MicroRNAs regulate *de novo* DNA methylation and histone mRNA 3' end formation in mammalian cells

Inauguraldissertation

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

von

Lasse Sinkkonen
aus Imatra, Finnland

Basel, 2008

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Professor Dr. Witold Filipowicz und Professor Dr. Mihaela Zavolan.

Professor Dr. Witold Filipowicz	Professor Dr. Mihaela Zavolan
(Referent)	(Koreferent)

Basel, 16.9.2008

Professor Dr. Eberhard Parlow
(Dekan)

Acknowledgements

First of all, I would like to thank Witold Filipowicz for giving me the opportunity to do my PhD studies under his supervision in a new and exciting field. Witek is a supportive, inspiring mentor and a great scientist with passion for his work.

I also wish to thank the members of my thesis committee, Helge Grosshans and Dirk Schübeler, for their critique and encouragement during our meetings as well as outside them.

Thank you to Mihaela Zavolan for her ideas and support, and for being the co-referee of this thesis. Together with her students, Philipp Berninger and Dimos Gaidatzis, Mihaela helped me to understand how much bioinformatics can do for us.

I wish to thank all the former and present members of the Filipowicz group. During the past 4 years, I have had the opportunity to work with more than 30 different group members with equally many backgrounds. You have all contributed to my studies and made it a unique experience.

Special thanks to Petr Svoboda, supervisor of my PhD studies. Petr showed amazing patience by tolerating my endless questions and correcting my rough drafts into the early morning. His mind can create more projects than one can ever undertake and he has constantly new ideas, especially after a visit to the PB.

Special thanks go also to Caroline Artus-Revel and Tabea Hugenschmidt. They have greatly helped me in all aspects in the lab and many aspects outside the lab. They have taught me several techniques and contributed to plenty of important experiments.

I would like to thank Fabio Mohn for sharing his reagents and expertise on studying epigenetics of embryonic stem cells.

I am thankful for the great facilities at the FMI. Especially I will remember the discussions with Ed Oakeley and the FACS expertise and tea offered by Hubertus Kohler.

I am grateful for my family for their continuous support and belief in me. Their encouragement has always been important for me.

Finally, I wish to thank Anne-Maria, for her support and understanding that have allowed me to pursue my ambitions, and most of all, for her love.

Abbreviations

ARE	AU-rich element
BS	bisulphite sequencing
CBC	Cap-binding complex
cDNA	complementary DNA
CDS	coding sequence
ChIP	chromatin immunoprecipiation
cpm	counts per minute
DNMT	DNA methyltransferase
dsRNA	double-stranded RNA
ESC	embryonic stem cell
FCS	fetal calf serum
GO	gene ontology
GSC	germ-line stem cell
H3K27me3	trimethylated lysine 27 of histone H3
H3K4me2	dimethylated lysine 4 of histone H3
H3K9me3	trimethylated lysine 9 of histone H3
HDE	histone downstream element
HIST	histone gene cluster
HMT	histone methyltransferase
ICM	inner cell mass
kb	kilobase
KD	knock-down
LIF	leukemia inhibitor factor
miRNA	microRNA
miRNP	micro-ribonucleoprotein
mRNA	messenger RNA
natsiRNA	natural-antisense transcript-derived siRNA
NELF	nuclear elongation factor
NP	neuronal precursor
nt	nucleotide
P-body	processing body
piRNA	Piwi-associated RNA
PRC	Polycomb group repressive complex
PRE	Polycomb response element
pre-miRNA	miRNA precursor
pri-miRNA	primary miRNA transcript
PTGS	post-transcriptional gene silencing
RA	retinoic acid

rasiRNA	repeat-associated siRNA
RISC	RNA-induced silencing complex
RNAi	RNA interference
RPA	RNase protection assay
RT-qPCR	real-time quantitative reverse transcription-PCR
shRNA	short haipin RNA
siRNA	short interfering RNA
snRNA	small nuclear RNA
ta-siRNA	trans-acting siRNA
TE	Tris-EDTA
TN	terminal neuron
tRNA	transfer RNA
TSS	transcription start site
UTP	uridine triphosphate
UTR	untranslated region

Table of Contents

1. Summary

MicroRNAs (miRNAs) are known to have many important functions in mammalian cells. They can influence the expression of their target genes and in this way regulate the function of not only their primary targets, but also of the pathways and mechanisms acting downstream of the primary targets. There are several key proteins that are required for the biogenesis of miRNAs and for mediating the repressive functions of miRNAs in mammals, the most critical being the ribonuclease (RNase) III enzyme Dicer. Since Dicer is required for generation of all known mammalian miRNAs, depletion of Dicer is an appealing strategy to identify and study the pathways under miRNA-mediated control.

Deletion of Dicer in mouse embryonic stem cells (ESCs) is rendering the cells to slow growth rate and inability to differentiate, and thus, to loose their most important feature i.e. pluripotency. We aimed to understand in further detail the causes behind these critical defects. We have performed transcriptional profiling of Dicer-deficient ESCs and through bioinformatic analysis we identified miRNAs of the ESC-specific miR-290 cluster to be functionally most important for mouse ESCs. These miRNAs were found to directly control the expression of several hundred primary targets and through their regulation influence many features of the ESCs. We found the miR-290 miRNAs to contribute to the growth rate of the ESCs and to influence also expression of many secondary target genes. Among their secondary targets we identified *de novo* DNA methyltrasferases (DNMT3s) that were significantly downregulated in Dicer-deficient mouse ESCs. The downregulation was due to an increased expression of Retinoblastoma-like2 (RBL2), a transcriptional repressor and primary target miR-290 miRNAs. As a consequence of lowered DNMT3 expression the cells were unable to methylate DNA at the promoter of pluripotency genes such as *Oct-4* (Octamer-binding transcription factor-4, also known as *Pou5f1* for POU-domain, class 5, transcription factor 1), and thus, incapable of fully silencing these genes during differentiation. Hence, regulation of DNMT3s by miR-290 miRNAs is contributing to the maintenance of mouse ESC pluripotency.

Further analysis of the promoter of primary miR-290 transcript (pri-miR-290) showed that the ESC specific expression and subsequent silencing of the transcript during

neuronal differentiation is regulated by the chromatin status of the promoter. During neuronal differentiation the pri-miR-290 promoter looses histone modifications characteristic of active genes and gains typical marks of silenced chromatin. This is followed by *de novo* DNA methylation of the pri-miR-290 promoter. It is likely that the silencing of pri-miR-290 depends on DNA methylation of its promoter, thus allowing an auto-regulatory loop between the miRNAs and DNMT3 enzymes.

In addition to Dicer-deficient mouse ESCs, we have studied the importance of Dicer as well as Argonaute proteins for the function of human cell lines by inducibly depleting these proteins in human HEK293T-REx cells. We observed that an intact RNA silencing pathway is needed for normal expression of many of the replication-dependent histone genes. We found up to 25% of all histone mRNAs to be upregulated upon loss of RNAi machinery and more detailed analysis of one of the histone genes, HIST1H3H, demonstrated that the upregulation was due to enhanced polyadenylation of the histone mRNA. This is in contrast to the normal 3' end processing of replication-dependent histone mRNAs that takes place at the 3' end-proximal stem-loop and is not followed by polyadenylation. The analysis of RNA from *Dicer*- or *Dgcr8*-deficient ESCs showed that this type of regulation of 3' end formation by RNA silencing pathway is conserved in mice and depends on the generation of miRNAs. Thus, miRNAs seem to regulate the 3' end processing of replication-dependent histone mRNAs. Future work will be needed to identify specific miRNAs and processing factors involved.

2. Introduction

2.1 Gene regulation by small RNAs

It has become evident that non-coding RNA molecules play pivotal regulatory roles in eukaryotic cells, indicating that these cells are more complex than would be expected simply based on the number of their protein coding genes. Our understanding of these regulatory phenomena has substantially increased during the past decade with the discovery and characterization of various classes of small regulatory RNAs (21- to 30-nt in length). The early work in plants had described post-transcriptional gene silencing (PTGS) where expression of a transgene was capable of suppressing other homologous sequences, suggesting a regulatory role for RNA (Napoli et al. 1990; Hobbs et al. 1993; Lindbo et al. 1993; English et al. 1996). But it was the experiments of Andrew Fire and Craig Mello showing double stranded RNA (dsRNA) as a potent inducer of gene silencing or RNA interference (RNAi) in nematode *Caenorhabditis elegans*, that provoked great interest into the regulatory function of RNA (Fire et al. 1998). Subsequent research in many different species has revealed that dsRNA is processed into short interfering RNAs (siRNAs, 21- to 25-nt in length) that guide the cleavage of their cognate target RNAs (Hamilton and Baulcombe 1999; Hammond et al. 2000; Zamore et al. 2000).

The discovery of siRNAs has been followed by identification of many other small regulatory RNAs. miRNAs were originally identified as non-coding developmental regulators in *C. elegans* and were later found to be evolutionary ancient, endogenously encoded, small RNAs (21- to 25-nt in length) capable of regulating the translation of their target mRNAs (Lee et al. 1993; Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). miRNAs are now known to play important roles in many cellular processes (see chapter 2.3). In addition to siRNAs produced from exogenous dsRNA, endogenously encoded siRNAs have been described in many different species. Plants have the biggest variety of endogenous siRNAs ranging from trans-acting siRNAs (ta-siRNAs) and natural-antisense transcript-derived siRNAs (natsiRNAs) to repeat-associated siRNAs (rasiRNAs), which differ from each other in both their biogenesis as

well as function (Vazquez 2006). Improved high-throughput sequencing technologies have allowed detection of endogenous siRNAs also in fission yeast *Schizosaccaromyces pombe*, nematode *C. elegans* and more recently also in fruit fly *Drosophila melanogaster* as well as mouse oocytes (Cam et al. 2005; Ruby et al. 2006; Czech et al. 2008; Tam et al. 2008; Watanabe et al. 2008). The main function of these siRNAs appears to be the repression of retrotransposons and other repetitive sequences. At least in plants and fission yeast the endogenous siRNAs can direct transcriptional silencing and chromatin condensation at the homologous sites of the genome (Wassenegger et al. 1994; Mette et al. 2000; Volpe et al. 2002; Verdel et al. 2004).

The most recently identified class of small regulatory RNAs is that of Piwi-associated RNAs (piRNAs). The biogenesis of piRNAs differs from siRNAs and miRNAs, which is reflected by their slightly longer length (24- to 30-nt), and piRNAs are produced from single-stranded precursors (Aravin et al. 2006; Grivna et al. 2006; Lau et al. 2006; Brennecke et al. 2007). piRNAs are specifically expressed in germ cells and seem to mediate their function through association with the Argonaute-related effector proteins called Piwi-proteins. The exact mechanisms of piRNA function through Piwi-proteins remains elusive but genetic studies in *D. melanogaster*, zebrafish and mice suggests that they are necessary for germline development and, similarly to endogenous siRNAs, needed for retrotransposon silencing (Cox et al. 1998; Deng and Lin 2002; Aravin et al. 2004; Kuramochi-Miyagawa et al. 2004; Carmell et al. 2007; Houwing et al. 2007). According to recent data, this silencing appears to be medaited by DNA methylation of the repeat sequences (Kuramochi-Miyagawa et al. 2004).

In the following chapters of the introduction I will be focusing on the biogenesis of siRNAs and miRNAs, mechanism of their function in RNA silencing - including the target recognition by miRNAs - and on the biological function of miRNAs in animals. Especially I will focus on the different cellular roles of miRNAs in mammals.

2.2 Mechanism of RNA silencing

Both siRNAs and miRNAs are processed from dsRNA precursors into mature 21- to 25-nt RNA duplexes by RNase III type enzyme called Dicer. Following this processing, they are loaded into a multiprotein-complex called RNA-induced silencing complex (RISC) (or micro-ribonucleoprotein (miRNP) complex in the case of miRNAs, see Figure 1).

This is considered the initiation phase of RNA silencing. It is followed by the effector phase where the mature siRNA or miRNA guides the RISC/miRNP to the correct target mRNA to induce its silencing.

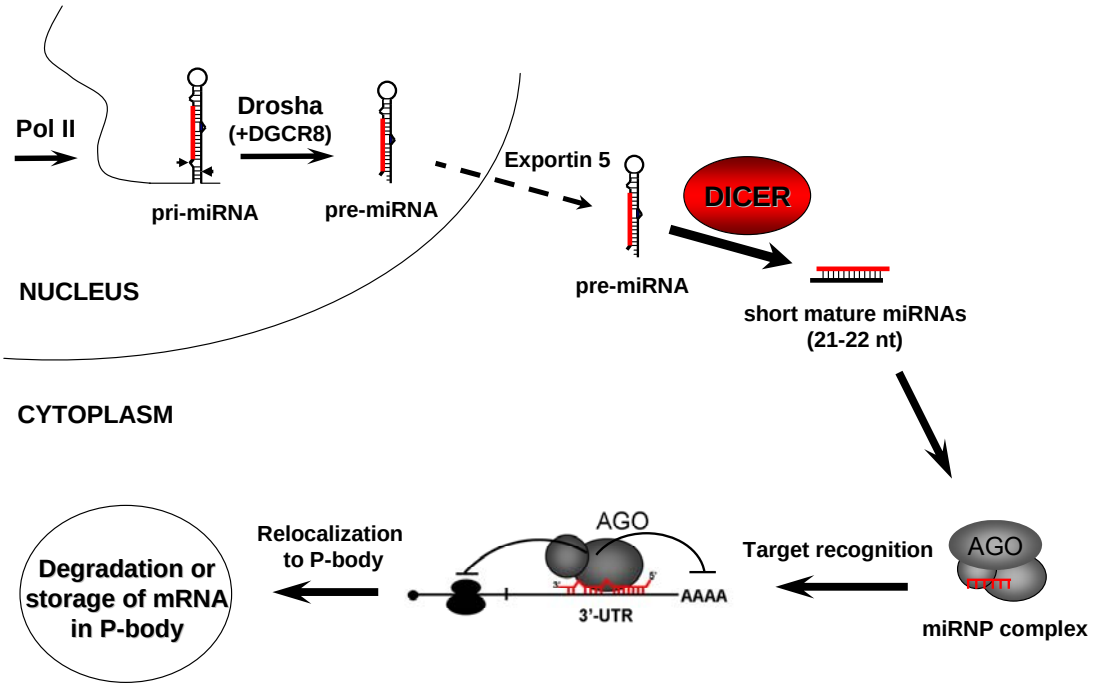


Figure 1. Biogenesis and function of miRNAs.

Primary miRNA transcripts are transcribed by RNA Pol II in the nucleus where they are processed by RNase III type enzyme Drosha and its dsRNA-binding partner DGCR8 into miRNA precursors. Precursor of the miRNA is exported to the cytoplasm where it is further processed by another RNase III type enzyme Dicer into a mature miRNA duplex. The strand with lower stability at its 5' end (in red) is selected to be loaded on to the miRNP complex. The miRNA guides the Argonaute protein and rest of the complex to the correct target mRNA that becomes translationally repressed and destabilized. This is accompanied by relocalization of the mRNA to a P-body.

