MicroRNAs in Early Embryonic Development: Dissecting the Role of *miR-290* through *miR-295* in the Mouse

by

Lucas M. Dennis

B.S., Biochemistry and Molecular Biology
Ursinus College, 2002

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JULY, 2008

Signature of Author: _____
Department of Biology
May 6th, 2008

Certified by: _____
Rudolf Jaenisch
Professor of Biology
Thesis Supervisor

Accepted by: _____
Stephen P. Bell
Professor of Biology
Chairman, Committee for Graduate Students

Acknowledgements

None of the work described in this thesis would have been completed without the support of my loving parents, Roger and Barbara and my beautiful wife, Shannon. Shannon especially has been with my through all the ups and downs of graduate school, for which I am profoundly grateful and deeply indebted.

Much of this work began as a result of a close collaboration with Hristo Houubaviy, formerly of the Sharp lab. His initial discovery of the miR-290 cluster and its biogenesis are essential underpinnings for this work. Characterization of the miR-290 cluster deletion mouse was done in close collaboration with Mark Gill and I thank him for his devotion to the project.

I am grateful for all the helpful comments, insights and aid from past and present members of the Jaenisch Lab, especially Grant Welstead who provided some much needed motivation while enduring countless miles of running with me.

I also wish to thank my classmates, BioGrad 2002, most notably my former roommates: Adam Saffer, Mark Gill, Mauro Calabrese and Rami Rahal with whom I shared some exceptional times.

I am appreciative of the members of my thesis committee, David Bartel, Tyler Jacks and Phil Sharp, whose guidance, encouragement and insight were key to my success at MIT. I would also like to thank Konrad Hochedlinger for generously agreeing to serve as the external reviewer of my thesis committee.

Finally, I would like to thank Rudolf for his tremendous support and guidance as a mentor. Without a doubt, none of this could have been achieved without him and I will be forever grateful for his willingness to standby me throughout my graduate school career.

MicroRNAs in Early Embryonic Development: Dissecting the Role of *miR-290* through *miR-295* in the Mouse

by
Lucas M. Dennis

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

Abstract

MicroRNAs mediate developmental regulation of gene expression via translational repression of target mRNAs. Targeted deletion of the miRNA biogenesis machinery in the mouse has demonstrated essential roles for miRNAs during early development. In this thesis, I have examined the role of a family of miRNAs, *miR-290* through *miR-295* (*miR-290* cluster), which are specifically expressed during early embryonic and germ cell development. This miRNA family is conserved only among mammals. *miR-290* cluster miRNAs are transcribed and processed from a common capped and polyadenylated primary transcript in the mouse. Deletion of the *miR-290* cluster in the mouse results in early embryonic lethality and misregulation of primordial germ cell migration, ultimately resulting in germ cell depletion, premature ovarian failure and infertility in the adult female. Loss of *miR-290-295* mediated repression results in significant changes in the gene expression profile of embryonic stem cells, allowing for the accumulation and precocious expression of many developmental regulators involved in differentiation. As such, we have shown that the *miR-290* cluster miRNAs are critical regulators of embryonic development.

Thesis Supervisor: Rudolf Jaenisch
Title: Professor of Biology

Table of Contents

Chapter 1:

Introduction

microRNAs in the Early Embryo: Small RNA Regulators of Development

Animal development begins with fertilization of the ovum, which initiates a cascade of gene expression changes resulting in rapid mitotic division, zygotic cleavage and ultimately cellular differentiation. For much of the past century, proteins have taken center stage as mediators of these dynamic developmental transitions through their ability to affect transcriptional output of the genome. Only since the discovery of the RNA-induced silencing phenomena, has the ability of an expanding class of small RNA species become appreciated to influence significant post-transcriptional modulation of gene expression. In particular, small, endogenous RNAs known as microRNAs (miRNAs) have been realized as key components of the animal developmental regulatory machinery through their ability to repress functional expression of target mRNAs by either repressing the translation of a target mRNA or directing mRNA degradation through targeted cleavage as part of the multi-protein RNA-induced silencing complex (RISC) (Hammond et al., 2000; Elbashir et al., 2001; Hutvágner et al., 2002).

microRNAs: Conserved small RNA regulators of development

The first miRNA, *lin-4*, was discovered in a *C. elegans* and was shown to post-transcriptionally regulate protein expression of the heterochronic gene *lin-14* via a RNA-RNA interaction between the small RNA and complementary sequences in the 3' UTR of the *lin-14* transcript (Lee et al., 1993, Wightman et al., 1993). In 2000, another *C. elegans* miRNA, *let-7*, was identified and shown to regulate developmental timing in the

nematode (Reinhart et al., 2000). However, unlike *lin-4*, which is conserved only in closely related species, *let-7* was found to be conserved and exhibit temporally controlled expression in a wide range of species, including mammals, (Pasquinelli et al., 2000). This finding demonstrated the potential for miRNA-mediated regulation of gene expression to be a pervasive phenomenon in animals.

Since then, numerous miRNAs have been cloned from several species, many of which exhibit spatially and/or temporally regulated patterns of gene expression (Lagos-Quintana et al., 2002, Aravin et al., 2003). For example, in zebrafish development, most miRNAs appear late in embryogenesis and reveal a tissue-specific distribution (Chen et al., 2005, Wienholds et al., 2005). Microarray based analysis of mouse miRNAs also demonstrate tissue specific expression of many miRNAs (Liu et al., 2004). While the founding member of this class of regulatory RNAs both were found to control developmental timing, other nematode miRNAs have been implicated in setting up neuronal asymmetry (Johnston, 2003) and chemosensory laterality (Chang, 2004). *Drosophila* miRNAs, *bantam* and *miR-14*, have been implicated in the control of cellular proliferation and reduce programmed cell death (Brennecke et al., 2003, Xu et al., 2003). *dme-miR-9a* is required for proper photoreceptor specification (Li et al., 2006) and the muscle-specific *dme-miR-1* regulates cardiac differentiation (Kwon et al., 2005). These observations have led to the notion that a principal function of miRNAs is to control cell differentiation and development.

miRNAs: Biogenesis

Mature miRNAs are an 18 to 24 nucleotide RNA species with 5'-phosphate and 3'-hydroxyl groups that are usually processed from long primary RNA (pri-miRNA) transcripts. Pri-mRNAs are either transcribed by RNA Polymerase II as an independent transcription unit or originate from spliced-out introns of a host gene (Lee et al., 2004a). Mature miRNA biogenesis begins in the nucleus where an enzymatic microprocessor complex comprised of the RNase III endonuclease *Drosha* and its counterpart *DGCR8* cleave pri-miRNAs into ~70nt precursor miRNAs (pre-miRNAs) with a characteristic stem-loop structure and 2-nucleotide 3' overhang (Lee et al., 2003, Denli et al., 2004) (Figure 1.1). MicroRNAs exhibit the highest evolutionary conservation in the stem region, with a variable degree of divergence within the loop (Lim et al., 2003a). Although conservation is markedly reduced in sequences flanking the miRNA hairpin, these regions are required for nuclear processing of the primary transcript (Chen et al., 2004).

The resulting pre-miRNAs are transported into the cytoplasm by *Exportin-5* in a Ran-GTP dependent manner (Yi et al., 2003, Lund et al., 2004). Once in the cytoplasm, another RNase III enzyme, *Dicer*, cleaves both strands of the pre-miRNA hairpin resulting in a ~22nt short RNA duplex containing the mature miRNA and the miRNA* (Hutvágner et al., 2001, Lau et al., 2001). The mature miRNA is then loaded onto a multi-protein enzyme complex, known as the RNA induced silencing complex (RISC) (Hutvágner et al., 2002). This miRNA loading is asymmetric, as the strand whose 5' end

is less tightly paired to its complement tends to be incorporated into the RISC (Khvorova et al., 2003, Schwarz et al., 2003).

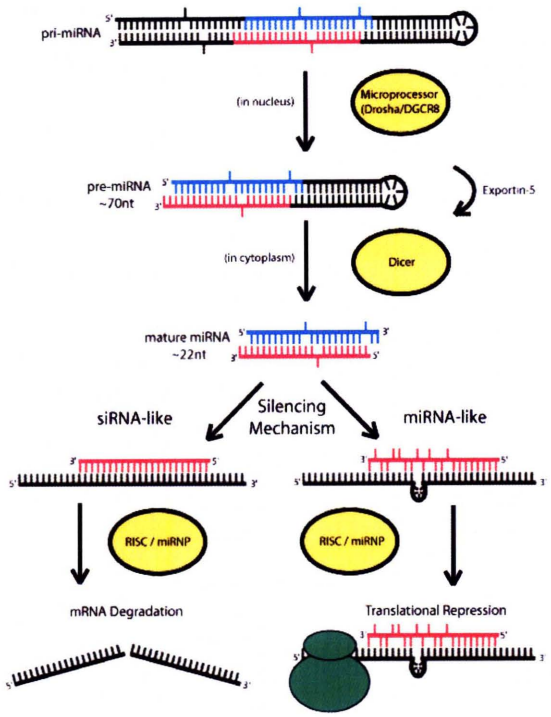

Figure 1.1. miRNA Biogenesis.

Mature miRNA-containing, or programmed, RISC promotes decreased gene expression of target mRNAs via translational repression or target mRNA degradation (Zeng et al., 2003). While several factors appear to be important for proper mRNA target recognition by miRNAs (Grimson et al., 2007), pairing between the 6-8nt "seed" sequence at the 5' end of the miRNA to a conserved site in the 3' UTR of the mRNA message is paramount (Lewis et al., 2003). Perfect or near-perfect complementarity tends to result in mRNA cleavage, while imperfect miRNA/mRNA complementarity results in translation suppression (Martinez et al., 2002, Doench et al., 2003). However, recent evidence demonstrates that at least some animal miRNAs can target mRNAs containing partially

complementary sequences in their 3' UTR for degradation (Lim et al., 2005, Yekta et al., 2004, Bagga et al., 2005). Interestingly, miRNA-mediated translational repression and/or mRNA degradation has been connected to discrete cytoplasmic foci called processing bodies (P-bodies) (Liu et al., 2005a, 2005b; Sen et al., 2005). The mRNA degradation machinery localizes to P-bodies, such that transcripts targeted by miRNAs concentrate in a miRNA-dependent manner. It is unclear whether this cytoplasmic relocation serves as a basis for translational repression. Many other mechanistic aspects of miRNA biogenesis, localization, assembly into RISC, and RISC function are also subjects of current study and await further elucidation.

Early work suggested that miRNAs commonly preserve target transcript levels and block protein translation. In one study, Lim *et al.* showed that after delivering a tissue specific miRNA into HeLa cells, a shift was observed in the expression profile toward that of tissue where the endogenous miRNA was most abundant (Lim et al., 2005). Notably, in each case, the levels of dozens of transcripts declined within 12 hours and the 3'-UTRs of affected mRNAs were enriched for sequences complementary to the 5'-ends of the cognate miRNAs. This raised the possibility that at least some of the down-regulated transcripts may have been targets for direct miRNA-mediated degradation, which was an effect distinct from the hypothesized main role of miRNAs in inhibiting protein translation. This implicated tissue specific miRNAs in aspects of gene regulation that are central to establishment and/or maintenance of cellular identity by means of regulating lineage specific gene expression.

miRNA Prediction and Target Identification

Gene-scanning algorithms developed to detect miRNA stem-loop precursors estimate that miRNAs compose approximately 1% of the coding genes within the genomes of worms, flies and mammals (Lim et al., 2003a; Lim et al., 2003b; Lai et al., 2003). Several animal genomes contain >500 genes encoding miRNAs, as well as thousands of transcripts that are potentially targets of miRNA regulation including many known developmental regulatory genes (Bartel et al., 2004a, Lim et al., 2005 and Lewis et al., 2005). Most cloned miRNAs show sequence conservation among closely related species and also across large evolutionary distances. Among the initially identified mouse and human miRNAs, over half have identifiable homologs in fish, and about half of these have homologs in flies or worms (Lagos-Quintana et al., 2003). This evolutionary conservation provides an important basis for computational prediction of miRNAs (Lim et al., 2003b). New techniques, such as deep sequencing, have also helped to confirm bioinformatic predictions and expand the list of known miRNAs (Ruby et al., 2007).

Most miRNAs share limited sequence complementarity with their mRNA targets, such that each miRNA can potentialy interact with hundreds of genes (Brennecke et al., 2005, Lewis et al., 2005). In an analysis of 3'-UTRs from genomes of human, mouse, rat and dog, Lewis *et al.* used the TargetScan algorithm to identify 13,000 putative regulatory relationships conserved among the 4 species, implicating over 5,000 human genes (30% of the tested gene set) as candidate miRNA targets in the process (Lewis et al., 2005). In the best studied examples in worms and mammals, miRNAs control key developmental

processes by inhibiting just one or a few mRNA targets. While some miRNAs may regulate a single gene, others are likely to exert a wider influence over gene expression programs associated with specific developmental states. Either way, it is likely that not all miRNAs function similarly and additional work will be needed to clarify bioinformatics predictions of miRNA targets.

Given that the typical miRNA can potentially regulate numerous genes and that most target mRNAs contain isolated miRNA recognition sites it has been proposed that miRNAs exert a selective pressure on many genes to avoid undesirable repression (Bartel et al., 2004b). Indeed, Stark and colleagues reported that genes that seem to avoid miRNA regulation, dubbed "antitargets", tend to function in processes common to all cells, whereas putative miRNA target genes are more frequently implicated in developmental processes (Stark et al., 2005). In both flies and mammals, several miRNAs targets have been shown to be either expressed in non-overlapping spatial/temporal distributions or at lower levels in tissues with cognate miRNAs (Stark et al., 2005, Farh et al., 2005). These observations suggest that miRNAs may have evolved to fine-tune gene expression programs and to eliminate gene products more rapidly than might occur by intrinsic decay. This particular role may be especially important in early embryonic development, where rapid lineage specification and differentiation must occur in a highly ordered manner.

miRNA mediated regulation of development

The maternal to zygotic transition

At fertilization the parental genomes come together in the oocyte cytoplasm to form the diploid zygote, a process marked by a number of critical gene regulatory events (Edwards, 2003). A growing mouse oocyte, arrested at diplotene of meiotic prophase, is transcriptionally and translationally active, although many of the synthesized mRNAs are not used for immediate translation, but rather are stored to support oocyte maturation and early development (Bachvarova, 1985; Wassarman et al., 1992). Oocytes arrest in metaphase of their second meiotic division, where transcription stops and translation of mRNA is reduced. Fertilization triggers the completion of meiosis, followed by the formation of a 1-cell embryo (zygote) containing haploid paternal and maternal pronuclei. Each pronucleus then undergoes DNA replication before entering the first mitosis to produce a 2-cell embryo containing two diploid zygotic nuclei.

Meiotic maturation triggers the degradation of maternal transcripts, which is ~90% complete by the 2-cell stage. Transcription from the newly formed zygotic genome, known as zygotic genome activation (ZGA) occurs in two phases: a minor activation (minor ZGA) before cleavage and a major activation (major ZGA) at the 2-cell stage. Initial zygotic transcription is weak and results in synthesis of only a small set of polypeptides that transiently increase at the 2-cell stage (Latham et al., 1991). The major ZGA promotes a dramatic reprogramming of gene expression pattern, coupled with the

generation of novel transcripts that are not expressed in oocytes. Thus, the genetic program governed by maternal transcripts/proteins must be switched to that dominated by transcripts/proteins from the newly formed zygotic genome (Kanka, 2003; Latham et al., 2001; Nothias et al., 1995; Schultz, 2002). This transition from maternally derived stores of mRNA to zygotically produced transcripts is known as the maternal to zygotic transition (MZT).

A role for miRNAs in regulation of the MZT was uncovered in 2006 when Giraldez *et al.* reported that the zebrafish *miR-430* family promotes turnover of maternal mRNAs (Giraldez et al., 2006). In previous work, they had identified miR-430 as an abundant early expressed miRNA in the developing zebrafish embryo (Giraldez et al., 2005), with subsequent cloning efforts indicating that *miR-430* was the only abundant miRNA in the first 4 to 8 hours of zebrafish development (Chen et al., 2005). *miR-430* is encoded by a large gene family (more than 90 members) that produces several, slightly different, forms of the mature miRNA. Given that these miRNAs are expressed at the same time and place and share very similar seed sequences, they are expected to have largely overlapping sets of targets. Expression of *miR-430* begins at the moment of transition from maternal to zygotic gene expression, and mature *miR-430* rapidly accumulates to high levels (Giraldez et al., 2006).

Given the complexity of the *miR-430* gene family and the fact that this miRNA was the only abundantly expressed miRNA in the early embryo, the authors produced adult female fish lacking germ-line *Dicer* activity to examine loss of maternal miR-430 in

Dicer deficient embryos. *Dicer*-free embryos exhibit subtle defects in gastrulation and brain morphogenesis, a phenotype that could mostly be suppressed by injecting the embryo with mature *miR-430* (Giraldez et al., 2005). By comparing expression profiles of *Dicer*-free embryos to those of embryos supplemented with *miR-430*, Giraldez *et al.* estimated that several hundred mRNAs were likely to be direct targets of destruction by *miR-430* in the early embryo. Target mRNA destruction was shown to be a result of *miR-430* promoted deadenylation of target mRNAs (Giraldez et al., 2006).

Conditional deletion of *Dicer* in the developing mouse oocyte confirmed that miRNA regulation of the MZT is a conserved regulatory mechanism during early embryonic development. Two groups demonstrated in 2007 that oocytes lacking *Dicer* exhibited defects in spindle organization and chromosomal alignment, thereby impairing proper meiotic maturation (Murchison et al., 2007, Tang et al., 2007). These studies also compared miRNA expression in the mature oocyte with expression in the zygote, finding little difference in the miRNA expression pattern in both of these cell types. This indicated that the miRNAs detected in the zygote are likely maternally inherited. Similarly to findings in the zebrafish, transcripts that are up-regulated in murine *Dicer* deficient oocytes are highly correlated with mRNAs that are degraded during oocyte maturation, suggesting that small RNAs may act to accelerate transcript turnover. However, in contrast to the MZT in zebrafish, small RNA regulation of maternal mRNAs during meiotic maturation in the mouse is more complex and is not primarily dependent upon the presence of a single miRNA.

Preimplantation development and pluripotent cells

Preimplantation development involves the approximately six cleavage divisions from just after the activation of the zygotic gene expression program by the two-cell stage to the formation of a blastocyst (Chazaud et al., 2006, Niwa et al., 2005 and Yamanaka et al., 2006). This critical developmental stage is unique to mammals. Blastocysts consist of approximately 60 cells, sequestered into two distinct populations: an inner cell mass (ICM) containing the primitive ectoderm cells and the outer trophectoderm cells that eventually give rise to extra-embryonic tissues. The cells of the ICM are pluripotent, i.e. able to serve as the foundation for all the tissue in adults, both somatic and germ. Following implantation, ICM cells develop to form the epiblast of the early egg cylinder, cells which are also pluripotent based upon the expression of pluripotent cell-specific genes such as Oct4. Deletion of the early embryonic transcription factors Oct4 and Nanog compromises development of the ICM and prohibits differentiation (Chambers et al., 2003, Mitsui et al., 2003 and Nichols et al., 1998). In addition to internal transcription factor cues, these pluripotent cells also respond to signals from the surrounding extra-embryonic tissues that direct efficient differentiation and initiation of gastrulation.

One of the earliest developmental events at the onset of gastrulation is the establishment of the primordial germ cells (PGCs). Germ cells are highly specialized, unipotential cells established by a specific transcriptional program that includes repression of the somatic cell fate (Surani et al., 2004). PGC specification begins as the pluripotent proximal

epiblast cells recieve signals from extraembryonic tissues and begin to express *fragilis/Ifitm3* (Saitou et al., 2002). At E6.5, within this *fragilis*-positive cell population approximately six cells in the prospective posterior proximal site of the embryo initiate expression of *Blimp1/Prdm1*, the earliest marker of the PGC cell fate (Ohinata et al., 2005). By E7.5, ~40 *Stella*-positive founder PGCs are observed at the posterior end of the primitive streak (McLaren and Lawson, 2005).

In mice, a unique germ cell–specific transcriptional network seems to regulate PGC specification. Analysis of gene expression in single cells shows the involvement of a molecular program during germ cell specification (Yabuta et al., 2006). This is the only lineage that exhibits expression of pluripotency-specific genes after gastrulation. In fact, both the epigenetic and transcriptional states of early germ cells are highly similar to the pluripotent cells of the early embryo and the end result of germ cell differentiation are cells with the potential to generate a new organism (Surani et al., 2007).

The ICM and PGCs are in turn the precursors of pluripotent embryonic stem (ES) and embryonic germ (EG) cells, respectively, which are derived and maintained only in tissue culture *in vitro* (Durcova-Hills et al., 2006, Matsui et al., 1992, Resnick et al., 1992 and Ying et al., 2003). Pluripotent stem cells have also been derived from spermatogonial stem cells, suggesting that the transcriptional network capable of supporting pluripotency may be maintained during germ cell development (Kanatsu-Shinohara et al., 2005). ES cells can exhibit a perpetual pluripotent state in vitro and are able to contribute to all cell lineages, including the germ-line (Prelle, 2002). While the pluripotent state of ES cells

likely corresponds to the transient pluripotent state of primitive ectoderm cells, it is not identical. While this limitation must be considered, the ability to study aspects of early embryogenesis such as pluripotency, differentiation and proliferation *in vitro* make ES cells an invaluable tool (Bradley, 1990).

miRNAs and early embryonic development

The overall role of miRNAs during early embryogenesis has been assessed by analyzing animals that are deficient in *Dicer*. Mutation of the C. elegans *Dicer*, *dcr-1*, results in heterochronic defects and sterility (Ketting et al., 2001, Knight et al., 2001, Grishok et al., 2001). In *Drosophila*, two Dicer isozymes exist: *Dicer-1* (*Dcr-1*), which is primarily required for miRNA biogenesis and Dicer-2, which is required for the siRNA biogenesis pathway (Lee et al., 2004b). While animals deficient for *Dicer-2* are viable and fertile (Lee et al., 2004b), work by Hatfield *et al.* suggests that *Dicer-1* is necessary for normal germline stem cell (GSC) self-renewal. *Dcr-1* mutant GSCs exhibit a delayed G1/S transition (Hatfield et al., 2005). In zebrafish, *Dicer* is also essential for normal embryonic development (Wienholds et al., 2003). *Dicer*-knockout and knockdown mutants exhibit a lack of miRNA production and arrest during embryogenesis. In agreement with studies from other model organisms, miRNA biogenesis is also required for embryogenesis in mice. Constitutive loss of *Dicer-1* leads to embryonic lethality by embryonic day (E) 7.5 (Bernstein et al., 2003), while loss of *Argonaute2* results in severely developmentally delayed embryos by E10.5 (Liu et al., 2004b).

The essential role of miRNAs in mouse early development is further supported by conditional knock-out studies of *Dicer* in ES cells. While *Dicer*-deficient ES cells are viable, they show a reduced proliferation rate and an inability to differentiate into embryoid bodies (Kanellopoulou et al., 2005). Markers of both endodermal and mesodermal differentiation were absent in these cells; whereas levels of the pluripotent marker *Oct4* remain high, even after 5 days of attempted differentiation. Additionally, *Dicer*-deficient cells exhibit alterations in DNA methylation of centromeric repeats, as well as, histone modifications implicating *Dicer*, and possibly mature miRNAs, in the maintenance of mammalian heterochromatin structure and centromeric silencing.

In a separate study, *Dicer*-deficient ES cells were found to proliferate more slowly initially, but over time, were able to recover and proliferate at rates similar to wild-type ES cells (Murchison et al., 2005). This rescue was hypothesized to be due to accumulation of secondary mutations and these *Dicer*-deficient ES cells were found to have increased G1 and G0 phases relative to their wild-type counterparts. Unlike the report from Kanellopoulou *et al.*, no apparent defect in DNA methylation was observed in the Murchison *et al. Dicer*-deficient ES cells.

While the studies of *Dicer*-conditional ES cells suggest an essential role for miRNAs in ES cell proliferation and differentiation, *Dicer*-deficient ES cells lack not only miRNAs, but also other small regulatory RNAs. In order to more directly assess the role of miRNA in ES cells, Wang *et al.* generated *DGCR8*-knockout ES cells (Wang et al., 2007). *DGCR8* is a dsRNA binding protein that constitutes the microprocessor complex

necessary for the proper maturation of pri-miRNAs to pre-miRNAs in the nucleus. Nuclear processing by *DGCR8* has been shown to be specific to miRNAs, such that loss of *DGCR8* reflects the specific loss of miRNA function (Denli et al., 2004).

DGCR8-knockout mES cells display a phenotype that is consistent with a global regulatory role for miRNAs in ES cell proliferation, cell cycle progression, and differentiation. *DGCR8*-deficient ES cells accumulate pri-miRNA transcripts and lack mature miRNAs. These ES cells exhibit a cell cycle defect, with cells accumulating in G1 and proliferating more slowly than wild-type ES cells. Although *DGCR8*-deficient cells are morphologically normal and express ES cell-specific markers like *Oct4*, they lack the ability to form embryoid bodies and have an abnormal expression profile of differentiation markers such as *FGF5*, *Hnf4* and *Sox1*. These observations are highly reminiscent of the *Dicer*-deficient ES cell phenotype, with the most significant difference being that *DGCR8*-knockout ES cells express some differentiation markers in embryoid body cultures and have a proliferation defect that is not rescued over time (Wang et al., 2007). This discrepancy between *DGCR8* and *Dicer* null phenotypes is interesting because it implies a distinct role for *Dicer* in the maintenance of normal methylation patterns that is separate from its well characterized involvement in miRNA biogenesis. Taken together, observations from both *DGCR8*- and *Dicer*-deficient ES cell lines clearly demonstrate that miRNAs are essential for ES cell differentiation.

Deletions of specific mouse miRNAs

While critical roles for miRNA biogenesis in the early embryo have been well established, roles for individual miRNAs have only recently been investigated in the mouse. Analysis of miRNA deletion mutants in the adult mouse has thus far focused on two distinct tissues: muscle and the hematopoietic lineage (see Table 1.1). As with all tissues, coordinated development is dependent upon the expression of lineage-specific genes and the suppression of genes that are specific to other cell lineages and tissues.

microRNA	Deletion phenotype	References
miR-1	Cardiac morphogenetic defects, cardiac electrophysiological defects, fatal with variable penetrance.	Zhao et al., 2007
miR-208	Absence of cardiac hypertrophy in stress conditions, failure to upregulate βMHC in stress conditions.	van Rooij et al, 2007
miR-155	Defective adaptive immunity, fibrosis and infiltration of the lung, defects in germinal centre reaction (decreased interleukin-2, interferon-γ), decreased production of immunoglobulin.	Rodriguez et al, 2007, Thai et al., 2007
miR-223	Expansion of the granulocytic compartment and spontaneous development of inflammatory lung pathology after an endotoxin challenge.	Johnnidas et al., 2008
miR-17~92	Inhibition of B cell development at the pro-B to pre-B transition and neonatal death.	Ventura et al., 2008

Table 1.1. Mouse miRNA deletion mutants

Blocking miRNA maturation specifically in the heart by deletion of *Dicer* led to heart failure at embryonic stages and poor development of the ventricular muscle (Zhao et al., 2007). *miR-1* is highly expressed in both skeletal and heart muscle. Consistent with a

crucial role for *miR-1* in the proper establishment and maintenance of these tissues, its expression is regulated by transcriptional regulators of myogenesis: *MEF* (myocyte-specific enhancer-binding factor) and *MYOD* (myoblast determination protein-1) (Zhao et al., 2005). Deletion of *mir-1-2*, one of the two genes encoding *miR-1* in the mouse genome, resulted in animals with enlarged hearts due to thickened walls (Zhao et al., 2007). Cardiomyocytes in the *mir-1-2* mutant mice failed to exit the cell cycle properly, resulting in hyperplasia. These defects led to the prenatal/early postnatal death of approximately half of the mutants. Surviving adult *mir-1-2* mutant animals displayed a complex array of electrophysiological defects that resulted in sudden death, similar to clinical observations in humans with abnormal electrocardiograph traces (Zhao et al., 2007). The electrophysiological effects of *miR-1* seem, at least in part, to be mediated by its control of the transcription factor *IRX5*, which in turn inhibits the expression of the potassium channel encoding gene, *KCND2*. Several additional genes involved in cell-cycle regulation, cardiac growth and cardiac differentiation have also been suggested as possible targets of *miR-1* (Zhao et al., 2007).

