non-redundant set of all exonic sequence from the mm9 and hg18 assemblies from the UCSC Genome Browser (Kent et al., 2002), for mouse and human, respectively. A non-redundant junction reference was constructed from all pairs of exons within a gene; unique pairs of genomic positions for donor and acceptor sites defined junctions. If a read mapped inside an exon or junction corresponding to a given transcript, the read was assigned to that transcript (in this way reads may be assigned to multiple transcripts).
Based on the read counts at each site within each transcript, endonuclease sites were identified for a given biological replicate as follows. First, within each transcript, the read counts were fit (max likelihood) to a negative binomial (modeling over-dispersed counts data) assuming any sites with read counts of 0 were latent data. We assigned p-values to each given site based on the negative binomial fit for each transcript containing that site. Finally, for each sample we applied FDR (< 0.05) to the complete set of site p-values, and kept each site with a p-value lower than the cutoff.
Drosha-dependent sites were required to be identified as significant in both wild-type replicates, but in neither Drosha Cre-out replicate. The set of Drosha-independent sites was obtained using the above procedure, but using pooled reads from sequencing runs for multiple biological replicates.

Plotting of RACE data

Individual sample libraries were aligned to a database of Ensembl transcripts including 1000nt of 5' and 3' flanking sequence. Matching reads were filtered for unique, sense, Ensembl mRNA-matching reads. Read counts were normalized to the total of such reads. Normalized libraries from biological replicates were combined and averaged.

QPCR of transcript levels

Changes in transcript levels were detected using the Taqman RNA-to-Ct 1-step kit (Applied Biosystems) with 18S probes as the internal control.

References

Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol 18, 758-762.

Azuma-Mukai, A., Oguri, H., Mituyama, T., Qian, Z. R., Asai, K., Siomi, H., and Siomi, M. C. (2008). Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. Proc Natl Acad Sci U S A 105, 7964-7969.

Barnes, T., Kim, W. C., Mantha, A. K., Kim, S. E., Izumi, T., Mitra, S., and Lee, C. H. (2009). Identification of Apurinic/apyrimidinic endonuclease 1 (APE1) as the endoribonuclease that cleaves c-myc mRNA. Nucleic Acids Res 37, 3946-3958.

Birney, E., Andrews, T. D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T., et al. (2004). An overview of Ensembl. Genome Res 14, 925-928.

Carthew, R. W., and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642-655.

Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol 15, 743-749.

Eberle, A. B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T. H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol 16, 49-55.

Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. Embo J 20, 6877-6888.

Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A human snoRNA with microRNA-like functions. Mol Cell 32, 519-528.

Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G. J., Kapranov, P., Foissac, S., Willingham, A. T., Duttagupta, R., Dumais, E., and Gingeras, T. R. (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature 457, 1028-1032.

German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L. A., Nobuta, K., German, R., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol 26, 941-946.

Glavan, F., Behm-Ansmant, I., Izaurralde, E., and Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. Embo J 25, 5117-5125.

Haley, B., and Zamore, P. D. (2004). Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol 11, 599-606.

Han, D., Lerner, A. G., Vande Walle, L., Upton, J. P., Xu, W., Hagen, A., Backes, B. J., Oakes, S. A., and Papa, F. R. (2009a). IRE1alpha kinase activation modes control alternate endoribonuclease outputs to determine divergent cell fates. Cell 138, 562-575.

Han, J., Pedersen, J. S., Kwon, S. C., Belair, C. D., Kim, Y. K., Yeom, K. H., Yang,W.Y.,Haussler,D.,Blelloch,R.,andKim,V.N.(2009b).Posttranscriptional crossregulation between Drosha and DGCR8. Cell 136, 75-84.
Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C., and Peterson, K. J. (2008). MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci U S A 105, 2946-2950.
Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. Rna 14, 2609-2617.
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.
Lebreton,A.,Tomecki,R.,Dziembowski,A.,andSeraphin,B.(2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature 456, 993-996. Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science 305, 1437-1441.
Martinez, J., and Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev 18, 975-980.
Otsuka, Y., Kedersha, N. L., and Schoenberg, D. R. (2009). Identification of a cytoplasmic complex that adds a cap onto 5'-monophosphate RNA. Mol Cell Biol 29, 2155-2167.
Schwarz, D. S., Tomari, Y., and Zamore, P. D. (2004). The RNA-induced silencing complex is a Mg2+-dependent endonuclease. Curr Biol 14, 787-791.
Seitz, H., Youngson, N., Lin, S. P., Dalbert, S., Paulsen, M., Bachellerie, J. P., Ferguson-Smith, A. C., and Cavaille, J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat Genet 34, 261-262.
Smalheiser, N. R., and Torvik, V. I. (2005). Mammalian microRNAs derived from genomic repeats. Trends Genet 21, 322-326.
Smith, A. D., Chung, W. Y., Hodges, E., Kendall, J., Hannon, G., Hicks, J., Xuan, Z., and Zhang, M. Q. (2009). Updates to the RMAP short-read mapping software. Bioinformatics 25, 2841-2842.
Tourriere, H., Gallouzi, I. E., Chebli, K., Capony, J. P., Mouaikel, J., van der Geer, P., and Tazi, J. (2001). RasGAP-associated endoribonuclease G3Bp: selective RNA degradation and phosphorylation-dependent localization. Mol Cell Biol 21, 7747-7760.
Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell 136, 669-687.
Wang, Z., and Kiledjian, M. (2000). Identification of an erythroid-enriched endoribonuclease activity involved in specific mRNA cleavage. Embo J 19, 295
Workman, C., and Krogh, A. (1999). No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. Nucleic Acids Res 27, 4816-4822.
Yekta, S., Shih, I. H., and Bartel, D. P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science 304, 594-596.