2.2.1 miRNA and siRNA biogenesis – Dicer as a key enzyme

The main difference between siRNAs and miRNAs is the source of their double-stranded precursor-RNA. The long dsRNA precursors of siRNAs can derive from antisense transcription, viral replication or for example transfection. miRNAs on the other hand are

RNA-polymerase II (RNA Pol II) transcripts of variable length that are 5' capped and polyadenylated (Cai et al. 2004). Still, the majority of miRNAs seem to arise from introns of protein coding genes (Kim and Kim 2007). These primary miRNA transcripts (pri-miRNAs) usually give rise to several different mature miRNAs. Such a group of co-transcribed miRNAs is called a miRNA cluster. They are processed in the nucleus by the Microprocessor complex containing RNase III enzyme Drosha and a double-stranded RNA binding protein DGCR8 (DiGeorge syndrome critical region gene-8 in vertebrates, Pasha in invertabrates) into around 70-nt imperfect hairpin structures called miRNA precursors (pre-miRNAs) (Denli et al. 2004; Gregory et al. 2004). Recent data has also indicated existence of so called mirtrons, miRNAs derived from introns through splicing, independently of Drosha and DGCR8 (Berezikov et al. 2007; Okamura et al. 2007; Ruby et al. 2007).

After the pre-miRNAs are exported into the cytoplasm by Exportin 5, like siRNAs, they are further cleaved by the RNase III enzyme Dicer (Hutvagner et al. 2001; Yi et al. 2003; Lund et al. 2004). Mammals and *C. elegans* have only one Dicer gene while *D. melanogaster* has two Dicers, Dcr-1 for miRNA production and Dcr-2 for siRNA production (Lee et al. 2004). Thus, in *D. melanogaster* miRNA and siRNA pathways are genetically diverged. Dicer measures approximately two helical turns from the Drosha cleavage site to produce 21- to 25-nt RNA duplex that has 2-nt 3'-overhangs, hallmarks of RNase III enzyme cleavage. Together with its interacting partner TRBP (TAR RNA binding protein), Dicer recruits one of the Argonaute proteins (AGO1 to AGO4 in mammals) to form a trimeric complex (Chendrimada et al. 2005; Haase et al. 2005). This initiates the formation of the RISC/miRNP (Gregory et al. 2005). Only one strand of the small RNA duplex, the guide strand, is loaded on to the RISC/miRNP and into the RNA binding pocket of the Argonaute protein, while the other strand, called the passenger strand, is degraded. The guide strand is selected based on the stability of the base-pairing at the 5' end of the RNA duplex so that the strand with lower stability is loaded on to the RISC/miRNP (Schwarz et al. 2003). Argonautes are considered to be the effector proteins of the RISC/miRNP. This is reflected for example by their ability to repress protein synthesis, when they are artificially tethered to the 3' untranslated region (3' UTR) of a reporter mRNA, independently of miRNAs (Pillai et al. 2004).

It has been shown that, in addition to transcriptional regulation, the biogenesis of miRNAs can be regulated both at the level of Drosha cleavage as well as at the level of Dicer cleavage (Obernosterer et al. 2006; Thomson et al. 2006; Davis et al. 2008; Newman et al. 2008; Rybak et al. 2008; Viswanathan et al. 2008). But the fact that all siRNAs and miRNAs require Dicer for their maturation makes Dicer the key enzyme necessary for RNA silencing.

2.2.2 The effector phase of RNAi and miRNA pathways

Once bound by the Argonaute protein of the RISC/miRNP, the siRNA or the miRNA can direct the complex to the correct target mRNA. This happens by basepairing between the guide RNA and the target mRNA, typically at the 3' UTR of the mRNA. When this interaction happens through perfect complementarity, a characteristic of siRNAs, it leads to endonucleolytic cleavage of the target mRNA in the middle of the interaction between positions 10 and 11 of the siRNA. This cleavage, referred to as slicing, can be mediated only by one of the mammalian Argonaute proteins, AGO2, and is catalyzed by the RNase H fold in the PIWI-domain of the protein (Liu et al. 2004; Meister et al. 2004). Only one miRNA has been shown to induce AGO2 mediated slicing (Yekta et al. 2004). However, animal miRNAs usually bind to their target mRNAs with partial complementarity and induce repression of protein synthesis.

The exact mechanism of repression of protein synthesis is still under debate and several different mechanisms have been proposed. Initial experiments aiming to address the mechanism of miRNA-mediated silencing showed that the cognate mRNAs of the original *C. elegans* miRNA lin-4 were associated with polyribosomes, arguing that repression by the miRNA takes place after the initiation of translation (Olsen and Ambros 1999; Seggerson et al. 2002). Degradation of the nascent polypeptide was suggested as one of the possible mechanisms. Later studies were able to confirm the association of the target mRNAs as well as of the miRNAs with polyribosomes in human cells but excluded peptide degradation as a possible mechanism of function (Maroney et al. 2006; Petersen et al. 2006). Instead, miRNAs were suggested to cause the ribosomes to drop off and prematurely terminate the translation of the repressed target mRNAs. This model is in conflict with the accumulating evidence for miRNA-mediated repression at the translational initiation. Experiments using reporter genes carrying let-7 binding sites in

their 3' UTRs have shown that m^7G-cap of the mRNA is necessary for translational repression (Humphreys et al. 2005; Pillai et al. 2005). This observation has been supported by several different *in vitro* assays using cell-free extracts from different species (Wang et al. 2006; Mathonnet et al. 2007; Thermann and Hentze 2007; Wakiyama et al. 2007). In addition to m^7G-cap, these studies also suggest a role for poly-A tail in miRNA-mediated repression. This is consistent with the model for inhibition of translational initiation, since poly-A-tail and the poly-A binding protein (PABP) are known to work in synergy with the m^7G-cap to regulate translational initiation (Kahvejian et al. 2005). Recently, a compromise to resolve the conflicting data supporting repression on initiation and repression on elongation was suggested. Kong et al. propose that the method of repression would be dependent on the promoter driving the expression of the target mRNA i.e. the nuclear history of the mRNA might determine its destiny in regard to miRNA-mediated repression (Kong et al. 2008).

Also additional proteins called GW182 proteins (GW182A to GW182C in mammals, GW182 in *D. melanogaster*) and their *C. elegans* homolog AIN-1 have been shown to be essential for miRNA-mediated repression (Ding et al. 2005; Liu et al. 2005; Eulalio et al. 2008). A direct interaction between GW182 and the Argonaute protein was found to be necessary for miRNA-induced repression, signifying that GW182 is mediating the repressive activity of the miRNA-bound Argonaute. This fits with the fact that the repressed mRNAs, miRNAs, as well as many components of the RNA silencing pathway, including Argonautes and GW182 proteins, accumulate in discrete cytoplasmic foci called GW-bodies or processing bodies (P-bodies) (Jakymiw et al. 2005; Liu et al. 2005; Pillai et al. 2005; Sen and Blau 2005; Bhattacharyya et al. 2006). Since the Argonaute proteins can be found distributed throughout the cytoplasm, in addition to their P-body localization, it is likely that they initiate the repression of the target mRNA in the cytoplasm outside of P-bodies, which is then later followed by accumulation into the P-bodies. The exact order of these events is still unknown. But interestingly, intact miRNA biogenesis and RNA silencing machinery are required for formation of P-bodies, supporting the idea that P-body accumulation of RISC/miRNP is a secondary effect of RNA silencing (Pauley et al. 2006; Eulalio et al. 2007). Because siRNA-loaded AGO2 can slice its target mRNA itself immediately after recognition, it would be reasonable to

suggest that P-body formation depends only on miRNA function. But curiously, also depletion of DCR-2 or AGO2, proteins specific for the RNAi pathway in *D. melanogaster*, is sufficient to disrupt P-bodies (Eulalio et al. 2007).

The P-bodies were originally identified as conserved sites of mRNA storage and degradation that contain a plethora of proteins required for different aspects of mRNA turnover such as decapping, deadenylation and exonucleolytic activity (reviewed in (Parker and Sheth 2007). Such colocalization of RNA silencing pathway and miRNAs with the mRNA decay machinery would argue for degradation of miRNAs targets, in addition to their translational inhibition. This indeed seems to be the case. Schmitter et al. showed that repression of reporter gene construct by endogenous let-7 is accompanied by mRNA degradation in human cells, more so in HEK293 than HeLa cells (Schmitter et al. 2006). In *C. elegans* the endogenous target mRNAs of miRNAs let-7 and lin-4, as well as transgene reporter mRNAs carrying response elements for these miRNAs, were shown to be downregulated in their translational efficiency as well as at the mRNA level, when the miRNAs were expressed (Bagga et al. 2005). Similarly, miR-125b was shown to target LIN28 during differentiation of mouse embryonal carcinoma cells and, in addition to downregulation of the protein, also the lin28 mRNA was reduced (Wu and Belasco 2005). This regulation too could be recapitulated using reporter gene constructs. Further analysis of miR-125b mediated silencing in human cell lines revealed that the mRNAs targeted by miR-125b were not cleaved at the miRNA binding site but were targeted for removal of their poly-A tail (Wu et al. 2006). Interestingly, replacement of the poly-A tail by histone 3' end stem-loop stabilized the mRNA but did not fully rescue the translation, indicating that the translational inhibition and mRNA decay are working in an additive manner. Observations supporting the role of miRNAs in target mRNA deadenylation have been also made in zebrafish where miR-430 has been shown to be responsible for deadenylation and removal of hundreds of maternal transcripts during early embryogenesis (Giraldez et al. 2006). The most detailed analysis of miRNA induced mRNA degradation was done with S2 cells of *D. melanogaster* (Behm-Ansmant et al. 2006). These experiments further strengthened the importance of GW182 in miRNA function by showing that tethered GW182 alone was sufficient to silence a reporter gene mRNA independently of the Argonaute protein or the miRNA. Notably, the GW182

induced mRNA decay was accompanied by deadenylation of the mRNA. And depletion of CCR4:NOT deadenylation complex or DCP1:DCP2 decapping complex, all of which are components of P-bodies, was sufficient to alleviate the mRNA degradation. Thus, miRNA-mediated RNA silencing seems to induce translational repression as well as mRNA degradation. Importantly, the fact that miRNAs affect their targets also at the mRNA level allows a genome-wide analysis of their impact on the transcriptome by the use of mRNA microarrays. Indeed, additional support for miRNA induced mRNA decay comes from microarray experiments (Lim et al. 2005; Behm-Ansmant et al. 2006; Rehwinkel et al. 2006; Schmitter et al. 2006; Wu et al. 2006). Overexpression or depletion of specific miRNAs is causing misregulation of transcripts enriched for respective miRNA binding sites in their 3' UTRs. And depletion of different components of the RNA silencing pathway seems to lead to similar misregulation at the transcriptome level, irrespective of which RNA silencing protein is depleted.

Yet several examples exist where miRNAs or tethering of RISC/miRNP components leads only to translational inhibition. In fact, in some special cellular conditions the repression by the miRNAs can be relieved (Bhattacharyya et al. 2006; Schratt et al. 2006; Kedde et al. 2007). This is consistent with the other function of P-bodies, the storage of repressed mRNAs. Some miRNA targets can become translationally silenced and stored in P-bodies until a specific cellular signal such as neuronal stimulation or cellular stress induces their rapid return to the translated pool. This relief of repression is mediated by additional translational regulators that bind to the 3' UTRs of the mRNAs targeted by the miRNA. The details of how certain miRNA targets are selected only for translational repression while others exhibit also mRNA decay remain to be solved. However, a very recent, large scale analysis for both proteomic and transcriptomic status of cells overexpressing or depleted of different miRNAs indicated that in most cases both protein as well as the mRNA level of the miRNA target are affected (Selbach et al. 2008).

Generally miRNAs and siRNAs are inducing repression and/or degradation of their target mRNAs. But some reports suggest that also the opposite i.e. RNA activation could be taking place under specific conditions. Vasudevan et al. were able to show that miR-369-3p can activate translation of TNFα (Tumor necrosis factor-α) mRNA through

binding to an AU-rich element (ARE) in its 3' UTR in cell cycle arrested, G0-stage human cells (Vasudevan and Steitz 2007; Vasudevan et al. 2007). This activation depended on the presence of AGO2 and an AGO2-interacting protein FXR1 (fragile-X-mental-retardation-related protein 1). The observation could be further extended also for regulation by other miRNAs like let-7 and a synthetic miRNA miRcxc4. For each of these miRNAs the selection between repression and activation of the target mRNA depended on the cell cycle conditions. Taken together, these and other reports imply that we have still a lot to learn about the exact mechanism of miRNA function.

2.2.3 miRNAs and recognition of their target mRNAs

The miRNA Registry (http://microrna.sanger.ac.uk) currently (release 11.0) enlists 678 human and 472 mouse miRNAs. The same number for both *C. elegans* and *D. melanogaster* is around 150 miRNAs each. These numbers of identified miRNAs have been steadily increasing over the past years and with the development of more sophisticated high-throughput sequencing methods, are expected to further increase. Considering that many of the miRNAs might be expressed in tissues and conditions that have not yet been analyzed, the total number of the mature miRNAs in mammals could rise to thousands. The largest analysis of miRNA expression profiles in mammals so far was conducted by Landgraf et al. (Landgraf et al. 2007). They cloned and sequenced small RNA sequences from 26 different organs and cell types from humans, mice and rats. This effort was able to confirm expression of 300-400 different miRNAs in each species with at least 70 different mature miRNAs expressed in each given cell type. Deep-sequencing of HeLa cells was able to identify more than 200 expressed miRNAs in this single cell type (Friedlander et al. 2008). However, approximately half of these miRNAs were expressed at fairly low levels that might not have a physiological significance. Landgraf et al. found several miRNAs to be expressed ubiquitously across the tested cell types while other miRNAs showed more specific expression patterns. A third of the miRNAs were expressed with high tissue specificity while only a few were restricted for certain cell type. The most ubiquitous and abundant miRNA turned out to be miR-16 while the highest exclusivity was conferred by the miRNAs expressed solely in embryo (Landgraf et al. 2007).