Analysis of another heart-specific miRNA, *miR-208*, demonstrated a role for miRNAs in postnatal regulation of hypertrophy. Van Rooij *et al* deleted *miR-208* and showed that mutant animals exhibited decreased contractility and expression of fast skeletal muscle-specific genes in the heart. In cardiac hypertrophy models, *mir-208*-null animals failed to show hypertrophy and induction of ß*MHC*, unlike wild-type animals. Reciprocally, overexpression of *miR-208* was sufficient to induce robust expression of ß*MHC*. These data suggested that *miR-208* aids in setting the threshold of induction of ß*MHC* in

response to stress and hypothyroidism, common causes of cardiac hypertrophy (van Rooij et al, 2007).

Recently, two groups have used genetic deletion of *mir-155* in mice to investigate its function *in vivo* during lymphogenesis (Thai et al., 2007, Rodriguez et al, 2007). Loss of *miR-155* results in a complex alteration of the immune response, as determined by tests for B-cell, T-cell and dendritic-cell function and by a failure to achieve protective immunity against a bacterial pathogen (Rodriguez et al, 2007). Furthermore, *mir-155*-null mutants showed a lung histopathology reminiscent of human autoimmune diseases, with diffuse fibrosis, increased collagen deposition and immune cells in the bronchioli (Rodriguez et al, 2007). *Mir-155*-null mutants displayed an altered equilibrium between the two classes of helper T lymphocytes, Th1 and Th2, with the balance shifted in favor of Th2 in *mir-155*-knockout mice. This result is, at least in part, explained by loss of *miR-155*-mediated inhibition of c-*MAF*, a transcription factor that promotes the expression of interleukin-4 (IL-4), one of the major outputs of Th2 cells. Additionally, B-lymphocytes in *mir-155*-null mice were decreased in the germinal centers, which are areas within lymph nodes where B lymphocytes divide, differentiate and begin immunoglobulin production (Thai et al., 2007) The phenotype of *mir-155*-null mice demonstrates a complex role for *miR-155* in various aspects of the adaptive immune response.

Loss of function studies in the adult mouse have also revealed roles for specific miRNAs in regulation of hematopoesis. The myeloid-specific microRNA-223 (miR-223)

negatively regulates progenitor proliferation and granulocyte differentiation (Johnnidas et al., 2008). *miR-223* mutant mice exhibit an expanded granulocytic compartment, resulting in an increased number of granulocyte progenitors, due at least in part to depression of the myeloid progenitor promoting transcription factor *Mef2c*. Additionally, granulocytes lacking *miR-223* are hypersensitive to activating stimuli leading *miR-223* mutant mice to spontaneously develop inflammatory lung pathology after an endotoxin challenge (Johnnidas et al., 2008). Another miRNA, *miR-17~92* is essential for B cell development. Absence of *miR-17~92* leads to increased levels of the proapoptotic protein *Bim* and inhibits B cell development at the pro-B to pre-B transition (Ventura et al., 2008). Overexpression of this miRNA cluster is also observed in human cancers and has oncogenic properties in a mouse model of B cell lymphoma (He et al., 2005, Xiao et al., 2008).

Identification of ES cell miRNAs

In 2003, Houbaviy *et al.* identified a set of ES cell-specific miRNAs via cloning of small RNAs from mouse ES cells grown in the presence of the pluripotency maintenance cytokine leukemia inhibitory factor (LIF) (Houbaviy et al., 2003). Houbaviy *et al.* cloned a total of 53 miRNAs, 15 of which were novel. Of the novel miRNAs, 6 miRNAs, *miR-290* through *miR-295*, shared near-identical 23-nucleotide sequences. This cluster of miRNAs also mapped to a single 2.2kb region of the mouse genome, suggesting that this group of miRNAs derived from a single primary transcript. EST database searches for entries homologous to the 2.2kb genomic sequence containing the miRNA cluster

identified cDNAs from only preimplantation embryos and ES cells. These observations led Houbivay *et al.* to propose that *miR-290* through *miR-295* were ES cell-specific and could potentially participate in early embryonic processes such as the maintenance of pluripotency.

In a similar study, Suh *et al.* cloned miRNAs from human ES cells, many of which were found to be specifically expressed in human ES cells and not in differentiated embryonic cells or adult human tissues (Suh et al., 2004). Of these miRNAs, *miR-296*, *miR-301*, and *miR-302*, were identical to mouse ES cell-specific miRNAs reported by Houbaviy *et al.* (Houbaviy et al., 2003). Similarly to the mouse ES-specific miRNAs, the human miRNAs were organized into clusters (Suh et al., 2004). Eight miRNA loci including the *miR-302* family and *miR-367* are located on chromosome 4, while *miR-371* through *miR-373* are clustered within a 1-kb region on chromosome 19. The *miR-371* cluster miRNAs are the human homologs of the mouse *miR-290* cluster miRNAs reported by Houbaviy *et al.* Expression of these human miRNA clusters are rapidly downregulated upon differentiation of ES cells into embryoid bodies, again suggesting a potential role in maintaining ES cell pluripotency for these miRNAs.

The *mir-290* through *miR-295* cluster

While critical roles for miRNA biogenesis in the early embryo have been established, the role of individual miRNAs during early mammalian embryogenesis is less clear. Of the numerous miRNAs cloned from mouse ES cells, the *miR-290* cluster miRNAs identified by Houbaviy *et al.* are among the most abundant (Calabrese et al., 2007). Additionally, the miRNAs from this cluster are the first embryonic miRNAs upregulated in the zygote,

with expression being initially observed in 4-cell stage embryos (Tang et al., 2007). The Surani group has also found *miR-290-295* to be expressed during embryonic germ cell development (Hayashi et al., 2008), making the expression pattern of this cluster highly similar to that of core pluripotency genes like Oct4 and Nanog. The *miR-290* cluster has been implicated in control of *de novo* DNA methylation in *Dicer* null ES cells (Sinkkonen et al., 2008, Benetti et al., 2008). All together, the expression data for *miR-290-295* suggests that these miRNAs may have a critical role in the regulation of early embryonic development.

This study seeks to specifically elucidate the function of the mammalian specific miRNAs, *miR-290-295*, via generation of a loss of function allele in the mouse. Specifically, this thesis will: 1) address the timing and location of *miR-290-295* expression. 2) characterize the phenotype of a *miR-290-295* knock-out mouse strain, 3) establish a homozygous null ES cells to test developmental potential by *in vitro* differentiation, and 4) assess molecular mechanism by examining changes in global gene expression caused by loss of the *miR-290* cluster and 5) identify potential mRNA targets of *miR-290-295* mediated repression. This information will offer insight into the scope of miRNA-mediated gene regulation in the developing mouse embryo and uncover roles in for *miR-290-295* in regulation of differentiation, pluripotency and developmental timing.

Acknowledgements

I would like to thank R. Jaenisch for critical comments.

References

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

Bachvarova, R. (1985). Gene expression during oogenesis and oocyte development in mammals. Dev Biol (N Y 1985) *1*, 453-524.

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

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

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

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. (2003). Dicer is essential for mouse development. Nature Genetics *35*, 215-217.

Bradley, A. (1990). Embryonic stem cells: proliferation and differentation. Curr. Opin. Cell Biol. *2*, 1013–1017.

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

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

Calabrese, J. M., Seila, A. C., Yeo, G. W., and Sharp, P. A. (2007). RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. Proc Natl Acad Sci U S A *104*, 18097-18102.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chang, S., Johnston, R.J., Frokjaer-Jensen, C., Lockery, S., and Hobert, O. (2004). MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature *430*, 785-789.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell *10*, 615-624.

Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83–86.

Chen, P. Y., Manninga, H., Slanchev, K., Chien, M., Russo, J. J., Ju, J., Sheridan, R., John, B., Marks, D. S., Gaidatzis, D., *et al.* (2005). The developmental miRNA profiles of zebrafish as determined by small RNA cloning. Genes Dev *19*, 1288-1293.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Doench, J.G., Peterson, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev. *17*, 438–442.

Durcova-Hills, G., Adams, I. R., Barton, S. C., Surani, M. A., and McLaren, A. (2006). The role of exogenous fibroblast growth factor-2 on the reprogramming of primordial germ cells into pluripotent stem cells. Stem Cells *24*, 1441-1449.

Edwards, R. G. (2003). Aspects of the molecular regulation of early mammalian development. Reprod Biomed Online *6*, 97-113.

Elbashir, S.M., Leneckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22- nucleotide RNAs. Genes Dev. *15*, 188–200.

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

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

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

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (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.

Hammond, S.C., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates posttranscriptional gene silencing in Drosophila cells. Nature *404*, 293–296.

Hatfield, S. D., Shcherbata, H. R., Fischer, K. A., Nakahara, K., Carthew, R. W., and Ruohola-Baker, H. (2005). Stem cell division is regulated by the microRNA pathway. Nature *435*, 974-978.

He, L., Thomson, J. M., Hemann, M. T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S. W., Hannon, G. J., and Hammond, S. M. (2005). A microRNA polycistron as a potential human oncogene. Nature *435*, 828-833.

Houbaviy, H.B., Murray, M.F., and Sharp, P.A. (2003). Embryonic stem cell-specific microRNAs. Dev. Cell *5*, 351–358.

Hutvágner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Hutvágner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056–2060.

Johnnidis, J. B., Harris, M. H., Wheeler, R. T., Stehling-Sun, S., Lam, M. H., Kirak, O., Brummelkamp, T. R., Fleming, M. D., and Camargo, F. D. (2008). Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. Nature *451*, 1125-1129.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature *426*, 845–849.

Kanatsu-Shinohara, M., Ogonuki, N., Iwano, T., Lee, J., Kazuki, Y., Inoue, K., Miki, H., Takehashi, M., Toyokuni, S., Shinkai, Y., *et al.* (2005). Genetic and epigenetic properties of mouse male germline stem cells during long-term culture. Development *132*, 4155-4163.

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

Kanka, J. (2003). Gene expression and chromatin structure in the pre-implantation embryo. Theriogenology *59*, 3-19.

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.

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

Knight, S. W., and Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269-2271.

Kwon, C., Han, Z., Olson, E. N., and Srivastava, D. (2005). MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. Proc Natl Acad Sci U S A *102*, 18986-18991.

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

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. Rna *9*, 175-179.

Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. (2003). Computational identification of Drosophila microRNA genes. Genome Biol *4:R42*, 1–20.

Latham, K. E., Garrels, J. I., Chang, C., and Solter, D. (1991). Quantitative analysis of protein synthesis in mouse embryos. I. Extensive reprogramming at the one- and two-cell stages. Development *112*, 921-932.

Latham, K. E., and Schultz, R. M. (2001). Embryonic genome activation. Front Biosci *6*, D748-759.

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

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

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. (2003). The nuclear Rnase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., and Kim, V. N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. Embo J *23*, 4051-4060.

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

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

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

Li, Y., Wang, F., Lee, J. A., and Gao, F. B. (2006). MicroRNA-9a ensures the precise specification of sensory organ precursors in Drosophila. Genes Dev *20*, 2793-2805.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003a). The microRNAs of *Caenorhabditis elegans*. Genes Dev. *17*, 991–1008.

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003b). Vertebrate microRNA genes. Science *299*, 1540.

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

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

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. (2004b). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-4141.

Liu, J., Rivas, F. V., Wohlschlegel, J., Yates, J. R., 3rd, Parker, R., and Hannon, G. J. (2005a). A role for the P-body component GW182 in microRNA function. Nat Cell Biol *7*, 1261-1266.

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

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

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563–574.

Matsui, Y., Zsebo, K., and Hogan, B. L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.

McLaren, A., and Lawson, K. A. (2005). How is the mouse germ-cell lineage established? Differentiation *73*, 435-437.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Murchison, E. P., Partridge, J. F., Tam, O. H., Cheloufi, S., and Hannon, G. J. (2005). Characterization of Dicer-deficient murine embryonic stem cells. Proc Natl Acad Sci U S A *102*, 12135-12140.

Murchison, E. P., Stein, P., Xuan, Z., Pan, H., Zhang, M. Q., Schultz, R. M., and Hannon, G. J. (2007). Critical roles for Dicer in the female germline. Genes Dev *21*, 682-693.

Nichols, J., Zevnik, B., Anastassiasdis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.

Nothias, J. Y., Majumder, S., Kaneko, K. J., and DePamphilis, M. L. (1995). Regulation of gene expression at the beginning of mammalian development. J Biol Chem *270*, 22077-22080.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S. C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., *et al.* (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207-213.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Muller, P., *et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86-89.

Prelle, K., Zink, N., and Wolf, E. (2002). Pluripotent stem cells--model of embryonic development, tool for gene targeting, and basis of cell therapy. Anat. Histol. Embryol. *31*, 169-86.

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

Resnick, J. L., Bixler, L. S., Cheng, L., and Donovan, P. J. (1992). Long-term proliferation of mouse primordial germ cells in culture. Nature *359*, 550-551.

Rodriguez, A., Vigorito, E., Clare, S., Warren, M. V., Couttet, P., Soond, D. R., van Dongen, S., Grocock, R. J., Das, P. P., Miska, E. A., *et al.* (2007). Requirement of bic/microRNA-155 for normal immune function. Science *316*, 608-611.

Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P., and Lai, E. C. (2007). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res *17*, 1850-1864.

Saitou, M., Barton, S. C., and Surani, M. A. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293-300.

Schultz, R. M. (2002). The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. Hum Reprod Update *8*, 323-331.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199–208.

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

Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila microRNA targets. PLOS Biol. *1*, E60.

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133-1146.

Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y., Kim, V.N., and Kim, K.S. (2004). Human embryonic stem cells express a unique set of microRNAs. Dev. Biol. *270*, 488-498.

Surani, M. A., Ancelin, K., Hajkova, P., Lange, U. C., Payer, B., Western, P., and Saitou, M. (2004). Mechanism of mouse germ cell specification: a genetic program regulating epigenetic reprogramming. Cold Spring Harb Symp Quant Biol *69*, 1-9.

Surani, M. A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. Cell *128*, 747-762.

Thai, T. H., Calado, D. P., Casola, S., Ansel, K. M., Xiao, C., Xue, Y., Murphy, A., Frendewey, D., Valenzuela, D., Kutok, J. L., *et al.* (2007). Regulation of the germinal center response by microRNA-155. Science *316*, 604-608.

van Rooij, E., Sutherland, L. B., Qi, X., Richardson, J. A., Hill, J., and Olson, E. N. (2007). Control of stress-dependent cardiac growth and gene expression by a microRNA. Science *316*, 575-579.

Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., Newman, J., Bronson, R. T., Crowley, D., Stone, J. R., *et al.* (2008). Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell *132*, 875-886.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nat Genet *39*, 380-385.

Wassarman, P. M., and Kinloch, R. A. (1992). Gene expression during oogenesis in mice. Mutat Res *296*, 3-15.

Wienholds, E., Koudijs, M. J., van Eeden, F. J., Cuppen, E., and Plasterk, R. H. (2003). The microRNA-producing enzyme Dicer1 is essential for zebrafish development. Nat Genet *35*, 217-218.

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

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the lin-14 3′–untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development. Genes Dev. *5*, 1813–1824.

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

Xiao, C., Srinivasan, L., Calado, D. P., Patterson, H. C., Zhang, B., Wang, J., Henderson, J. M., Kutok, J. L., and Rajewsky, K. (2008). Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat Immunol *9*, 405-414.

Xu, P., Vernooy, S. Y., Guo, M., and Hay, B. A. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.

Yabuta, Y., Kurimoto, K., Ohinata, Y., Seki, Y., and Saitou, M. (2006). Gene expression dynamics during germline specification in mice identified by quantitative single-cell gene expression profiling. Biol Reprod *75*, 705-716.

Yamanaka, Y., Ralston, A., Stephenson, R. O., and Rossant, J. (2006). Cell and molecular regulation of the mouse blastocyst. Dev Dyn *235*, 2301-2314.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

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

Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

Zhao, Y., Ransom, J. F., Li, A., Vedantham, V., von Drehle, M., Muth, A. N., Tsuchihashi, T., McManus, M. T., Schwartz, R. J., and Srivastava, D. (2007). Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. Cell *129*, 303-317.

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

Chapter 2:

Characterization of a highly variable eutherian microRNA gene

Hristo Houbaviy, Lucas Dennis, Rudolf Jaenisch, Phillip Sharp

Personal Contribution

This chapter is the result of a close collaboration with Hristo Houbaviy. It should be noted that the Early Embryonic miRNA Cluster (EEmiRC) referred to in this chapter is the mouse miRNA cluster *mIR-290-295*. The nomenclature was standardized after the preparation of this manuscript for publication in the journal *RNA*. Hristo was primarily responsible for preparation of the manuscript and performed the majority of the experiments discussed in this chapter. I prepared a plasmid construct, pESmiRC, utilized in this study as well as aided in experimental design and data interpretation.

Summary

Mouse microRNAs (miRNAs) miR-290–miR295 are encoded by a cluster of partially homologous pre-miRNA hairpins and are likely to be functionally important in embryonic stem (ES) cells and preimplantation embryos. We present evidence that a spliced, capped, and polyadenylated primary transcript spans this entire Early Embryonic microRNA Cluster (EEmiRC). Partial Drosha processing yields additional large nuclear RNA intermediates. A conserved promoter element containing a TATA-box directs EEmiRC transcription. Sequence analysis shows that the EEmiRC transcription unit is remarkably variable and can only be identified bioinformatically in placental (eutherian) mammals. Consistent with eutherian-specific function, EEmiRC is expressed in trophoblastic stem (TS) cells. When analyzing evolutionary and functional relationships, the organization of the entire miRNA loci should be considered in addition to the mature miRNA sequences. Application of this concept suggests that EEmiRC is a recently acquired rapidly evolving gene important for eutherian development.

Introduction

microRNAs (miRNAs) are a family of single-stranded 21–25-nt-long RNA species that target cognate mRNAs for degradation or translational repression (For review, see Bartel 2004). The founding members, *lin-4* and *let-7*, were identified genetically in *Caenorhabditis elegans* and are important developmental regulators (Lee et al. 1993; Reinhart et al. 2000). Numerous miRNAs were subsequently discovered in worms, flies, and human cells (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). Presently the miRNA database contains several hundred miRNAs cloned from a variety of sources (Ambros et al. 2003; Griffiths-Jones 2004). Bioinformatic analyses suggest that the majority of animal miRNAs have been discovered (Lim et al. 2003a,b) and that miRNAs regulate a wide array of biological phenomena that are probably not restricted to developmental transitions (Lewis et al. 2003, 2005).

Mature miRNAs are embedded in characteristic hairpin folds within their primary transcript (pri-miRNA) (Lee et al. 1993, 2002; Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). This feature distinguishes miRNAs from the related short interfering RNAs (siRNAs), which are double-stranded 21–25-nt species with 2-nt 3' overhangs produced from long double-stranded RNA precursors (Hamilton and Baulcombe 1999; Zamore et al. 2000; Elbashir et al. 2001).

Some miRNAs are present in the sense orientation within introns and may, therefore, be cotranscribed with protein coding mRNAs. However, since most miRNAs are located in intergenic regions or within introns but in the antisense orientation, dedicated miRNA transcription units must exist (Lagos-Quintana et al. 2001, 2002; Lau et al. 2001; Lee and

Ambros 2001; Aravin et al. 2003). Indeed, capped and polyadenylated pri-miRNAs were recently characterized and dedicated miRNA promoter regions were identified, suggesting that most if not all miRNAs are transcribed by RNA polymerase II (Bracht et al. 2004; Cai et al. 2004; Lee et al. 2004).

Drosha, a member of the RNAse III family of nucleases, excises the hairpin stem-loop from the nuclear pri-miRNA and generates one end of the mature miRNA (Lee et al. 2003). This event depends on additional protein cofactors that, together with Drosha, form the so-called Microprocessor complex (Denli et al. 2004; Gregory et al. 2004; Han et al. 2004). The ~65-nt pre-miRNA hairpin is subsequently exported to the cytoplasm by exportin-5 and Ran/GTP (Yi et al. 2003; Bohnsack et al. 2004; Lund et al. 2004). A double-stranded break within the pre-miRNA catalyzed by Dicer, another RNAse III family member, results in the generation of a transient double-stranded siRNA-like intermediate (Grishok et al. 2001; Hutvagner et al. 2001; Ketting et al. 2001). While subsequent miRNA processing has not been studied directly, it is thought to be very similar to siRNA processing (Hutvagner and Zamore 2002). A multiprotein complex measures the thermodynamic stability of the siRNA and loads the strand whose 5' end is less tightly paired to its complement into the effector RNA-induced silencing complex (RISC) (Caudy et al. 2002; Schwarz et al. 2003; Tomari et al. 2004). mRNA target recognition results in cleavage or translational repression depending on the degree of miRNA-mRNA complementarity and the type of Argonaute (PAZ/PIWI) family member present within RISC (Hutvagner and Zamore 2002; Doench et al. 2003; Liu et al. 2004; Song et al. 2004).

Previously, we identified a cluster of six pre-miRNA precursors that have similar sequences and encode a family of miRNAs that appear to be embryonic stem (ES) cell or early embryo specific by four criteria (Houbaviy et al. 2003): (1) their sequences are distinct from those of previously described miRNAs, including miRNAs cloned from adult mouse organs; (2) they cannot be detected in adult mouse organs by Northern analyses; (3) they are repressed during ES cell differentiation in vitro; and (4) all ESTs that map within the cluster are derived from ES cells or preimplantation embryos. This Early Embryonic microRNA Cluster (EEmiRC) may have a role in the maintenance of the pluripotent cell state and in the regulation of early mammalian development.

Here, we investigate the initial events that lead to EEmiRC expression. Bioinformatic analysis and ectopic overexpression identify a TATA box containing minimal promoter and a transcription unit that displays considerable variation between different mammalian orders. Northern analyses detect several large RNAs originating from EEmiRC. These RNAs are down-regulated upon ES cell differentiation, suggesting that EEmiRC expression is transcriptionally regulated. The EEmiRC pre-miRNA hairpins are transcribed within a common capped and polyadenylated primary transcript. The remaining large EEmiRC RNAs result from partial Drosha processing.

The RNA species from which the nuclease Drosha excises the pre-miRNA hairpin precursors has been designated pri-miRNA (from primary microRNA) because it was thought to be identical to the primary transcript in which the miRNAs are initially incorporated. In this study, we show that the EEmiRC primary transcript is spliced and,

therefore, it need not be the direct Drosha substrate. Thus, we use the term pri-miRNA to designate a Drosha substrate regardless of its splicing status.

Results

Comparative bioinformatics of EEmiRC

Our previous in silico analysis suggested the presence of an EEmiRC counterpart in the human genome (Houbaviy et al. 2003). Recent cloning of the corresponding microRNA homologs from human ES cells provided experimental support for this prediction (Suh et al. 2004).

While individual mouse and human EEmiRC miRNAs are sufficiently different from each other to warrant different numerical designations (Ambros et al. 2003), multiple sequence alignment reveals that they have related sequences both within each individual cluster and across species. Furthermore, sequence conservation extends beyond the mature miRNAs to the entire pre-miRNA hairpin sequences. Thus, an EEmiRC homolog is a locus that contains one or more pre-miRNA hairpins that match the multiple sequence alignment of experimentally identified (at this time mouse and human) EEmiRC pre-miRNA hairpins.

BLAST searches with the entire mouse and human EEmiRC pre-miRNAs identified at least one homolog in the chimpanzee, rat, dog, and cow genomes. Scanning the flanking genomic sequences with the program HMMER (Eddy et al. 1995) for matches to the human/mouse EEmiRC pre-miRNA consensus identified additional homologous sequences that could be folded into pre-miRNA hairpins (Fig.2.1). Systematic folding did not identify additional hairpins. Interestingly, HMMER found a seventh hairpin in the mouse locus (mm-X), which we had not identified in our original analysis.

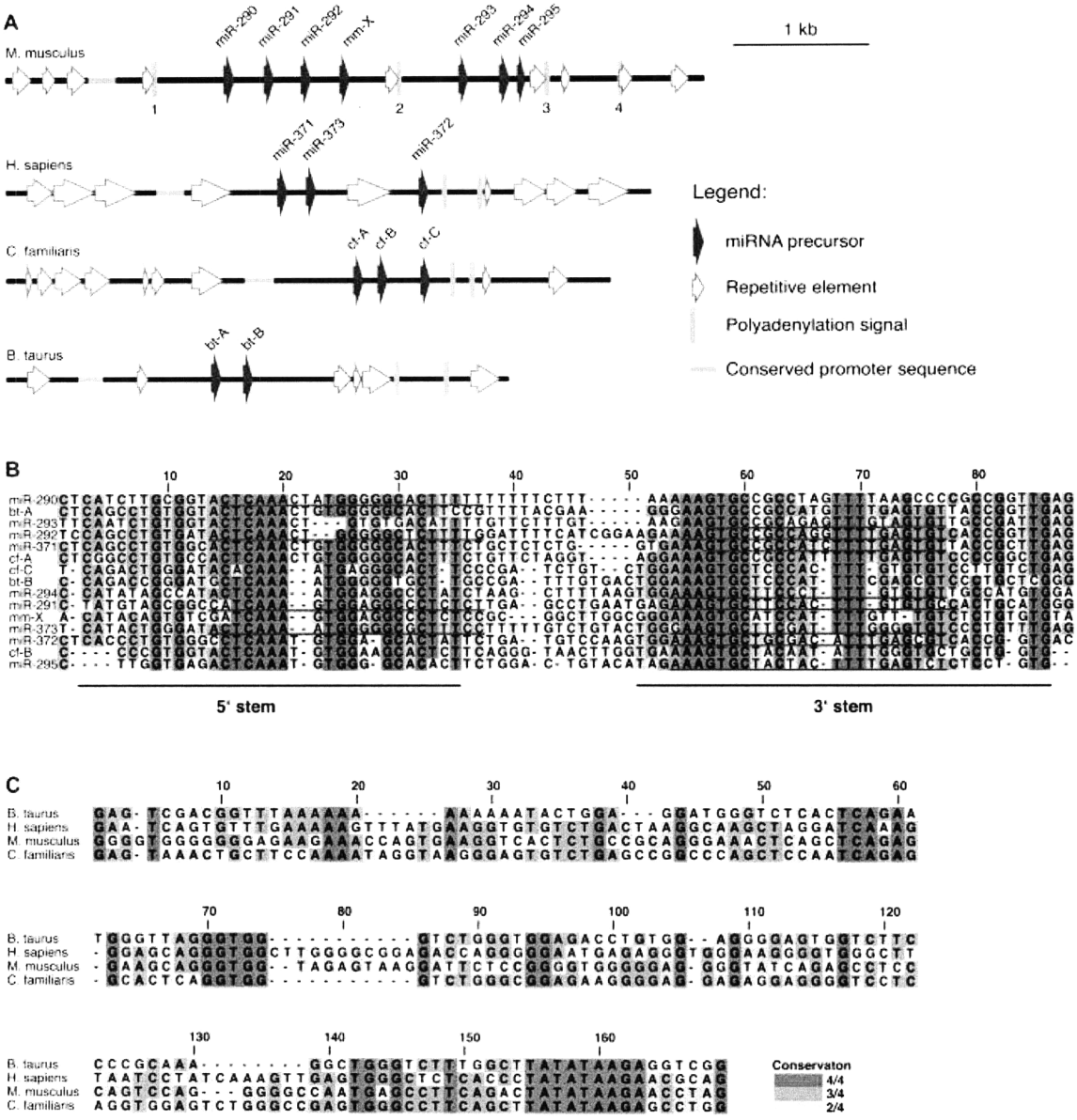

Figure 2.1. Bioinformatic analysis of EEmiRC.