Appendix4: A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity

Daniel Cifuentes[1], Huiling Xue[1], David, W. Taylor[3] , Heather Patnode[1], Yuichiro Mishima[1, 4], **Sihem Cheloufi[5, 6]**,Enbo Ma[10] Shrikant Mane[9], Gregory J. Hannon[5], Nathan Lawson[7], Scot Wolfe[7,8], Antonio J. Giraldez[1,2‡]

[1]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510

[2]Yale Stem Cell Center, Yale University School of Medicine, New Haven, CT 06520

[3]Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT
06510

[4]Department of Biology, Graduate School of Science, Kobe University1-1 Rokkodaicho Nadaku, Kobe 657-
8501, Japan

[5]Cold Spring Harbor Laboratory. Watson School of Biological Sciences Howard Hughes Medical Institute
Cold Spring Harbor,NY 11724

[6]Program in Genetics, Stony Brook University, Stony Brook, NY 11794

[7]Program in Gene Function and Expression, [8]Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School, Worcester

[9]Yale Center for Genome Analysis, Yale West Campus, Orange, CT 06477

[10]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720.

‡To whom correspondence should be addressed
E-mail: antonio.giraldez@yale.edu
Tel: 203.785.5423
Fax: 203.785.4415

Abstract

Dicer is a central enzyme in miRNA processing. Here, we identify a Dicer-independent miRNA biogenesis pathway that employs Argonaute2 (Ago2) slicer catalytic activity. In contrast to other miRNAs, miR-451 levels were refractory to *dicer* loss-of-function but were reduced in MZ*ago2* mutants. We show that pre- miR-451 processing requires Ago2 catalytic activity in vivo. MZ*ago2* mutants showed delayed erythropoiesis that could be rescued by wild-type Ago2 or miR-451-duplex but not catalytically dead Ago2. Changing the secondary structure of dicer-dependent miRNAs to mimic that of pre-miR-451 restored miRNA function and rescued developmental defects in MZ*dicer* mutants, indicating that the pre- miRNA secondary structure determines the processing pathway *in vivo*. We propose that Ago2-mediated cleavage of pre-miRNAs followed by uridylation and trimming generates functional miRNAs independently of Dicer.

microRNAs (miRNAs) are ~22nt small RNAs that regulate deadenylation, translation, and decay of their target mRNAs (*1*, *2*). In animals, most miRNAs are processed from a primary transcript (termed pri-miRNA) by two RNase III enzymes, Drosha and Dicer. Recent studies have identified several miRNA classes that bypass Drosha-mediated processing namely miRtrons, tRNAZ and snoRNA (*2-6*). In contrast to Drosha, Dicer has been viewed as a central processing enzyme in the maturation of small RNAs (*2*). But are there functional miRNAs that bypass Dicer? To identify pathways that might process miRNAs in a Dicer-independent manner, we have sequenced small RNAs (19-36nt) in wild type and maternal-zygotic *dicer* mutants (MZ*dicer*) (*7*). We analyzed 48h-old embryos in two wild type replicates and two *dicer* mutant alleles (*8*), *dicer*^{hu715} and *dicer*^{hu896} (fig. S1). Of the ~2 millions reads/sample, 69-82% mapped to known 5'- or 3'-derived miRNAs in wild-type, while 4-9% mapped to miRNAs in the MZ*dicer* mutants (fig. S2). Several miRNAs appeared refractory to dicer loss- of-function, notably miR-451-5', miR-2190-5', miR-2190-3' and miR-735-5' (Fig.1A, fig. S3 and S4). Based on read-frequency, reproducibility, and evolutionary conservation, we focused subsequent analysis on miR-451. miR-451 differs from other "canonical" miRNAs. First, it is encoded in a conserved 42nt hairpin (fig. S5) with a 17nt stem, while Dicer requires >19nt-long stem for efficient processing (*9*). Second, miR-451 has a defined 5'-end but a variable 3'-end that extends over the loop region and ranges between 20-30nt (Fig.1C and D). Third, reads stopped at nucleotide #30, and longer reads carried 1-5 non-templated uridines, with nucleotide #31 mostly being a non-templated U (Fig.1D). The final templated base pairs with nucleotide #10 of the mature miRNA (Fig. 1C, fig. S1), a site where slicer activity cleaves the passenger strand in siRNAs (*10-12*). These observations lead us to hypothesize that Ago2 slicer activity could participate in miRNA maturation (fig. S1).