The number of miRNA targets varies depending on the miRNA and the more conserved miRNAs seem to have the highest number of targets (Lewis et al. 2003). Computational predictions based on miRNA binding sites in the 3' UTRs of mRNAs imply that an average vertebrate miRNA has more than 200 putative targets and, at least in humans, more than 20% of the transcriptome could be regulated by miRNAs (Lewis et al. 2003; Krek et al. 2005; Xie et al. 2005). Yet these predictions may be underestimates as they do not take into consideration the evolutionary new, non-conserved binding sites. On the other hand, many mRNAs and miRNAs might never interact with each other in physiological conditions since they can be expressed in different tissues or developmental stages. The predicted numbers of targets have received some validation from microarray experiments monitoring the transcriptomes of cells overexpressing or depleted of individual miRNAs (Krutzfeldt et al. 2005; Lim et al. 2005; Linsley et al. 2007). Depletion of endogenous miR-122 from mouse liver by use of antagomirs induced upregulation of 363 transcripts (Krutzfeldt et al. 2005). Consistent with direct miRNA-mediated regulation, these transcripts were enriched for binding sites for miR-122 in their 3' UTRs. Similarly, transfection of miR-1 or miR-124 to HeLa cells led to downregulation of 96 and 174 mRNAs, respectively (Lim et al. 2005). Consistently with their specific endogenous expression in skeletal muscle (for miR-1) and in brain (for miR-124), their transfection shifted the transcriptome of HeLa cells towards that of the aforementioned tissues. That is to say that genes downregulated by miR-1 or mir-124 are ones that are expressed at low levels in skeletal muscle or brain, respectively. This is in keeping with the analyses of expression profiles of predicted miRNA targets (Farh et al. 2005; Stark et al. 2005; Sood et al. 2006). These analyses show that a miRNA and its putative targets are often expressed in the same tissues but the levels of the target mRNAs are very low compared to other tissues not expressing the miRNA. In addition, the mRNAs that are expressed at high levels in a tissue with a given miRNA, especially the ubiquitously expressed mRNAs of housekeeping genes, have evolved to avoid miRNA binding sites in their 3' UTRs (Farh et al. 2005).

miRNAs recognize their target mRNAs by basepairing to the complementary binding sites in the target mRNA. Several reports have described universal and conserved rules for miRNA target recognition in animals (Doench and Sharp 2004; Kloosterman et

al. 2004; Brennecke et al. 2005; Gaidatzis et al. 2007; Grimson et al. 2007). The binding sites for miRNAs are usually located in the 3' UTRs of the target mRNAs but an insertion of a binding site to the 5' UTR or even the coding sequence (CDS) is also capable of inducing silencing. In the long 3' UTRs (> 1300-nt) the binding sites seem to localize to the 5' and 3'ends of the 3' UTR rather than the center. Still, the binding site should be further than 15 nt from a stop codon. Number of miRNA binding sites appears to be attributable to the extent of silencing observed and a close proximity of binding sites in the 3' UTR seems to enhance the silencing. This is true for two binding sites for the same miRNA as well as binding sites for two different miRNAs. In addition, miRNA binding sites reside preferentially near AU-rich sequences supporting the idea of interplay between miRNA regulation and regulatory proteins binding to AREs. While siRNAs bind their targets with perfect complementarity, miRNAs show imperfect basepairing. The computational analysis of microarray data as well as reporter gene assays utilizing point mutations have demonstrated that the 5' end of the miRNA is most important for the miRNA:mRNA interaction. Especially the positions 2-8 of the miRNA appear to be critical for efficient target repression. This region has been termed the seed region of the miRNA. Yet, there are cases where imperfect base-pairing or weaker G-U base-pairing at the seed can still stimulate effective silencing. This is usually due to an increased base-pairing in the 3' half of the miRNA, especially at the positions 13 to 16.

Defining rules for miRNA:mRNA interaction has been vital for generation of different tools for predicting miRNA targets. Currently most prediction programs rely on the presence and conservation of an intact complement for the seed sequence in the target mRNA. In their proteome and transcriptome wide analysis of miRNA-mediated regulation, Selbach et al. compared the accuracy of different prediction programs (Selbach et al. 2008). This comparison, together with other aforementioned genome-wide analyses, suggests that in general the seed sequence is the most critical determinant of miRNA target recognition. But it is likely that many special cases exist where the seed does not play a crucial role.

Many of the mature miRNAs are conserved across animal species, particularly at their seed regions. In addition to their homologs in other species, the miRNAs can also have multiple paralogous miRNAs expressed from within the same genome. These

related miRNAs can derive from the same primary transcript or from separate transcripts and have probably been generated through gene duplications during the evolution. The miRNAs with similar sequences at their seed region as well as beyond it form miRNA families. Members of miRNA families are often functionally redundant, meaning that they can regulate the same target mRNAs and the removal of a single member of a family is often not sufficient to cause major regulatory defects. This type of additive regulation has been demonstrated for example by genetic studies of miRNA families in *C. elegans* and mouse (Abbott et al. 2005; Miska et al. 2007; Ventura et al. 2008). The redundancy between miRNAs allows multicellular organisms an additional level of regulation by altering the number of miRNA family members expressed in a given tissue but further complicates our effort to understand the miRNA-mediated regulation.

2.3 Biological role of miRNAs in animals

Gene ontology (GO) analysis of predicted miRNA targets revealed gene categories related to developmental processes as the most significant categories under miRNA control in the tissues of *Drosophila* (Stark et al. 2005). This prediction is now supported across the animal kingdom by vast body of literature that relies on different approaches from complete depletion of miRNAs to analysis of effects of single miRNAs. miRNAs appear to fine-tune and support the transition from one transcriptional program to another during development. Still, miRNAs have biological functions beyond just development and they have been implicated in processes as variable as immune defense and metabolism (Esau et al. 2006; Vigorito et al. 2007). In the following chapters (2.3.1 and 2.3.2) I will focus on few main biological roles of miRNAs that are also interconnected, their function in cell cycle and in development.

2.3.1 miRNAs in proliferation and cell cycle control

Proliferation is a critical part of successful development and defects in differentiation can often be attributed to malfunctioning cell cycle control. During differentiation from a stem cell or a progenitor to a terminally differentiated cell type, the cells usually have to orchestrate an exit from the cell cycle, and occasionally, re-enter it. miRNAs are known to be necessary for proliferation and proper cell cycle control in many species. Grishok and Sharp studied the nuclear divisions in *C. elegans* intestine and discovered that knock-

down (KD) of Argonaute proteins of *C. elegans* (ALG-1 and ALG-2) or Dicer (DCR-1) resulted in slight increase in the number of divisions (Grishok and Sharp 2005). And when these KDs were carried out in the absence of LIN-35 (*C. elegans* homolog of retinoblastoma (RB) protein), the increase was even greater than that in *Lin35* knock-out alone. One of the reasons for increased divisions was found to be upregulation of *cyclin E* expression. These data suggest a synergistic function of RNAi pathway and RB pathway in the control of cell cycle, although miRNAs were not directly implicated. Similarly, the analysis of germ-line stem cells (GSCs) in *D. melanogaster* showed that loss of DCR-1, the *Drosophila* Dicer required for miRNA processing, triggered a delay in G1- to S-phase transition (Hatfield et al. 2005). This delay was found to be specific for stem cells. Also here the phenotype was accompanied by increased *cyclin E* expression that interestingly depended on upregulation of cyclin-dependent kinase inhibitor Dacapo (Dap, homolog of mammalian cyclin-dependent kinase inhibitors CDKN1A/CDKN1B or p21/p27). The role of miRNAs in cell cycle control is not a specialty of invertebrates. Loss of Dicer and miRNAs in both mouse ESCs as well as mouse chondrocytes leads to drastically decreased growth rate (Kanellopoulou et al. 2005; Murchison et al. 2005; Kobayashi et al. 2008). Very similar proliferation defect was observed also in mouse ESCs lacking DGCR8, arguing that this defect is due to loss of Drosha and Dicer generated miRNAs (Wang et al. 2007). Consistent with these observations, inducible human HEK293 Dicer- and AGO2-KD cells lines show significantly decreased growth rate upon loss of Dicer or AGO2 (Schmitter et al., unpublished results). Reduced cell division is also true for chicken-human DT40 hybrid cells that have been depleted for Dicer (Fukagawa et al. 2004). These cells accumulate in the G2/M-phase of the cell cycle but in this case the growth defect was suggested to be due to premature sister chromatid separation in mitosis, possibly caused by improper heterochromatin formation.

Since loss of miRNAs seems to cause decreased proliferation in so many different cell types and species, it is tempting to speculate that there are miRNAs that can inhibit some conserved pathways responsible for stalling the cell cycle progression. Indeed, such miRNAs have been described. One of the first miRNAs to have a function described to was *bantam* miRNA of *D. melanogaster*. *bantam* null mutants are lethal and Brennecke et al. showed that *bantam* was necessary for growth of imaginal discs through regulation

of cell proliferation (Brennecke et al. 2003). Consistently, cells overexpressing *bantam* show a strong increase in growth rate (Thompson and Cohen 2006). In addition, *bantam* has also some anti-apoptotic activity. The above discussed growth defect involving Dap (CDKN1A/CDKN1B homolog) overexpression upon loss of miRNAs in *D. melanogaster* has been further dissected in human cells. Several groups have shown that two miRNAs with the same seed sequence, miR-221 and miR-222, are able to induce proliferation of human cancer cells by repressing the translation of human CDKN1B (Galardi et al. 2007; Gillies and Lorimer 2007; le Sage et al. 2007). The repression happens through two miR-221/222 binding sites in the 3' UTR of the Cdkn1b mRNA and removal of miR-221 and miR-222 or points mutations in their binding sites were sufficient to reduce the growth rate of the cells. Another similar case of miRNA-mediated proliferation control comes from investigation of role of miR-21 in cancer cells *in vivo* and *in vitro* (Si et al. 2007). miR-21 was found to be necessary for fast proliferation and inhibition of miR-21 using antagomirs led to slower growth rate. The observation was reproduced by many groups and several targets mediating the activity of miR-21 have been identified (Frankel et al. 2008). One of the best studied miRNA clusters with a role in cell cycle control in mammals is that of miR-17-92. miR-17-92 is overexpressed in many rapidly dividing cancers and its overexpression has been shown to induce faster proliferation also in other cells (Hayashita et al. 2005; He et al. 2005; Lu et al. 2007). In fact, miR-17-92 is also called Oncomir-1. Expression of miR-17-92 is regulated by c-Myc, a transcription factor equally upregulated in many human cancers (O'Donnell et al. 2005). It gives rise to 6 mature miRNAs and has two paralogs, miR-106a-363 cluster and miR-106b-25 cluster, which transcribe additional 9 mature miRNAs. miR-17-92 and miR-106b-25 are expressed fairly ubiquitously with highest expression in embryos and ESCs while tissues expressing miR-106a-363 are unknown (Ventura et al. 2008). Experiments with mice lacking these miRNAs suggest that they play important roles in many biological processes in a redundant manner (Ventura et al. 2008). The mature miRNAs from these clusters can be divided into four miRNA families based on their seed sequence. Most functional data on these miRNAs deals with the six miRNAs forming the miRNA family that shares a common seed sequence AAAGUGC, namely miR-17, miR-20a, miR-20b, miR-106a, miR-106b, and miR-93. Recent reports have identified some targets for these

miRNAs and elucidated the mechanisms that allow them to accelerate the cell cycle. miR-17 and miR-20a can silence mRNAs encoding transcription factors E2F1, E2F2 and E2F3 (O'Donnell et al. 2005; Sylvestre et al. 2007). All of these transcription factors were found to regulate the expression of miR-17-92, creating a self-regulatory loop. In addition, the members of this miRNA family were discovered to control the translation of mRNAs encoding for RBL2 (or p130) in different tissues (Lu et al. 2007; Wang et al. 2008). This is interesting since RBL2 is a transcriptional repressor that represses expression of E2F target genes by binding to some E2F proteins at the target gene promoters during G1-phase of the cell cycle and, in this way, regulates the decision between cell cycling and cell cycle exit (Litovchick et al. 2007). Finally, miR-106b was lately found to inhibit translation of CDKN1A, a cyclin-dependent kinase inhibitor related to CDKN1B and *D. melanogaster* Dap and an upstream regulator of RB pathway (Ivanovska et al. 2008). In addition to proliferation control, the AAAGUGC-seeded miRNAs are known to have anti-apoptotic activity and this activity is at least in part mediated through inhibition of proapoptotic factor BIM (Matsubara et al. 2007; Ventura et al. 2008). Some human miRNAs have also been implicated as oncogenes in testicular germ cell tumors (Voorhoeve et al. 2006). Both human miR-372 and miR-373 can induce proliferation and tumorigenesis of primary human cells. Remarkably, these miRNAs have the same core hexamer (AAGUGC) in their seed sequence as miR-17 and the related miRNAs discussed above, suggesting further redundancy.

As we have seen, many miRNAs can increase cell proliferation and act as oncogenes, and the net outcome of total loss of miRNAs appears to be slower growth rate. But there are also some miRNAs that can do the opposite i.e. inhibit cell cycle progression and in this way function as tumor suppressors rather than oncogenes. One of the first miRNAs to be identified as a potential growth repressor was also one of the first known miRNAs: let-7 and miR-84, a member of let-7 miRNA family, were shown to regulate protein levels of RAS, a kinase signaling protein and a known oncogene, both in *C. elegans* and in humans (Johnson et al. 2005). RAS and let-7 showed inverse expression patterns in lung cancer cells, and consistently, increased expression of let-7 was sufficient to decrease proliferation of these cells. Lee et al. were able to reproduce the effect on lung cancer proliferation and proposed HMGA2 as another oncogene that is

a primary target of let-7 and could contribute to the phenotype (Lee and Dutta 2007). Further follow-up of the original discovery of RAS regulation in lung cancer showed that also proliferation of human liver cancer cells could be reduced by let-7 expression and that any of the let-7 family members could trigger this reduction (Johnson et al. 2007). The growth defect was suggested to be mediated by delaying G1- to S-phase transtition. This work was accompanied by microarray analysis to identify transcripts targeted by let-7 in both types of cancer cells and found a number of cell cycle regulators to be inhibited by let-7. These included for example cyclin-dependent kinase 6 and cyclin D. Although well studied, let-7 is not the only miRNA to restrain cell cycle progression. Linsley et al. screened 24 miRNAs for transcriptomic changes induced by their overexpression (Linsley et al. 2007). They found that miRNAs sharing similar seed sequences were causing similar transcriptomic changes. For one miRNA family (formed by miR-15, miR-16 and miR-103) a significant enrichment for cell cycle regulating genes was found among the downregulated transcripts. miR-16 was confirmed to be able to cause accumulation of cells to G0/G1-phase of the cell cycle and this phenotype could be reversed by using anti-miR-16 oligonucleotides. Several primary miR-16 targets were tested by siRNA induced KDs and were found to be able to partially phenocopy miR-16 overexpression. But it is likely that the strong effect of miR-16 on cell cycle comes, as often with miRNAs, from synergistic effect of inhibiting several different targets.

In some cases miRNAs have been described as an important part of signaling cascades. TP53 (Tumor protein p53) is a DNA-binding transcription factor that responds to various cellular stress conditions such as DNA damage by activation of numerous target genes that can, for example, induce apoptosis and stall cell cycle progression. Several laboratories have reported miRNAs of the miRNA family of miR-34 to be conserved target genes of TP53 (Bommer et al. 2007; Chang et al. 2007; He et al. 2007; Raver-Shapira et al. 2007). There are two primary transcripts giving rise to miR-34 miRNAs, one for miR-34a and one for miR-34b and miR-34c. TP53 was shown to bind to conserved binding sites in the promoters of both of these miRNA genes and upregulate their transcription. Increased expression of miR-34 miRNAs was leading to altered expression of various genes functionally related to TP53 target genes (cell cycle, apoptosis, DNA repair etc.). Importantly, blocking of miR-34a function by anti-miR-34a

was sufficient to significantly reduce apoptotic response to TP53 activation, arguing that miR-34a mediates a major fraction of TP53 signaling and, together with miR-34b-c, is an important tumor suppressor.