(A) Schematic representation of the murine, human, canine, and bovine EEmiRC loci. Sequence features are given according to the legend. (B) Multiple sequence alignment of the murine (experimentally identified miR-290–295 and predicted mm-X), human (experimentally identified miR-371–373), canine (predicted cf-A–C), and bovine (predicted bt-A,B) pre-miRNA hairpins. The positions of cloned mature miRNAs and the two strands of the hairpin stem are underlined. (C) Multiple sequence alignment of the four conserved EEmiRC core promoter sequences.

Scanning the entire zebra fish, puffer fish, and chick genomes with HMMER failed to identify strong matches to the EEmiRC pre-miRNA consensus. HMMER also failed to identify EEmiRC paralogs within the mouse genome. Preliminary data suggest that EEmiRC is also absent in marsupials. Thus, EEmiRC is likely to be a uniquely eutherian locus present at one copy per haploid genome.

Having identified six mammalian EEmiRC loci we tried to find conservation beyond the pre-miRNA hairpins. The chimpanzee and rat EEmiRC loci are so similar to their respective human and mouse counterparts that they were excluded from further consideration so as not to bias the analysis toward the primate and rodent lineages. The remaining loci are distributed among four distinct mammalian orders (*Rodentia*, *Primates*, *Carnivora*, and *Artiodactyla*). Smith–Waterman local alignments revealed a statistically significant region of homology, ~150 bp long and positioned <1 kb upstream of the first pre-miRNA in each cluster (Fig. 2.1A,C).

The most obvious feature of this conserved sequence element is a TATATAAGA motif, which contains a canonical TATA box, centered at alignment position 159, that is conserved in all four species. Other highly conserved regions are a TCANAN(N)G motif at position 59, a GGTGG motif at position 72, and a TGNG motif at position 143, all of unknown function. The presence of a conserved TATA box, taken together with the recent evidence that RNA polymerase II is responsible for miRNA transcription (Lee et al. 2004), strongly suggests that the conserved region described above constitutes the core promoter responsible for the transcription of EEmiRC. This putative promoter element is

the only sequence motif conserved between the four EEmiRC loci, apart from the pre-miRNAs and some homologous repetitive elements.

Consistent with transcription by RNA polymerase II, there are four polyadenylation motifs in the murine EEmiRC (Fig. 2.1A, vertical bars labeled 1–4). Of the seven ESTs that map within the mouse EEmiRC (Houbaviy et al. 2003) the 3' ends of four align adjacent to polyadenylation signal 3 immediately downstream of miR-295, strongly suggesting that this is the bona fide polyadenylation site. Thus, bioinformatic analysis predicts that the mouse EEmiRC primary transcript is ~3.1 kb long. Polyadenylation signals were also found in the remaining mammalian EEmiRC loci (Fig. 2.1A).

The seed sequence, consisting of nucleotides 2–7 at the miRNA 5' end, is the major determinant of miRNA target recognition (Lewis et al. 2003, 2005). Five of the experimentally identified EEmiRC miRNAs (miR-291-as, miR-294, miR-295, miR-372, and miR-373-as) share seeds with miR-302a–d and miR-93, whereas the remaining seven (miR-292-as, miR-293, and miR-371, the 5' ends of which are shifted, and miR-290, miR-291-s, miR-292-s, and miR-373-s, which are excised from the opposite strand of the hairpin stem) have unique seeds. Together with pre-miR-367, pre-miR-302a–d form a cluster, which, like EEmiRC is expressed in ES cells (Suh et al. 2004). Neither miR-93, which is not ES cell specific, nor the cluster consisting of miR-302a–d and miR-367, can be considered EEmiRC paralogs. miR-93 is excised from the 5' stem of its pre-miRNA hairpin and the sequence of pre-miR-93 does not match the EEmiRC pre-miRNA consensus sequence. While multiple sequence alignment reveals similarities between the EEmiRC pre-miRNAs and the individual hairpins of pre-miRNAs 302a–302d, the miR-

302a–302d–367 cluster is not an EEmiRC paralog because it does not contain sequences corresponding to the EEmiRC promoter and because pre-miR-367 is not homologous to any sequences within EEmiRC.

Large RNAs from the mouse EEmiRC and identification of the pri-miRNA

To obtain experimental support for our bioinformatic predictions, we looked for large RNAs originating from EEmiRC. Northern analyses of ES cell total RNA readily detected four major large RNA species, all of which had the same polarity as the mature miRNAs (Fig. 2.2A–C, bands A–D). All RNAs were down-regulated upon differentiation of ES cells into embryoid bodies or upon induction of differentiation with retinoic acid in monolayer (Fig. 2.2A–C, cf. lanes 1 and 2 with lanes 3–5). This is consistent with the previously determined expression pattern of the mature EEmiRC miRNAs (Houbaviy et al. 2003). Band A has an apparent length consistent with that predicted for the EEmiRC pri-miRNA. The remaining bands may be EEmiRC pri-miRNAs corresponding to alternative initiation and/or termination events or they may be processing products of band A.

The EEmiRC miRNAs were initially discovered in ES cells, suggesting a function in the blastocyst inner cell mass (ICM) (Houbaviy et al. 2003). To address a potential function of EEmiRC in the other blastocyst lineage, the trophoblast, we looked for EEmiRC expression in trophoblastic stem (TS) cells (Tanaka et al. 1998). Northern analyses detected both the large EEmiRC RNA intermediates and the mature EEmiRC miRNAs in TS cells. (Fig. 2.2E,F). The latter were ~10 times more abundant in TS cells than ES cells.

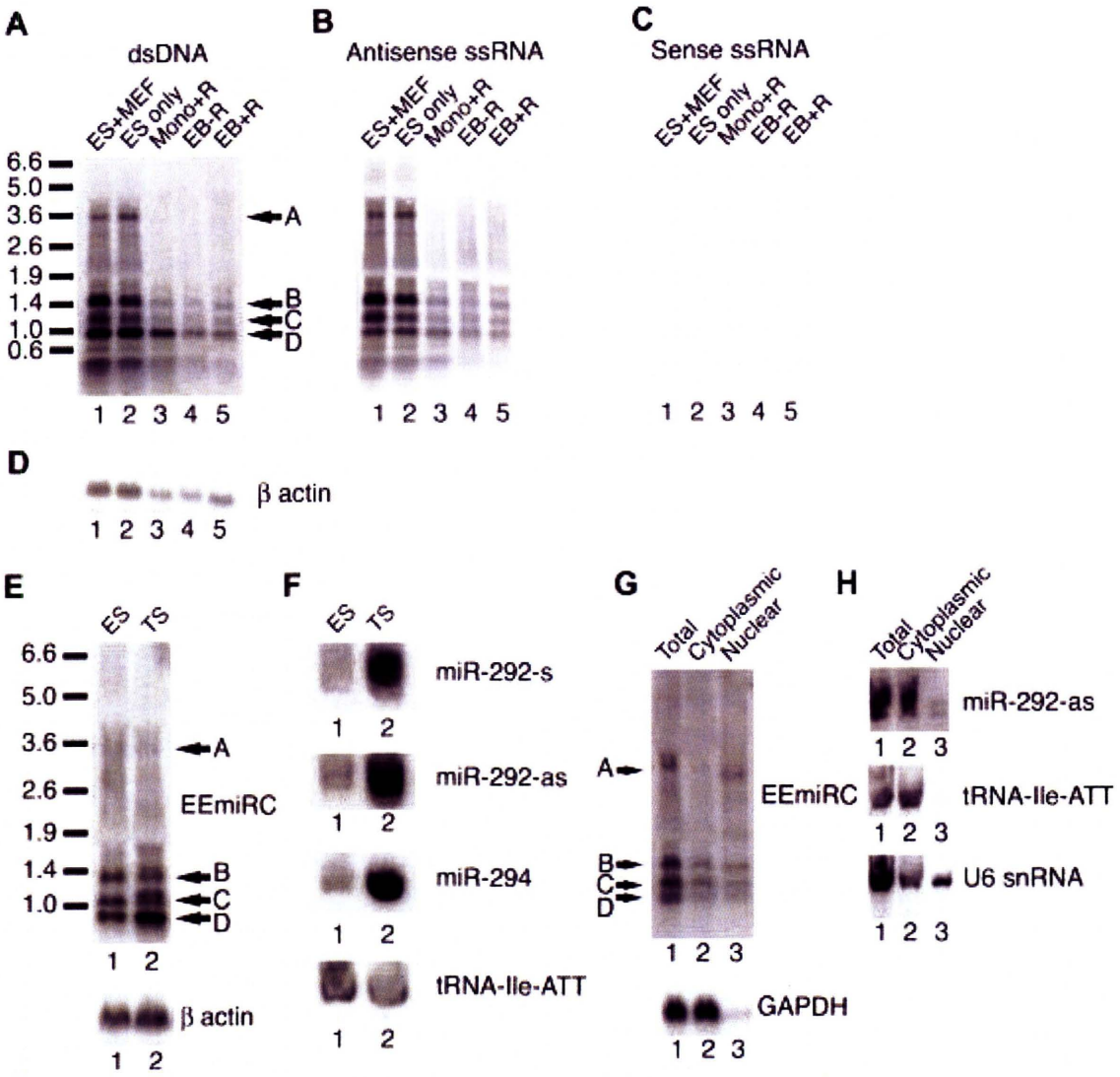

Figure 2.2. Large RNAs from the EEmiRC locus.

(A–C) Northern analysis of total RNA from ES cells grown on feeders (lane 1), ES cells grown without feeders (lane 2), ES cells differentiated with retinoic acid in monolayer (lane 3), and ES cells differentiated into embryoid bodies without (lane 4) and with (lane 5) retinoic acid. Hybridization was performed with a random primed DNA probe (A) or with single-stranded RNA probes antisense (B) or sense (C) to the mature microRNAs. The β-actin loading control is given in D. (E) Northern analysis of the large EEmiRC RNA intermediates in ES (lane 1) and TS (lane 2) cells is shown in the top panel. The β-actin mRNA (bottom panel) serves as a loading control. The corresponding analyses of the mature miR-292-s, miR-292-as, miR-294, and the tRNA-Ile-ATT loading control are shown in F. (G) Subcellular localization of the large EEmiRC RNAs (top panel) or the GAPDH mRNA control (bottom panel). Northern analysis of total (lane 1), cytoplasmic (lane 2), and nuclear (lane 3) RNA was performed as in A. The apparent change in the mobility of band A RNA in lanes 1 and 3 is due to differences in the total RNA amounts loaded. (H) Short RNA Northern analysis of the subcellular localization of mature miR-292-as (top), tRNA-Ile-ATT (middle), and U6 snRNA (bottom). Total, cytoplasmic, and nuclear RNAs are analyzed in lanes 1,2,3, respectively.

Thus, EEmiRC is probably expressed in both the epiblast and the trophoblast and may be involved in the development of the placenta.

Initial processing of pri-miRNAs into pre-miRNAs by the nuclease Drosha takes place in the nucleus (Lee et al. 2002, 2003). Thus, we sought to determine the subcellular distribution of the EEmiRC RNA species. Northern analysis of total, nuclear, and cytoplasmic ES cell RNA preparations revealed that band A was predominantly nuclear whereas bands B, C, and D were approximately evenly distributed between the nucleus and the cytoplasm (Fig. 2.2G, top panel). As expected, the mature miRNAs were predominantly cytoplasmic as were the tRNA and GAPDH controls (Fig. 2.2G,H). A significant fraction of the U6 snRNA was detected in the cytoplasm, suggesting leakage of nuclear components (Fig. 2.2H, bottom panel). Thus, the presence of bands B–D in the cytoplasm could be an artifact of the nuclear isolation procedure.

To unravel the relationship between EEmiRC RNA bands A–D we performed Northern blot hybridizations of ES cell RNA with oligonucleotide probes spanning the EEmiRC locus (Fig. 2.3A). As a negative control, total RNA from NIH/3T3 cells was used. Due to the much lower specific activity of radiolabeled oligonucleotides as compared to the random primed probes, gels had to be overloaded in order to obtain robust detection of band A (Fig. 2.3A, top panels). This band was detected by probes d–m but not by probes a–c or n and o. Thus, the 5' end of the RNA corresponding to band A is between the positions of probes c and d or immediately downstream of the conserved putative core promoter region, whereas the 3' end of band A is between probes m and n, which flank

polyadenylation signal 3. Therefore, band A most likely corresponds to the theoretically predicted EEmiRC pri-miRNA.

The relationships between EEmiRC RNAs A–D, revealed by the above analysis, are illustrated in Figure 2.3B. The mapping data suggest that RNAs B–D are processing intermediates of the RNA corresponding to band A, which is the only EEmiRC primary transcript. Band C RNA could be a product of band B RNA and it appears, at this level of resolution, that a single event produces the 3' end of band D RNA and the 5' ends of bands B and C RNAs.

It was recently shown that many pri-miRNAs are capped and polyadenylated RNA polymerase II transcripts (Lee et al. 2004). Because the EEmiRC miRNAs are flanked by a TATA-box and a polyadenylation signal we expected its pri-miRNA to be capped and polyadenylated as well. Thus, we looked for the presence of a poly(A) tail and a cap structure in the large EEmiRC RNAs. Fractionation of total ES cell RNA on an oligo(dT) column showed that the putative pri-miRNA corresponding to band A is polyadenylated whereas RNAs corresponding to bands B, C, and D are not (Fig. 2.3C).

To test for the presence of a cap structure at the 5' end of the EEmiRC pri-miRNA and to map the transcription start site at single nucleotide resolution, we used a modified 5'-RACE protocol that specifically amplifies capped RNAs. A similar procedure was used to characterize a capped plant pri-miRNA (Aukerman and Sakai 2003). To amplify capped RNAs by 5' RACE the cap structure must be converted to a 5' phosphate. This is achieved by treating the RNA with tobacco acid pyrophosphatase (TAP). When 5' RACE was performed with dephosphorylated and then TAP treated ES cell RNA a strong band

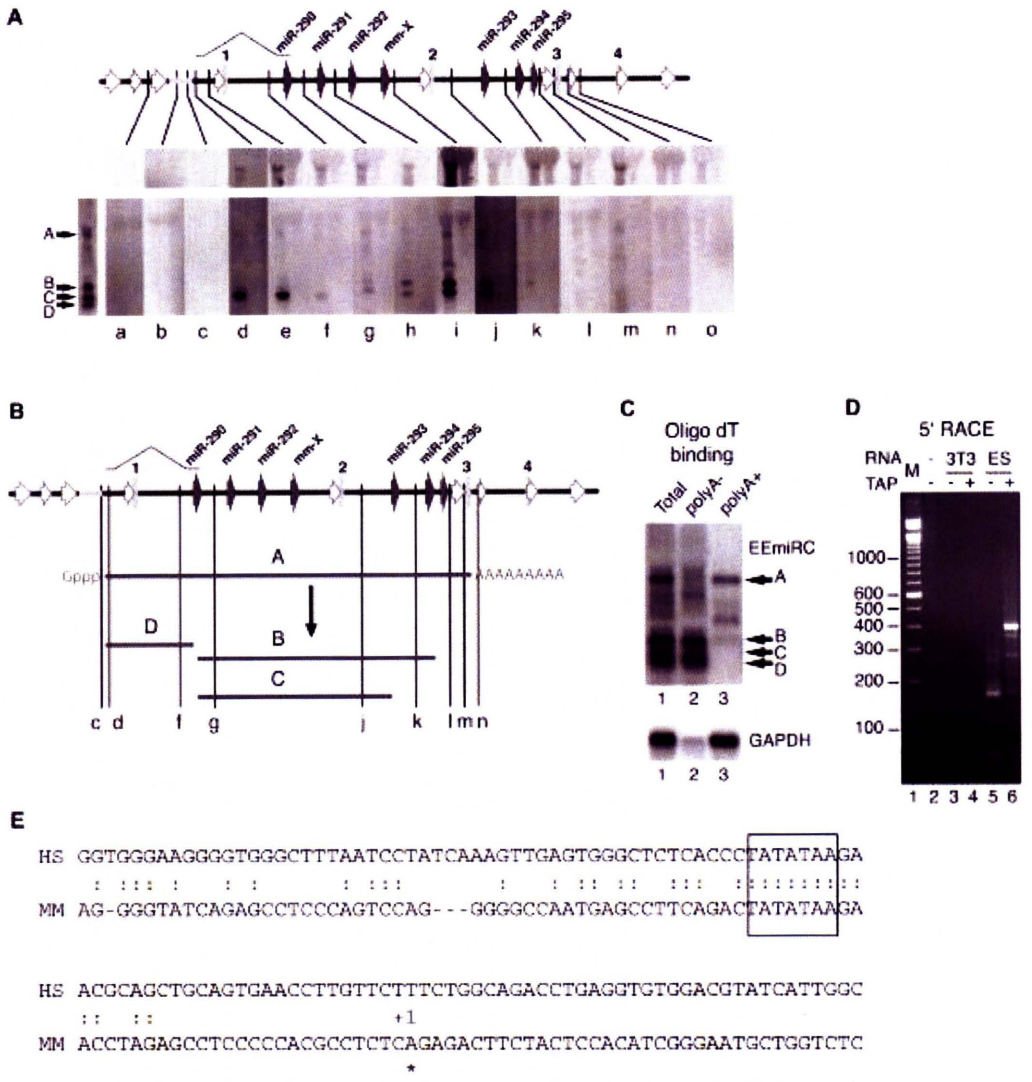

Figure 2.3. Mapping of the large EEmiRC RNAs.

(A) Mapping of the EEmiRC RNAs with oligonucleotide probes. Northern hybridizations with oligonucleotide probes (Table 1) corresponding to the indicated positions within the EEmiRC locus are shown. In each individual panel, the left lane contains total RNA from ES cells and the right lane contains total RNA from NIH/3T3 cells. Hybridization with a random primed probe and the positions of the four major RNA species are given on the left. Robust detection of band A can only be achieved after overloading the membranes (top panels). The EEmiRC locus representation is according to Figure 1A and includes a schematic of the intron preceding pre-miR-290. (B) Schematic representation of the RNA species deduced from A. (C) Polyadenylation of the EEmiRC transcripts. Northern analysis of total ES cell RNA (lane 1), ES cell RNA not bound to oligo(dT) cellulose (lane 2), and RNA bound to oligo(dT) cellulose (lane 3). Analysis of the EEmiRC transcripts and the polyadenylated GAPDH mRNA control are shown in the top and bottom panels, respectively. (D) Mapping of the 5′ end of the EEmiRC pri-miRNA. Final PCR amplification of the 5′ RACE products from NIH/3T3 (lanes 3,4) or ES (lanes 5,6) cell total RNA without (lanes 3,5) or with (lanes 4,6) TAP treatment. No 5′ RACE template was added to the PCR reaction shown in lane 2. The molecular weight marker is shown in lane 1. (E) The transcription start site (asterisk, +1) determined from the experiment shown in D is superimposed onto the alignment of the putative mouse (MM) and human (HS) EEmiRC promoter regions. The box highlights the conserved TATA element.

migrating near the 400-bp marker was observed (Fig. 2.3D, lane 6). This band constituted at least 90% of the final PCR product and was not amplified from dephosphorylated TAP treated NIH/3T3 RNA or ES cell RNA that was dephosphorylated but not treated with TAP (Fig. 2.3D, cf. lane 6 with lanes 4 and 5, respectively). The 5' RACE adaptor adds 38 bp to the final PCR products. Therefore, the 3' PCR primer maps ~360 bp downstream of the transcription start site. Sequencing of the major 5' RACE product from TAP treated ES cell RNA positioned the transcription start site at an A residue 35 bp downstream of the first T residue of the conserved TATA box (Fig. 2.3E). The 3' PCR primer spans positions +341 to +359 relative to the transcription start site, consistent with the size of the 5' RACE product. Note that 35 bp is the canonical distance between the TATA box and the transcription start site.

To consider the observed EEmiRC RNA intermediates in the context of the pri-miRNA processing pathway, we performed RNAi mediated Drosha knockdowns in ES cells. Transfection of a Drosha siRNA resulted in the dramatic up-regulation of a band migrating similar to band A, a modest increase of band B, and up-regulation of an additional minor band and a smear of material migrating between bands A and B (Fig. 2.4A). A GFP control siRNA did not change the levels of the EEmiRC RNAs (Fig. 2.4A, cf. the odd with the even lanes). The effect of the Drosha knockdown was evident as early as 24-h post-transfection, reached a maximum at 48 h, and began to decrease by 72 h. Retransfecting the cells sustained the increase in EEmiRC RNA levels over a period of 6 d. No significant changes were observed in the levels of bands C and D at any time point.

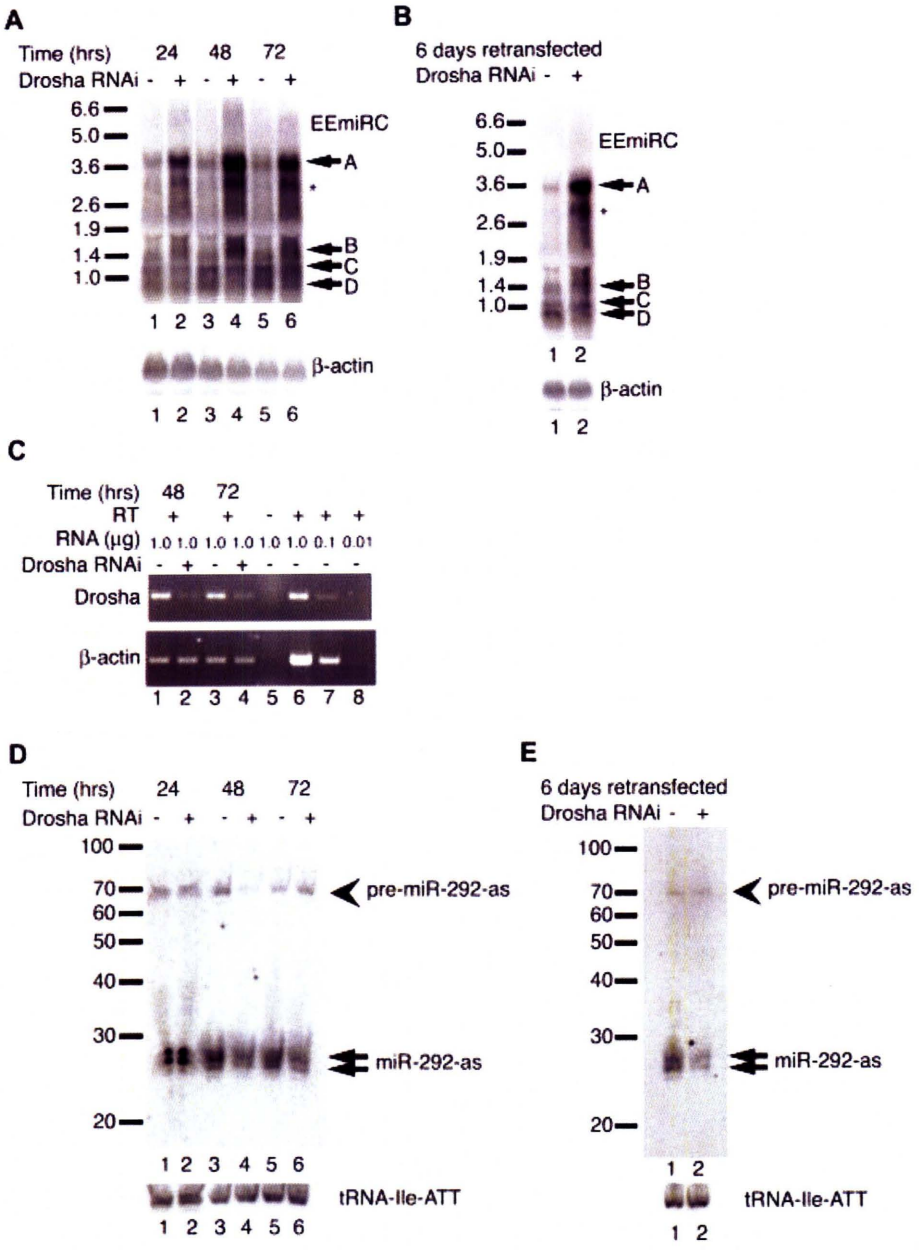

Figure 2.4. Effects of Drosha knockdown on EEmiRC RNA expression.

(A) ES cells were transfected with a Drosha siRNA (even lanes) or an EGFP control siRNA (odd lanes). Large EEmiRC RNAs (top panel) were monitored by Northern analysis at 24, 48, and 72 h post-transfection. The positions of bands A–D are indicated. The position of an additional RNA that was up-regulated in response to Drosha knockdown is shown by an asterisk. The β-actin loading control is given in the bottom panel. (B) ES cells were transfected twice with an EGFP siRNA (lane 1) or a Drosha siRNA (lane 2) and the large EEmiRC RNAs were monitored by Northern analysis at 6 d after the first transfection (top panel). The β-actin loading control is shown in the bottom panel. (C) Drosha mRNA levels (top panel) in cells transfected with an EGFP siRNA (lanes 1,3) or a Drosha siRNA (lanes 2,4) monitored by RT-PCR at 48 or 72 h post-transfection. Amplification of the β-actin mRNA is shown in the bottom panel. No reverse transcriptase was added to the reactions analyzed in lanes 5. Lanes 1–5 contained 1 μg total RNA. The titration of 1, 0.1, and 0.01 μg total RNA in lanes 6–8 shows that the assay is in the linear range. (D,E) Short RNA Northern analyses of miR-292-as and pre-miR-292-as (top panels) or tRNA-Ile-ATT (bottom panels) in the experiments shown in panels A and B, respectively.

The above experiments imply that RNAs migrating as bands A and B are Drosha substrates. Importantly, these observations reinforce our conclusion that RNA migrating as band A is an EEmiRC pri-miRNA whereas RNAs corresponding to bands B–D are processing intermediates. However, while it is not surprising that the level of band D RNA, which does not contain any pre-miRNA hairpins, did not depend on Drosha, it was surprising that the level of band C did not change since it contains at least three pre-miRNAs (Fig. 2.3A,B) and is, therefore, expected to be a Drosha substrate.

The changes in EEmiRC RNA levels correlated inversely with the levels of Drosha mRNA (Fig. 2.4C). Approximately 10% of the initial Drosha mRNA remained in cells subjected to Drosha RNAi, suggesting that some residual Drosha activity was still present. Changes in the levels of the mature miRNAs and pre-miRNAs were consistent with a role for Drosha in their production—mature miR-292-as decreased in cells transfected with Drosha siRNA and reached a minimum level in retransfected cells (Fig. 2.4D,E). Interestingly, pre-miR-292-as showed only a transient decrease at 48 h post-transfection as well as in retransfected cells. Its levels remained constant at all other time points (Fig. 2.4D,E). This behavior is consistent with reestablishment of the normal steady-state rate of pre-miRNA production at a decreased concentration of Drosha compensated by an increased concentration of Drosha substrate. Thus, the apparent insensitivity of the levels of band C RNA to the Drosha knockdown could be explained by similar kinetic effects. For example, if band C RNA is produced by Drosha cleavage of band A RNA, then a decrease in Drosha accompanied by an increase in band A RNA levels could, by saturating the available Drosha, possibly generate a constant level of band C RNA. A

similar argument may also explain why changes in the intensity of band B were much less pronounced than the increase in RNAs migrating as band A.

Ectopic expression of EEmiRC

The above results strongly suggest that the conserved sequence upstream of the EEmiRC miRNAs constitutes a core promoter. To obtain further experimental evidence for this conclusion, we investigated the activity of this DNA element in transient transfections.