To determine whether Ago2 participates in miRNA maturation, we generated a deletion in the Piwi domain of the *ago2* gene (*ago2*$^{!\ 90}$) using zinc finger nucleases (*13-15*) (Fig. 1E, fig. S1). Because argonaute genes are maternally expressed (fig. S6), we generated maternal-zygotic *ago2* mutants (MZ*ago2*). Indeed, slicer cleavage of an mRNA with perfectly complementary targets to miR-1 was severely reduced in MZ*ago2* but not Z*ago2* compared to wild-type embryos (Fig. 1F, fig. S1).

To investigate the role of Ago2 in miRNA processing, we sequenced small RNAs (19-36nt) from 48h-old MZ*ago2* mutant embryos. Comparing the normalized read frequency for each 5'- and 3'-mature miRNA between wild type and MZ*ago2* mutants, revealed a reduction in the number of reads that mapped to miR-451 (Fig.1B and D). In contrast, other miRNAs remained largely unchanged. miR-451 and miR-144 are co-expressed from a common primary transcript in the erythroid lineage (*16, 17*) (Fig. 1C). While miR-451 accumulated in the absence of Dicer (~3 fold higher), miR-144 reads were

strongly reduced in MZ*dicer* mutants (>200 fold) (Fig.1A, and SOM).