As apparent from aforementioned instances, many of the examples for miRNA controlled proliferation come from study of cancer cells. This is reasonable since it is cancer where the miRNAs are often misregulated, making pinpointing of their role in cell cycle much easier. In fact, miRNA expression analysis has become increasingly useful diagnostic tool for classification of tumours (Rosenfeld et al. 2008). And the misexpression of miRNAs is often a major contributer to the abnormal behaviour of a cancerous cell: miRNA genes are repeatedly located at fragile genomic sites that undergo amplifications or deletions in different cancers (Calin et al. 2004). For example, miR-21 and miR-17-92 cluster are amplified in neuroblastoma and follicular lymphoma, respectively, while many let-7 family members, miR-34a and miR15a/miR-16 cluster have been deleted in diverse cancers. The significance of miRNA-mediated regulation for cancer simply highlights the importance of miRNAs in control of endogenous processes, coordinating the balance between proliferation and differentiation, and allowing normal development of an organism.

2.3.2 miRNAs in development and differentiation

The development from one totipotent cell to a functioning, multicellular organism requires numerous coordinated cell divisions that are followed by differentiation from one cell type to another. At molecular level the difference between the various cell types is determined by the transcriptome and the proteome expressed by the cells. And any failure in accomplishing this specific expression profile can challenge the normal development. It has now become clear that miRNAs are needed to adjust these expression profiles and to support the transcriptional regulation in a range of developmental processes in all studied animal species. Below I will discuss a few examples where miRNAs are known to contribute to regulation of development

Clear evidence for the importance of miRNAs for development comes from animals lacking the protein components indispensable for miRNA biogenesis. Screens for RNAi-resistant mutants in *C. elegans* demonstrated that deletion of *dcr-1* or the Argonaute genes *alg-1* and *alg-2* leads to several defects in larval development including

a classical loss of let-7 phenotype, burst vulva. (Grishok et al. 2001; Ketting et al. 2001; Knight and Bass 2001). In *D. melanogaster*, AGO1 and AGO2 are known to have overlapping functions and double, but not single, mutations of *ago1* and *ago2* as well as of *ago1* and *dcr-1* lead to segmentation defects in the embryo (Meyer et al. 2006). For zebrafish the loss of Dicer is leading to a growth arrest one week after fertilization and by two weeks most fish die (Wienholds et al. 2003). The relatively long survival time was shown to be due to presence of maternal Dicer in the embryos and later Giraldez et al. created zebrafish depleted of both maternal and zygotic Dicer (Giraldez et al. 2005). Also in these fully Dicer-deficient fish many parts of the early development were unaffected but processes like gastrulation and heart and brain development were strongly perturbed. Interestingly, another family of miRNAs with an AAGUGC-sequence in their seed region, the miR-430 family of zebrafish, was found to be able to rescue large part of the brain development defect. In mouse the loss of *Dicer* or loss of *Ago2* are embryonic lethal but the details of the phenotype vary between reports (Bernstein et al. 2003; Liu et al. 2004; Yang et al. 2005; Morita et al. 2007). Bernstein et al. reported that *Dicer* knock-out mice show morphological abnormalities by embryonic day 7.5, die already before embryonic day 8.5 and the embryos do not have stem cells. Yang et al. created *Dicer* knock-out mice that survived somewhat longer until embryonic day 12.5 and the death was accompanied by impaired blood vessel formation. Similarly to *Dicer*-depleted mice of Bernstein et al., *Ago2*-deficient mice produced by Morita et al. are dying by embryonic day 7.5 but many developmental markers absent in *Dicer* knock-outs were present after the loss of *Ago2*. Again the phenotype of another *Ago2* knock-out was less severe and embryos survived 3 days longer (Liu et al. 2004). It is curious that depletion of AGO2 is embryonic lethal although at least AGO1 and AGO3 are expressed in embryos and should be able to compensate for AGO2. It is possible that AGO2 is normally expressed at very high levels and other AGOs can not match this expression level. Another possibility is that, since AGO2 is the only mammalian Argonaute able to cleave its target mRNA, some developmental processes require this cleavage activity for example to degrade targets of endo-siRNAs (Liu et al. 2004).

miRNAs are also important for proper germ cell development and meiosis. As mentioned above, *dcr-1* null *C. elegans* are sterile, and their oocytes are abnormal and

divide (Ketting et al. 2001). The fertility of these worms can be restored by expression of transgenic *dcr-1*. In *D. melanogaster*, Loquacious, a dsRNA-binding partner of Dicer required for pre-miRNA processing, was shown to be necessary for oogenesis and fertility (Forstemann et al. 2005). The mutant flies had small ovaries and appeared to be unable to maintain GSCs. This is reminiscent of the results of Hatfield et al. that were discussed above and suggested a role for miRNAs in proliferation control of GSCs (Hatfield et al. 2005). Indeed, analysis GSCs in *ago1* mutant flies further confirmed that miRNAs are needed for division and self-renewal, rather than survival of GSCs in *D. melanogaster* (Yang et al. 2007). In mice the miRNAs with AAGUGC-seed sequence are highly expressed in primordial germ cells and conditional deletion of Dicer from these cells, similarly to *D. melanogaster*, causes defective proliferation and leads to an early arrest in spermatogenesis (Hayashi et al. 2008). Interestingly, conditional knock-out of *Ago2* does not show a similar defect. Furthermore, conditional *Dicer* knock-out oocytes have been described (Murchison et al. 2007; Tang et al. 2007). They arrest in meiosis due to spindle formation defects that prevent normal chromosome segregation. It is unclear whether this defect is a result of loss of miRNAs or some other function of Dicer. Tang et al. observed similar fault in Dicer knock-out oocytes' spindle formation and additionally reported that maternal miRNAs of the oocyte are present in the zygote still after fertilization, suggesting that they have a role in the first moments of the embryonic development (Tang et al. 2007). Indeed, mice lacking maternal miRNAs are infertile and unable to proceed through the first cell divisions.

Another conserved function for miRNAs in early embryonic development has been described in *D. melanogaster* and zebrafish. When zygotic transcription takes place soon after fertilization, many of the maternally contributed mRNAs get degraded fairly rapidly in order to make way for establishment of a new transcriptional profile. Giraldez et al. demonstrated that miR-430, a miRNA family expressed at high levels in zebrafish development after the onset of zygotic transcription, is needed for degradation of many of the maternal mRNAs (Giraldez et al. 2006). Similarly, miRNAs of miR-309 cluster, also expressed after the onset of zygotic transcription, are necessary for maternal mRNA degradation in *D. melanogaster* (Bushati et al. 2008). Interestingly, miRNAs of the miR-

309 cluster of *D. melanogaster* are not related to the miR-430 family of zebrafish in their sequence.

One of the extensively studied processes of cell differentiation and lineage commitment in mammals is that of hematopoiesis where hematopoietic stem cells give rise to a variety of progenitor cells that further differentiate to mature blood cells. Hematopoiesis also serves as a valuable model system for studying miRNAs in differentiation. Hematopoietic cells express more than one hundred different miRNAs, five of which are fairly specific for the hematopoietic cells (Chen et al. 2004; Landgraf et al. 2007; Neilson et al. 2007). These are miR-142, -144, -150, -155 and -223. In addition, miR-181 is expressed at very high levels in these cells. Detailed analysis of miRNA expression during T-lymphocyte development shows that expression of most of these as well as many other miRNAs, such as members of miR-17-92 cluster, varies between differentiation stages (Neilson et al. 2007). A change in expression of certain miRNAs like miR-181 was accompanied by altered levels of mRNAs that have their 3' UTRs enriched for sequences complementary to the seed sequence of the respective miRNA. Targets of miR-181 included for example the mRNA for T-cell receptor-α. miR-181 has a role in lineage selection as overexpression of miR-181 in hematopoietic progenitors can increase the number of cells differentiating to B-lymphocyte lineage (Chen et al. 2004). In contrast, overexpression of miR-142 or miR-223 can lead to an increase in cells that differentiate to T-lymphocytes. Similarly, overexpression of miR-150 in hematopoietic stem cells can block the differentiation of B-lymphocytes without affecting development of other lineages (Zhou et al. 2007). The importance of miRNAs for T-cell differentiation has been substantiated by conditional deletion of Dicer at different stages of T-lymphocyte development (Cobb et al. 2005; Muljo et al. 2005; Neilson et al. 2007). The loss of Dicer and the subsequent loss of miRNAs affect different aspects of T-cell biology and cause a decrease in the number of differentiated T-cells, at least in part, through an increase in apoptosis.

Several miRNAs might contribute to the apoptosis control in lymphocytes. miR-181 was shown to inhibit pro-apoptotic protein B-cell CLL/lymphoma 2 (BCL2). Another pro-apoptotic protein, BCL2-like 11 or BCL2-interacting protein (BIM), is repressed by members of miR-17-92 cluster (Ventura et al. 2008). Consistently, deletion

of miR-17-92 cluster from hematopoietic cells leads to significant reduction in the number of B-cells and increased apoptosis of early B-cell progenitors. The necessity of miRNAs for B-lymphocyte development is further supported by the effects of *Ago2* deletion in bone marrow progenitor cells, which impairs differentiation beyond pro-B cell stage (O'Carroll et al. 2007). In addition, *Ago2*-deficient bone marrow cells are unable to produce functional red blood cells implying that miRNAs are essential also for erythropoiesis. Remarkably, the slicing activity of AGO2 is not vital for the abovementioned processes.

Another developmental process where miRNAs, and especially the miR-17-92 cluster, have a fundamental function is lung development. Mice with conditional deletion of Dicer in their lungs show defects in lung branching and increased cell death in lung epithelium (Harris et al. 2006). Overexpression of miR-17-92 cluster in lung epithelium increases the proliferation of the epithelial progenitor cells and inhibits their differentiation (Lu et al. 2007). Consistently, the mice lacking miR-17-92 cluster die immediately after birth, largely due to underdeveloped lungs (Ventura et al. 2008). It remains to be seen whether also other miRNAs, in addition to miR-17-92 cluster, contribute to the lung development.

In order to find out whether miRNAs regulate morphogenesis or patterning of vertebrate limbs, Harfe et al. created a conditional deletion of Dicer in mouse limb mesoderm (Harfe et al. 2005). The limbs of the knock-out mice showed impaired morphogenesis and were smaller than those of the control mice. The morphogenesis defect was accompanied by increased cell death. Interestingly, the differentiation of the limb cells was not affected as all normal limb cell types could be found in the Dicer knock-out mice. A specific role for miRNAs in limb development has been described by Hornstein and collegues (Hornstein et al. 2005). Expression of the signaling gene *Shh* (*Sonic hedgehog*) is an important determinant of anterior-posterior polarity of fore- and hindlimbs in mice. The forelimb-specific induction of *Shh* is mediated by Hox protein HOXB8 (Homeobox B8). Hornstein et al. demonstrated that the inhibition of *Shh* induction in hindlimbs is due to specific expression of miR-196, which in turn can regulate HOXB8 levels by mediating cleavage of its mRNA (Yekta et al. 2004; Hornstein et al. 2005).

In addition to the aforementioned examples, miRNAs are now known to be important for many other developmental processes such as skin morphogenesis, hair follicle formation and development of heart and muscle in mice (Zhao et al. 2005; Andl et al. 2006; Yi et al. 2008). And without a doubt a plethora of additional functions for miRNAs will be discovered in the coming years. miRNAs seem to contribute to development by regulating the balance between proliferation and differentiation, by suppressing cell death and by serving as switches for lineage selection. Also they are needed for maintaining the potential of stem cells and progenitors to differentiate into a variety of cell types. In fact, one of the key questions for understanding developmental processes is to determine how this pluripotency (of stem cells) or multipotency (of progenitors) is maintained and how it is lost in a controlled manner during differentiation.

2.4 Epigenetics of embryonic stem cells and their differentiation

ESCs are derived from the inner cell mass of blastocysts and are capable of differentiating into any type of cell or tissue of an organism i.e. they are pluripotent (Figure 2) (reviewed in (Smith 2001). They can be maintained in culture in their undifferentiated state for prolonged periods under appropriate culturing conditions, either in the presence of so called feeder cells or in the presence of a cytokine produced by these cells called leukemia inhibitor factor (LIF). LIF acts via gp130 receptor to induce JAK/STAT (Janus kinase/Signal transducer and transcription activator) signaling cascade that enforces the ESCs into continuous self-renewal. Upon removal of LIF the cells will continue to proliferate but begin to differentiate. This differentiation can be directed into a desired cell type by addition of further factors like retinoic acid (RA). Understanding the molecular basis of pluripotency and differentiation is of great interest. Research of recent years has started to recognize that ESCs are epigenetically very unique and the correct epigenetic regulation could be underlying the “stemness” of ESCs.

2.4.1 Transcriptional core circuitry of ESCs

In addition to the external signaling initiated by LIF, intrinsic regulation of self-renewal also takes place. Several transcription factors have been discovered to contribute or to be essential for pluripotency and self-renewal of ESCs. The best characterized of these factors is OCT-4. Deletion of *Oct-4* prohibits the development of pluripotent stem cells in

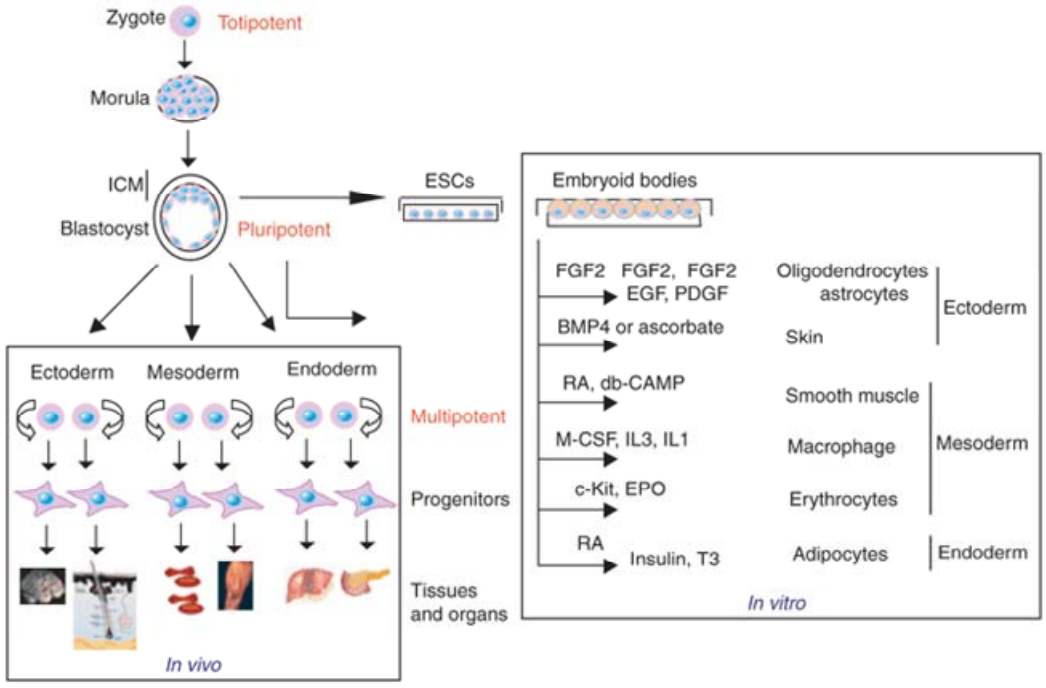


Figure 2. ESCs are pluripotent cells isolated from blastocyst stage embryo.