The genomic fragment spanning positions –2003 to +4893 relative to the EEmiRC pri-miRNA transcription start site was subcloned into pBluescript, resulting in plasmid pEEmiRC (Fig. 2.5A). Next, the region from +790 to +3017 of pEEmiRC was replaced with a DNA fragment encoding EGFP. In the resulting plasmid, designated pEEmiRC::EGFP, all pre-miRNA hairpins are eliminated, but the putative promoter and polyadenylation sites are intact. Transfection of pEEmiRC::EGFP into ES cells resulted in reproducible expression of EGFP as measured by FACS (Fig. 2.5B, green traces). Deleting DNA between positions –199 and +29 (plas-mid pEEmiRC::EGFPΔTATA; for reference the conserved mouse sequence element shown in Fig. 1C spans positions –181 to –20) resulted in a significant decrease in EGFP expression but did not eliminate it completely (Fig. 2.5B, red traces; compare with the blue trace, which corresponds to a transfection of a LacZ expressing construct). Similar results were obtained when the plasmids were transfected in HEK-293 cells, although deletion of the putative core promoter had a less pronounced effect on EGFP expression (Fig. 2.5C).

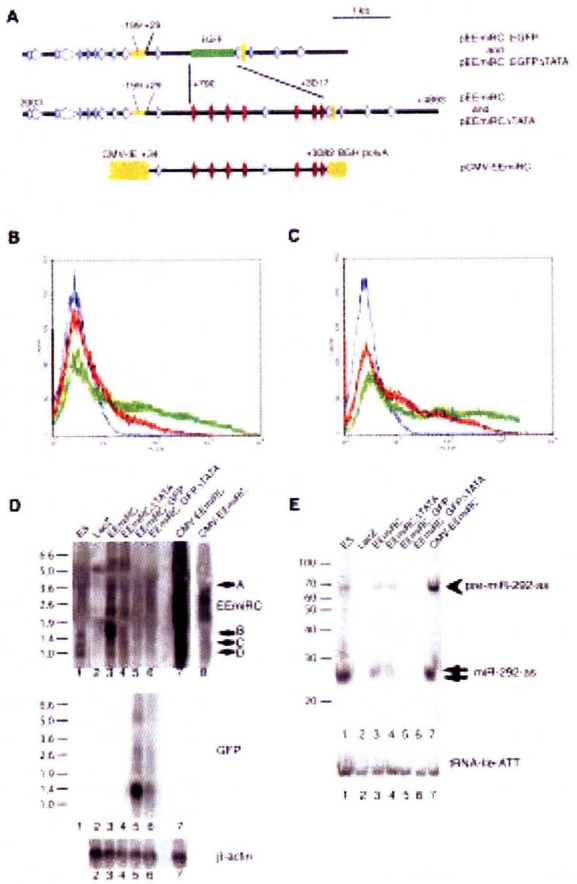

Figure 2.5. Expression of EEmiRC in transient transfections.
(A) Maps of the transfected constructs. pEEmiRC contains a genomic fragment spanning positions −2003 to +4893 relative to the EEmiRC transcription start site. In pEEmiRC::EGFP positions +790 to +3017 are replaced by the EGFP coding sequence. In plasmids pEEmiRC(ΔTATA) and pEEmiR-C::EGFP(ΔTATA) positions −199 to +29 have been deleted. In pCMV-EEmiRC a fragment spanning the region from +34 to +3082 is placed between the cytomegalovirus immediate early promoter and the bovine growth hormone polyadenylation signal. (B) EGFP expression in ES cells monitored by FACS. Cells were transfected with pEEmiRC::EGFP (green traces, superposition of three independent transfections) or pEEmiRC::EGFP(ΔTATA) (red traces, superposition of three independent transfections). The blue trace shows the background fluorescence of cells transfected with a LacZ expressing construct. (C) Transfection of HEK-293 cells as in B. (D) Northern analysis of the expression of large EEmiRC RNAs (top panel) or EGFP mRNA (middle panel) in HEK-293 cells transfected with a LacZ expressing construct, pEEmiRC, pEE-miRC(ΔTATA), pEEmiRC::EGFP, pEEmiRC::EGFP(ΔTATA), or pCMV-EEmiRC (lanes 2–7, respectively). RNA from untransfected ES cells is shown in lane 1. Lane 8 is a shorter exposure of lane 7. The β-actin loading control is given in the bottom panel. (E) The levels of miR-292-as and pre-miR-292-as (top panel) and the tRNA-Ile-ATT loading control (bottom panel) in the samples from D are monitored by short RNA Northern analysis.

Is the low level of transcription observed from pEEmiRC::EGFPΔTATA physiologically relevant? To address this question, we assayed the EGFP mRNA levels in transfected HEK-293 cells by Northern analysis. A random primed EGFP probe detected a single RNA species in cells transfected with pEEmiRC::EGFP but not in cells transfected with pEEmiRC::EGFPΔTATA or untransfected cells (Fig. 2.5D, middle panel, cf. lanes 5 and 6). This RNA migrated slightly faster than expected for the full-length EGFP mRNA (apparent size ~1400 nt; predicted size 1678 nt). An identical result was obtained in ES cells (data not shown). Failure to detect defined transcripts in cells transfected with pEEmiRC::EGFPΔTATA argues that the low level of EGFP expression may be an artifact due to nonspecific transcription initiation and termination at sites not normally used in vivo. We note that several repetitive elements are present in pEEmiRC::EGFPΔ TATA. Some of them are LINEs and, therefore, potentially contain RNA polymerase II promoters. The detection of EEmiRC promoter activity in transiently transfected HEK-293 cells suggests that ES cell specific factors are not required for basal transcription of EEmiRC and that silencing of EEmiRC expression in somatic tissue may occur via heterochromatinization of the locus.

The endogenous EEmiRC locus is not expressed in HEK-293 cells (Fig. 2.5D, top panel, lane 2, and Fig. 2.5E, lane 2). Transfection of pEEmiRC into HEK-293 cells resulted in the production of all four large EEmiRC RNAs (bands A–D) similar to those observed in ES cells plus additional RNA species (Fig. 2.5D, top panel, cf. lane 3 to lane 1). Interestingly, the abundance of band B relative to bands A, C, and D in transfected HEK-293 cells was much higher than in untransfected ES cells. Deletion of the putative

promoter element (pEEmiRCDTATA, deletion identical to pEEmiRC::EGFPΔTATA) abolished expression of all large EEmiRC RNAs (Fig. 2.5, top panel, lane 4).

Consistent with expression of the large EEmiRC RNAs, both miR-292-as and pre-miR-292-as were detected by short RNA Northern analysis in HEK-293 cells transfected with pEEmiRC (Fig. 5E). Deletion of the putative promoter element, however, had almost no effect on the levels of miR-292-as and pre-miR-292-as. Again, we attribute this to nonspecific transcription that generates heterogeneous transcripts that direct the production of mature miRNAs. Ectopic expression from pEEmiRC in HEK-293 cells resulted in miRNA and pre-miRNA levels similar to the endogenous levels in ES cells. Furthermore, EEmiRC expression from the strong CMV promoter (plasmid pCMV-EEmiRC, Fig. 2.5A) caused only a modest increase in miRNA and pre-miRNA production (Fig. 2.5E, cf. lanes 7 and 3) even though the levels of some large EEmiRC RNAs increased by at least one order of magnitude (Fig. 2.5D, top panel, lane 7). Thus, in transiently transfected HEK-293 cells, the EEmiRC miRNA synthesis pathway is probably at or close to saturation, regardless of promoter strength. Even the low-level nonspecific transcription from pEEmiRCDTATA may be enough to saturate the miRNA synthesis machinery, which would explain why deletion of the putative promoter had little effect on the miRNA and pre-miRNA levels.

Neither band D RNA nor the EGFP mRNA could be detected by an EEmiRC probe in cells transfected with pEEmiRC::EGFP even though the primary EGFP transcript should contain sequences complementary to the probe (Fig. 2.5D, top panel, lane 5). Thus, consistent with the detection of a single mRNA band by the EGFP probe (Fig. 2.5D,

middle panel), the proposed single cleavage that generates the 3' end of band D and the 5' ends of bands B and C in the native EEmiRC pri-miRNA does not occur when the pre-miRNA hairpins are replaced by EGFP. Furthermore, the 1.4-kb EGFP mRNA expressed from pEEmiRC::EGFP is probably not colinear with the plasmid sequence from the initiation to the polyadenylation site.

Splicing of the EEmiRC pri-miRNA and Drosha processing at pre-miR-290

The simplest explanation for the above discrepancy is the presence of an intron spanning most of the primary transcript sequence between the transcription start site and pre-miR-290. To obtain evidence for splicing of the native EEmiRC pri-miRNA, we performed RNAse protection assays with RNA probes that spanned the expected splice junctions (Fig. 2.6, probes rpA and prB, respectively).

Two RNAse resistant fragments, designated pA1 and pA2, were obtained when a probe spanning the transcription start site (rpA) was hybridized with total ES cell RNA. (Fig. 2.6B). No protected bands were seen when NIH/3T3 cell RNA was used in the assay, confirming that both protected fragments were EEmiRC specific (data not shown). Probe rpA extends from –125 to +195 relative to the transcription start site. Thus, pA1, which is ~190 nt long, likely corresponds to the 5' end of the primary transcript. Similarly, pA2 (~82 nt) probably corresponds to an exon extending from +1 to ~+82 relative to the transcription start site. Interestingly, Drosha knockdown by RNAi caused a significant increase in pA2 but only a modest increase in pA1 (Fig. 2.6B cf. lanes 2 and 3). Thus, it appears that the spliced pri-miRNA is up-regulated upon Drosha depletion, and this is

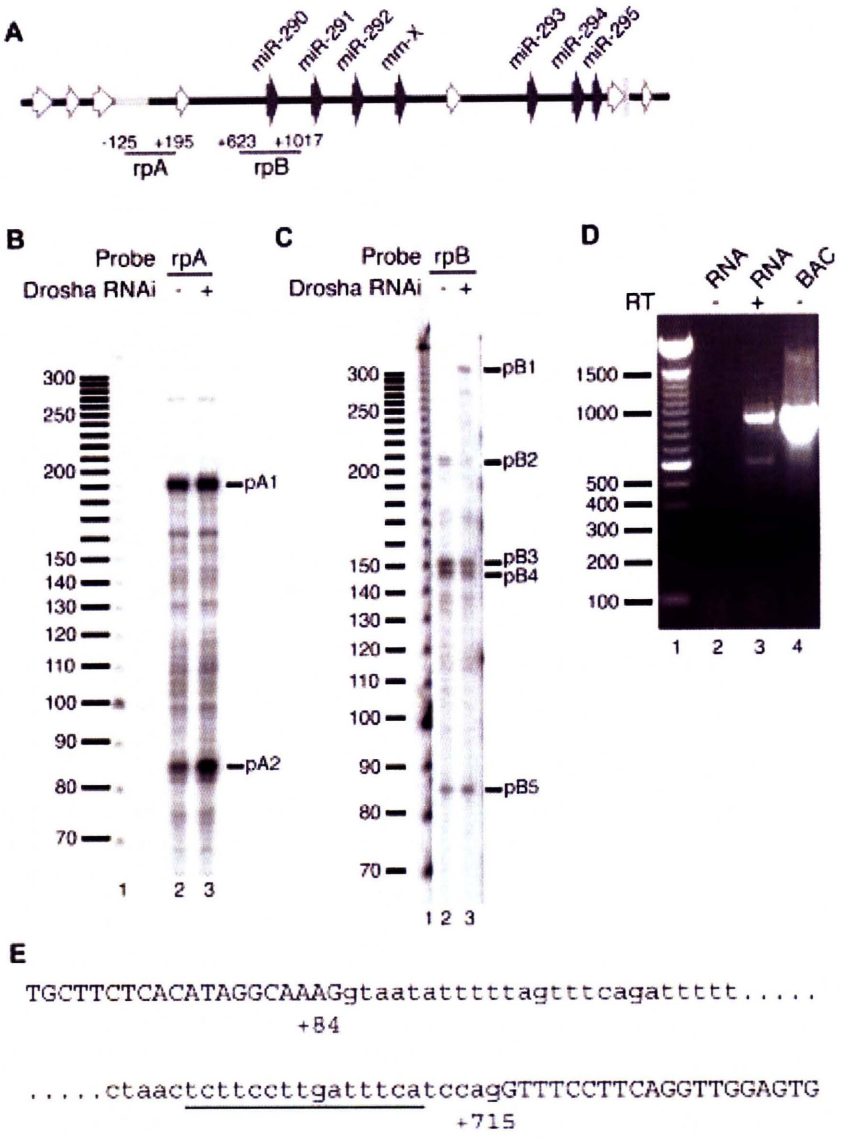

Figure 2.6. Splicing of the EEmiRC primary transcript and Drosha processing at pre-miR-290.

(A) The positions of the RNAse protection probes (rpA and rpB) superimposed onto the schematic representation of the EEmiRC locus as in Figure 1. (B,C) Total RNA from ES cells transfected with an EGFP control siRNA (lane 2) or a Drosha siRNA (lane 3) was hybridized with probe rpA (B) or rpB (C) and subjected to RNAse protection analysis. The molecular weight marker is shown in lane 1. (D) The RT-PCR products obtained from total ES cell RNA with probes flanking the predicted intron are shown in lane 3. The reaction analyzed in lane 2 did not contain reverse transcriptase. The molecular weight marker and PCR of the genomic EEmiRC BAC are shown in lanes 1 and 4, respectively. (E) Positions of the splice sites deduced from D. Exon sequences are given in capital letters; the intron is in lowercase. The polypyrimidine tract is underlined.

probably the band A RNA that accumulated in Figure 2.4A,B after treatment with Drosha siRNA.

Probe prB, which spans the region between positions +623 and +1017 relative to the transcription start site, yielded four protected fragments (pB2-pB5) when hybridized to ES cell total RNA (Fig. 2.6C, lane 2). Drosha knockdown resulted in a decrease of protected fragments pB2–pB4 and in the appearance of an additional fragment designated pB1 (~310 nt; Fig. 2.6C, cf. lanes 2 and 3). The level of fragment pB5 (~85 nt) remained unchanged. Thus, it is likely that pB2–pB4 are generated by Drosha cleavages. Indeed, the sizes of pB3 (~155 nt) and pB4 (~150 nt) are consistent with RNA species whose 5' ends are within the 3' stem of the pre-miR-290 hairpin, and extend from approximately position +867 to the 5' end of probe rpB at position +1017. Likewise, the length of pB2 (~210 nt) implies that it has a 3' end resulting from cleavage within the 5' stem of the pre-miR-290 hairpin. This fragment would then correspond to a protected RNA species between positions +623 and +828. Thus, the apparent single event that was postulated to produce bands B, C, and D from band A (Fig. 2.3A,B) is, in fact, Drosha processing of the pre-miR-290 hairpin. This conclusion is further supported by the fact that band D was not detected in HEK-293 cells transfected with pEEmiRC::EGFP, which lacks all pre-miRNA hairpins (cf. lanes 5 and 6 in Fig. 2.5, top and middle panels).

Consistent with the up-regulation of protected fragment pA2, fragment pB1, which is also upregulated upon Drosha knockdown, likely corresponds to a 5' exon junction around position +707 and extends to the 5' end of the probe at position +1014. The nature of fragment pB5 is presently unclear. Since its length and the length of fragment pB1 add up

approximately to the size of probe rpB (310 + 85 = 395 nt vs. 356 nt), it may represent the predicted intron, consistent with the constant levels of pB5 upon knockdown of Drosha.

The above data strongly suggest the presence of an intron between positions +82 and +707 of the EEmiRC primary transcript. Thus, we performed RT-PCR with primers flanking these positions (Fig. 2.6D). Three DNA species with apparent lengths of ~300 bp, 600 bp, and 950 bp were amplified from total ES cell RNA in a complete RT-PCR reaction but not when reverse transcriptase was omitted (Fig. 2.6, cf. lanes 2 and 3). Only the 950-bp band was amplified in a PCR reaction that used the genomic DNA as a template (Fig. 2.6, lane 4). This is consistent with the 950-bp cDNA product representing the unspliced primary EEmiRC transcript. Both the 950-bp and 300-bp RT-PCR products could be reamplified by PCR whereas reamplification of the 600-bp band yielded a mixture of the 950-bp, 600-bp, and 300-bp products. Thus, the 300-bp band likely corresponds to the spliced EEmiRC pri-miRNA whereas the 600-bp band is probably a heteroduplex of spliced and unspliced cDNA. Sequencing of the 300-bp cDNA product and comparison with the genomic EEmiRC sequence revealed a splice junction of positions +84 and +715 relative to the transcription start site (Fig. 2.6E). This is consistent, within experimental error, with the predictions of the RNAse protection mapping.

Note that band D is unlikely to correspond to the excised intron RNA, since its production depends on the EEmiRC hairpins (Fig. 2.5D) and since it can be detected by oligonucleotide probe d, which maps within the first EEmiRC exon (Fig. 2.3A).

Discussion

EEmiRC shows a remarkable degree of evolutionary variation. The only conserved regions within the locus are the pre-miRNA hairpins and the putative minimal promoter. The number and precise sequences of the pre-miRNAs, their distance from the promoter and polyadenylation sites, the regions flanking the hairpins, and the types, positions, and numbers of repetitive element insertions vary in species belonging to different mammalian orders. Such variation is not a general feature of miRNA loci. The pre-miR-302a–d hairpins, which, as pointed out above, are related to the EEmiRC hairpins, show a much greater degree of sequence conservation than the pre-miRNAs within EEmiRC (data not shown). The number and spacing of the individual pre-miR-302 hairpins is conserved, and homology beyond the pre-miRNAs can be easily detected between the chick and mouse loci (data not shown). The microRNA cluster miR-17/91-miR-18-miR-19a-miR-20-miR-19b-1-miR-92 is also very well conserved in mammals (Tanzer and Stadler 2004). Sequences up to 200 nt upstream of most *C. elegans* pre-miRNA hairpins share significant homology with their *C. briggsae* counterparts (Ohler et al. 2004).

The common seed shared between some EEmiRC miRNAs, miR-302a–d and miR-93, implies regulation of common targets. Our analysis suggests that these three classes of miRNAs have evolved independently and might have both overlapping and nonoverlapping functions. For example, they may regulate some common targets in different developmental contexts. The variability exhibited by EEmiRC suggests that this gene may have appeared relatively recently in the mammalian lineage and may have mammalian-specific functions. Together with previous evidence on its expression in ES

cells and preimplantation embryos, the finding that EEmiRC is expressed in the trophoblast supports the above conclusion.

Several recent studies have reported the characterization of miRNA transcription units (Bracht et al. 2004; Cai et al. 2004; Lee et al. 2004). However, because of their low abundance, the detection and characterization of pri-miRNAs relied extensively on Drosha knockdowns and/or RT-PCR amplification of fractionated RNA species. In contrast, the large EEmiRC RNA precursors reported here are relatively abundant and can be easily detected by standard Northern analyses. Thus, we have analyzed the nucleocytoplasmic distribution, polyadenylation status, and origin of the different EEmiRC RNA species directly. Experimental mapping of the EEmiRC primary transcript agrees with the results of comparative bioinformatic analysis of the homologous EEmiRC loci. The conserved region upstream of the mapped transcription start site contains a canonical TATA-box and likely represents the core EEmiRC promoter since it showed promoter activity that correlated with the presence of the TATA motif.

The bulk of the long RNAs synthesized from the EEmiRC locus consists of cleavage products of the pri-miRNA. Oligonucleotide probe hybridizations suggest that the molar ratio of the pri-miRNA to the shorter RNA species is <5%. The observed EEmiRC pri-miRNA processing products are clearly not an artifact of the RNA isolation procedure. The pri-miRNA is converted into well-defined species that can be resolved on formaldehyde-aga-rose gels and detected by Northern hybridizations. The same pattern of large EEmiRC RNAs is observed in independent Northern analyses of different ES cell

RNA preparations as well as in TS cells and can be qualitatively reproduced upon overexpression from transiently transfected constructs in HEK-293 cells.

How do the RNAs corresponding to bands B–D fit within an EEmiRC miRNA processing pathway? The nuclease Drosha is thought to excise the pre-miRNA stem-loop precursor from the pri-miRNA (Lee et al. 2003). The flanking RNA fragments are stable byproducts of this process (Han et al. 2004). Thus, Drosha cleavage at only some of the seven EEmiRC hairpins would result in partial cleavage of the pri-miRNA and the generation of long RNA intermediates such as the band B–D RNAs. Our observations support this scenario. First, the appearance of an additional EEmiRC RNA band, migrating between bands A and B, upon Drosha knockdown is consistent with partial cleavage. Second, RNAse protection analysis of the region surrounding pre-miR-290 detects only Drosha cleavages. Thus, the 3' end of band D RNA and the 5' ends of band B and C RNAs are likely due to Drosha processing of the pre-miR-290 hairpin. Similarly, Drosha processing near pre-miR-293 and pre-miR-294 probably generates the 3' ends of RNA species corresponding to band B and C RNAs respectively. Consistent with its being a Drosha reaction byproduct, band D RNA was no longer generated when the EEmiRC pre-miRNA hairpins were replaced by the EGFP coding sequence.

RT-PCR and RNAse protection assays showed that the EEmiRC primary transcript is spliced between positions +82 and +715. The changes in the RNAse protection patterns upon Drosha knockdown suggest that the spliced pri-miRNA is preferentially cleaved by Drosha. The unspliced primary transcript increased only slightly when Drosha was depleted, whereas RNAs containing the spliced first and second exons increased

dramatically. Thus, splicing may be required for, or may enhance, Drosha processing. Indeed, *trans*-splicing of the let-7 pri-miRNA appears to be required for let-7 expression (Bracht et al. 2004). However, splicing is not absolutely necessary for Drosha processing of the EEmiRC pri-miRNA since band D RNA is clearly produced in a Drosha-dependent manner from the unspliced primary transcript. Further studies are needed to elucidate the role of splicing in EEmiRC miRNA expression.

Attempts to assign function to miRNAs have focused mainly on the analysis of the mature miRNA sequences. Here, we show that additional useful information can be obtained by studying the organization of the entire miRNA genes. The remarkable variability of EEmiRC and its presence only in placental mammals reinforce previous conclusions that EEmiRC is likely to have interesting biology (Houbaviy et al. 2003; Suh et al. 2004). Definition of the EEmiRC transcription unit and the characterization of large primary transcript processing intermediates provide a foundation for understanding how this miRNA cluster is regulated during development.

Material and Methods

Bioinformatics

Genomic data were downloaded from the ENSEMBL server (http://www.ensembl.org). Multiple sequence alignments, Smith–Waterman local alignments, HMM searches, and RNA folding were performed with the MacOS X implementations of CLUS-TALW, SSEARCH, HMMER, and MFOLD (Higgins and Sharp 1988; Pearson 1991; Eddy et al. 1995; Zuker et al. 1999).

Cell culture, plasmids, and transfections

Fragments of the EEmiRC locus were PCR amplified from mouse genomic DNA. Detailed information on plasmid construction is available upon request. Synthetic siRNAs against GFP [(NN) GCACCAUCUUCUUCAAGGA(GC)] and Drosha [(NN)CAAC AGUCAUAGAAUAUGA(GC)] were obtained from Dharmacon (option A2) and processed as suggested by the manufacturer. ES cells were differentiated as described previously (Wutz and Jaenisch 2000). Plasmid DNA and siRNAs were transfected into ES cells and HEK-293 cells with Lipofectamine 2000 (Invitrogen) and Fugene 6 (Roche), respectively, as suggested by the manufacturers.

RNA analysis

Total RNA was extracted as described previously (Houbaviy et al. 2003). To prepare nuclear and cytoplasmic RNA, ES cells were lysed in 50 mM HEPES (pH 7.5), 15 mM NaCl, 60 mM KCl, 0.34 M sucrose, 0.5 mM spermidine-HCl, 0.15 mM spermine-HCl, 10 mM DTT, 5 mM EDTA, 0.5 mM EGTA, 1000 u/mL RNAsin, 0.5% Nonidet P40, and the

nuclei were separated from the cytoplasm by centrifugation for 5 min at 500*g*. RT-PCR of Drosha mRNA was done with primers 5'-ACTCGGAGGTGTTCGATGTC-3' and 5'-CATGTTGGCAATCTCCTCCT-3'. The ß-actin control primers were 5'-TGTTACCAACTGGGACGACA-3' and 5'-AAGGAAGGCT GGAAAAGAGC-3'. RT-PCR mapping of the EEmiRC intron was performed with primers 5'-TCTGCGGTCTTCAGGGATAC-3' and 5'-GTACTCACCACGCTGCAGTT-3'. Random primed EEmiRC DNA hybridization probes were synthesized from a PCR template amplified with primers 5'-TCTGCGGTCTTCAGGGATAC-3' and 5'-TCCAGGAAACCTTCATCTGG-3' from genomic DNA. The same primers, but with appended T3 or T7 phage promoter sequences, were used to amplify the in vitro transcription templates from which single-stranded RNA probes were synthesized with the MAXIscript kit (Ambion). Hybridizations with the above probes were performed in the presence of 10 µg/mL mouse Cot-1 DNA (Invitrogen) to suppress repetitive sequences. A random primed DNA probe that did not contain repetitive elements was synthesized from a mixture of PCR products amplified with primer pairs 5'-CGGTTTGGCTGGGTTTACTA-3'/5'-TAGACTCACCACCCCTGG AC-3' and 5'-GTTGGACTGATGGTTGTGAGTC-3'/5'-GAAAGCA GCCGACCTGTG-3'. Starfire probes (Integrated DNA Technologies) with the sequences shown in Table 1 were used for oligonucleotide Northern hybridization mapping of EEmiRC RNAs. GAPDH and ß-actin random primed probes were synthesized from commercially available templates (DECAtemplate, Ambion). The random primed GFP probe was synthesized from a XhoI–NotI fragment of pEGFP-N1 (Clontech). Short RNA Northern analyses were done as described previously (Hamilton and Baulcombe 1999; Houbaviy et al. 2003). The U6

oligonucleotide probe had the sequence 5'-GGGCCATGC TAATCTTCTCTGT-3'. RNAse protection probes were synthesized by in vitro transcription of templates amplified from genomic DNA with the primers shown in Table 1. T7 promoter sequences were appended to the reverse primers. 5' RACE was performed with the FirstChoice RLM-RACE kit (Ambion). The inner and outer gene-specific PCR primers were 5'-GAGCGAGGAAGGCTGAGTT-3' and 5'-ACATAGGCTCGTTCCTCCCT-3', respectively. Oligo-dT RNA fractionation was performed with the MicroPoly(A) purist kit (Ambion).

Acknowledgements

We thank Zhongde Wang for providing TS cells, John Doench for advice on the manuscript, and members of the Sharp and Jaensich laboratories for discussions.

This work was supported by United States Public Health Service MERIT Award R37-GM34277 from the National Institutes of Health, PO1-CA42063 and U19 AI056900 from the National Cancer Institute to PAS, and partially by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute.

References

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. 2003. A uniform system for microRNA annotation. *RNA* **9:** 277–279.

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

Aukerman, M.J. and Sakai, H. 2003. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15:** 2730–2741.

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

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10:** 185–191.

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

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

Caudy, A.A., Myers, M., Hannon, G.J., and Hammond, S.M. 2002. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes & Dev.* **16:** 2491–2496.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432:** 231–235.

Doench, J.G., Petersen, C.P., and Sharp, P.A. 2003. siRNAs can function as miRNAs. *Genes & Dev.* **17:** 438–442.

Eddy, S.R., Mitchison, G., and Durbin, R. 1995. Maximum discrimination hidden Markov models of sequence consensus. *J. Comput. Biol.* **2:** 9–23.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. 2001. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Dev.* **15:** 188–200.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432:** 235–240.

Griffiths-Jones, S. 2004. The microRNA registry. *Nucleic Acids Res.* **32:** D109–111.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. 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.

Hamilton, A.J. and Baulcombe, D.C. 1999. A species of small anti-sense RNA in posttranscriptional gene silencing in plants. *Science* **286:** 950–952.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes & Dev.* **18:** 3016–3027.

Higgins, D.G. and Sharp, P.M. 1988. CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73:** 237–244.

Houbaviy, H.B., Murray, M.F., and Sharp, P.A. 2003. Embryonic stem cell-specific microRNAs. *Dev. Cell* **5:** 351–358.

Hutvagner, G. and Zamore, P.D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297:** 2056–2060.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293:** 834–838.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans. Genes & Dev.* **15:** 2654–2659.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294:** 853–858.