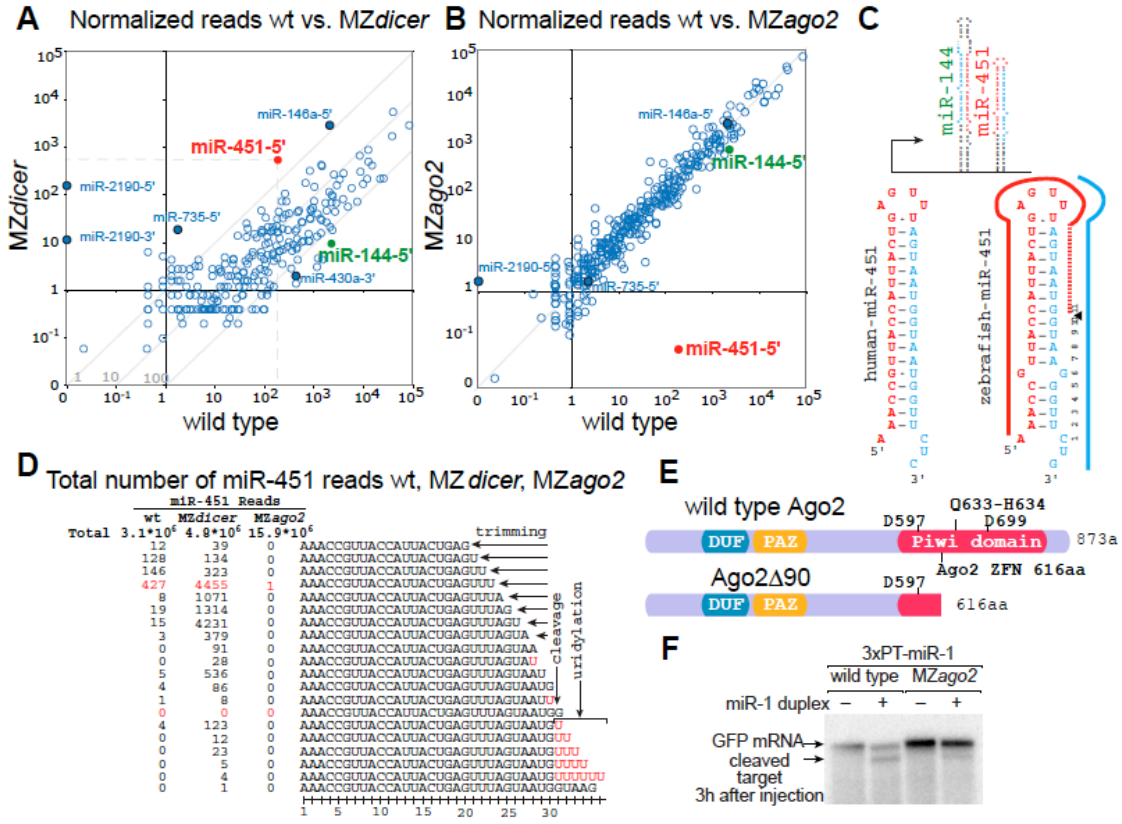

Figure 1. microRNA analysis in wild type, MZ*dicer* and MZ*ago2*. (A, B) Normalized reads from wild type vs. MZ*dicer* (A) or MZ*ago2* (B) libraries for all annotated zebrafish miRNAs. Some miRNAs are shown as a reference for enhanced and reduced miRNAs (filled circles). miR-144-5'(green) and miR-451-5'(red) are expressed in the same pri-miRNA.**(C)** Scheme of miR-144/miR-451 genomic loci and predicted secondary structure of both human and zebrafish pre- miR-451 (mature miRNA red) **(D)** Total number of reads that match miR-451 in wild type (wt), MZ*dicer* and MZ*ago2*. Non-templated uridines are shown in red. **(E)** Domain organization of Ago2. The 90nt deletion (**!** 90) results in a predicted truncated protein lacking two of three catalytic residues. **(F)** Northern blot of embryos to detect slicer cleavage of an injected GFP target mRNA with three complementary targets to miR-1 (3xPT-miR-1) in the presence (+) or absence (–) of miR-1 (*7*). Slicer activity is indicated by the higher mobility product (fig. S1).

Conversely, *ago2* loss-of- function did not affect the read frequency of miR-144 (Fig.1B) but did reduce miR-451 levels >8000 fold. Taken together, these results indicate Ago2 regulates miR-451 levels post-transcriptionally by affecting either its processing or stability.

Recent studies suggest that Ago2 binds pre-miRNAs and miRNA:miRNA* duplexes (*19-21*). Ago2 interacted with radiolabeled synthetic pre-miR-451 *in vitro* (fig. S7). Co-expression of Flag-mouse-Ago2 (mAgo2) with pre-miR-451 or a mutant pre-miR-451$^{mm10\text{-}11}$ (with two mismatches in

the predicted slicer cleavage site) followed by Ago2 immunoprecipitation showed that Ago2 bound to both mature miR-451 and pre-miR-451$^{mm10-11}$ (Fig. 2A). Incubation of hAgo2 with pre-miR-451 but not pre-miR-430 resulted in a sharp 30nt band corresponding with the predicted slicer cleavage product of miR-451 (Fig. 2B). Conversely, recombinant Dicer bound both pre-miRNAs (fig. S7), but could only process pre- miR-430 (Fig. 2B). To investigate whether Ago2 processes miR-451, unlabeled or 5'-P32-labeled pre-miRNAs were injected into one-cell stage embryos. Synthetic and endogenous pre-miR-451 hairpins were processed into ~30nt intermediates and a ~22-26nt mature miR-451 in wild type and MZ*dicer*- but not in MZ*ago2*-mutant embryos (Fig.2D and F). In contrast, a canonical mature miR-430 was processed in both wild type and MZ*ago2* mutant embryos but not in MZ*dicer* (Fig. 2F). Based on the sequencing results we hypothesized that Ago2- processed hairpin might undergo nucleolytic trimming at the 3'-end (Fig. 1D). We observed that Ago2 protected the ~30nt slicer-cleaved intermediate from RNase1 in vitro, resulting in a ~20-26nt 3'-end trimmed product (Fig. 2C), similar to the mature miRNAs observed *in vivo* (Fig 2D-F). Ago2 slicer activity depends on its catalytic triad (DDH) and the pairing between the guide and the target mRNA(*22-24*). Expressing wild type but not catalytically dead(D669A) mAgo2 in MZ*ago2* mutants rescued pre-miR-451 processing *in vivo* (Fig. 2E). Furthermore, a hairpin with mismatches that disrupt pairing in the predicted slicer-cleavage, was bound by Ago2 (fig. S7) but was inefficiently processed into mature miR-451 (Fig.3E). These results indicate that Ago2 binds and cleaves pre-miR-451 in a process that requires the slicer catalytic activity, and is independent of Dicer.