After 3.5 days of mouse development or 5 days of human development, the fertilized oocyte or zygote has developed into a blastocyst. The cells in the inner cell mass (ICM) of the blastocyst are considered pluripotent as they have the potential to give rise to all three primary germ layers: ectoderm, mesoderm and endoderm. These in turn develop into the tissues and organs of the body. ESCs are isolated from the ICM and can be cultured indefinitely *in vitro* or differentiated into variety of cell types by using correct culturing condition. Modified from (Guasch and Fuchs 2005).

mouse blastocyst and KD of OCT-4 in mouse or human ESCs leads to their differentiation (Nichols et al. 1998; Hay et al. 2004). The exact level of OCT-4 expression is critical since already a mild overexpression of OCT-4 can induce differentiation towards endoderm and mesoderm (Niwa et al. 2000). Similarly, depletion of another transcription factor, NANOG ("Tir Na Nog" or "land of the ever young" in Celtic mythology), induces ESC differentiation (Chambers et al. 2003; Mitsui et al. 2003). The strength of the intrinsic self-renewal pathway is reflected by the fact that overexpression of NANOG is sufficient to maintain ESC self-renewal in the absence of LIF induced external signals. Due to their specific expression in pluripotent cells, transcription factors like OCT-4 and NANOG are often used as markers for pluripotency

of ESCs. OCT-4 and NANOG can both repress and activate their target genes which they regulate through binding to the DNA at the target gene promoters. The decision between activation and repression depends on the interacting transcription factors at the promoter. One of the interacting partners of OCT-4 is SOX-2 (SRY box-2) that heterodimerizes with OCT-4 to regulate common target genes (Yuan et al. 1995).

In order to understand the means by which OCT-4, NANOG and SOX-2 can confer pluripotency and to identify their target genes, Boyer et al. performed chromatin immunoprecipitation of these factors coupled to microarray analysis (ChIP-chip) of thousands of promoters in human ESCs (Boyer et al. 2005). Each factor was found to be associated with hundreds of promoters of both active and inactive genes. Interestingly, over 90 % of promoters occupied by OCT-4 and SOX-2 were also occupied by NANOG. Many active genes among the targets were previously associated with pluripotent state while the inactive targets included many genes driving developmental processes. OCT-4, NANOG and SOX2 were suggested to form a core transcriptional network that can drive self-renewal of ESCs and inhibit their differentiation. Also, the three transcription factors were all shown to regulate their own expression, forming an autoregulatory circuit that can enforce the pluripotent status as well as to allow its rapid silencing.

Although critical for stemness of ESCs, OCT-4, NANOG and SOX-2 are not the only important regulators and many other transcription factors have been implicated. For example, Krüppel-like factors KLF-2, KLF-4 and KLF-5 were recently shown to be essential for maintenance of pluripotent status (Jiang et al. 2008). Depletion of all three factors induces differentiation and misregulation of *Nanog* expression. In addition, many targets of KLFs are also targeted by NANOG. The reason that KLFs were not previously found to be critical for ESC maintenance is mainly due to the fact they are redundant and a loss of a single factor is not sufficient to induce a phenotype.

The most promising application of the knowledge concerning the transcriptional circuitry governing ESC pluripotency is the reprogramming of differentiated cells back to the pluripotent status. The first successful reprogramming by using simple expression of critical transcription factors was performed by Takahashi and Yamanaka who reprogrammd mouse fibroblasts to pluripotent cells by ectopically expressing *Klf-4*, *Oct-4*, *Sox-2*, and *c-Myc* (Takahashi and Yamanaka 2006). Also other combinations of

transgenes (such as *OCT-4*, *SOX-2*, *NANOG* and *LIN28*) have been able to reprogram human somatic cells into pluripotent cells (Yu et al. 2007). This further underlines the importance of these few regulators for ESC self-renewal.

The proper silencing of the self-renewal promoting transcriptional network and its components such as *Oct-4* and *Nanog* is one of the key steps in successful differentiation. It is initiated by activation of transcriptional repressors, such as GCNF (Germ cell nuclear factor), that target *Oct-4*, *Nanog*, and other genes (Gu et al. 2005). This leads to complete silencing of the targeted genes by formation of condensed chromatin structure as well as methylation of the promoter DNA. In the chapters 2.4.2 and 2.4.3 I will shortly discuss the details of these processes in ESCs before discussing the roles of miRNAs in ESCs in chapter 2.4.4.

2.4.2 Histone modifications in ESCs

Nuclear eukaryotic DNA is packaged and wrapped around protein structures called nucleosomes that are formed by an equimolar octamer of four histone proteins: histones H2A, H2B, H3 and H4. The level of packaging of DNA into this chromatin structure is known to be affected by post-translational covalent modifications of these histones. Addition and removal of histone modifications are catalyzed by a number of enzymes specific for a given modification and position. By modulating the packaging of DNA, the histone modifications can affect the accessibility of DNA for replication, transcription and DNA repair. In addition to altering the accessibility of DNA through changes in the interaction between DNA and the nucleosome, histone modifications can serve as binding sites for many regulatory proteins, such as transcriptional activators and repressors. Different combinations of histone modifications have been suggested to form a so called histone code, which can be interpreted by different histone-interacting proteins, leading to a correct output, e.g. decreased transcription (Jenuwein and Allis 2001). For example, trimethylation of histone H3 at lysine 9 of its N-terminal tail (H3K9me3) by histone methyltransferase (HMT) SUV39H1 (Suppressor of variegation 3-9 homolog 1) can serve as a binding site for HP1 (Heterochromatin protein-1). HP1 can recruit further SUV39H1 proteins to induce the same modification in the surrounding nucleosomes, allowing additional HP1 proteins to bind. These HP1 proteins can then dimerize in order to form silenced and condensed heterochromatin. Many different histone modifications

have been identified, including methylation, acetylation, phosphorylation, and ubiquitination (Turner 2002; Kouzarides 2007). And each modification can take place at many different positions of each histone. Most modifications take place at the N-terminal tails of H3 and H4 but also H2AB has several sites for covalent modifications. Generally, acetylation of a histone is associated with accessible and open euchromatin that is transcriptionally active while methylation of many lysine residues is linked to both transcriptional activity as well as to inaccessible and condensed heterochromatin that is transcriptionally silent (also depending on whether mono-, di-, or trimethylation is taking place). As notable modifications related to transcriptional activity, di-, and trimethylation of H3K4 and H3K79, and of H3K36 are known to occur at the transcription start sites (TSS) and at the body of highly transcribed genes, respectively (Shilatifard 2008; Steger et al. 2008). Similarly, methylation of arginine residues has been implicated with transcriptional activity (Bauer et al. 2002).

Chromatin in ESCs is different from that in somatic and differentiated cells. In ESCs heterochromatin is localized only in few large domains and is less condensed than in differentiated cells (Kobayakawa et al. 2007). Differentiation leads to an increase in smaller, highly condensed foci that vary from one cell type to another. This increase in heterochromatin condensation can also be observed by ChIP analysis of specific repressive histone modifications. RA-induced differentiation of mouse ESCs is accompanied by a notable change in the modifications of histones at various repetitive regions (Martens et al. 2005). While the undifferentiated cells do not carry any methylation of H3K9, H3K27 or H4K20 at their transposons, satellite repeats, or ribosomal DNA, methylation of all these residues accumulates soon after the beginning of differentiation (with exception of methylation of H3K9 at satellite repeats where it is constantly present). Although ESCs are lacking most repressive histone modifications, they are very abundant in the markers of open chromatin structure such as acetylation of H4 and H3K9 as well as methylation of H3K4 (Azuara et al. 2006). Also, the number of late-replicating genes is lower in ESCs than in lineage-committed cells. This is consistent with the open chromatin structure of ESCs since late-replication timing is usually associated with condensed, more difficully accessible chromatin (Schubeler et al. 2002). It has been suggested that the ESC genome is in a permissive default state from where it

can differentiate into any given cell type through selective silencing of different parts of the genome (see for example (Niwa 2007)). The evidence for accessible chromatin structure of ESCs has been further substantiated by Efroni et al. (Efroni et al. 2008). They compared ESCs and neuronal precursor cells (NPCs) using several approaches and could confirm the more relaxed chromatin organization and higher levels of transcription associated histone modifications in ESCs. Intruigingly, the open chromatin structure was also accompanied by significantly higher global transcription levels, including transcription of genomic regions not usually expressed. The transcriptional hyperactivity was also coupled with higher expression of general transcription factors.

Is the lack of repressive histone modifications essential for pluripotency and self-renewal of ESC? Recent reports suggest this indeed could be the case. Loh and collegues demonstrated that OCT-4 controls the expression of two histone demethylases, JMJD1A and JMJD2C (Jumonji-containing proteins), and allows their high expression on mouse ESCs (Loh et al. 2007). JMJD1A and JMJD2C catalyze the removal of di- and trimethylation of H3K9 at promoters of many ESC specific genes, respectively. Thus, these enzymes are contributing to the active chromatin structure of ESCs. Importantly, KD of JMJD1A and JMJD2C induces differentiation and loss of ESC pluripotency through misregulation of expression of transcription factors such as NANOG.

Methylation of H3K4 is found at the promoters of almost all active genes (Kim et al. 2005) Interestingly, many lineage specific genes and developmental transcription factors are not expressed in ESCs at any significant levels despite having high levels of H3K4 methylation at their promoter regions. Their expression is repressed by the presence of another histone modification, methylation of H3K27 (Azuara et al. 2006; Bernstein et al. 2006). These promoters are considered to have a so called bivalent state and are silent in ESCs but have the potential to become active in response to correct development cues. The methylation of H3K27 is catalyzed by a multiprotein complex called Polycomb group repressive complex 2 (PRC2) (Cao et al. 2002; Muller et al. 2002; Kirmizis et al. 2004). In *D. melanogaster* PRC2 identifies its target genes through binding to a Polycomb response element (PRE) at the DNA adjacent to the target gene but how PRC2 selects its targets in mammals it is still not fully understood. The trimethylated H3K27 (H3K27me3) in turn serves as a binding site for another complex

named Polycomb group repressive complex 1 (PRC1) that induces repression of the target gene through compaction of chromatin that blocks the access for chromatin remodeling complexes (Shao et al. 2000; Francis et al. 2004).

Upon differentiation of ESCs many lineage-specific genes loose their H3K27 methylation, gain H3K4 methylation, increase their RNA Pol II occupancy and become transcriptionally active (Boyer et al. 2006). Simultaneously, most pluripotency genes such as *Oct-4* or *Nanog* are epigenetically targeted for silencing that is accompanied by loss of H3K4 methylation and loss of RNA Pol II occupancy as well as increase in methylation of histone H3 at positions H3K9 and H3K27 (Mikkelsen et al. 2007; Pan et al. 2007; Mohn et al. 2008). One of the better studied examples is silencing of *Oct-4* during differentiation of ESCs. In undifferentiated ESCs *Oct-4* promoter, like much of the ESC genome, is marked for high transcriptional activity by histone modifications like acetylation of H3 and H4 as well as methylation of H3K4 (Hattori et al. 2004; Feldman et al. 2006). Within few days of differentiation these modification are completely lost at the promoter and are replaced by high levels of methylation of H3K9 and H3K27 (Feldman et al. 2006; Pan et al. 2007; Mohn et al. 2008). Following methylation of H3K9 by two days of RA-induced differentiation, HP1 is recruited to the promoter to facilitate condensation of the chromatin. While at many other loci like centromeric satellite regions H3K9 methylation is catalyzed by SUV39H1, at *Oct-4* promoter H3K9 methylation is catalyzed by another HMT called G9a. Although the more condensed chromatin structure created by these local changes in histone modifications can largely repress *Oct-4* expression they are not sufficient to maintain it. Subsequent methylation of DNA at the *Oct-4* promoter by *de novo* DNA methyltransferases is needed for irreversible silencing of the gene.

2.4.3 DNA methylation in ESCs

Around 5 % of mammalian cytosines are methylated at CpG-dinucleotides. DNA methylation is usually considered to be associated with silent chromatin structure and is thought to mediate repression either via blocking of transcription factor binding or recruitment of proteins carrying methylated DNA binding (MDB) domains (Watt and Molloy 1988; Nan et al. 1997; Hendrich and Bird 1998).

DNA methylation is established by enzymes called DNA methyltransferases (DNMTs). Five DNMTs have been identified in mammals. These are DNMT1, DNMT2, DNMT3A, DNMT3B and DNMT3L. Lately it has been shown that DNMT2 is in fact methylating RNA, not DNA, *in vivo* (Jurkowski et al. 2008). The remaining four DNMTs can be divided into two functional classes. DNMT1 localizes at the DNA replication foci and is necessary for copying the methylation patterns during DNA replication from the old DNA strand into the new DNA strand (Leonhardt et al. 1992). Thus, DNMT1 is maintaining the DNA methylation through cell divisions. *Dnmt1* knock-out mice are embryonic lethal but the ESCs are viable and, although reduced in DNA methylation, still exhibit some DNA methylation (Li et al. 1992). This is because the DNMT3 enzymes also possess some maintenance activity and co-operate with DNMT1 to achieve an efficient maintanance of DNA methylation (Liang et al. 2002; Chen et al. 2003). On the contrary, DNMT1 is unable to induce new DNA methylation *de novo*. Instead, *de novo* DNA methylation of previously unmethylated CpGs is established by DNMT3 enzymes (Okano et al. 1999). Function of DNMT3A and DNMT3B is necessary for normal development as well as for differentiation of ESCs (Okano et al. 1999; Li et al. 2007). Unlike 3A and 3B, DNMT3L does not have a catalytic domain and is dispensible for normal development with the exception of germline development. Instead DNMT3L is known to enhance the activity of 3A and 3B and interacts with them to form a functional complex in ESCs (Chedin et al. 2002; Li et al. 2007; Ooi et al. 2007). DNMT3L can recognize nucleosomes with unmethylated H3K4 and recruit the other DNMT3 enzymes to exercise their catalytic activity. Both DNMT3A and DNMT3B are mutually stimulative and needed for robust DNA methylation of target gene promoters (Li et al. 2007).

Consistently with its function, *Dnmt1* is expressed in all proliferating cells, especially in rapidly dividing cells of embryos and in ESCs. Similarly, *Dnmt3b* (more specifically the splicing variants *3b1* and *3b6*) and *Dnmt3L* have highest expression levels in undifferentiated ESCs and during the early phases of differentiation. *Dnmt3a* has two major transcriptional variants: *Dnmt3a1* is expressed ubiquitously at low levels while *Dnmt3a2* has similar expression pattern to *Dnmt3L* with specific expression at high levels in ESCs and germ cells (Chen et al. 2002; Su et al. 2002). Upon differentiation of ESCs *Dnmt3a2/3b/3L* are initially upregulated, which is later followed by strong downregulation of all three enzymes. At least Dnmt1 and Dnmt3b are also fluctuating during cell cycle with their expression peaking during the S-phase (Robertson et al. 2000).