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

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

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

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

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. 2002. MicroRNA maturation: Stepwise processing and subcellular localization. *EMBO J.* **21:** 4663–4670.

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.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23:** 4051–4060.

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

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

Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. 2003a. Vertebrate microRNA genes. *Science* **299:** 1540.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. 2003b. The micro-RNAs of *Caenorhabditis elegans. Genes & Dev.* **17:** 991–1008.

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.

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

Ohler, U., Yekta, S., Lim, L.P., Bartel, D.P., and Burge, C.B. 2004. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA* **10:** 1309–1322.

Pearson, W.R. 1991. Searching protein sequence libraries: Comparison of the sensitivity and selectivity of the Smith–Waterman and FASTA algorithms. *Genomics* **11:** 635–650.

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

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115:** 199–208.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305:** 1434–1437.

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

Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. 1998. Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282:** 2072–2075.

Tanzer, A. and Stadler, P.F. 2004. Molecular evolution of a microRNA cluster. *J. Mol. Biol.* **339:** 327–335.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P.D. 2004. A protein sensor for siRNA asymmetry. *Science* **306:** 1377–1380.

Wutz, A. and Jaenisch, R. 2000. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5:** 695–705.

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

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. 2000. RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101:** 25–33.

Zuker, M., Mathews, D.H., and Turner, D.H. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In *RNA biochemistry and biotechnology* (eds. J. Barciszewski and B.F.C. Clark), pp. 11–43. Kluwer Academic, Dordrecht/Boston/London.

Chapter 3:

The miRNA cluster, *miR-290* through *miR-295*, is an important regulator of early embryogenesis and germ cell development.

Lucas M. Dennis, Mark E. Gill, Hristo B. Houbaviy, Styliani Markoulaki, Grace X.Y. Zheng, Amy C. Seila, Lea A. Medeiros, Dong Dong Fu, Phillip A. Sharp, David C. Page, Rudolf Jaenisch

Personal Contributions

This chapter is the result of a collaboration with members of the Sharp and Page labs, with contributions by Mark Gill of note. For the study presented in this chapter, I designed and constructed the mouse mutation and performed all breeding experiments. I was also responsible for overseeing the experimental design and interpretation and prepared the manuscript. Mark Gill performed the immunohistochemistry and gonad dissections.

Summary

The *miR-290* through *miR-295* (*miR-290-295*) miRNA cluster comprises a mammalian-specific miRNA family expressed during early embryogenesis and germ cell development in the mouse. Here we show that most *miR-290-295* deficient embryos die prior to implantation. Embryonic germ cell number is decreased in both mutant XX and XY E11.5 gonads due to impaired primordial germ cell migration. Germ cell number recovers in mutant XY animals, which are fertile and show no abnormal phenotype. Embryonic germ cell depletion in mutant XX animals fails to recover and ultimately resulted in premature ovarian failure in the adult mouse. ES cells lacking *miR-290-295* showed misregulation of numerous transcripts, including many developmental regulators. These results suggest that members of the miR-290 family play a critical role in the developmental control of the early embryo and primordial germ cell.

Introduction

Early embryonic mammalian development is facilitated by a complex set of spatio-temporal signals that induce changes in gene expression, ultimately restricting cell potential and promoting establishment of differentiated cell types. The ability of cell lines derived from the early embryo, i.e. embryonic stem (ES) cells and embryonic germ (EG) cells, to maintain their developmental potential relies upon transcriptional networks set up by key regulatory genes such as *Oct4*, *Nanog* and *Sox2* (Boyer et al. 2005, Loh et al. 2006). Evidence for the involvement of miRNA mediated post-transcriptional regulation during these processes has been demonstrated by both the early embryonic lethality of mice deficient in miRNA biogenesis (Bernstein et al. 2003, Liu et al. 2004) and the differentiation defects displayed in both *Dicer* and *Dgcr8* null ES cells (Kanellopoulou et al. 2005, Murchison et al. 2005, Wang et al. 2007).

In the mouse, over 500 miRNAs have been annotated (Griffiths-Jones et al. 2007). With each miRNA potentially able to regulate scores of mRNA targets (Lewis et al. 2005), miRNA-mediated translational repression is likely a widespread regulatory mechanism of developmental processes in the mouse. The first loss-of-function studies for individual miRNAs have implicated *miR-1-2* and *miR-208* in cardiogenesis and cardiac function (Zhao et al. 2007, van Rooij et al. 2007) and *miR-155* as integral to normal immune function (Rodriguez et al. 2007). More recent studies have defined roles for the myeloid-specific *miR-223* in negative regulation of hematopoeitic progenitor proliferation and granulocyte differentiation (Johnnidis et al. 2008) and the *miR-17* through 92 family in B-cell lymphopoiesis and lung development (Ventura et al. 2008). Additionally,

phenotypes of lineage specific *Dicer* deletion support key developmental roles for miRNAs at many stages during mouse development (Cobb et al. 2005, Harfe et al. 2005, Yang et al. 2005, Andl et al. 2006, Yi et al. 2006, Murchison et al. 2007, Chen et al. 2008, Hayashi et al. 2008).

While critical roles for miRNA biogenesis in the early embryo have been established, the role of individual miRNAs during early mammalian embryogenesis remains unclear. Six miRNA families comprise the majority of miRNA species cloned from mouse ES cells, with miRNAs from the *miR-290* cluster, *miR-290* through *miR-295* (*miR-290-295*), being the most abundant (Calabrese et al. 2007). Members of this cluster are the first embryonic miRNAs upregulated in the zygote (Tang et al. 2007). Previously, we have shown that the *miR-290* cluster miRNAs are processed from a single primary transcript (Houbaviy et al. 2005) and possess highly similar seed sequences (Houbaviy et al. 2003). Recently, *miR-290-295* have been shown to be expressed during normal embryonic germ cell development (Hayashi et al. 2008) and the cluster has been implicated in control of *de novo* DNA methylation in *Dicer* null ES cells (Sinkkonen et al. 2008, Benetti et al. 2008). The human homologs of the *miR-290* cluster, *hsa-miR-371-373*, are upregulated in some testicular cancers (Voorhoeve et al. 2006), providing the first link between these miRNAs and human disease.

In this study, we examine the consequence of targeted deletion of the *miR-290* cluster in the developing mouse. We demonstrate that *miR-290-295* have a significant impact on embryonic germ cells that ultimately leads to premature ovarian failure in homozygous

females. Our loss-of-function model also demonstrates that *miR-290-295* are important regulators of early embryonic development, as most miR-290-295 deficient embryos are lost prior to implantation.

Results

miR-290-295 is specifically expressed in the early embryo and embryonic germ cells

Previously, we have shown that *miR-290-295* is expressed in ES cells and not in adult somatic tissues (Houbaviy et al. 2003). To address the timing of the *miR-290* cluster expression, we performed RT-PCR for the *miR-290-295* primary transcript throughout early embryonic development on pools of embryos. We observed onset of expression of the primary transcript at the 4-cell stage (Figure 3.1A), consistent with the finding that expression of these miRNAs is upregulated post-zygotically (Tang et al. 2007). Expression is maintained until gastrulation and locked nucleic acid (LNA) *in situ* hybridization shows widespread expression throughout the developing embryo during this time period (Figure 3.2).

Recently, *miR-290-295* has been shown to be expressed in embryonic germ cells (Hayashi et al. 2008). In order to more closely examine the expression of the *miR-290* cluster miRNAs after gastrulation, we performed RT-PCR for the primary transcript on a panel of tissues from E14.5 embryos. Expression of the miR-290 cluster was only detectable in the embryonic testis and not in any other tissue examined (Figure 3.1B). To further analyze expression of the *miR-290* cluster in germ cells during embryonic development, total RNA was prepared from both male and female E12.5 to E15.5 gonads. Expression of the *miR-290* cluster was observed in gonads of both sexes at E12.5. The *pri-miR-290-295* transcript was down-regulated in female gonads between E13.5 and E14.5, while in male gonads expression persisted through E14.5 and became undetectable by E15.5 (Figure 3.1C). Notably, the timing of *miR-290-295* downregulation was

undetectable by E15.5 (Figure 3.1C). Notably, the timing of *miR-290-295* downregulation was

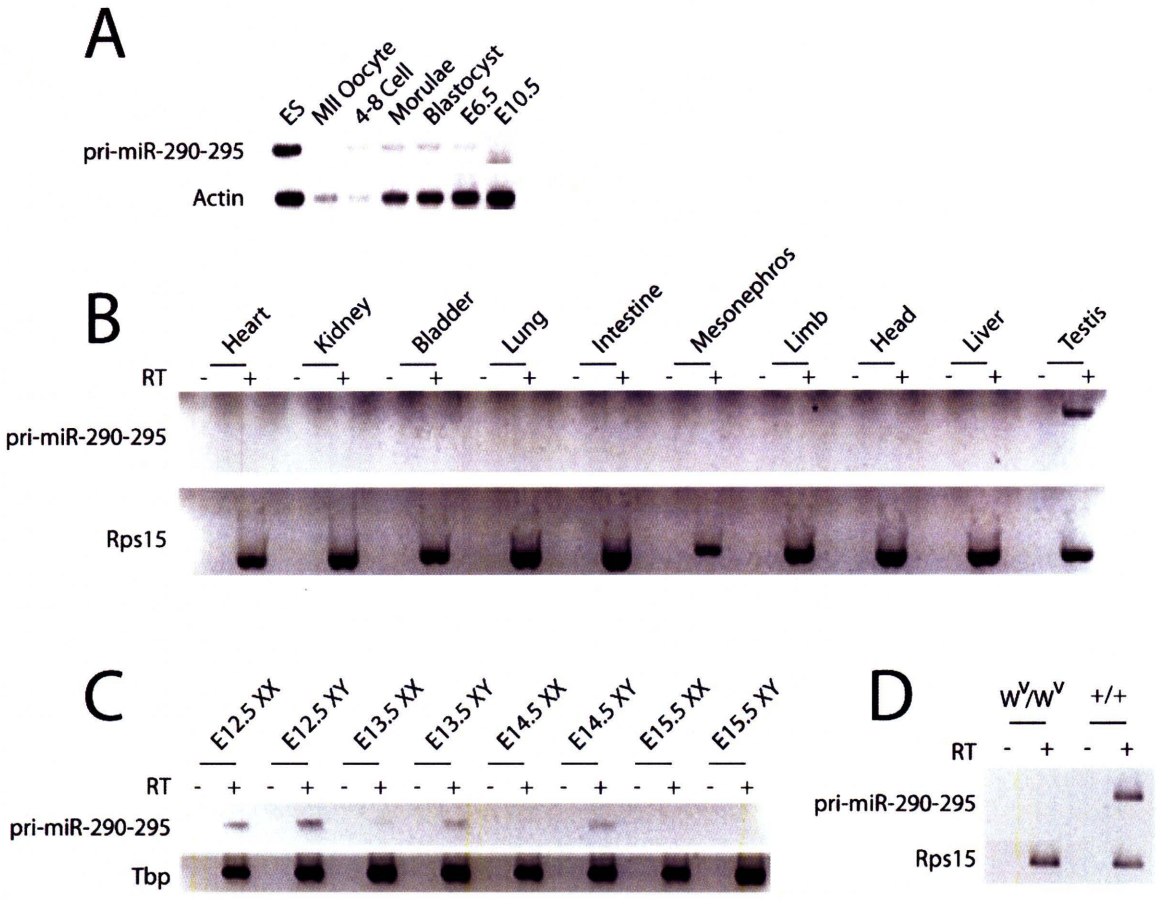

Figure 3.1. Expression of the miR-290-295 cluster.

(A) RT-PCR of pri-miR-290-295 and Actin control in early embryos.
(B) RT-PCR of pri-miR-290-295 and Rps15 control in a tissue panel from E14.5 embryos.
(C) RT-PCR of pri-miR-290-295 and Tbp control in E12.5 – E15.5 embryonic gonads from males and females as indicated.
(D) RT-PCR of pri-miR-290-295 and Rps15 control in E14.5 testis isolated from wild-type and homozygous c-kit (Wv) mutants.

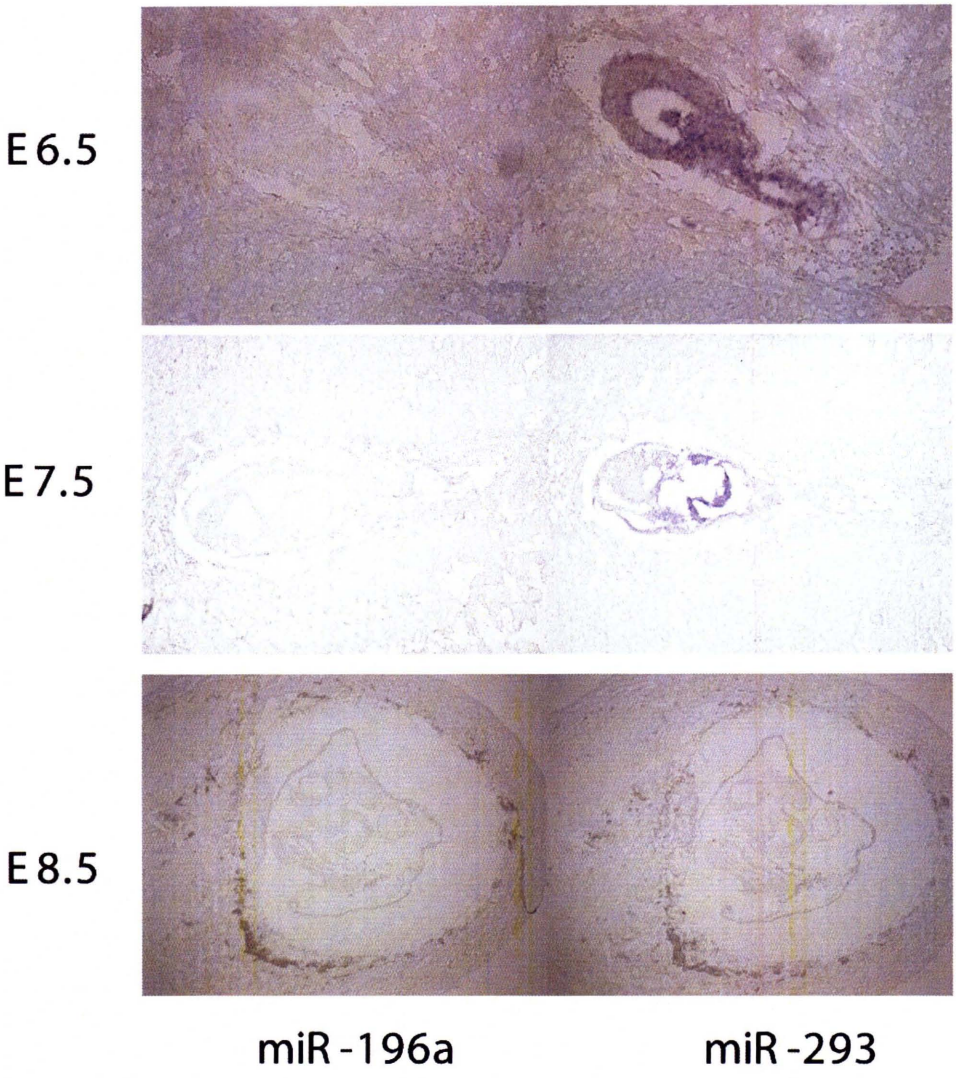

Figure 3.2. miR-293 expression in the early embryo.
LNA in situ on frozen sections containing wild-type embryos and surrounding decidua. miR-293 expression can be seen throughout the embryo at E6.5 and E7.5, but becomes undetectable at E8.5. A probe for miR-196a is shown as a control.

coincident with meiotic initiation in female (XX) germ cells and mitotic arrest in male (XY) germ cells. The female expression pattern was similar to that of key developmental regulatory genes, like *Oct4*, and consistent with previously observed *miR-290-295* expression in embryonic germ cells (Hayashi et al. 2008).

The embryonic gonad is comprised of both somatic and germ cell populations. Embryonic germ cells continue to express early embryonic pluripotency markers like *Oct4* and *Nanog* that are silenced in somatic cells prior to gastrulation. We examined expression in E14.5 W^v/W^v gonads to test whether *miR-290-295* expression in the E14.5 testis was dependent upon the presence of germ cells. W^v homozygotes harbor a mutation in the *c-kit* gene that impairs primordial germ cell (PGC) migration to the gonad, resulting in loss of germ cells prior to E14.5 (Nocka et al. 1990). *pri-miR-290-295* was undetectable in embryonic W^v/W^v testes, suggesting that *miR-290-295* expression in embryonic gonads is restricted to the germ cells (Figure 3.1D).

Non-Mendelian inheritance of the *miR-290-295*$^{\Delta}$ allele.

In order to determine the function of *miR-290-295 in vivo*, we generated mice deficient for the 2kb locus containing these miRNAs by targeted ablation in ES cells. For this a 6.3kb targeting vector was constructed in which 2.1kb of the *miR-290-295* locus encompassing the mature *miR-290-295* sequences was replaced by a neomycin resistance (NeoR) selection cassette (Figure 3.3A). Correct targeting of the locus was confirmed by Southern blot analysis and resulted in generation of two independent *miR-290-295*$^{+/\Delta}$ ES lines that were subsequently injected into B6D2F1 host blastocysts in order to produce

chimeric animals. Transmission of the targeted allele (*miR-290-295*$^{\Delta}$) through the male germ-line was confirmed by Southern blot and PCR (Figure 3.3B).

To confirm that deletion of the *miR-290-295* locus eliminated expression of *miR-290*-family miRNAs, we derived homozygous mutant ES cell lines to assess expression of miRNAs from the cluster. Blastocysts from heterozygous matings were explanted and grown on a feeder layer, resulting in the establishment of 41 independent ES cell lines. Of these 41 lines, 2 were identified to be *miR-290-295*$^{\Delta/\Delta}$ by Southern analysis (data not shown). Northern blot analysis (Figure 3.3C) and real-time RT PCR (data not shown) using probes to detect mature *miR-290-295* species failed to detect any transcripts from the mutated locus, indicating that the *miR-290-295*$^{\Delta}$ allele is a null mutation.

Mice heterozygous for *miR-290-295* deletion were fertile and indistinguishable from wild-type littermates and were maintained on a mixed genetic background (129Sv/J x C57BL/6) for this study. Homozygous *miR-290-295*$^{\Delta/\Delta}$ progeny were produced from *miR-290-295*$^{+/\Delta}$ intercrosses, and observed at a lower frequency than the expected 1:2:1 frequency predicted by Mendelian segregation (Table 3.1). Only 36 of 537 postnatal progeny from *miR-290-295*$^{+/\Delta}$ matings were *miR-290-295*$^{\Delta/\Delta}$, indicating that ~80% of the homozygous mutants were lost during development. Equal numbers of male and female progeny were observed from these crosses (data not shown).

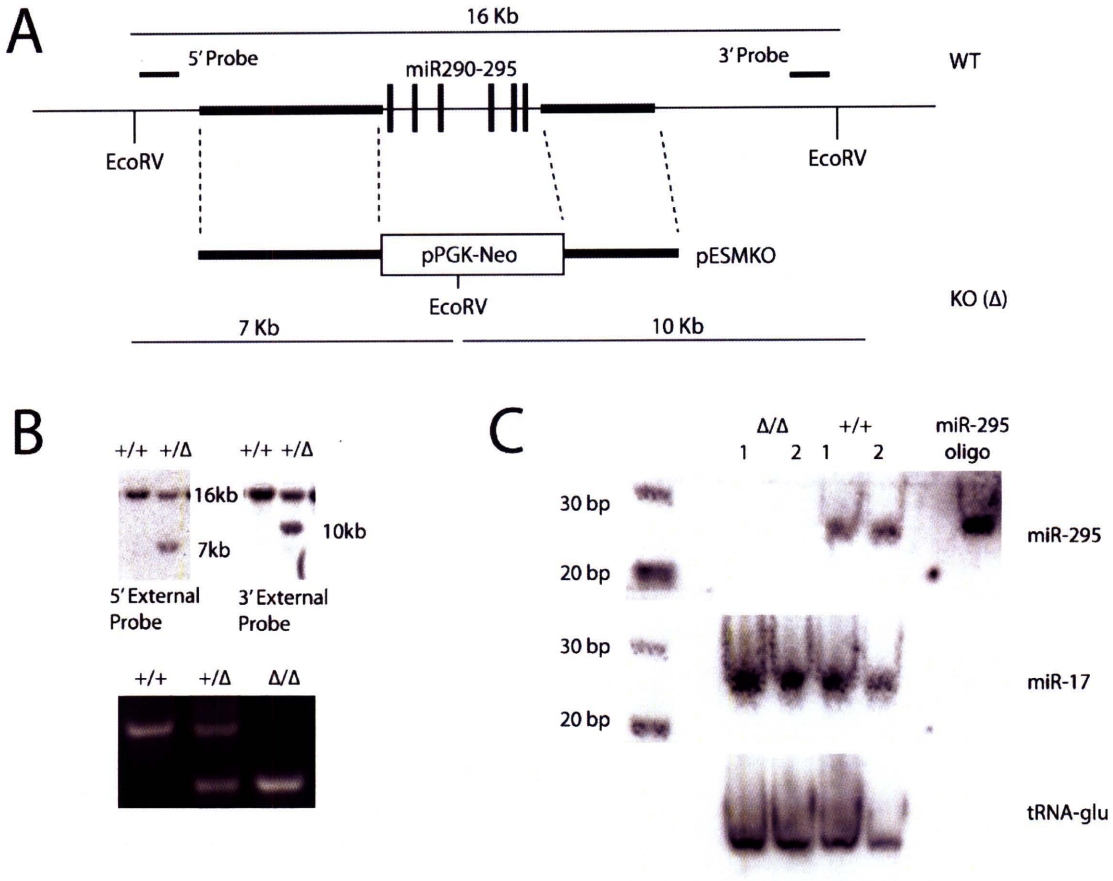

Figure 3.3. Generation of targeted deletion of miR-290-295.

(A) Targeting scheme for the deletion of miR-290-295. Location of DNA probes used to detect targeting shown as black bars.
(B) Genotyping Southern blot analysis on ES cell DNA showing correct targeting of the miR-290-295 allele. Genotyping PCR showing germline transmission of the targeted miR-290-295 allele.
(C) Northern blot analysis using an LNA probe specific for miR-295 on total RNA prepared from miR-290-295 homozygous null ES cells (lanes 1 and 2) and wild-type ES cells (lanes 3 and 4). A synthetic miR-295 oligo (lane 5) was used as positive control. Northern blots with probes specific for miR-17 and the glutamine tRNA are shown as controls.

Developmental Stage	Age	miR290-295+/+	miR290-295+/Δ	miR290-295Δ/Δ	Mendelian Deviation	n
Pre-implantation	E3.5	14	29	3		46
Mid-gestation	E9.5	5	14	6		25
Mid-gestation	E11.5	24	52	5		81
Mid-gestation	E12.5	7	13	2		22
Mid-gestation	E13.5	12	33	5		50
Post-natal	P3	27	54	4		85
Post-natal	3+ weeks	168	252	32		452
Totals		**257**	**447**	**57**	**4.5** chi sq. = 128.4	**761**

Table 3.1. Genotyping of embryonic and postnatal progeny from *miR-290-295* heterozygous intercrosses.

To investigate when *miR-290-295*$^{\Delta/\Delta}$ animals were lost during gestation, embryos were isolated from heterozygous intercrosses between E3.5 and E13.5. Genotyping of the embryos revealed a significant deviation from expected Mendelian ratios as early as the blastocyst stage (Table 3.1), indicating that loss of *miR-290-295*$^{\Delta/\Delta}$ embryos occurs during early pre-implantation development. No morphological differences were observed between surviving *miR-290-295*$^{\Delta/\Delta}$ and their *miR-290-295*$^{+/\Delta}$ or *miR-290-295*$^{+/+}$ littermates at any time-point analyzed. Postnatal *miR-290-295*$^{\Delta/\Delta}$ mice were viable and indistinguishable from wild-type littermates.

Premature ovarian failure in *miR-290-295*$^{\Delta/\Delta}$ females.

None of six homozygous females produced any offspring when caged with fertile C57BL/6 males for up to 12 months. In contrast, *miR-290-295*$^{\Delta/\Delta}$ males fathered litters with both C57BL/6 and *miR-290-295*$^{+/\Delta}$ females, indicating a female specific fertility defect in *miR-290-295*$^{\Delta/\Delta}$ animals. To investigate whether homozygous females cycle normally and come into estrus, we checked for vaginal plugs and took vaginal swabs

from homozygous mutant females caged with C57BL/6 males. No plugs were detected indicating a failure of these animals to mate. Swabs from *miR-290-295*$^{\Delta/\Delta}$ females were entirely composed of non-cornified epithelial cells, consistent with mice in an anestrus state, suggesting the absence of active cycling (data not shown).

Because the *miR-290* family knock-out females exhibited a defect in estrus cycling, we examined serum gonadotropin levels in *miR-290-295*$^{\Delta/\Delta}$ females. The gonadotropins follicle stimulating hormone (FSH) and luteinizing hormone (LH) normally oscillate in females of reproductive age, acting in concert to stimulate follicle maturation and release of the mature oocyte from the ovary. Serum from *miR-290-295*$^{\Delta/\Delta}$ females and both heterozygous and wild-type controls was collected and assayed for FSH and LH by ELISA. Serum from *miR-290-295*$^{\Delta/\Delta}$ females contained about 10-20 fold elevated levels of both LH and FSH (Figures 3.4A and B). This significant increase in serum gonadotropin levels is observed in post-menopausal women and patients presenting with premature ovarian failure (Mason et al. 1992), suggestive of a defect in ovarian function in *miR-290-295*$^{\Delta/\Delta}$ females.

Ovaries from adult (5-12 weeks) *miR-290-295*$^{\Delta/\Delta}$ females were small, having a volume <20% of wild-type and *miR-290-295*$^{+/\Delta}$ mice (Figure 3.4C-E). Ovaries from homozygous mutants older than 10 weeks contained no observable follicular structures unlike ovaries from wild-type and heterozygous females (Figure 3.4F-H). In contrast, testes from age-matched *miR-290-295*$^{\Delta/\Delta}$ males were indistinguishable from control males and contained

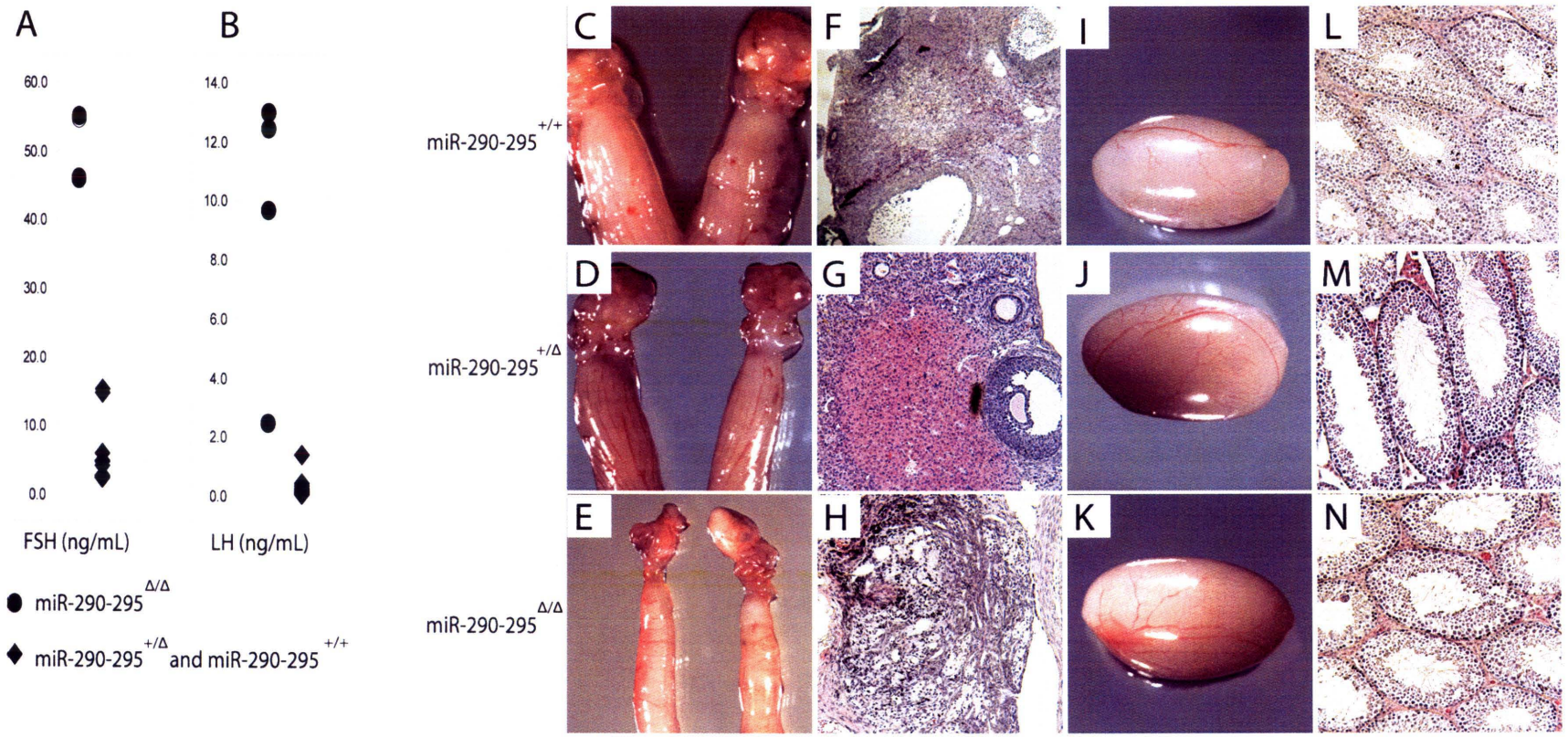

Figure 3.4. Serum gonadotropin levels and postnatal gonad morphology/histology.