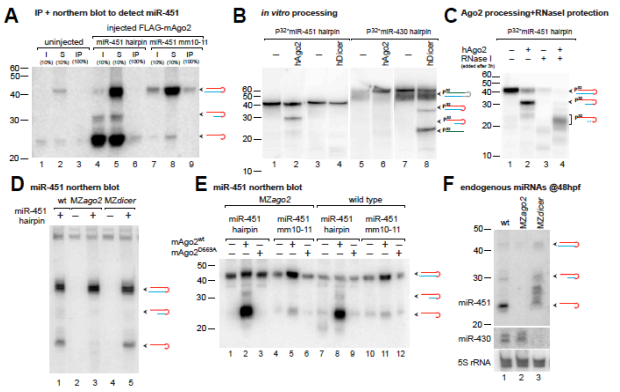

Figure 2: Ago2 binds and processes pre-miR-451.

(**A**) Immunoprecipitation of FLAG-mAgo2 in wild-type and mutant embryos injected with pre-miR-451 followed by northern blot analysis to detect bound miR-451. Input (I), supernatant (S), and immunoprecipitate (IP) are indicated. **(B, C)** *In vitro* cleavage assay using hAgo2 or hDicer protein and 5'-radiolabeled pre- miR-430 or pre-miR-451. (C) Ago2 processing reactions were treated with(+) or without(-) RNAse1 to assay protection of the processed hairpin by Ago2. **(D, E, F)** Northern blot analyses to detect mature miR-451 after injection with pre-miR-451 (+) (D, E) or endogenous miR-451 and miR-430 (F). Injection of wild-type mAgo2 but not a catalytically dead mAgo2^{D669A} rescues pre-miR-451 processing *in vivo* (E). The processing of miR-451$^{mm10-11}$ is strongly reduced. Endogenous pre-miR-451 at 48hpf is processed in wild type and MZ*dicer* but not in MZ*ago2* mutants. Diagrams for predicted hairpins, cleavage intermediates, mature miR-451 (red), miR-430 (green) and miRNA* (blue) are shown on the left. P^{32} indicates that injected hairpins were radiolabeled.

MZ*ago2* mutant embryos displayed normal morphogenesis during gastrulation, brain development, and heart development(fig S8). Ago2 is maternally expressed, and later in development, it acquires tissue-specific

expression in the brain and intermediate cell mass (ICM) (Fig. 3C; fig. S6). The ICM corresponds to the hematopoietic precursors, and overlaps with the expression domain of miR-451 (*16*), which plays an important role in erythrocyte maturation in zebrafish (*16, 17*). Consistent with the Ago2-dependent processing of miR-451, MZ*ago2* but not MZ*dicer* mutants showed a reduction in the number of hemoglobinized erythrocytes (Fig. 3A, fig. S8). In zebrafish, erythrocyte maturation can be monitored by changes in erythrocyte morphology and reduced nuclear:cytoplasmic (N:C) ratio (*17*, *25, 26*). Erythrocyte maturation was delayed MZ*ago2*, manifested by a significant increase in N:C ratio at 60hpf ($P<10^{-15}$) (Fig.3D, E). Providing back wild-type mAgo2 or mature miR-451-duplex but not catalytically dead mAgo2(D669A) rescued erythrocyte maturation in MZago2 (Fig. 3D, E). Thus, Ago2 catalytic function plays an important role during
erythrocyte maturation.