Most DNA methylation in primary cells is taking place outside of regulatory DNA regions at intergenic DNA and repetitive sequences (Weber et al. 2005). At these regions, DNA methylation is needed, for example, for repression of retrotransposon expression and maintenance of genomic integrity (Walsh et al. 1998; Karpf and Matsui 2005). Still, DNA methylation can also take place at regulatory DNA elements like promoters. Level of promoter methylation and its functional impact seems to depend on the CpG density of the promoter (Weber et al. 2007). Most mammalian genes have dense CpG islands in their promoters that appear to avoid DNA methylation. The few CpG islands that are methylated are usually transcriptionally inactive. In contrast, the promoters with only a few CpG-dinucleotides are often methylated but this methylation does not correlate with expression. Importantly, some promoters contain an intermediate number of CpGs. Usually these promoters carry methylated DNA and this methylation is mostly associated with low levels of expression. Especially germ-line specific genes appear to be silenced via DNA methylation.

In ESCs, only a small number of promoters is methylated in the undifferentiated state (Mohn et al. 2008). But upon differentiation into neuronal precursors more than 300 promoters gain *de novo* DNA methylation which is accompanied by transcriptional silencing of the respective genes. The promoters gaining DNA methylation are highly enriched for genes like *Oct-4* and *Nanog* that are required for pluripotency. As mentioned before, DNA methylation seems to be required rather for maintenance of transcriptional silencing. For example, silencing of *Oct-4* upon differentiation can be reversed in

Dnmt3a/3b double knock-out ESCs by returning the cells to pluripotent culturing conditions (Feldman et al. 2006). Further experiments have shown that silencing of Oct-4 in ESCs lacking DNMT3 enzymes is incomplete and can become reversed even when the cells are kept in the differentiation inducing conditions (Li et al. 2007). Another target gene, *Nanog*, is known to behave in the same manner. Similarly, silencing and heterochromatinization of an integrated reporter transgene can take place independently of CpG-dinucleotides but maintenance of this silencing requires the presence of methylated CpGs (Feng et al. 2006). These observations suggest that DNA methylation serves as an epigenetic memory that prevents accidental reactivation of targeted genes later in development where it could be detrimental for the organism.

2.4.4 miRNAs in ESCs

As mentioned already in previous chapters, miRNAs are important for normal function of mouse ESCs. Depletion of all miRNAs from ESCs by knock-out of *Dicer* or *Dgcr8* leads to several different defects. The knock-out cells are viable and continue to self-renew but their growth rate is significantly slower that that of wild-type or heterozygous control cells (Kanellopoulou et al. 2005; Murchison et al. 2005; Wang et al. 2007). The cells exhibit features of pluripotent ESCs such as colony formation and expression of pluripotency markers like *Oct-4* and *Nanog*. But the cells fail to differentiate *in vivo* and *in vitro* showing that, in fact, they are not pluripotent. The initiation of differentiation occurs normally in miRNA depleted ESCs but many pluripotency markers do not become fully silenced while different lineage-specific genes fail to become activated or are delayed in their expression. In addition, *Dicer*-deficient ESCs show increased expression of centromeric transcripts, suggesting that Dicer would be necessary for silencing of these repetitive sequences. The phenotypes of *Dicer* knock-out and *Dgcr8* knock-out ESCs are very similar arguing that loss of miRNAs is the major cause for these defects.

Many miRNAs have been found to be expressed in ESCs and some of them appear to be specific for ESCs and embryonic development. Houbaviy et al. described expression of miRNAs of the miR-290 cluster in mouse ESCs (Houbaviy et al. 2003). The miR-290 cluster consists of six homologous miRNA hairpins, produced from a single primary transcript of 2.2 kb in length, and which is expressed specifically in ESCs and preimplantation embryos. For most of the miR-290 cluster miRNAs the mature

miRNA comes from the descending strand (-3p) of the hairpin except of miR-290 where the ascending strand (-5p) is dominant. The mature miR-290 miRNAs become strongly downregulated upon ESC differentiation and their expression has not been detected in other tested tissues. Curiously, the downregulation of miR-290 cluster miRNAs during RA induced differentiation has been shown to depend on GCNF mediated repression of *Oct-4* and *Nanog* (Gu et al. 2008). Recently it has been reported that, in addition to ESCs, miR-290 miRNAs are also expressed in primordial germ cells (Hayashi et al. 2008). The follow-up experiments by Houbaviy et al. showed that the pri-miR-290 transcript is spliced, capped and polyadenylated as any mRNA and is under the regulation by a conserved TATA-box in its promoter (Houbaviy et al. 2005). Additionally, they predicted *in silico* the existence of homologous miRNA clusters in other eutherians. In human ESCs, the expression of the miR-371 cluster has been confirmed to be ESC specific similarly to the murine miR-290 cluster (Suh et al. 2004). Also another related miRNA cluster, the miR-302 cluster, is highly expressed in human ESCs and to a lesser extent in mouse ESCs. Still, the miR-302 clusters appear to have slightly less restricted expression patterns than miR-290/miR-371 clusters. Since all of these miRNA clusters have specific embryonal expression profile, further analysis of their transcriptional regulation would be of interest. Curiously, genome-wide ChIP analysis has suggested that the miR-302 cluster in mouse might be under the control of transcription factors OCT-4 and NANOG (Loh et al. 2006).

The very intriguing feature of all of these ESC specifc miRNAs is that they are very highly related, especially in their seed sequence, to each other as well as to some members of miR-17-92, miR-106a-363 and miR-106b-25 clusters, suggesting functional redundancy between these miRNAs (Figure 3). All of the miRNAs share the common hexamer AAGUGC within their seed sequence. Recently, deep sequencing aimed to identify all miRNAs expressed in mouse ESCs described 126 different mature miRNAs expressed at least at the level of 50 molecules per cell (Calabrese et al. 2007). Still, 40 % of all mature miRNA molecules in these cells originated from miR-290 and miR-17-92 clusters, indicating that they might play an important role in maintaining the pluripotency of ESCs. Other highly expressed miRNAs included for example miR-15/16 and miR-21.

In contrast to downregulation of ESC specific miRNAs, several miRNAs also become upregulated upon differentiation. In later stages of neuronal differentiation, brain-specific miR-124 and miR-9 are robustly upregulated and contribute to neurogenesis (Krichevsky et al. 2006). Curiously, NANOG was found to bind in the proximity of genomic loci encoding both of these miRNAs, suggesting that they could be under NANOG-mediated repression in ESCs (Loh et al. 2006). Another miRNA that becomes upregulated upon RA-induced differentiation is miR-134 (Tay et al. 2008). Overexpression of miR-134 in mouse ESCs was found to be sufficient to induce differentiation and miR-134 was suggested to exhibit its activity by directly targeting

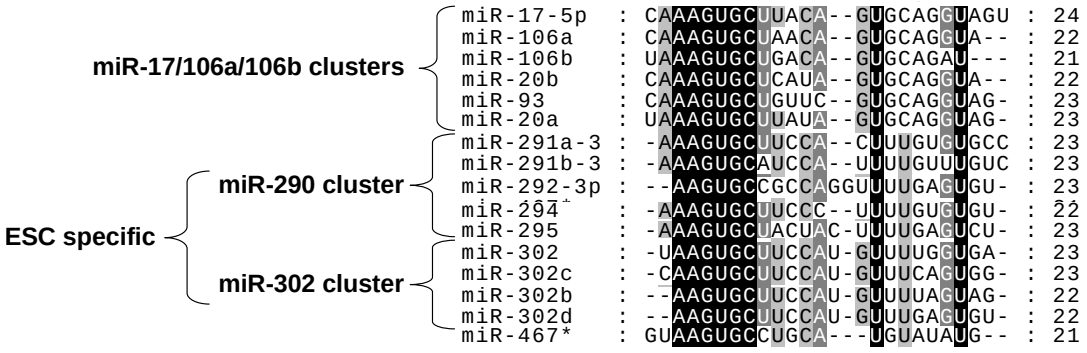


Figure 3. Many related miRNAs are highly expressed in mouse ESCs.

Most of the mature miRNAs from ESC specific miRNA clusters miR-290 and miR-302, miR-467* as well as many members of miR-17/106a/106b clusters are highly related. The sequence alignment reveals that all of the miRNAs share a common AAGUGC hexamer within their 5' most nucleotides. On top of this, many members of miR-290 and miR-17/106a/106b clusters have a common adenosine in the beginning of their seed sequence. In addition, most of the miRNAs have shared GU-dinucleotides in their 3' half.

transcription factors such as NANOG and LRH1 (Liver receptor homolog 1). Interestingly, mature let-7 is not expressed in ESCs but becomes upregulated upon differentiation while the precursor of let-7 is constantly expressed also in ESCs (Newman et al. 2008; Rybak et al. 2008; Viswanathan et al. 2008). In ESCs the processing of pre-let-7 into the mature miRNA by Dicer is blocked by LIN28, another marker of pluripotency that binds specifically to let-7 pre-miRNA. As ESCs differentiate, LIN28 becomes downregulated and allows processing and upregulation of mature let-7. The downregulation of LIN28 is mediated by another upregulated miRNA, miR-125b, as well as by let-7 itself, creating a self-regulatory feedback-loop (Wu and Belasco 2005; Rybak

et al. 2008). These interactions between miRNAs and pluripotency regulating factors imply that miRNAs might contribute more to the core circuitry of pluripotency than previously appreciated.

2.5 Replication-dependent histone genes

In metazoans there are two different families of histone genes, replication-independent and replication-dependent histone genes. Replication-independent histone genes are continuously expressed at constant level to provide histones for example for chromatin repair. In mammals replication-independent histone genes are located across the genome as single genes. On the contrary, replication-dependent histone genes are located in three separate histone gene cluster that are located in the chromosomes 6 (*HIST1* cluster) and 1 (*HIST2* and *HIST3* clusters) in humans (Marzluff et al. 2002). The largest cluster, *HIST1* contains 55 separate histone genes and there are close to 70 replication-dependent histone genes in human genome altogether. Thus, there are 10-20 different genes for each of the four core histone proteins as well as for the linker histone H1. Replication-dependent histone genes fluctuate in their expression during the cell cycle with a more than 30-fold upregulation in histone protein production in the S-phase (Harris et al. 1991). This large pool of new histone proteins is needed to cover the newly synthesized DNA after DNA replication. In fact, high expression of histones outside the S-phase is very toxic for the cell (Osley 1991). The remarkable upregulation of histone expression during the S-phase is achieved at two different levels.

First, transcription of histone genes is coordinately increased two to five-fold as the cells enter the S-phase (Heintz et al. 1983). No common transcription factor responsible for the regulation of all histone genes has been identified. But Oct-1 and its coactivators, for example, are contributing to the upregulation of all *H2B* genes (LaBella et al. 1988; Zheng et al. 2003). Also, phosphorylation of NPAT (Nuclear protein, ataxia-telangiectasia locus) by cyclin E-cyclin dependent kinase 2 complex can enhance transcription of many histones (Ma et al. 2000). On the other hand, the G1-phase specific transcriptional repressor RBL2 is binding to promoters of a number of *HIST1* cluster genes (Litovchick et al. 2007).

Second and more important, the stability and translation of replication-dependent histone mRNAs is robustly increased when DNA synthesis takes place during the S-

phase. The half-life of an histone mRNA increases from 10 minutes outside the S-phase up to one hour in the S-phase (Harris et al. 1991). This specific regulation of stability and translation is due to the very unique structure and processing of histone mRNAs. Unlike all other metazoan mRNAs, the replication-dependent histone mRNAs do not have poly-A tails and only few of them have any introns. Instead of polyadenylation, the histone mRNAs undergo an alterative 3' end processing (Figure 4). Their short 3' UTR contains a highly conserved sequence of 26-nt, 16 of which form a perfect stem-loop structure. Several nucleotides downstream of the stem-loop structure another conserved sequence element called the histone downstream element or HDE is found. The stem-loop of a newly synthesized pre-mRNA is bound by a protein called SLBP (stem-loop binding protein, sometimes also named HBP for hairpin binding protein). At the same time, small nuclear RNA (snRNA) U7 basepairs with its complementary sequence to the HDE and, brings its interacting proteins to the site. SLBP helps to stabilize the interaction between the pre-mRNA and the U7-protein complex via a bridging protein called ZFP100 (Zinc finger protein 100) (Dominski et al. 2002). Once bound to the mRNA, U7 snRNP recruits further proteins such as endonuclease CPSF-73 (Cleaveage/polyadenylation specificity factor-73) (Dominski et al. 2005). CPSF-73 cleaves the pre-mRNA 5 nucleotides (ACCCA) immediately downstream of the stem-loop. Curiously, CPSF-73 is the same endonuclease that cleaves all other mRNAs before their polyadenylation. Also other components are shared between the 3' end processing of histone mRNAs and that of other mRNAs (Kolev and Steitz 2005; Friend et al. 2007). And, for example, the U2 snRNP splicing complex has been shown to stimulate the cleavage of histone mRNAs (Friend et al. 2007). Once processed, the mature mRNA stays bound by SLBP as it becomes exported to the cytoplasm. During translation of the message, SLBP seems to function as a replacement for PABP of polyadenylated mRNAs and the presence of the stem-loop as well as SLBP are necessary for efficient histone translation (Gallie et al. 1996; Sanchez and Marzluff 2002).

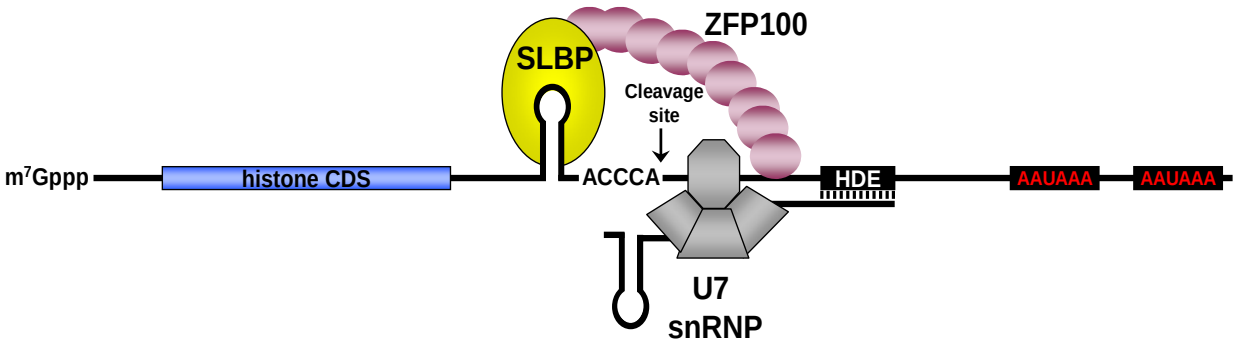


Figure 4. 3' end processing of a replication-dependent histone mRNA by SLBP and U7 snRNP.
U7 snRNP basepairs with the HDE of the histone mRNA downstream of the stem-loop structure to measure the correct cleavage site. Binding of SLBP to the stem-loop stabilizes U7 binding via an interacting bridging protein ZFP100. U7 recruits further proteins to cleave the mRNA after the ACCCA sequence. In the absence of U7 or SLBP the downstream polyadenylation signals (AAUAAA) are used.