(A) Follicle stimulating hormone (FSH) levels in peripheral blood serum (ng/mL) collected from miR-290-295 null females (filled circles) and littermate controls (filled diamonds).
(B) Luteinizing hormone (LH) levels in peripheral blood serum (ng/mL) collected from miR-290-295 null females (filled circles) and littermate controls (filled diamonds).
(C-E) Reproductive tracts from adult littermates of the indicated genotype.
(F-H) Hematoxylin-eosin (H-E) staining of ovaries in (C-E).
(I-K) Testes from adult littermates of the indicated genotype.
(L-N) H-E staining of testes in (I-K).

cells from all stages of spermatogenesis (Figure 3.4I-N). Sperm counts did not differ between *miR-290-295*$^{\Delta/\Delta}$ and control males (data not shown).

While no follicles were found in >10 week old mutant females, small numbers of mature follicles were observed in ovarian sections of younger (5-8 week old) individuals (data not shown). Additionally, postnatal day 3 (P3) *miR-290-295*$^{\Delta/\Delta}$ ovaries contained >80% fewer oocytes than wild-type and heterozygous ovaries as ascertained by morphological analysis and immunostaining for MVH (mouse vasa homolog), a germ cell marker (Figure 3.5). The few germ cells present in mutant embryos appeared histologically identical to germ cells from age-matched wild-type controls.

To determine if the observed follicles could produce functional oocytes, we attempted to isolate mature oocytes from homozygous mutant animals by super-ovulation. Of three *miR-290-295*$^{\Delta/\Delta}$ hormone-induced females, we were only able to retrieve oocytes from a single female. The total number of oocytes retrieved from this female was markedly reduced relative to littermate controls (Table 3.2). Oocytes from both homozygous mutant and control females fertilized by intra-cytoplasmic sperm injection (ICSI) of *miR-290-295*$^{\Delta/\Delta}$ sperm formed pro-nuclei and developed to 2-cell embryos (Figure 3.6), after which half failed to develop to the morulae stage. These data demonstrate that oocytes from homozygous mutants are viable and functionally indistinguishable from wild-type oocytes.

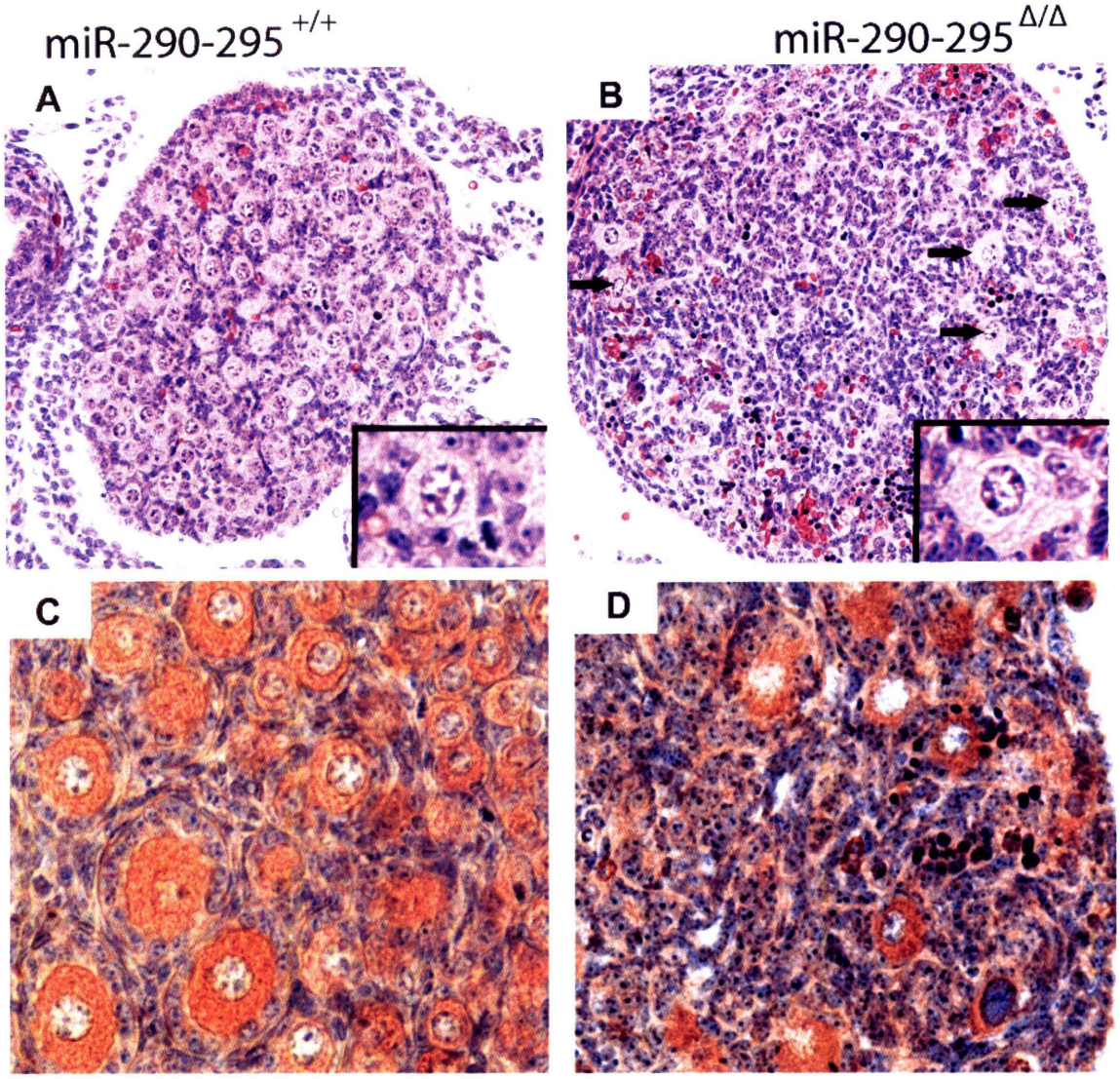

Figure 3.5. Postnatal miR-290-295Δ/Δ ovaries contain fewer primordial follicles.

(A-B) Hematoxylin-eosin staining of comparable transverse sections through the ovaries of P3 wild-type (A) and miR-290-295Δ/Δ (B) animals. Arrows point to primordial follicles in the knock-out. Insets show a representative follicle from each section at increased magnification.

(C-D) Immunohistochemical staining with an antibody against MVH on sections from wild-type (C) and miR-290-295Δ/Δ (D) postnatal day 3 (P3) ovaries. MVH positive cells appear with brown cytoplasm.

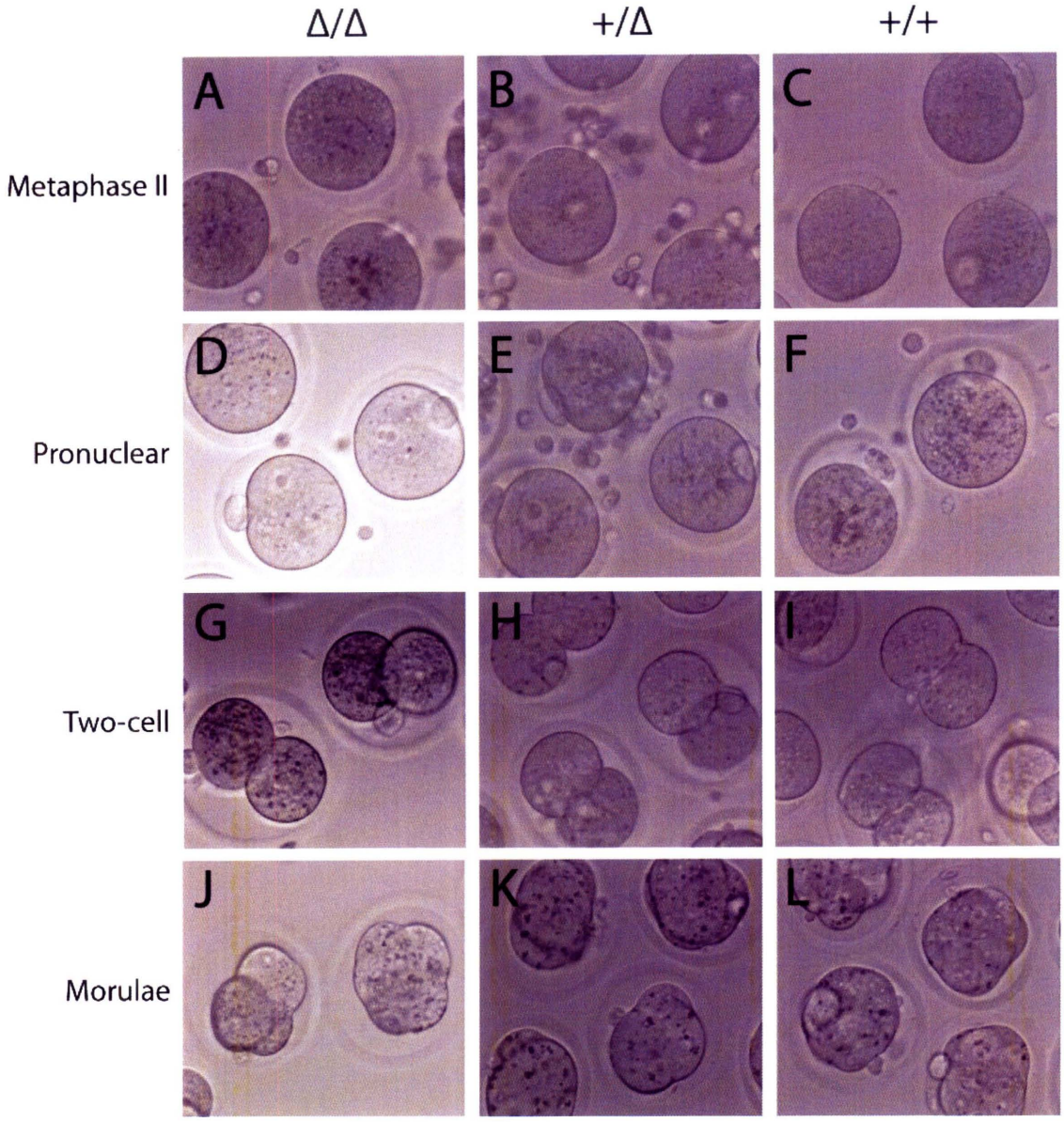

Figure 3.6. Oocytes from knock-out females appear normal and develop similarly to controls after fertilization by ICSI.

(A-C) Metaphase II stage eggs from knock-out (A), heterozygote (B) and wild type (C) females.
(D-F) Pronuclear stage embryos 6 hours after ICSI in knock-out (D), heterozygote (E), and wild type (F) embryos.
(H-I) Two-cell stage embryos 24 hours after ICSI: knock-out (G), heterozygote (H), and wild type (I).
(J-L) Morulae-stage embryos 72 hours after ICSI: knock-out (J), heterozygote (K), and wild type (L).

Genotype	Eggs	Injected	Pro Nuclei	2 Cell	Morulae
WT	17	12	12	12	12
Het	18	13	13	8	8
KO	5	5	4	4	2

Table 3.2. Oocytes from knock-out females are competent for fertilization (Quantification of experiment shown in Supplemental Figure 3).

Our results are consistent with premature ovarian failure (POF), a human disease clinically defined as the early onset of menopause. POF etiology is complex, potentially arising from genetic, immune or other idiopathic defects (Mahmoud et al. 2007). Our findings suggest that premature ovarian failure in *miR-290-295*$^{\Delta/\Delta}$ females may be caused by an insufficient stock of primordial follicles to permit normal cycling and fertility in sexually mature mice, even though young females harbor small numbers of potentially functional follicles.

Decreased germ cell number in the embryonic gonad and impairment of primordial germ cell migration in *mir-290-295*$^{\Delta/\Delta}$ embryos.

The reduced postnatal oocyte population in *miR-290-295*$^{\Delta/\Delta}$ animals suggested a defect in embryonic germ cell development. As miR-290-295 is expressed in XX germ cells until E13.5 - E14.5, we sought to examine early embryonic gonads from *miR-290-295*$^{\Delta/\Delta}$ and control animals in order to determine the status of germ cell development. MVH staining of E13.5 XX gonads detected a dramatic reduction (>20-fold) in germ cell number in homozygous mutants relative to the number of germ cells detected in wild-type gonads (Figure 3.7A-B and I). To ensure that loss of expression of MVH was indicative of decreased germ cell number, rather than a gene specific defect of *miR-290-295* deletion, we examined expression of two other markers of germ cells, DPPA3 and GCNA, by

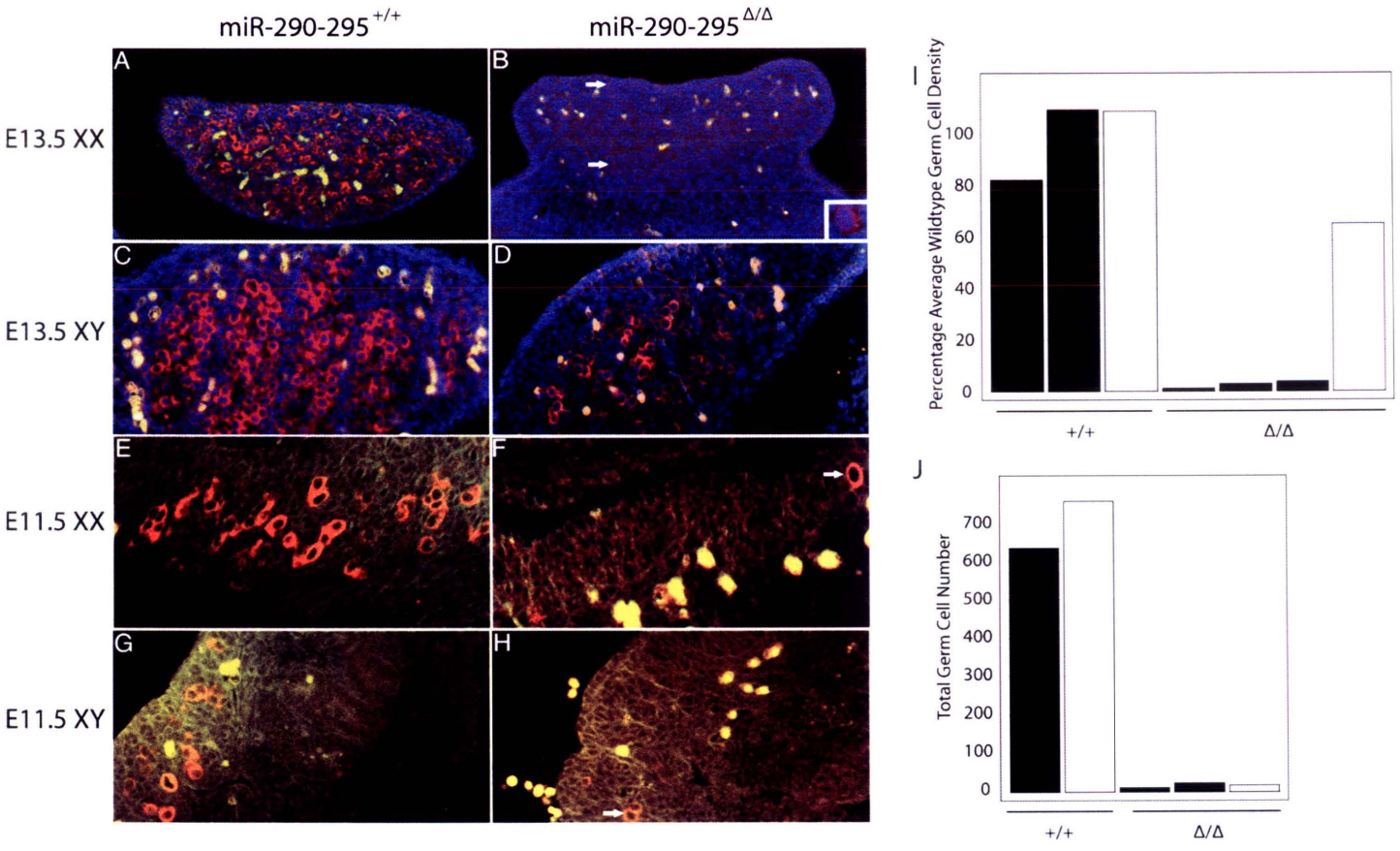

Figure 3.7. Immunohistological detection of germ cells in neonatal and embryonic gonads.

(A-F) Immunohistochemical staining with an antibody against mouse vasa homolog (MVH) on sections from wild-type and miR-290-295Δ/Δ E13.5 XX (A and B) and XY (C and D) and E11.5 XX (E and F) and XY (G and H) gonads. MVH positive cells appear as red cytoplasm. Auto-fluorescent cells appear yellow and DAPI is in blue. MVH positive cells in the E13.5 miR-290-295Δ/Δ ovary are indicated with arrows and one example is shown at increased magnification in inset.

(I) Quantification of germ cell density (germ cells/section area) for representative sections stained with MVH from E13.5 gonads. Each bar represents a single embryo with black bars denoting XX embryos and white bars XY embryos. Embryo genotypes are indicated along the x-axis as wildtype (+/+) and knock-out (Δ/Δ).

(J) Quantification of total germ cell number in E11.5 gonads as determined by serial sectioning and MVH staining. Each bar represents a single embryo with embryo sex and genotype indicated as in (I).

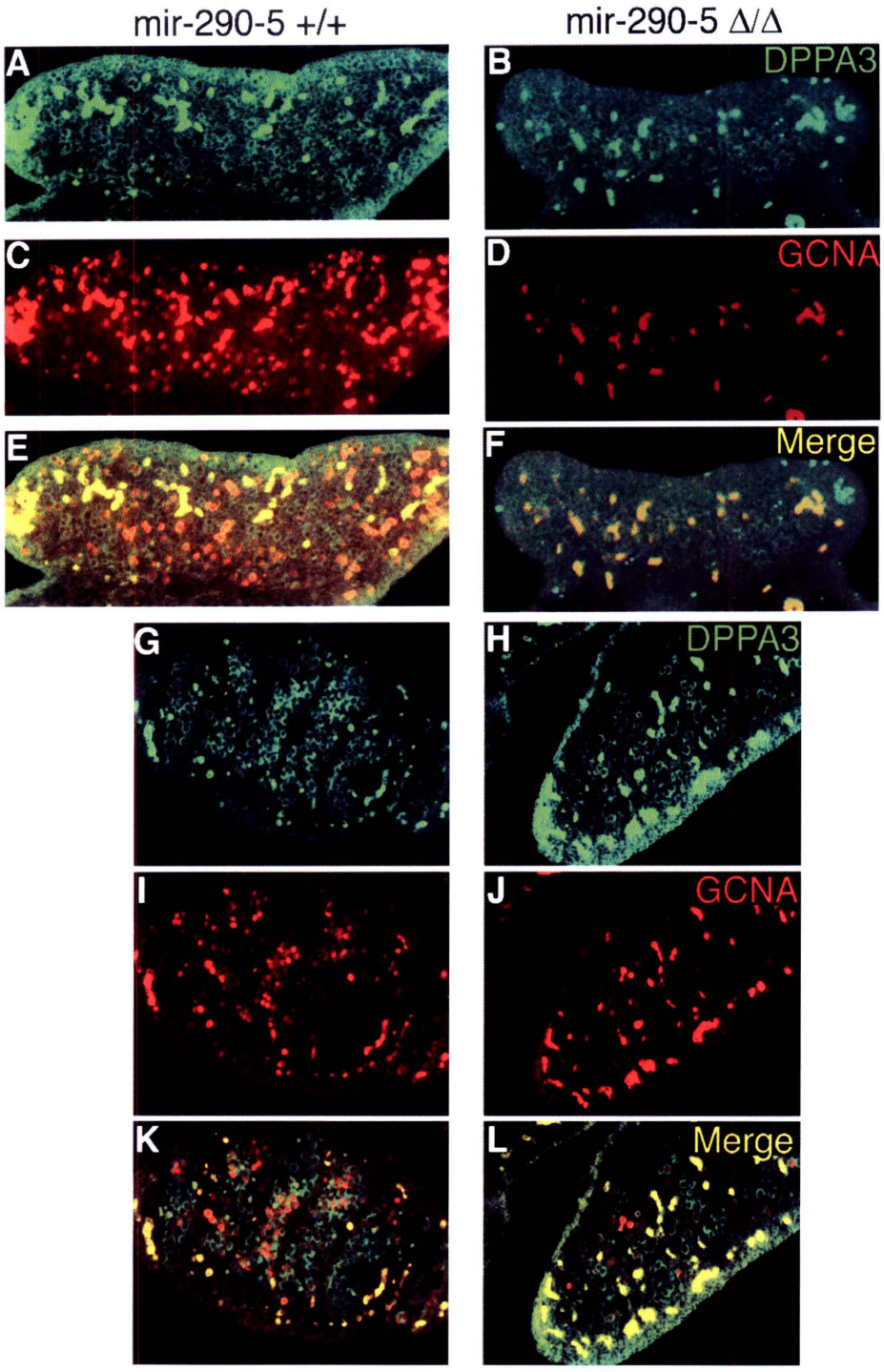

Figure 3.8. Decreased expression of other germ cell markers in E13.5 miR-290-5 knock-out gonads.

(A-B) Immunostaining with antibody to DPPA3 on wild-type (A) and miR-290-5Δ/Δ (B) E13.5 ovary.
(C-D) Immunostaining with antibody to GCNA on wild-type (C) and miR-290-5Δ/Δ (D) E13.5 ovary.
(E) Merge of (A and C). (F) Merge of (B and D).
(G-H) Immunostaining with antibody to DPPA3 on wild-type (G) and miR-290-5Δ/Δ (H) E13.5 testis.
(I-J) Immunostaining with antibody to GCNA on wild-type (I) and miR-290-5Δ/Δ (J) E13.5 testis.
(K) Merge of (A and C). (L) Merge of (B and D).

immunostaining. Similar results were observed with these markers as with MVH (Figure 3.8). MVH immunostaining of E13.5 testes also revealed a decreased germ cell number in *mir-290-295*$^{\Delta/\Delta}$ gonads, which was surprising given the lack of an abnormal phenotype in the postnatal testis (Figure 3.7C and D). The loss of germ cells in E13.5 mutant XY gonads was, however, much less severe than in XX gonads with mutant XY gonads containing ~60% of the wild-type number of germ cells (Figure 3.7I). Both mutant XX (Figure 3.7E-F) and mutant XY (Figure 3.7G-H) gonads at E11.5 exhibited less than 5% wild-type germ cell number as assessed by MVH immunostaining (Figure 3.7J). The embryonic loss of germ cells in both XX and XY gonads by E11.5 and normal histology of postnatal XY gonads suggests that XY germ cells are able to recover in number sufficient to fully populate the postnatal testis.

The observation of decreased germ cell number in E11.5 gonads suggested that loss of *miR-290-295* is likely to occur during either germ cell specification or migration of primordial germ cells (PGCs) to the gonad. To distinguish between these two possibilities, we assessed the location of migrating PGCs in E9.5 embryos by whole-mount alkaline phosphatase (AP) staining. In wild-type animals at E9.5, most PGCs are located in the hind-gut (Figure 3.9A), having migrated from the posterior end of the embryo (Figure 3.9B). AP staining of *mir-290-295*$^{\Delta/\Delta}$ embryos showed a decreased number of PGCs in the hindgut and an increased number near the base of the allantois (Figure 3.9C-E). *miR-290-295*$^{\Delta/\Delta}$ PGCs were also diffusely organized within the hindgut, unlike wild-type PGCs, which were found in a narrow band (Figure 3.9A and C). This migration defect occurred in both XX and XY embryos (Figure 3.9E). Altogether, these

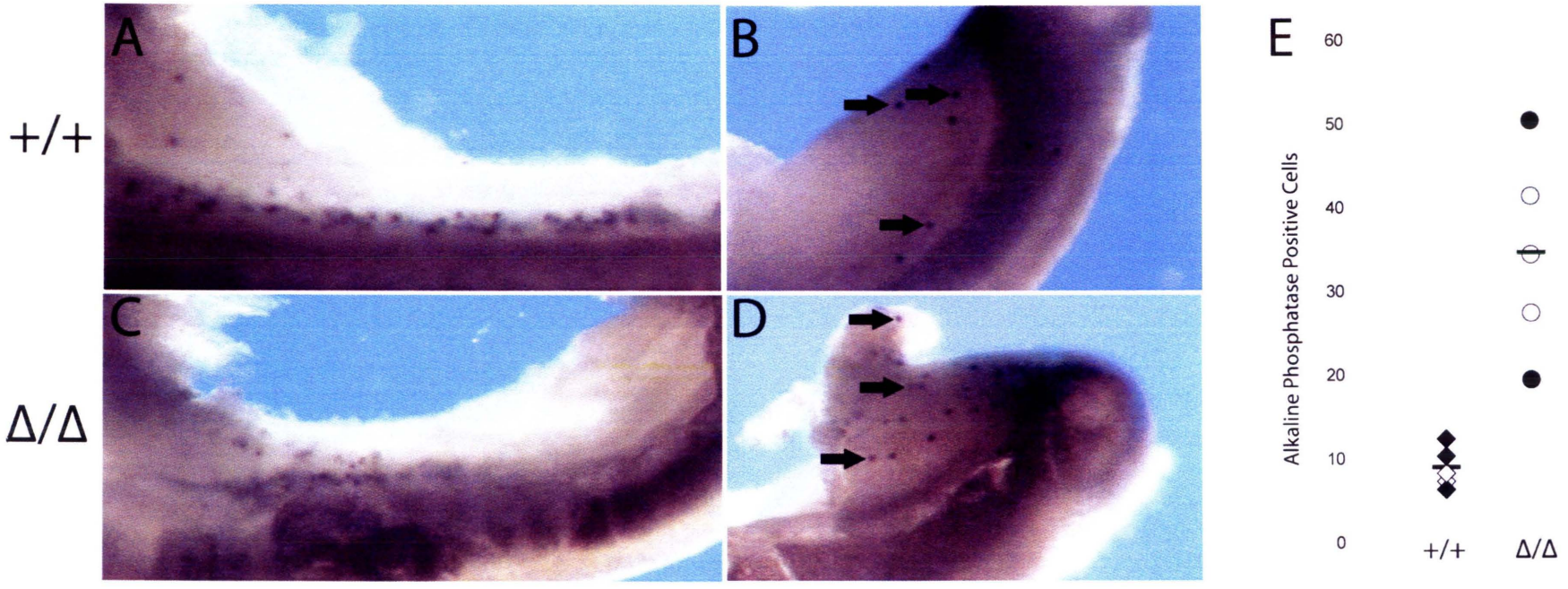

Figure 3.9. Migration defect in miR-290-295Δ/Δ E9.5 embryos.