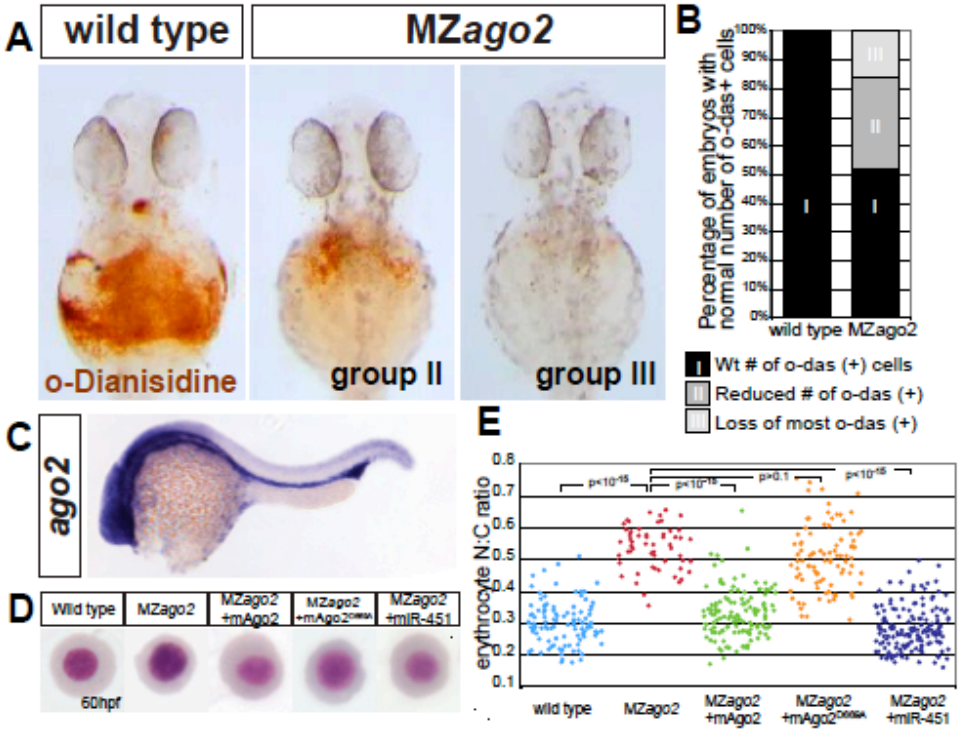

Figure 3. MZ*ago2* mutants show reduced erythropoiesis.

(A) Expression of hemoglobin (brown) visualized by the oxidation of o- dianisidine (o-das) at 48hpf in the ducts of Cuvier. Hemoglobinized cells accumulate in wild type but are reduced in MZ*ago2* mutants (Group II mild and group III severe reduction of o-das positive cells). **(B)** Percentage of embryos with hemoglobinized cells in MZ*ago2* mutants (n=61) compared to wild-type embryos (n=200); (light gray) strongly reduced (group III) and partially reduced (gray, group II) number of o-das (+) cells, (Chi square test P<0.001). **(C)** Whole- mount *in situ* hybridization of *ago2* expression at 24hpf. **(D)** May- Grünwald/Giemsa stain of erythrocytes from wild type, MZ*ago2* mutants and MZ*ago2* injected at one-cell stage with various RNAs as shown (+). Erythrocytes are representative of the mean for each group. **(E)** Scatter plot of the nuclear cytoplasmic ratio (N:C) for each genotype in (D) as a read-out of erythrocyte maturation (*17*). Distributions of the N:C ratios in wild type compared to MZ*ago2* differed significantly (Wilcoxon rank-sum test after Bonferroni correction $P<10^{-15}$). Erythrocyte maturation is rescued by miR-451-duplex (MZago2 and MZago2+miR-451 $P<10^{-15}$) and wild-type mAgo2 (MZago2 and MZago2+mAgo, $P<10^{-15}$) but not catalytically dead mAgo2 (D669A) (MZago2 and MZago2+mAgo^{D669A}, P>0.1).

Whereas miR-451 is a 42nt miRNA hairpin, canonical vertebrate miRNAs are ~60nt and unlike most miRNAs, mature miR-451 extends into the loop of the hairpin where it overlaps with the miRNA* (Fig. 4A and fig. S5). We hypothesized that selection of the processing pathway may be determined by structural differences or specific sequence motifs. To distinguish between these two scenarios, we modified the sequence of pre-miR-451 to encode a Dicer- dependent miRNA (miR-430c or miR-1) mimicking pre-miR-451 structure and length (pre-miRNA$^{ago2\text{-}hairpin}$) (Fig. 4A, fig. S10). miR-430c is a member of a zygotically expressed miRNA family that regulates maternal mRNA clearance, gastrulation and brain morphogenesis (*7, 27*). These processes are disrupted in MZ*dicer* mutants, but can be rescued by injection of a Dicer-independent miR-430-duplex (*7, 27*). Three lines of evidence indicate that pre-miR-430c$^{ago2\text{-}hairpin}$ is processed and functional independently of Dicer. First, synthetic pre- miRNA$^{ago2\text{-}hairpin}$ was processed into a ~23nt mature miRNA *in vivo* (Fig. 4C, fig S10) and processed by recombinant hAgo2 but not hDicer *in vitro*(Fig. 4D). Second, injection of miR-430c$^{ago2\text{-}hairpin}$ into MZ*dicer* embryos repressed translation of a GFP-miR-430-reporter compared to a dsRed control (Fig. 4B). Third, injection of pre-miR-430c$^{ago2\text{-}hairpin}$ into MZ*dicer* mutants rescued the gastrulation and brain morphogenesis defects similarly to a miR-430-duplex (Fig.4E). In contrast, equimolar levels of the endogenous Dicer-dependent pre-miR-430c did not rescue the MZ*dicer* phenotype (Fig. 4E). A second engineered miRNA (miR-1$^{ago2\ hairpin}$) was also processed independently of Dicer and downregulated a GFP reporter in vivo (fig. S10). These results support a model in which the secondary structure of the hairpin determines whether a pre-miRNA is processed by Ago2 to form a physiologically functional dicer-independent miRNA.