The stem-loop of a histone mRNA can also be bound by 3' hExo (3' human exonuclease), simultaneously with SLBP, forming a ternary complex (Dominski et al. 2003). The presence of SLBP prevents 3' hExo from degrading the mRNA during DNA replication. But, as the DNA replication is completed at the end of S-phase, SLBP becomes phosphorylated and released from the stem-loop. This in turn allows initiation of the histone mRNA degradation by 3' hExo and explains the significant drop in histone mRNA half-life at the end of S-phase. The expression of SLBP is regulated in the cell cycle with a 10-20 fold upregulation taking place during the S-phase (Whitfield et al. 2000). Thus, SLBP is the key player responsible for the increased stability and translation of histone mRNAs in the S-phase.

The proper processing of histone mRNAs is critical for survivial and normal development of an organism. Deletion of *SLBP* in *D. melanogaster* is lethal and mutations in U7 snRNA disturb proper oogenesis, rendering the flies sterile (Godfrey et al. 2006). Consistently, maternal *SLBP* is required for early embryonic development of mouse (Arnold et al. 2008). The disruption of histone mRNA processing leads to expression of longer, polyadenylated histone mRNAs. This is due to the use of, often multiple, polyadenylation signals located downstream of the normal processing site. Reason for the presence of these polyadenylation signals is unclear but it has been

suggested that they are needed to prevent transcriptional read-through to the following genes (Lanzotti et al. 2002). In *D. melanogaster*, all histone genes have downstream polyadenylation signals while in mammals some of the histone genes lack them. Recently, two protein complexes, NELF (Nuclear elongation factor) and CBC (Cap-binding complex), which are involved in transcription elongation and the coupled mRNA processing of many RNA Pol II transcripts, were found to be involved in the selection between normal histone mRNA 3' end processing or their polyadenylation (Narita et al. 2007). By constructing several different reporter genes, Narita et al. were able to show that the location of the histone stem-loop in relation to the poly-A signal does not matter for the selection between the two modes of processing. Thus, under normal conditions most histone mRNAs would be cleaved after their stem-loop structure even if it would be preceded by a polyadenylation signal. Although processing of histone mRNAs after their stem-loop is the favored pathway in mammals, there seems to be some leakage to histone mRNA polyadenylation even under normal conditions. This is reflected by the existence of a number of clones for polyadenylated replication-dependent histone genes in human and mouse EST databases. Genome-wide RNAi screen has now identified a number of genes needed for proper histone mRNA 3' end processing in *D. melanogaster* (Wagner et al. 2007).Although miRNAs have been suggested to regulate a number of basic cellular processes, miRNAs and RNAi machinery have not yet been implicated to have a role in the regulation of expression and processing of histone mRNAs.

2.6 References

Abbott, A. L., E. Alvarez-Saavedra, et al. (2005). "The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in Caenorhabditis elegans." Dev Cell **9**(3): 403-14.

Andl, T., E. P. Murchison, et al. (2006). "The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles." Curr Biol **16**(10): 1041-9.

Aravin, A., D. Gaidatzis, et al. (2006). "A novel class of small RNAs bind to MILI protein in mouse testes." Nature **442**(7099): 203-7.

Aravin, A. A., M. S. Klenov, et al. (2004). "Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line." Molecular and Cellular Biology **24**(15): 6742-6750.

Arnold, D. R., P. Francon, et al. (2008). "Stem-loop binding protein expressed in growing oocytes is required for accumulation of mRNAs encoding histones H3 and H4 and for early embryonic development in the mouse." Dev Biol **313**(1): 347-58.

Azuara, V., P. Perry, et al. (2006). "Chromatin signatures of pluripotent cell lines." Nat Cell Biol **8**(5): 532-8.

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

Bauer, U. M., S. Daujat, et al. (2002). "Methylation at arginine 17 of histone H3 is linked to gene activation." EMBO Rep **3**(1): 39-44.

Behm-Ansmant, I., J. Rehwinkel, et al. (2006). "MRNA degradation by miRNAs and GW182 requires both CCR4 : NOT deadenylase and DCP1 : DCP2 decapping complexes." Genes & Development **20**(14): 1885-1898.

Berezikov, E., W. J. Chung, et al. (2007). "Mammalian mirtron genes." Molecular Cell **28**(2): 328-336.

Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-26.

Bernstein, E., S. Y. Kim, et al. (2003). "Dicer is essential for mouse development." Nat Genet **35**(3): 215-7.

Bhattacharyya, S. N., R. Habermacher, et al. (2006). "Relief of microRNA-mediated translational repression in human cells subjected to stress." Cell **125**(6): 1111-24.

Bommer, G. T., I. Gerin, et al. (2007). "p53-mediated activation of miRNA34 candidate tumor-suppressor genes." Curr Biol **17**(15): 1298-307.

Boyer, L. A., T. I. Lee, et al. (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-56.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature **441**(7091): 349-353.

Brennecke, J., A. A. Aravin, et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell **128**(6): 1089-1103.

Brennecke, J., D. R. Hipfner, et al. (2003). "bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila." Cell **113**(1): 25-36.

Brennecke, J., A. Stark, et al. (2005). "Principles of microRNA-target recognition." PLoS Biol **3**(3): e85.
Bushati, N., A. Stark, et al. (2008). "Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila." Curr Biol **18**(7): 501-6.
Cai, X. Z., C. H. Hagedorn, et al. (2004). "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." Rna-a Publication of the Rna Society **10**(12): 1957-1966.
Calabrese, J. M., A. C. Seila, et al. (2007). "RNA sequence analysis defines Dicer's role in mouse embryonic stem cells." Proc Natl Acad Sci U S A **104**(46): 18097-102.
Calin, G. A., C. Sevignani, et al. (2004). "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers." Proc Natl Acad Sci U S A **101**(9): 2999-3004.
Cam, H. P., T. Sugiyama, et al. (2005). "Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome." Nat Genet **37**(8): 809-19.
Cao, R., L. J. Wang, et al. (2002). "Role of histone H3 lysine 27 methylation in polycomb-group silencing." Science **298**(5595): 1039-1043.
Carmell, M. A., A. Girard, et al. (2007). "MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline." Developmental Cell **12**(4): 503-514.
Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.
Chang, T. C., E. A. Wentzel, et al. (2007). "Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis." Mol Cell **26**(5): 745-52.
Chedin, F., M. R. Lieber, et al. (2002). "The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a." Proc Natl Acad Sci U S A **99**(26): 16916-21.
Chen, C. Z., L. Li, et al. (2004). "MicroRNAs modulate hematopoietic lineage differentiation." Science **303**(5654): 83-6.
Chen, T., Y. Ueda, et al. (2003). "Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b." Mol Cell Biol **23**(16): 5594-605.
Chen, T., Y. Ueda, et al. (2002). "A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation." J Biol Chem **277**(41): 38746-54.
Chendrimada, T. P., R. I. Gregory, et al. (2005). "TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing." Nature **436**(7051): 740-4.
Cobb, B. S., T. B. Nesterova, et al. (2005). "T cell lineage choice and differentiation in the absence of the RNase III enzyme Dicer." J Exp Med **201**(9): 1367-73.
Cox, D. N., A. Chao, et al. (1998). "A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal." Genes & Development **12**(23): 3715-3727.
Czech, B., C. D. Malone, et al. (2008). "An endogenous small interfering RNA pathway in Drosophila." Nature **453**(7196): 798-802.
Davis, B. N., A. C. Hilyard, et al. (2008). "SMAD proteins control DROSHA-mediated microRNA maturation." Nature.

Deng, W. and H. F. Lin (2002). "miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis." Developmental Cell **2**(6): 819-830.
Denli, A. M., B. B. J. Tops, et al. (2004). "Processing of primary microRNAs by the Microprocessor complex." Nature **432**(7014): 231-235.
Ding, L., A. Spencer, et al. (2005). "The developmental timing regulator AIN-1 interacts with miRISCs and may target the argonaute protein ALG-1 to cytoplasmic P bodies in C. elegans." Mol Cell **19**(4): 437-47.
Doench, J. G. and P. A. Sharp (2004). "Specificity of microRNA target selection in translational repression." Genes Dev **18**(5): 504-11.
Dominski, Z., J. A. Erkmann, et al. (2002). "A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing." Genes Dev **16**(1): 58-71.
Dominski, Z., X. C. Yang, et al. (2003). "A 3' exonuclease that specifically interacts with the 3' end of histone mRNA." Mol Cell **12**(2): 295-305.
Dominski, Z., X. C. Yang, et al. (2005). "The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing." Cell **123**(1): 37-48.
Efroni, S., R. Duttagupta, et al. (2008). "Global transcription in pluripotent embryonic stem cells." Cell Stem Cell **2**(5): 437-47.
English, J. J., E. Mueller, et al. (1996). "Suppression of Virus Accumulation in Transgenic Plants Exhibiting Silencing of Nuclear Genes." Plant Cell **8**(2): 179-188.
Esau, C., S. Davis, et al. (2006). "miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting." Cell Metab **3**(2): 87-98.
Eulalio, A., I. Behm-Ansmant, et al. (2007). "P-body formation is a consequence, not the cause, of RNA-mediated gene silencing." Mol Cell Biol **27**(11): 3970-81.
Eulalio, A., E. Huntzinger, et al. (2008). "GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay." Nat Struct Mol Biol **15**(4): 346-53.
Farh, K. K., A. Grimson, et al. (2005). "The widespread impact of mammalian MicroRNAs on mRNA repression and evolution." Science **310**(5755): 1817-21.
Feldman, N., A. Gerson, et al. (2006). "G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis." Nat Cell Biol **8**(2): 188-94.
Feng, Y. Q., R. Desprat, et al. (2006). "DNA methylation supports intrinsic epigenetic memory in mammalian cells." PLoS Genet **2**(4): e65.
Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-11.
Forstemann, K., Y. Tomari, et al. (2005). "Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein." PLoS Biol **3**(7): e236.
Francis, N. J., R. E. Kingston, et al. (2004). "Chromatin compaction by a polycomb group protein complex." Science **306**(5701): 1574-1577.
Frankel, L. B., N. R. Christoffersen, et al. (2008). "Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells." J Biol Chem **283**(2): 1026-33.
Friedlander, M. R., W. Chen, et al. (2008). "Discovering microRNAs from deep sequencing data using miRDeep." Nat Biotechnol **26**(4): 407-15.

Friend, K., A. F. Lovejoy, et al. (2007). "U2 snRNP binds intronless histone pre-mRNAs to facilitate U7-snRNP-dependent 3' end formation." Mol Cell **28**(2): 240-52.
Fukagawa, T., M. Nogami, et al. (2004). "Dicer is essential for formation of the heterochromatin structure in vertebrate cells." Nat Cell Biol **6**(8): 784-91.
Gaidatzis, D., E. van Nimwegen, et al. (2007). "Inference of miRNA targets using evolutionary conservation and pathway analysis." BMC Bioinformatics **8**: 69.
Galardi, S., N. Mercatelli, et al. (2007). "miR-221 and miR-222 expression affects the proliferation potential of human prostate carcinoma cell lines by targeting p27Kip1." J Biol Chem **282**(32): 23716-24.
Gallie, D. R., N. J. Lewis, et al. (1996). "The histone 3'-terminal stem-loop is necessary for translation in Chinese hamster ovary cells." Nucleic Acids Res **24**(10): 1954-62.
Gillies, J. K. and I. A. Lorimer (2007). "Regulation of p27Kip1 by miRNA 221/222 in glioblastoma." Cell Cycle **6**(16): 2005-9.
Giraldez, A. J., R. M. Cinalli, et al. (2005). "MicroRNAs regulate brain morphogenesis in zebrafish." Science **308**(5723): 833-8.
Giraldez, A. J., Y. Mishima, et al. (2006). "Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs." Science **312**(5770): 75-9.
Godfrey, A. C., J. M. Kupsco, et al. (2006). "U7 snRNA mutations in Drosophila block histone pre-mRNA processing and disrupt oogenesis." Rna **12**(3): 396-409.
Gregory, R. I., T. P. Chendrimada, et al. (2005). "Human RISC couples microRNA biogenesis and posttranscriptional gene silencing." Cell **123**(4): 631-640.
Gregory, R. I., K. P. Yan, et al. (2004). "The Microprocessor complex mediates the genesis of microRNAs." Nature **432**(7014): 235-240.
Grimson, A., K. K. Farh, et al. (2007). "MicroRNA targeting specificity in mammals: determinants beyond seed pairing." Mol Cell **27**(1): 91-105.
Grishok, A., A. E. Pasquinelli, et al. (2001). "Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing." Cell **106**(1): 23-34.
Grishok, A. and P. A. Sharp (2005). "Negative regulation of nuclear divisions in Caenorhabditis elegans by retinoblastoma and RNA interference-related genes." Proc Natl Acad Sci U S A **102**(48): 17360-5.
Grivna, S. T., E. Beyret, et al. (2006). "A novel class of small RNAs in mouse spermatogenic cells." Genes & Development **20**(13): 1709-1714.
Gu, P., D. LeMenuet, et al. (2005). "Orphan nuclear receptor GCNF is required for the repression of pluripotency genes during retinoic acid-induced embryonic stem cell differentiation." Mol Cell Biol **25**(19): 8507-19.
Gu, P., J. G. Reid, et al. (2008). "Novel microRNA candidates and miRNA-mRNA pairs in embryonic stem (ES) cells." PLoS ONE **3**(7): e2548.
Guasch, G. and E. Fuchs (2005). "Mice in the world of stem cell biology." Nat Genet **37**(11): 1201-6.
Haase, A. D., L. Jaskiewicz, et al. (2005). "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing." EMBO Rep **6**(10): 961-7.
Hamilton, A. J. and D. C. Baulcombe (1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants." Science **286**(5441): 950-2.