(A-D) Whole-mount alkaline phosphatase staining of hindgut (A and C) and posterior end of (B and D) wild-type and knock-out ED9.5 embryos. Examples of alkaline phosphatase positive germ cells are indicated with arrows.

(E) Quantification of the number of alkaline positive cells on the surface of the allantois at the posterior end of the embryo from whole-mount in situs of E9.5 wild-type (diamonds) and knock-out (circles) embryos. Each data point represents a single XX (black) or XY (white) E9.5 embryo Averages germ cell number for each genotype is indicated as a black bar.

observations suggest that impaired migration of PGCs may be responsible for reduced germ cell number in miR-290 cluster deficient gonads.

Loss of *mir-290-295* in ES cells results in derepression of developmental regulators. Given the observed defect in early embryogenesis, we sought to determine whether a population of mRNAs was differently regulated upon loss of the miR-290 cluster miRNAs in ES cells. To this end, we examined global gene expression in wild-type and *miR-290-295*$^{\Delta/\Delta}$ ES cell lines using Affymetrix Mouse Genome 430 2.0 microarrays. Of the >21000 genes on the array, more than 600 (~400 up- and ~200 down-regulated) were found to be significantly dysregulated in knock-out ES cells (data not shown). Loss of *miR-290-295* resulted in significant upregulation in levels of many key developmental regulators, including: *Foxd1*, *Hoxb13*, *Lefty1*, *Sox1*, and *Sox21* (Figure 3.10A). No change in mRNA level was observed for any of the core regulators of ES cell pluripotency: *Oct4*, *Nanog* and *Sox2*. Array results were verified by qRT-PCR (Figure 3.10A).

Pairing between the 6-8nt seed sequence at the 5' end of the miRNA to a conserved site in the 3' UTR (seed match) of the mRNA message has been demonstrated to be the most critical determinant of miRNA target specificity (Lewis et al. 2003). To test if mRNAs containing *miR-290* cluster seed matches were enriched in the upregulated messages, we first determined the abundance of genes containing *miR-290* family seed matches in their 3'-UTR on the array. We found 4279 (19.9%) predicted M8 7-mer targets (mRNAs containing a match to nucleotides 2-8 of a miR-290 cluster miRNA) and 1765 (8.2%) predicted M8-A1 8-mer targets (mRNAs containing a match to nucleotides 2-8 of a miR-

290 cluster miRNA followed by an adenosine) among the 21465 transcripts on the mouse whole genome array (see Methods). Of the 3197 transcripts upregulated >=1.2 fold, 1394 (~32.6 %) contained a 7-mer match to miR290-295 miRNAs, while 614 (34.8 %) contained an 8-mer match to *miR290-295* miRNAs, suggesting a significant enrichment for *miR-290* family targets in this population. This enrichment can be seen as a shift to the right in a plot of the cumulative density of *miR-290-295* seed match containing transcripts versus $\log_2$ of fold change relative to control transcripts (Figure 3.10B) likely representing the accumulation of derepressed target messages in *miR-290-295*$^{\Delta/\Delta}$ ES cell lines as has been previously observed in other systems (Lim et al. 2005, Giraldez et al. 2006). No enrichment for *mir-290-295* targets was observed among transcripts found to be downregulated in *miR-290-295*$^{\Delta/\Delta}$ ES cells.

The majority of cumulative log fold density change between *miR-290* cluster seed containing transcripts and controls was centered around 0.5, indicating the predominant effect of *miR-290-295* loss was a modest (1.4-fold) upregulation of seed containing transcripts. While some gene expression changes observed between *miR-290-295*$^{\Delta/\Delta}$ and wildtype ES cells are likely to be due to indirect effects of impaired *miR-290-295*-mediated repression, a subset of the upregulated transcripts may be direct miRNA targets. These direct targets would represent those mRNAs regulated at the level of mRNA stability (Valencia-Sanchez et al. 2006) and would be expected to be enriched in miRNA knock-out cells (Zhao et al. 2007). All together, these observations suggest that *miR-290-295* regulates a number of transcripts in ES cells, primarily by fine-tuning gene

expression regulating genes associated with differentiation, such as *Foxd1*, *Hoxb13*, *Sox1*, and *Sox21*.

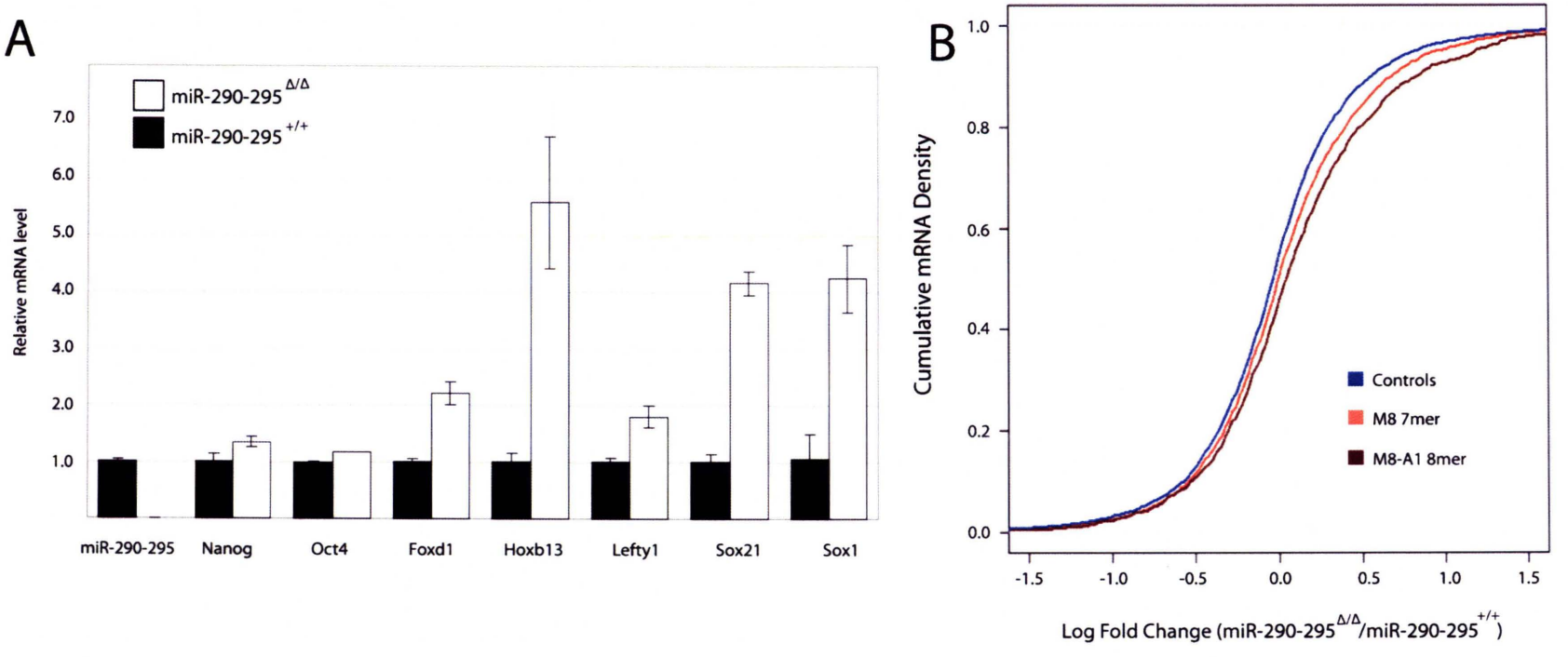

Figure 3.10. Loss of miR-290-295 in ES cells results in derepression of developmental regulators.

(A) mRNA levels as determined by qRT-PCR for select genes in wild-type (black bars) and miR-290-295Δ/Δ (white bars) ES cells. Values for each gene represent the average of three experiments with error bars indicating standard deviation.

(B) Cumulative mRNA density versus Log(2) fold change of expression in miR-290-295Δ/Δ and wild-type ES cell microarray replicates. Log(2) fold change is defined as the log base 2 of expression in miR-290-295Δ/Δ ES cells over that in wild-type ES cells. The set of control transcripts were defined as the set of transcripts capable of being targeted by a heptameric sequence of similar nucleotide composition as the miR-290-295 seeds (See Methods).

Discussion

In this study, we investigated the biological function of the *miR-290* cluster by targeted deletion in the mouse. We found that *miR-290-295* has a significant role in both the early embryo and during germ cell development. Approximately 80% of *miR-290-295* deficient embryos were lost during embryonic development and *miR-290-295* deficient females exhibited premature ovarian failure caused by depletion of the embryonic germ cell population. Deletion of *miR-290-295* resulted in significant changes in gene expression, with hundreds of transcripts harboring *miR-290* family seed matches being upregulated, including many developmentally regulated transcription factors.

Role of *miR-290-295* in the early embryo.

The *miR-290* cluster miRNAs are not expressed at an appreciable level in either the mature ovary or mature oocytes (Houbaviy et al. 2003). In contrast, *miR-290-295* expression was upregulated at the 4-cell stage. Significant loss of both maternal miRNAs and mRNAs occurs between the one- and two-cell stages of development and oocytes lacking all miRNAs failed to proceed through the first cell division (Tang et al. 2007). The ability of *miR-290-295*$^{\Delta/\Delta}$ oocytes to divide to 2-cell embryos after fertilization is consistent with these observations and demonstrates that defects in *miR-290-295*$^{\Delta/\Delta}$ embryos likely occur post-zygotically, although we can not rule out a maternal effect.

Disruption of miRNA biogenesis by deletion of *Dicer* resulted in lethality of all embryos by E7.5, with ~50% having been lost earlier (Bernstein et al. 2003). In addition, a significant role for miRNAs in the regulation of proliferation, cell cycle progression and differentiation has also been demonstrated in ES cells (Kanellopoulou et al. 2005,

Murchison et al. 2005, Wang et al. 2007). While the precise mechanism of early embryonic lethality in *miR-290-295*$^{\Delta/\Delta}$ embryos is presently unclear, knockout ES cells showed dysregulation in the levels of many genes that are important during early development. Misexpression of developmental regulators like *Foxd1*, *Sox1*, *Sox21*, and *Hoxb13* that help to establish the transcriptional state of more differentiated cell types may result in increased developmental instability that ultimately reduces embryo viability. It is possible that stochastic differences in gene expression between embryos and/or compensation by other miRNAs with similar seed sequences may permit some embryos to complete normal development. Also, the loss of a significant percentage of *miR-290-295*$^{\Delta/\Delta}$ embryos by the blastocyst stage suggests that at least some of the embryonic lethality observed in *Dicer* deficient mice may be due to a lack of *miR-290* cluster miRNA expression.

Role of *miR-290-295* in germ cell development.

Typically, germ cells emerge from the posterior primitive streak region of the gastrula at E7.5 and move into the region of the embryonic endoderm fated to become the hindgut (Anderson et al. 2000). By E9.5, PGCs begin migrating dorsally out of the hindgut towards the mesenchyme of the dorsal body wall eventually reaching the bilateral genital ridges by E11.5 (Molyneaux et al. 2001; Molyneaux et al. 2003). Primordial germ cells that have failed to correctly migrate and are improperly located in the midline at E10.5 begin apoptosis due to lack of Steel factor (Runyan et al. 2006). PGC location in E9.5 *miR-290-295*$^{\Delta/\Delta}$ embryos was abnormal, with many PGCs remaining in the posterior region of the embryo or along the midline. As such, germ cells in the mutant E11.5 gonads likely represent the few PGCs that correctly migrated and escaped apoptosis. The

reduced viability of *miR-290-295*$^{\Delta/\Delta}$ embryos and limited understanding of genes known to be involved in PGC migration impede the examination of a molecular mechanism for the PGC migration defect in animals lacking *miR-290* cluster miRNAs. Given that embryonic germ cells and ES cells have highly overlapping expression profiles, the transcripts misregulated in mutant ES cells may provide a useful source for further mechanistic insight into germ cell migration.

The recovery of germ cell number in *miR-290-295* deficient XY gonads that is initiated between E11.5 and E13.5 is striking as most mutations affecting early germ cell development result in sex non-specific sterility (Kehler et al. 2004, Ohinata et al. 2005, Youngren et al. 2005, Runyan et al. 2006). Deletion of the zinc finger protein ZFX results in a similar depletion of germ cells at E11.5 with recovery in postnatal males, but not females (Luoh et al. 1997), however, the cause of the germ cell loss and recovery in XY gonads in these animals is unknown. Sexually dimorphic development of germ cells is generally considered to occur after E11.5 in response to somatic cues from the gonad (McLaren, 2003). Onset of germ cell recovery in *miR-290-295*$^{\Delta/\Delta}$ XY gonads is coincident with this initiation of sexually dimorphic development, suggesting that either the XY soma or an intrinsic XY germ cell factor promotes germ cell recovery in mutant males. Given that XX germ cells enter into meiotic arrest shortly after arriving in the gonad, the failure of germ cells to recover in females likely reflects their intrinsic inability to proliferate after entry into meiosis.

In conclusion, we have demonstrated a complex role in the regulation of early embryogenesis and germ cell development for *miR-290-295*. These miRNAs promote embryonic viability and are necessary for proper migration of primordial germ cells to the embryonic gonad. Aberrant PGC migration leads to a reduced population of germ cells colonizing the embryonic gonad that ultimately manifests as premature ovarian failure in adult females. The status and regulation of this locus will be interesting to explore in POF patients as whether an embryonic etiology of human premature ovarian failure exists is unclear.

Methods

RT-PCR, qPCR, Northern blot analysis and miRNA in-situ.

RNA samples were isolated by homogenizing tissue or cells in Trizol (Invitrogen) following manufacturers suggested protocol. Five micrograms of total RNA was DNase I treated using the DNA-Free RNA Kit (Zymo Research). One microgram of DNase I-treated RNA was reverse transcribed using a First Strand Synthesis Kit (Invitrogen). Quantitative real-time PCR was performed using 1/80 of the reverse transcription reaction in an ABI 7900 HT Fast Real-Time PCR machine (Applied Biosystems) with Platinum SYBR Green qPCR SuperMix-UDG with ROX (Invitrogen). Sequences of primers used for PCR are listed in Table 3.3. Relative quantitations were calculated using the comparative ($\Delta\Delta C_T$) method. All data were normalized to GAPDH and expressed relative to wild type ES cells.

Gene	Primer Name	Sequence (5'--> 3')
miR-290-295	miR290-F1	AGA GGC GAA AGT AGA TCC AG
(genotyping)	miR290-R1	AAG TTA GTG GGG CACT TTG C
	neo-R2	AGA GGC CAC TTG TGT AGC GC
miR-290-295	miR290-F1_2	GAA CCT CAC GGG AAG TGA CC
(real time)	miR290-R1_2	TGC CCA CAG GAG AGA CTC AA
Foxd1	Foxd1-F1	GAA GCC GCC CTA CTC GTA CA
(real time)	Foxd1-R1	GCT GCT GAT GAA CTC GCA GAT
Hoxb13	Hoxb13-F1	CCA CCA GGT CCA TTC TGG AA
(real time)	Hoxb13-R1	GCT ATA GGG AAT GCG TTT TTT GC
HPRT	HPRT-F1	TCA GTC AAC GGG GGA CAT AAA
(real time)	HPRT-R1	GGG GCT GTA CTG CTT AAC CAG
Lefty1	Lefty1-F1	CCT CGC ACG TGA GGA CTC A
(real time)	Lefty1-R1	CCT GCC ACC TCT CGA AGG T
Nanog	Nanog-F1	AGG GTC TGC TAC TGA GAT GCT CTG
(real time)	Nanog-R1	TAG CTC AGG TTC AGA ATG GAG GAG AG
Oct-4	Oct4-F1	ACA TCG CCA ATC AGC TTG G
(real time)	Oct4-R1	AGA ACC ATA CTC GAA CCA CAT CC
Sox1	Sox1-F1	GTC AAG CGG CCC ATG AAC
(real time)	Sox1-R1	GCT TGC TGA TCT CCG AGT TGT
Sox21	Sox21-F1	CTT GAA AGG CCG AGC CTT C
(real time)	Sox21-R1	GAA TGA TGC TGC GAA ATC CAC

Table 3.3. Primers used in this study.

For the Northern blot analysis, 30ug of total RNA was diluted in 3x volume of formamide loading buffer, incubated at 95°C for 5minutes and loaded onto a 15% denaturing 10mM MOPS pH 7.0 after pre-running for 30minutes. The samples were electrophoresed for 45 minutes at 35W. Afterwards, the RNA was transferred to Hybond-NX nylon membrane (Amersham Biosciences) at 18V in a semi-dry transfer apparatus at 4°C for 1.5 hours. The RNA was cross-linked and incubated with an LNA probe as previously described (Pall et al. 2007, Valoczi, A. et al. 2004). Hybridization took place overnight after which the membrane was washed 2 times for 2 minutes with 2xSSC, 0.1% SDS at room temperature followed by 2 30 minute washes with 2x SSC,

0.1% SDS at 37°C. The membrane was wrapped in Saran Wrap and exposed to a phosphorimager screen for 1-3 days depending on the probe. The probes were synthesized by IDT; (1) mir-290, A+AAG+TGC+CCC+CAT+AGT+TTG+AGT (2) mir-17, C+TAC+CTG+CAC+TGT+AAG+CAC+TTT+G, (3) mir-295, A+GAC+TCA+AAA+GTA+GTA+GCA+CTT+T, and (4) tRNA glu, TGGAGGTTCCACCGAGAT, where + indicates that the following nucleotide is an LNA.

For miRNA *in situ*, E6.5 to E8.5 embryos were fixed, frozen sections prepared and hybridized as previously described (Obernosterer et al. 2007) except that the formamide was omitted from the hybridization solution and hybridization was done at 42 degrees Celsius. LNA for miR-293 and miR-196a were produced by the manufacturer (Exiqon) and labeled with digoxigenin (Roche) with terminal transferase.

Generation of *miR-290-295*$^{\Delta/\Delta}$ mice.

In order to generate mice deficient for *miR-290-295*, a null *miR-290-295* allele was introduced into ES cells via homologous recombination. Upstream and downstream arms were PCR amplified from a RPCI-23-222D1 BAC DNA template, resulting in a construct in which 2.1kb of the *miR-290-295* locus containing the mature miRNA sequences was replaced by the 1.6kb neomycin resistance selection cassette. The targeting construct was electroporated into V6.5 ES cells and subjected to selection in media containing G418. After 10 days of culture in selective conditions, G418 resistant clones were analyzed by Southern blotting with external probes. Clones exhibiting correct targeting were injected into B6D2F1 recipient blastocysts for subsequent chimera generation.

ES cell culture, manipulations and derivation.

ES cells were cultivated on irradiated MEFs in DMEM containing 15% FBS, leukemia inhibiting factor (LIF), penicillin/streptomycin, L-glutamine and nonessential amino acids. For ES cell derivation, the zona pellucida was removed using acidic tyrode (AT) solution and blastocysts were explanted on irradiated feeders in ES medium plus MEK1 inhibitor (PD98059). For ES cell differentiation experiments, ES cells were grown in either ES medium without LIF (EB media) or EB media supplemented with 1μM retinoic acid.

Estrous cycle determination and measurement of serum gonadotropin levels.

Assessment of estrous cycle stages was made by histological analysis of vaginal smears taken daily for a period of 21 days. Peripheral blood samples were taken from 2 to 6-month-old cycling mice by retro-orbital bleeding. Serum was prepared and stored at -20°C until assayed. Serum LH and FSH levels were assayed by the University of Virginia Ligand Assay and Analysis Core Laboratory as described on their website: http://www.healthsystem.virginia.edu/internet/crr/methodspage.cfm.

Intra-cytoplasmic sperm injection.

Metaphase II stage eggs for ICSI were collected from knock out, heterozygote and wild type super-ovulated females that were 7-8 weeks of age. As described before (Markoulaki et al. 2007), superovulation was induced by i.p. injections of PMSG and hCG. ICSI was carried out as described previously (Markoulaki et al. 2007). Briefly, injections were performed in HEPES-CZB medium. Sperm were released by squeezing the cauda epididymides of one mouse in 1 ml of injection buffer (75 mM KCl and 20 mM

HEPES, pH 7.0). After allowing the sperm to disperse for a few minutes, one volume of suspension was mixed with one volume injection buffer containing 12% polyvinyl pyrrolidone (PVP, M.W. 360 kDa; Sigma). A sperm head was first aspirated into a 5 μm pipette (Humagen Fertility Diagnostics) attached to a piezo micropipette-driving unit and the head was separated from the tail by applying a few piezo pulses. Then, the sperm head was delivered into the ooplasm, after drilling through the zona pellucida and subsequently the plasma membrane by applying a few piezo pulses. After injection, the eggs were transferred into KSOM medium (Chemicon) and cultured in a 5% CO_2, humidified, incubator at 37°C.

Antibodies, histological and immunohistochemical analysis of adult and embryonic gonads.

For immunohistochemistry and histology, gonads were fixed in Bouin's fixative overnight, embedded in paraffin and sectioned. For histology, slides were stained with hematoxylin and eosin. For immunohistochemistry, slides were de-waxed, rehydrated and steamed in 10 mM Sodium Citrate buffer, pH 6.0 for 10 minutes. Rabbit polyclonal anti-MVH (Abcam) was used at a dilution of 1:200, rabbit polyclonal anti-DPPA3 (Abcam) was used at 1:200, rat anti-GCNA1 (a gift from George Enders) was used as undiluted supernatant. For postnatal analysis ImmPress Anti-Rabbit IG (VectorLabs) kit was used for secondary antibody incubation and detected using DAB (VectorLabs). Slides were then counterstained with hematoxylin, dehydrated and mounted with Permount (VectorLabs). For embryonic analysis Texas Red- or FITC-conjugated anti-rabbit secondary antibody (Jackson ImmunoResearch Laboratories Inc.) or Texas Red-

conjugated anti-rat was used at a dilution of 1:200 and mounted with VectorShield Mounting Media with DAPI (VectorLabs).

Microarray and bioinformatic analysis.

15 μg of total RNA was hybridized to Mouse Genome 430 2.0 microarrays (Affymetrix) following manufacturers recommended protocol by the Whitehead Institute Center for Microarray Technology. Data was analyzed using DChip. Probes on the Affymetrix Mouse 430 2.0 Array were mapped to Refseq transcripts using annotation files provided by the Affymetrix. Refseq transcript expression levels were then calculated using GCRMA using default settings. Genes with normalized log2 intensity below 4 were excluded from analysis.

Mouse 3' UTRs used in this study were obtained from UCSC (mm8, Feb 2007). The 3' UTR sequences were searched for non-overlapping seed matched to relevant miRNAs. M8 7mer seed is defined as nucleotides 2 to 8 of a miRNA sequence. mRNAs that contain a complementary match to M8 7mer are termed M8 7mer targets. The mRNA targets with an adenosine following the M8 7mer match are called M8-A1 8mer targets. The log fold change (LFC) of a target is defined as the expression difference between the mRNA in miR-290-295$^{\Delta/\Delta}$ ES cells and wild-type ES cells. LFCs of miR290-295 targets are compared against that of control targets to assess the statistical significance of the change. Control targets are mRNAs that contain seed matched to heptamers/octamers that are similar in abundance and sequence composition as miR290-295 seeds. Control targets and miR290-295 targets are not necessarily exclusive, as one mRNA can be targeted by both the control seed as well as miR290-295 seeds.

Acknowledgements

The authors thank Jessica Dausman and Ruth Flannery for their help with the generation of chimeras and animal husbandry; the Whitehead Institute Center for Microarray Technology for microarray hybridizations, George Enders for GCNA anti-sera and members of the Jaenisch, Page and Sharp labs for critical discussions. This work was supported by NIH grants 5-F32-HD051190 (to A.S), RO1-GM34277 (to P.S.), 5R37CA084198 and 5R01-HD045022 (to R.J), National Cancer Institute grant PO1-CA42063 (to P.S.) and core grant P30-CA14051 (to Koch Institute) and the Howard Hughes Medical Institute (to D.P.).

References

Anderson, R., Copeland, T.K., Scholer, H., Heasman, J., and Wylie, C. 2000. The onset of germ cell migration in the mouse embryo. *Mech Dev.* **91**: 61-8.

Andl, T., Murchison, E.P., Liu, F., Zhang, Y., Yunta-Gonzalez, M., Tobias, J.W., Andl, C.D., Seykora, J.T., Hannon, G.J., and Millar, S.E. 2006. The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles. *Curr Biol.* **16**: 1041-9.

Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P. et al. 2008. A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat Struct Mol Biol.* **15**: 268-79.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. 2003. Dicer is essential for mouse development. *Nat Genet.* **35**: 215-7.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G. et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell.* **122**: 947-56.

Calabrese, J.M., Seila, A.C., Yeo, G.W., and Sharp, P.A. 2007. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc Natl Acad Sci U S A.* **104**: 18097-102.

Chen, J.F., Murchison, E.P., Tang, R., Callis, T.E., Tatsuguchi, M., Deng, Z., Rojas, M., Hammond, S.M., Schneider, M.D., Selzman, C.H. 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**: 2111-6.

Cobb, B.S., Nesterova, T.B., Thompson, E., Hertweck, A., O'Connor, E., Godwin, J., Wilson, C.B., Brockdorff, N., Fisher, A.G., Smale, S.T. et al. 2005. T cell lineage choice and differentiation in the absence of the RNase III enzyme Dicer. *J Exp Med.* **201**: 1367-73.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science.* **312**: 75-9.

Griffiths-Jones, S., Saini, H.K., Dongen, S.V., and Enright, A.J. 2007. miRBase: tools for microRNA genomics. *Nucleic Acids Res.*

Harfe, B.D., McManus, M.T., Mansfield, J.H., Hornstein, E., and Tabin, C.J. 2005. The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb. *Proc Natl Acad Sci U S A.* **102**: 10898-903.

Hatfield, S.D., Shcherbata, H.R., Fischer, K.A., Nakahara, K., Carthew, R.W., and Ruohola-Baker, H. 2005. Stem cell division is regulated by the microRNA pathway. *Nature.* **435**: 974-8.

Hayashi, K., Chuva de Sousa Lopes, S.M., Kaneda, M., Tang, F., Hajkova, P., Lao, K., O'Carroll, D., Das, P.P., Tarakhovsky, A., Miska, E.A. et al. 2008. MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis. *PLoS ONE.* **3**: e1738.

Houbaviy, H.B., Dennis, L., Jaenisch, R., and Sharp, P.A. 2005. Characterization of a highly variable eutherian microRNA gene. *Rna.* **11**: 1245-57.

Houbaviy, H.B., Murray, M.F., and Sharp, P.A. 2003. Embryonic stem cell-specific MicroRNAs. *Dev Cell.* **5**: 351-8.

Johnnidis, J.B., Harris, M.H., Wheeler, R.T., Stehling-Sun, S., Lam, M.H., Kirak, O., Brummelkamp, T.R., Fleming, M.D., and Camargo, F.D. 2008. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature.* **451**: 1125-9.

Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M., and Rajewsky, K. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* **19**: 489-501.

Kehler, J., Tolkunova, E., Koschorz, B., Pesce, M., Gentile, L., Boiani, M., Lomeli, H., Nagy, A., McLaughlin, K.J., Scholer, H.R. et al. 2004. Oct4 is required for primordial germ cell survival. *EMBO Rep.* **5**: 1078-83.

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

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. 2003. Prediction of mammalian microRNA targets. *Cell.* **115**: 787-98.

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**: 1437-41.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J. et al. 2006. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet.* **38**: 431-40.