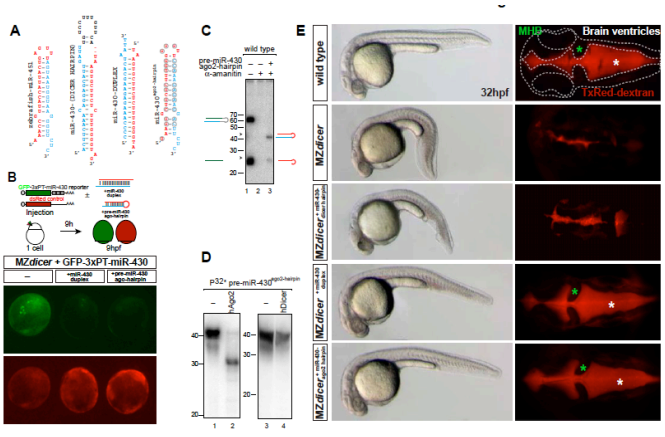

Figure 4: A Dicer-independent miRNA

(A) Zebrafish pre-miRNAs and duplexes as indicated. pre-miR-430$^{ago2\text{-}hairpin}$ is a miR-430c hairpin that has been mutated and shortened to form a 42nt hairpin mimicking pre-miR-451 (ago2-hairpin).**(B)** GFP-reporter mRNA (green) was co- injected at the one-cell stage with control dsRed mRNA (red). The GFP reporter contains three complementary target sites to miR-430 in its 3'UTR. **(C)** Northern blot to detect miR-430 in wild-type embryos injected with hairpins as indicated. **"**- amanitin was co-injected to inhibit transcription of endogenous pri-miR-430. **(D)** Northern blot to detect 5'-radiolabeled pre-miR-430$^{ago2\text{-}hairpin}$ after *in vitro* processing by recombinant hAgo2 and hDicer. **(E)** *In vivo* assay to rescue miR-430 function in MZ*dicer* mutants. Bright field and fluorescent images of the dorsal view of the brain following injection of TxRed dextran in the ventricles (right) in 32hpf embryos. Brain outline(dashed line), midhindbrain boundary (green asterisk) and ventricles (red, white asterisk) are shown. Morphogenesis defects are rescued by injection of a Dicer-independent pre-miR-430c$^{ago2\text{-}hairpin}$ or a miR-430-duplex but not a Dicer-dependent pre-miR-430c

Our study defines a Dicer-independent pathway for miRNA processing that is dependent on Ago2 catalytic activity. We propose a model whereby Ago2 binds the pre-miRNA and cleaves the paired miRNA* passenger strand 10 nucleotides away from the 5'-end of the Ago2-bound miRNA guide strand. Based on our small-RNA sequencing, this intermediate would undergo polyuridylation and nuclease-mediated removal of uridines and templated nucleotides not protected to Ago2 to generate the mature miRNA (fig. S11). Previous studies have shown that the terminal uridylyl transferase (TUT4) is recruited by lin-28 to uridylate pre-let7 (*28*), which blocks miRNA maturation and accelerates its degradation. While we cannot exclude that miR-451 uridylated intermediates are targeted for complete degradation, our model favors a scenario where uridylated Ago2-cleaved pre-miRNAs are trimmed by a cellular nuclease to generate mature miRNA sequences protected by Ago2.

Ago2 has been reported to cleave siRNAs and pre-miRNAs (*10-12, 19, 20*). Ago2-cleaved precursors (ac-pre-miRNAs) can serve as Dicer substrates, but their physiological functions remain unclear (*20*). Here, we show that Ago2- cleavage is necessary for the generation of a functional miRNA (Fig. 1, 2, 4). The identification of a miRNA-processing pathway that bypasses Dicer function might have wide implications for the processing of canonical miRNAs. Our study provides a biological context in which Ago2 slicer activity is needed to process a blood-specific miRNA miR-451. While it is likely that Ago2 has additional roles in the cell by cleaving perfectly complementary targets(*1*), the strong conservation of the sequence and secondary structure of miR-451 across vertebrates suggests that significant constraints are in place to maintain this Ago2-mediated miRNA processing pathway through evolution (See Supplementary Discussion).