Luoh, S.W., Bain, P.A., Polakiewicz, R.D., Goodheart, M.L., Gardner, H., Jaenisch, R., and Page, D.C. 1997. Zfx mutation results in small animal size and reduced germ cell number in male and female mice. *Development.* **124**: 2275-84.

Mahmoud, M.S., Merhi, Z.O., and Yelian, F.D. 2007. Mechanisms of premature ovarian failure: reappraisal and overview. *J Reprod Med.* **52**: 623-9.

Markoulaki, S., Kurokawa, M., Yoon, S.Y., Matson, S., Ducibella, T., and Fissore, R. 2007. Comparison of Ca2+ and CaMKII responses in IVF and ICSI in the mouse. *Mol Hum Reprod.* **13**: 265-72.

Mason, M., Fonseca, E., Ruiz, J.E., Moran, C., and Zarate, A. 1992. Distribution of follicle-stimulating hormone and luteinizing hormone isoforms in sera from women with primary ovarian failure compared with that of normal reproductive and postmenopausal women. *Fertil Steril.* **58**: 60-5.

McLaren, A. 2003. Primordial germ cells in the mouse. *Dev Biol.* **262**: 1-15.

Molyneaux, K.A., Stallock, J., Schaible, K., and Wylie, C. 2001. Time-lapse analysis of living mouse germ cell migration. *Dev Biol.* **240**: 488-98.

Molyneaux, K.A., Zinszner, H., Kunwar, P.S., Schaible, K., Stebler, J., Sunshine, M.J., O'Brien, W., Raz, E., Littman, D., Wylie, C. et al. 2003. The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. *Development.* **130**: 4279-86.

Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S., and Hannon, G.J. 2005. Characterization of Dicer-deficient murine embryonic stem cells. *Proc Natl Acad Sci U S A.* **102**: 12135-40.

Murchison, E.P., Stein, P., Xuan, Z., Pan, H., Zhang, M.Q., Schultz, R.M., and Hannon, G.J. 2007. Critical roles for Dicer in the female germline. *Genes Dev.* **21**: 682-93.

Nocka, K., Tan, J.C., Chiu, E., Chu, T.Y., Ray, P., Traktman, P., and Besmer, P. 1990. Molecular bases of dominant negative and loss of function mutations at the murine c-kit/white spotting locus: W37, Wv, W41 and W. *Embo J.* **9**: 1805-13.

Obernosterer, G., Martinez, J., and Alenius, M. 2007. Locked nucleic acid-based in situ detection of microRNAs in mouse tissue sections. *Nat Protoc.* **2**: 1508-14.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A. et al. 2005. Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature.* **436**: 207-13.

Rodriguez, A., Vigorito, E., Clare, S., Warren, M.V., Couttet, P., Soond, D.R., van Dongen, S., Grocock, R.J., Das, P.P., Miska, E.A. et al. 2007. Requirement of bic/microRNA-155 for normal immune function. *Science.* **316**: 608-11.

Runyan, C., Schaible, K., Molyneaux, K., Wang, Z., Levin, L., and Wylie, C. 2006. Steel factor controls midline cell death of primordial germ cells and is essential for their normal proliferation and migration. *Development.* **133**: 4861-9.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C.G., Zavolan, M., Svoboda, P., and Filipowicz, W. 2008. MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat Struct Mol Biol.* **15**: 259-67.

Tang, F., Kaneda, M., O'Carroll, D., Hajkova, P., Barton, S.C., Sun, Y.A., Lee, C., Tarakhovsky, A., Lao, K., and Surani, M.A. 2007. Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* **21**: 644-8.

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

Ventura, A., Young, A.G., Winslow, M.M., Lintault, L., Meissner, A., Erkeland, S.J., Newman, J., Bronson, R.T., Crowley, D., Stone, J.R. et al. 2008. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell.* **132**: 875-86.

Voorhoeve, P.M., le Sage, C., Schrier, M., Gillis, A.J., Stoop, H., Nagel, R., Liu, Y.P., van Duijse, J., Drost, J., Griekspoor, A. et al. 2006. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Cell.* **124**: 1169-81.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. 2007. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet.* **39**: 380-5.

Yang, W.J., Yang, D.D., Na, S., Sandusky, G.E., Zhang, Q., and Zhao, G. 2005. Dicer is required for embryonic angiogenesis during mouse development. *J Biol Chem.* **280**: 9330-5.

Yi, R., O'Carroll, D., Pasolli, H.A., Zhang, Z., Dietrich, F.S., Tarakhovsky, A., and Fuchs, E. 2006. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat Genet.* **38**: 356-62.

Youngren, K.K., Coveney, D., Peng, X., Bhattacharya, C., Schmidt, L.S., Nickerson, M.L., Lamb, B.T., Deng, J.M., Behringer, R.R., Capel, B. et al. 2005. The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. *Nature.* **435**: 360-4.

Zhao, Y., Ransom, J.F., Li, A., Vedantham, V., von Drehle, M., Muth, A.N., Tsuchihashi, T., McManus, M.T., Schwartz, R.J., and Srivastava, D. 2007. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell.* **129**: 303-17.

Chapter 4:

Conclusions and Future Directions

MicroRNAs mediate developmental regulation of gene expression via translational repression of target mRNAs. In this thesis, I have examined the role of a family of miRNAs, miR-290 through miR-295, which are specifically expressed during early embryonic and germ cell development. Members of the miRNA family have only been observed in mammals, with the total number of miRNAs varying widely in individual species. We have shown that members of this miRNA family are transcribed and processed from a common capped and polyadenylated primary transcript in the mouse. Deletion of the miR-290 family in the mouse results in an incompletely penetrant early embryonic lethality and primordial germ cell depletion in the female gonad. This female-specific germ cell depletion ultimately results in premature ovarian failure and infertility. We have also demonstrated that loss of miR-290-295 mediated repression results in significant changes in the gene expression profile of embryonic stem cells, allowing for the accumulation and precocious expression of many developmental regulators involved in differentiation, cell survival and proliferation. As such, we have shown that miR-290-295 are critical regulators of early development.

Target validation and implications for miRNA-mediated silencing.

To date, analysis of mammalian miRNA knock-outs have focused on miRNA regulation of a small number of tissue-restricted transcripts, often a single target encoding a transcription factor that controls a specific aspect of development, cell differentiation, or physiology. For example, deletion of *miR-223* results in a neutrophilic phenotype caused by increased progenitor proliferation and decreased granulocyte differentiation. Ablation of *Mef2c*, a transcription factor that promotes myeloid progenitor proliferation and is a

target of miR-223, suppresses progenitor expansion and corrects the neutrophilic phenotype in miR-223 null mice (Johnnidis et al., 2008). Findings such as these suggest that the primary mechanism by which miRNAs mediate developmental regulation is by the suppression of lineage specific developmental regulators.

However, the typical miRNA can potentially regulate numerous target genes (Lim et al., 2005). As such, miRNAs have been proposed to fine-tune gene expression programs and eliminate transcripts more rapidly than might occur by intrinsic decay (Farh et al., 2005). Transcripts that function in processes common to all cells tend to avoid miRNA regulation, whereas miRNA target genes are more frequently implicated in developmental processes (Stark et al., 2005). Ectopic expression of lineage specific miRNAs has been shown to shift the transcriptional profile of HeLa cells toward that of tissue where the endogenous miRNA was most abundant (Lim et al., 2005). These observations suggest that specific suppression of sets of tissue-specific transcripts may be important in mediating developmental transitions.

This raises a central question concerning miRNA-mediated regulation of gene expression, namely: Are the developmental phenotypes observed in miRNA deletion mutants caused by derepression of a small number of key genes (one miRNA, one target model) or by a global upregulation of transcripts bearing miRNA-specific seed matches (miRNA manager model)? Testing of these two models has been hampered by the inefficiency with which miRNA targets are identified and the difficulty of recapitulating *in vivo* developmental transitions in tissue culture. Currently bioinformatically predicted

miRNA/mRNA interactions are validated by 3'UTR binding assays and confirmed by observing changes in expression at the protein level, a process that likely instills a bias toward characterization of single targets. We have generated a list of potential miRNA targets and shown an enrichment for miR-290-295 seed matches in many upregulated transcripts (Chapter 3 of this thesis), including numerous developmental transcription factors that are normally silent (or expressed at low levels) in wild-type ES cells. It is currently unclear how these transcripts are regulated at the protein level in ES cells and whether or not specific dysregulation of expression of one (or many) is sufficient to recapitulate the precocious differentiation observed in miR-290 cluster knock-out ES cells. These studies will be critical to unmasking the molecular mechanism of the early embryonic lethality of miR-290-295$^{\Delta/\Delta}$ mice.

Examination of methylation defects in miR-290-295$^{\Delta/\Delta}$ ES cells.

Expression of the mir-290 cluster miRNAs is upregulated post-zygotically, ultimately becoming one of the most abundantly expressed miRNA families in the early embryo. Inactivating mutations of members of the miRNA biogenesis pathway exhibit a wide range of developmental phenotypes. Both miR-290 cluster and Dicer knock-outs are early embryonic lethal, with the Dicer null phenotype exhibiting complete penetrance by gastrulation. This observation suggests that some of the Dicer null phenotype may be explained by loss of miR-290 cluster repression during embryogenesis.

Examination of the contribution of miR-290-295 loss to the Dicer null phenotype should help to ascertain whether or not Dicer has a miRNA-biogenesis independent role in the

regulation of methylation status. Initial reports in Dicer-null embryonic stem cell lines indicated a global defect in the establishment of methylation in these cells, while subsequent reports have failed to find a substantial methylation defect (see Chapter I of this thesis). More recently, the *miR-290* cluster has been thrust into this debate, as two studies have implicated these miRNAs in indirect control of the *de novo* methyltransferases through an Rbl2 dependent mechanism (Sinkkonen et al., 2008, Benetti et al., 2008). These studies also reported a reduction in AcH3K9-associated and sub-telomere methylation in Dicer null ES cells, which was partially rescued by reintroduction of miR-290 cluster miRNAs.

If miR-290-295 play a significant role in the regulation of methylation via regulation of the Rb family proteins, one expectation would be that levels of the *de novo* methyltransferases are altered in mIR-290-295$^{\Delta/\Delta}$ ES cells. We have preliminary evidence that this is true, as microarray data suggest that Dnmt3 is found at lower levels in miR-290 cluster knock-out ES cells. Whether or not this is a direct effect of Rbl2 is presently unclear. Establishment of ES cell lines containing conditionally responsive knock-downs of the methyltransferases will be critical in assessing the miR-290 cluster's influence on global methylation patterns.

Timing of germ cell loss in embryonic germ cells

Normally, a small number of primordial germ cells (PGCs) are set apart from somatic cells early in development at the posterior end of the developing embryo and migrate toward the future gonad. As PGCs arrive in the gonad they begin differentiation into

germ-line stem cells capable of forming the gametes (McLaren, 2003). Impairment at any step in the process, i.e. specification, migration or proliferation during migration is detrimental to the embryonic germ cell population. Assessment of these phenomena is complicated by both the small size of the starting population of primordial germ cells and the period during development at which germ cells are specified and then migrate to the gonad.

One striking observation in miR-290-295 mutants is the sex-specific recovery of germ cell number in the male gonad. Given that germ cell loss in *miR-290-295* deficient mice seems to occur before primordial germ cells reach the gonad suggests that a defect in germ cell specification and/or migration exists. The timing of germ cell loss raises an interesting set of questions that can be addressed by the generation of tools that will aid in the identification of small numbers of germ cells in early embryos. For example, the role of *miR-290* cluster miRNAs in both specification and migration could be assessed by crossing an *Oct4*-GFP allele into the *miR-290-295*$^{\Delta/\Delta}$ strain. This would enable germ cells to be more easily followed, as *Oct4* remains actively expressed in the non-somatic cell population at this time during development. These experiments would also aid in understanding whether or not the recovery of embryonic germ cells observed in males is due to an intrinsic defect of XX germ cells, such as a problem in X-inactivation or simply due to the ability of XY PGCs to proliferate for a longer period of time.

RNA binding proteins may affect miR-290-295 regulation in germ cells

Recent evidence suggests that the processing of primary miRNA transcripts may be blocked post-transcriptionally by some RNA binding proteins (RNAbp). For example, the RNAbp Lin28 selectively blocks the processing of pri-let-7 miRNAs in ES cells, implicating it as a negative regulator of miRNA biogenesis and suggesting that it may inhibit miRNA-mediated repression of differentiation in stem cells (Viswanathan et al., 2008). The RNAbp *Dead End* (*Dnd1*) has been shown to supress function of *miR-430* in zebrafish, at least in part through blocking miRNA accessibility by binding to U-rich mRNA regions (Kedde et al., 2007). This finding is interesting, as zebrafish *miR-430* has a seed sequence that is highly similar to the mouse *miR-290* and human *miR-373* family microRNAs, suggesting that *Dnd1* may regulate function of these miRNAs.

Several RNA binding proteins, including *Dead end* are expressed in germ cells. *Dead end* (*Dnd1*) has been shown to be essential for PGC motility in mice, as mice with a naturally occurring truncation mutation in the *Dnd1* gene (the *Ter* mutation) lose most PGCs (Youngren et al, 2005). In one genetic background (129-strain), this Dnd1 mutation results in testicular tumor formation, with tumors arising from the few surviving germ cells. Inhibition of *Dnd1* expression in human testicular germ cell tumor cell lines resulted in a depletion of *miR-373* family targets, suggesting a balance between *miR-373* family function and *Dnd1* in germ cells (Kedde et al., 2007).

It would be interesting to determine whether *Dnd1*-mediated abrogation of miR-290-295 activity is critical to germ cell migration and proliferation. Expression of *Dnd1* has been

observed in germ cells during specification and is known to be differentially expressed in male and female PGCs by E12.5, where it is upregulated in the XY gonad and downregulated in the XX gonad. During this period of development, XX germ cells begin entry into meiosis, whereas XY germ cells continue proliferation and populate the testis cords. Given that the human homologs of this cluster, *miR-371-373*, are upregulated in some testicular cancers (Voorhoeve et al, 2007), it is possible that Dnd1 may be acting to limit a proliferative stimulus provided by miR-290-295 expression. Establishment of miR-290-295 deficient embryonic germ cell lines may provide a means of exploring this regulation.

Acknowledgements

I would like to thank R. Jaenisch for critical comments.

References

Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P., *et al.* (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. Nat Struct Mol Biol *15*, 268-279.

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

Johnnidis, J. B., Harris, M. H., Wheeler, R. T., Stehling-Sun, S., Lam, M. H., Kirak, O., Brummelkamp, T. R., Fleming, M. D., and Camargo, F. D. (2008). Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. Nature *451*, 1125-1129.

Kedde, M., Strasser, M. J., Boldajipour, B., Vrielink, J. A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P. M., van Duijse, J., Orom, U. A., *et al.* (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273-1286.

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

McLaren, A. (2003). Primordial germ cells in the mouse. Dev Biol *262*, 1-15.

Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nat Struct Mol Biol *15*, 259-267.

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133-1146.

Viswanathan, S. R., Daley, G. Q., and Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. Science *320*, 97-100.

Voorhoeve, P. M., le Sage, C., Schrier, M., Gillis, A. J., Stoop, H., Nagel, R., Liu, Y. P., van Duijse, J., Drost, J., Griekspoor, A., *et al.* (2007). A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Adv Exp Med Biol *604*, 17-46.

Youngren, K. K., Coveney, D., Peng, X., Bhattacharya, C., Schmidt, L. S., Nickerson, M. L., Lamb, B. T., Deng, J. M., Behringer, R. R., Capel, B., *et al.* (2005). The Ter mutation

in the dead end gene causes germ cell loss and testicular germ cell tumours. Nature *435*, 360-364.

Appendix I.

Generation and characterization of *miR-290-295* hypomorph.

Personal Contributions

This appendix is the result of a collaboration with Hristo Houbaviy. For the study presented in this appendix, I designed and constructed the mouse mutation and performed all the breeding experiments. I was also responsible for overseeing the experimental design and interpretation and prepared the manuscript. Hristo Houbaviy performed the Northern blot analysis.

Introduction

In order to study *miR-290-295* function, I sought to examine the consequence of targeted deletion of the *miR-290* cluster in the developing mouse (see Chapter 3 of this thesis). In addition to the null *miR-290-295*$^{\Delta/\Delta}$ allele previously described, a hypomorphic *miR-290-295* allele (*miR-290-295*^{hyp}) was also generated. This appendix examines the molecular characterization of the *miR-290* cluster locus in hypomorphic animals, expression from the hypomorphic locus and describes the embryonic lethality of *miR-290-295*$^{hyp/hyp}$ animals.

Results

miR-290-295^{hyp} generation.

The *miR-290-295*^{hyp} allele arose from inappropriate recombination of the *miR-290* cluster targeting construct with genomic DNA in V6.5 ES cells. Based upon Southern blot analysis with several restriction enzymes and DNA probes external to the targeting construct, it was determined that one end of the targeting construct had undergone homologous recombination, whereas the other end had inserted into the genome, resulting in a tandem duplication of some genomic sequences. This arrangement is outlined in Figure A1.1.

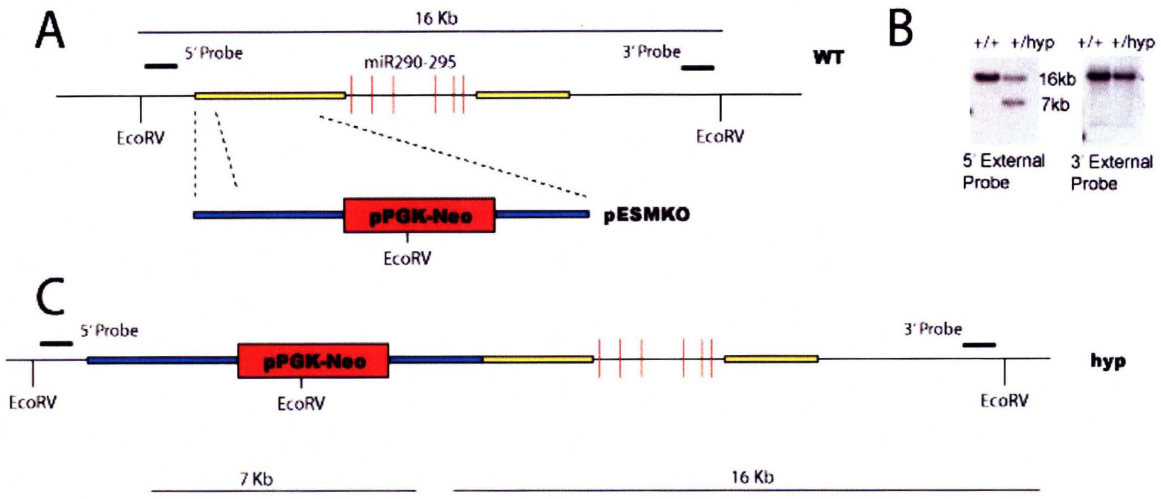

Figure A1.1. Generation of the miR-290-295 hypomorphic allele.
(A) pESMKO targeting vector recombined at the 5' end, while its 3' end inserted downstream of the 5' recombination site.
(B) Southern blot analysis of genomic DNA from wild-type and hypomorph heterozygous ES cells. EcoRV digestion.
(C) Graphical depiction of the miR-290-295 hypomorphic allele.

Analysis of expression from the hypomorphic locus.

miR-290-295 expression was assessed in homozygous *miR-290-295*$^{hyp/hyp}$ ES cell lines. Levels of both the precursor and mature miRNAs from the cluster were downregulated 5-10 fold in *miR-290-295*$^{hyp/hyp}$ ES cells compared to controls (Figure A1.2). This suggests that reduced expression of miRNA locus is the result of the insertion of the tagerting construct. The mechanism by which miR-290 cluster transcription has been reduced remains unclear, but could include disruption of the promoter or upstream enhancer elements.

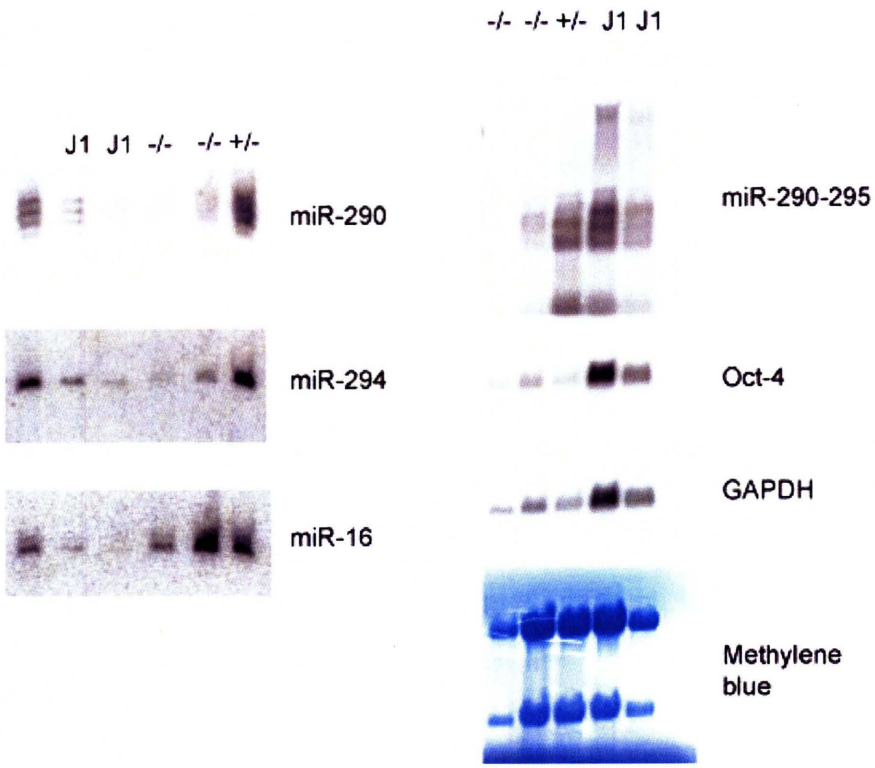

Figure A1.2. miR-290-295 expression in hypomorphic ES cells.
Northern blot analysis of the miR-290 cluster in hypomorphic ES cells. Total RNA from J1 ES cells are shown as a control. Total expression from the hypomorphic allele is reduced 5-10 fold. Oct4, GAPDH, miR-16 and Methylene blue are shown as loading controls.

Characterization of *miR-290-295*$^{hyp/hyp}$ mice.

Mice heterozygous for hypomorphic allele were fertile and indistinguishable from wild-type littermates and were maintained on a mixed genetic background (129Sv/J x C57BL/6). Homozygous *miR-290-295*$^{hyp/hyp}$ progeny were produced from *miR-290-295*$^{+/hyp}$ intercrosses, and observed at a significantly lower frequency than the expected 1:2:1 frequency predicted by Mendelian segregation (Figure A1.3A and B). Only 43 of 170 postnatal progeny from *miR-290-295*$^{+/hyp}$ matings were *miR-290-295*$^{hyp/hyp}$, indicating that ~50% of the homozygous mutants were lost during development. The embryonic loss of hypomorphic homozygotes is similar to that observed for the null allele, although

at a decreased severity (see Chapter 3). This suggests that *miR-290-295* loss-of-function mediated embryonic lethality is dependent upon the expression level of the *miR-290* cluster miRNAs.

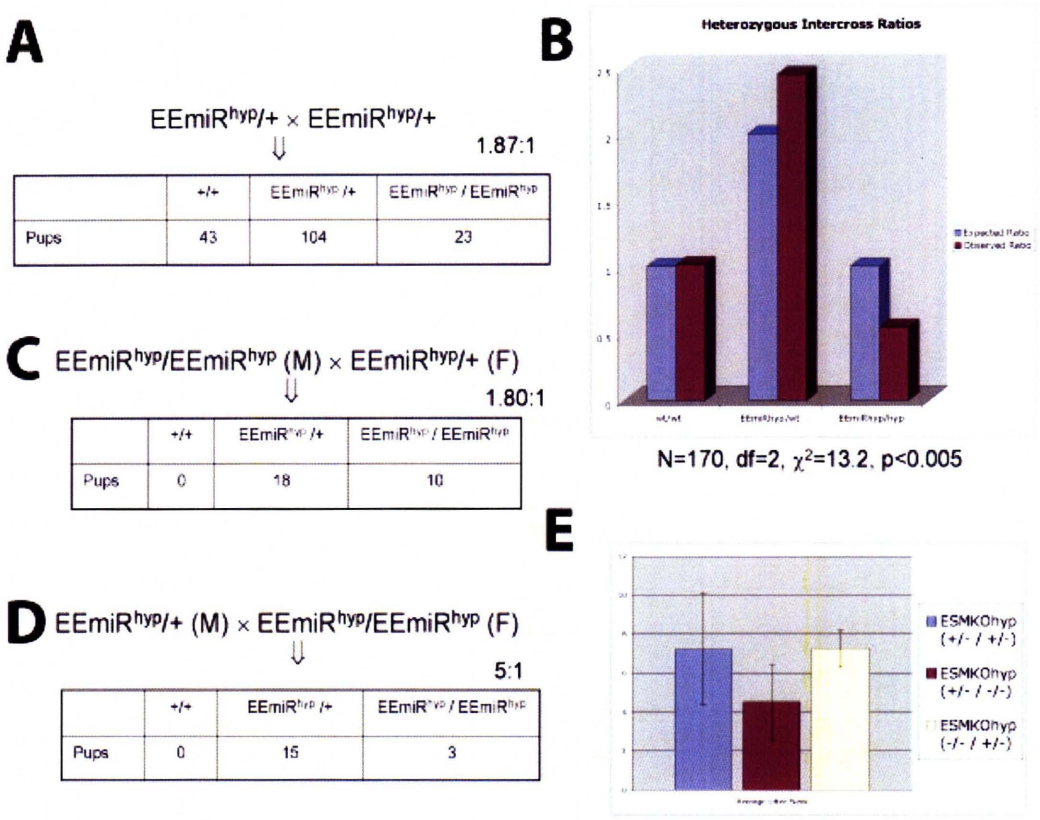

Figure A1.3. miR-290-295 hypomorph breeding data

(A) Postnatal progeny from miR-290-295 hypomorph heterozygous intercrosses. Ratio in upper left is (+/+ over hyp/hyp), indicating that ~50% of hypomorphic homozygotes are lost during gestation.

(B) Graphical depiction of data in (A). Chi-square analysis indicates that genotypic skewing is statistically signficant.

(C-D) Postnatal progeny from the indicated crosses, indicating that hypomorphic males (B) and females (C) are fertile.

(E) Average litter size from the indicated cosses.

Both *miR-290-295*$^{hyp/hyp}$ males and females were fertile (Figure A1.3C and D). However, only 2 of 9 hypomorphic females produced litters. Each female only produced a single

litter, despite consistent caging with a fertile male. This suggests *miR-290-295*$^{hyp/hyp}$ females are sub-fertile and is consistent with observed infertility and premature ovarian failure in *miR-290-295* null females (see Chapter 3). Average litter size was not significantly different among the three crosses illustrated (Figure A1.3E), indicating that the reduced number of *miR-290-295*$^{hyp/hyp}$ progeny from hypomorphic females may be a result of the small sample size.

Conclusions

Targeting of the *miR-290* cluster resulted in the generation of a partial loss-of-function allele that expresses 10-20% of wild-type levels of mature *miR-290-295* miRNAs. Approximately, 50% of homozygous hypomorphic animals die during gestation, likely in a similar manner to *miR-290-295* null animals. While hypomorphic females are able to produce progeny unlike null females, they are sub-fertile, consistent with a reduced reproductive capacity.

The reduced severity of both the embryonic lethality and fertility phenotypes in hypomorphic versus null animals is striking and indicates that *miR-290-295* expression levels are critical during development. If these miRNAs are important for silencing of developmental regulators (see Chapter 3), these data suggest that high concentrations of miR-290-205 miRNAs are necessary for proper regulation. This suggests a model in which high levels of miR-290 cluster miRNAs act to buffer cells against precocious differentiation and promote efficient transition from one cell identity to another.