References

1. D. P. Bartel, *Cell* **136**, 215 (Jan, 2009).
2. R. W. Carthew, E. J. Sontheimer, *Cell* **136**, 642 (Feb, 2009).
3. J. E. Babiarz, J. G. Ruby, Y. Wang, D. P. Bartel, R. Blelloch, *Genes Dev* **22**, 2773 (Oct, 2008).
4. E. Berezikov, W. J. Chung, J. Willis, E. Cuppen, E. C. Lai, *Mol Cell* **28**, 328 (Oct,2007).
5. K. Okamura, J. W. Hagen, H. Duan, D. M. Tyler, E. C. Lai, *Cell* **130**, 89 (Jul,2007).
6. J. G. Ruby, C. H. Jan, D. P. Bartel, *Nature* **448**, 83 (Jul, 2007).
7. A. J. Giraldez *et al.*, *Science* **308**, 833 (May, 2005).
8. E. Wienholds, M. J. Koudijs, F. J. van Eeden, E. Cuppen, R. H. Plasterk, *Nat Genet* **35**, 217 (Nov, 2003).
9. D. Siolas *et al.*, *Nat Biotechnol* **23**, 227 (Feb, 2005).
10. C. Matranga, Y. Tomari, C. Shin, D. P. Bartel, P. D. Zamore, *Cell* **123**, 607 (Nov,2005).
11. J. Martinez, A. Patkaniowska, H. Urlaub, R. Luhrmann, T. Tuschl, *Cell*

110, 563 (Sep, 2002).
12. B. Czech *et al.*, *Mol Cell* **36**, 445 (Nov, 2009).
13. Y. Doyon *et al.*, *Nat Biotechnol* **26**, 702 (Jun, 2008).
14. M. L. Maeder *et al.*, *Mol Cell* **31**, 294 (Jul, 2008).
15. X. Meng, M. B. Noyes, L. J. Zhu, N. D. Lawson, S. A. Wolfe, *Nat Biotechnol* **26**, 695 (Jun, 2008).
16. L. C. Dore *et al.*, *Proc Natl Acad Sci U S A* **105**, 3333 (Mar, 2008).
17. L. Pase *et al.*, *Blood* **113**, 1794 (Feb, 2009).
18. S. M. Hammond, S. Boettcher, A. A. Caudy, R. Kobayashi, G. J. Hannon, *Science* **293**, 1146 (Aug, 2001).
19. B. Wang *et al.*, *Nat Struct Mol Biol* **16**, 1259 (Dec, 2009).
20. S. Diederichs, D. A. Haber, *Cell* **131**, 1097 (Dec, 2007).
21. G. S. Tan *et al.*, *Nucleic Acids Res* **37**, 7533 (Dec, 2009).
22. J. Liu *et al.*, *Science* **305**, 1437 (Sep, 2004).
23. J. J. Song, S. K. Smith, G. J. Hannon, L. Joshua-Tor, *Science* **305**, 1434 (Sep, 2004).
24. N. H. Tolia, L. Joshua-Tor, *Nat Chem Biol* **3**, 36 (Jan, 2007).
25. B. M. Weinstein *et al.*, *Development* **123**, 303 (Dec, 1996).
26. F. Qian *et al.*, *PLoS Biol* **5**, e132 (May, 2007).
27. A. J. Giraldez *et al.*, *Science* **312**, 75 (Apr, 2006).
28. Z. S. Kai, A. E. Pasquinelli, *Nat Struct Mol Biol* **17**, 5 (Jan).

Acknowledgements:

Jennifer Doudna, Donal O'Carroll, Len Zon, Scott Lacadie, Graham Lieschke, Diane Krausse and Stephanie Halene for reagents and protocols. Anton Enright, Cei Abreu-Goodger and Julius Brennecke for initial small RNAanalysis. Beth Schachter, Carter Takacs, Valentina Greco and Demian Cazallafor discussions and manuscript editing. DC and HX are supported by the Fundación Ramón Areces and the Human Frontier Science Fellowshiprespectively. This project was supported by the Yale Scholar program, the Pew Scholars Program and NIH R01GM081602-03/03S1 AJG.

Contributions: DC, AJG designed and performed experiments, HX performed computational analysis. DC and DWT performed the in vitro assays, HP performed *in situs*, YM, GJHand SC helped with initial small RNA library sequencing and discussion, SW andNL designed the ZFN. AJG wrote the manuscript. Sequencing data accession number GSE21503 (GEO)