Hammond, S. M., E. Bernstein, et al. (2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells." Nature **404**(6775): 293-6.
Harfe, B. D., M. T. McManus, et al. (2005). "The RNaseIII enzyme Dicer is required for morphogenesis but not patterning of the vertebrate limb." Proc Natl Acad Sci U S A **102**(31): 10898-903.
Harris, K. S., Z. Zhang, et al. (2006). "Dicer function is essential for lung epithelium morphogenesis." Proc Natl Acad Sci U S A **103**(7): 2208-13.
Harris, M. E., R. Bohni, et al. (1991). "Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps." Mol Cell Biol **11**(5): 2416-24.
Hatfield, S. D., H. R. Shcherbata, et al. (2005). "Stem cell division is regulated by the microRNA pathway." Nature **435**(7044): 974-8.
Hattori, N., K. Nishino, et al. (2004). "Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells." J Biol Chem **279**(17): 17063-9.
Hay, D. C., L. Sutherland, et al. (2004). "Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells." Stem Cells **22**(2): 225-35.
Hayashi, K., S. M. Chuva de Sousa Lopes, et al. (2008). "MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis." PLoS ONE **3**(3): e1738.
Hayashita, Y., H. Osada, et al. (2005). "A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation." Cancer Res **65**(21): 9628-32.
He, L., X. He, et al. (2007). "A microRNA component of the p53 tumour suppressor network." Nature **447**(7148): 1130-4.
He, L., J. M. Thomson, et al. (2005). "A microRNA polycistron as a potential human oncogene." Nature **435**(7043): 828-33.
Heintz, N., H. L. Sive, et al. (1983). "Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle." Mol Cell Biol **3**(4): 539-50.
Hendrich, B. and A. Bird (1998). "Identification and characterization of a family of mammalian methyl-CpG binding proteins." Mol Cell Biol **18**(11): 6538-47.
Hobbs, S. L., T. D. Warkentin, et al. (1993). "Transgene copy number can be positively or negatively associated with transgene expression." Plant Mol Biol **21**(1): 17-26.
Hornstein, E., J. H. Mansfield, et al. (2005). "The microRNA miR-196 acts upstream of Hoxb8 and Shh in limb development." Nature **438**(7068): 671-4.
Houbaviy, H. B., L. Dennis, et al. (2005). "Characterization of a highly variable eutherian microRNA gene." Rna **11**(8): 1245-57.
Houbaviy, H. B., M. F. Murray, et al. (2003). "Embryonic stem cell-specific MicroRNAs." Dev Cell **5**(2): 351-8.
Houwing, S., L. M. Kamminga, et al. (2007). "A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish." Cell **129**(1): 69-82.

Humphreys, D. T., B. J. Westman, et al. (2005). "MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function." Proc Natl Acad Sci U S A **102**(47): 16961-6.
Hutvagner, G., J. McLachlan, et al. (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." Science **293**(5531): 834-838.
Ivanovska, I., A. S. Ball, et al. (2008). "MicroRNAs in the miR-106b family regulate p21/CDKN1A and promote cell cycle progression." Mol Cell Biol **28**(7): 2167-74.
Jakymiw, A., S. Lian, et al. (2005). "Disruption of GW bodies impairs mammalian RNA interference." Nat Cell Biol **7**(12): 1267-74.
Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-80.
Jiang, J., Y. S. Chan, et al. (2008). "A core Klf circuitry regulates self-renewal of embryonic stem cells." Nat Cell Biol **10**(3): 353-60.
Johnson, C. D., A. Esquela-Kerscher, et al. (2007). "The let-7 microRNA represses cell proliferation pathways in human cells." Cancer Res **67**(16): 7713-22.
Johnson, S. M., H. Grosshans, et al. (2005). "RAS is regulated by the let-7 microRNA family." Cell **120**(5): 635-47.
Jurkowski, T. P., M. Meusburger, et al. (2008). "Human DNMT2 methylates tRNAAsp molecules using a DNA methyltransferase-like catalytic mechanism." Rna.
Kahvejian, A., Y. V. Svitkin, et al. (2005). "Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms." Genes Dev **19**(1): 104-13.
Kanellopoulou, C., S. A. Muljo, et al. (2005). "Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing." Genes Dev **19**(4): 489-501.
Karpf, A. R. and S. Matsui (2005). "Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells." Cancer Res **65**(19): 8635-9.
Kedde, M., M. J. Strasser, et al. (2007). "RNA-binding protein Dnd1 inhibits microRNA access to target mRNA." Cell **131**(7): 1273-86.
Ketting, R. F., S. E. Fischer, et al. (2001). "Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans." Genes Dev **15**(20): 2654-9.
Kim, T. H., L. O. Barrera, et al. (2005). "A high-resolution map of active promoters in the human genome." Nature **436**(7052): 876-80.
Kim, Y. K. and V. N. Kim (2007). "Processing of intronic microRNAs." Embo Journal **26**(3): 775-783.
Kirmizis, A., S. M. Bartley, et al. (2004). "Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27." Genes & Development **18**(13): 1592-1605.
Kloosterman, W. P., E. Wienholds, et al. (2004). "Substrate requirements for let-7 function in the developing zebrafish embryo." Nucleic Acids Res **32**(21): 6284-91.
Knight, S. W. and B. L. Bass (2001). "A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans." Science **293**(5538): 2269-71.

Kobayakawa, S., K. Miike, et al. (2007). "Dynamic changes in the epigenomic state and nuclear organization of differentiating mouse embryonic stem cells." Genes Cells **12**(4): 447-60.
Kobayashi, T., J. Lu, et al. (2008). "Dicer-dependent pathways regulate chondrocyte proliferation and differentiation." Proc Natl Acad Sci U S A **105**(6): 1949-54.
Kolev, N. G. and J. A. Steitz (2005). "Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs." Genes Dev **19**(21): 2583-92.
Kong, Y. W., I. G. Cannell, et al. (2008). "The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene." Proc Natl Acad Sci U S A.
Kouzarides, T. (2007). "Chromatin modifications and their function." Cell **128**(4): 693-705.
Krek, A., D. Grun, et al. (2005). "Combinatorial microRNA target predictions." Nat Genet **37**(5): 495-500.
Krichevsky, A. M., K. C. Sonntag, et al. (2006). "Specific microRNAs modulate embryonic stem cell-derived neurogenesis." Stem Cells **24**(4): 857-64.
Krutzfeldt, J., N. Rajewsky, et al. (2005). "Silencing of microRNAs in vivo with 'antagomirs'." Nature **438**(7068): 685-9.
Kuramochi-Miyagawa, S., T. Kimura, et al. (2004). "Mili, a mammalian member of piwi family gene, is essential for spermatogenesis." Development **131**(4): 839-849.
LaBella, F., H. L. Sive, et al. (1988). "Cell-cycle regulation of a human histone H2b gene is mediated by the H2b subtype-specific consensus element." Genes Dev **2**(1): 32-9.
Lagos-Quintana, M., R. Rauhut, et al. (2001). "Identification of novel genes coding for small expressed RNAs." Science **294**(5543): 853-8.
Landgraf, P., M. Rusu, et al. (2007). "A mammalian microRNA expression atlas based on small RNA library sequencing." Cell **129**(7): 1401-14.
Lanzotti, D. J., H. Kaygun, et al. (2002). "Developmental control of histone mRNA and dSLBP synthesis during Drosophila embryogenesis and the role of dSLBP in histone mRNA 3' end processing in vivo." Mol Cell Biol **22**(7): 2267-82.
Lau, N. C., L. P. Lim, et al. (2001). "An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans." Science **294**(5543): 858-62.
Lau, N. C., A. G. Seto, et al. (2006). "Characterization of the piRNA complex from rat testes." Science **313**(5785): 363-367.
le Sage, C., R. Nagel, et al. (2007). "Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation." Embo J **26**(15): 3699-708.
Lee, R. C. and V. Ambros (2001). "An extensive class of small RNAs in Caenorhabditis elegans." Science **294**(5543): 862-4.
Lee, R. C., R. L. Feinbaum, et al. (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-54.
Lee, Y. S. and A. Dutta (2007). "The tumor suppressor microRNA let-7 represses the HMGA2 oncogene." Genes Dev **21**(9): 1025-30.

Lee, Y. S., K. Nakahara, et al. (2004). "Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways." Cell **117**(1): 69-81.
Leonhardt, H., A. W. Page, et al. (1992). "A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei." Cell **71**(5): 865-73.
Lewis, B. P., I. H. Shih, et al. (2003). "Prediction of mammalian microRNA targets." Cell **115**(7): 787-98.
Li, E., T. H. Bestor, et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-26.
Li, J. Y., M. T. Pu, et al. (2007). "Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog." Mol Cell Biol **27**(24): 8748-59.
Liang, G., M. F. Chan, et al. (2002). "Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements." Mol Cell Biol **22**(2): 480-91.
Lim, L. P., N. C. Lau, et al. (2005). "Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs." Nature **433**(7027): 769-773.
Lindbo, J. A., L. Silva-Rosales, et al. (1993). "Induction of a Highly Specific Antiviral State in Transgenic Plants: Implications for Regulation of Gene Expression and Virus Resistance." Plant Cell **5**(12): 1749-1759.
Linsley, P. S., J. Schelter, et al. (2007). "Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression." Mol Cell Biol **27**(6): 2240-52.
Litovchick, L., S. Sadasivam, et al. (2007). "Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence." Mol Cell **26**(4): 539-51.
Liu, J., F. V. Rivas, et al. (2005). "A role for the P-body component GW182 in microRNA function." Nat Cell Biol **7**(12): 1261-6.
Liu, J., M. A. Valencia-Sanchez, et al. (2005). "MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies." Nat Cell Biol **7**(7): 719-23.
Liu, J. D., M. A. Carmell, et al. (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science **305**(5689): 1437-1441.
Loh, Y. H., Q. Wu, et al. (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nat Genet **38**(4): 431-40.
Loh, Y. H., W. Zhang, et al. (2007). "Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells." Genes Dev **21**(20): 2545-57.
Lu, Y., J. M. Thomson, et al. (2007). "Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells." Dev Biol **310**(2): 442-53.
Lund, E., S. Guttinger, et al. (2004). "Nuclear export of microRNA precursors." Science **303**(5654): 95-98.
Ma, T., B. A. Van Tine, et al. (2000). "Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription." Genes Dev **14**(18): 2298-313.
Maroney, P. A., Y. Yu, et al. (2006). "Evidence that microRNAs are associated with translating messenger RNAs in human cells." Nat Struct Mol Biol **13**(12): 1102-7.
Martens, J. H., R. J. O'Sullivan, et al. (2005). "The profile of repeat-associated histone lysine methylation states in the mouse epigenome." Embo J **24**(4): 800-12.

Marzluff, W. F., P. Gongidi, et al. (2002). "The human and mouse replication-dependent histone genes." Genomics **80**(5): 487-98.
Mathonnet, G., M. R. Fabian, et al. (2007). "MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F." Science **317**(5845): 1764-7.
Matsubara, H., T. Takeuchi, et al. (2007). "Apoptosis induction by antisense oligonucleotides against miR-17-5p and miR-20a in lung cancers overexpressing miR-17-92." Oncogene **26**(41): 6099-105.
Meister, G., M. Landthaler, et al. (2004). "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs." Molecular Cell **15**(2): 185-197.
Mette, M. F., W. Aufsatz, et al. (2000). "Transcriptional silencing and promoter methylation triggered by double-stranded RNA." Embo J **19**(19): 5194-201.
Meyer, W. J., S. Schreiber, et al. (2006). "Overlapping functions of argonaute proteins in patterning and morphogenesis of Drosophila embryos." PLoS Genet **2**(8): e134.
Mikkelsen, T. S., M. C. Ku, et al. (2007). "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells." Nature **448**(7153): 553-U2.
Miska, E. A., E. Alvarez-Saavedra, et al. (2007). "Most Caenorhabditis elegans microRNAs are individually not essential for development or viability." PLoS Genet **3**(12): e215.
Mitsui, K., Y. Tokuzawa, et al. (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-42.
Mohn, F., M. Weber, et al. (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." Mol Cell **30**(6): 755-66.
Morita, S., T. Horii, et al. (2007). "One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation." Genomics **89**(6): 687-96.
Muljo, S. A., K. M. Ansel, et al. (2005). "Aberrant T cell differentiation in the absence of Dicer." J Exp Med **202**(2): 261-9.
Muller, J., C. M. Hart, et al. (2002). "Histone methyltransferase activity of a Drosophila polycomb group repressor complex." Cell **111**(2): 197-208.
Murchison, E. P., J. F. Partridge, et al. (2005). "Characterization of Dicer-deficient murine embryonic stem cells." Proc Natl Acad Sci U S A **102**(34): 12135-40.
Murchison, E. P., P. Stein, et al. (2007). "Critical roles for Dicer in the female germline." Genes Dev **21**(6): 682-93.
Nan, X., F. J. Campoy, et al. (1997). "MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin." Cell **88**(4): 471-81.
Napoli, C., C. Lemieux, et al. (1990). "Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans." Plant Cell **2**(4): 279-289.
Narita, T., T. M. Yung, et al. (2007). "NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs." Mol Cell **26**(3): 349-65.
Neilson, J. R., G. X. Zheng, et al. (2007). "Dynamic regulation of miRNA expression in ordered stages of cellular development." Genes Dev **21**(5): 578-89.
Newman, M. A., J. M. Thomson, et al. (2008). "Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing." Rna.

Nichols, J., B. Zevnik, et al. (1998). "Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4." Cell **95**(3): 379-91.
Niwa, H. (2007). "Open conformation chromatin and pluripotency." Genes Dev **21**(21): 2671-6.
Niwa, H., J. Miyazaki, et al. (2000). "Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells." Nat Genet **24**(4): 372-6.
O'Carroll, D., I. Mecklenbrauker, et al. (2007). "A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway." Genes Dev **21**(16): 1999-2004.
O'Donnell, K. A., E. A. Wentzel, et al. (2005). "c-Myc-regulated microRNAs modulate E2F1 expression." Nature **435**(7043): 839-43.
Obernosterer, G., P. J. F. Leuschner, et al. (2006). "Post-transcriptional regulation of microRNA expression." Rna-a Publication of the Rna Society **12**(7): 1161-1167.
Okamura, K., J. W. Hagen, et al. (2007). "The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila." Cell **130**(1): 89-100.
Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-57.
Olsen, P. H. and V. Ambros (1999). "The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation." Dev Biol **216**(2): 671-80.
Ooi, S. K., C. Qiu, et al. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-7.
Osley, M. A. (1991). "The regulation of histone synthesis in the cell cycle." Annu Rev Biochem **60**: 827-61.
Pan, G., S. Tian, et al. (2007). "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells." Cell Stem Cell **1**(3): 299-312.
Parker, R. and U. Sheth (2007). "P bodies and the control of mRNA translation and degradation." Mol Cell **25**(5): 635-46.
Pauley, K. M., T. Eystathioy, et al. (2006). "Formation of GW bodies is a consequence of microRNA genesis." EMBO Rep **7**(9): 904-10.
Petersen, C. P., M. E. Bordeleau, et al. (2006). "Short RNAs repress translation after initiation in mammalian cells." Mol Cell **21**(4): 533-42.
Pillai, R. S., C. G. Artus, et al. (2004). "Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis." Rna **10**(10): 1518-25.
Pillai, R. S., S. N. Bhattacharyya, et al. (2005). "Inhibition of translational initiation by Let-7 MicroRNA in human cells." Science **309**(5740): 1573-6.
Raver-Shapira, N., E. Marciano, et al. (2007). "Transcriptional activation of miR-34a contributes to p53-mediated apoptosis." Mol Cell **26**(5): 731-43.
Rehwinkel, J., P. Natalin, et al. (2006). "Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster." Molecular and Cellular Biology **26**(8): 2965-2975.

Robertson, K. D., K. Keyomarsi, et al. (2000). "Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells." Nucleic Acids Res **28**(10): 2108-13.
Rosenfeld, N., R. Aharonov, et al. (2008). "MicroRNAs accurately identify cancer tissue origin." Nat Biotechnol **26**(4): 462-9.
Ruby, J. G., C. Jan, et al. (2006). "Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans." Cell **127**(6): 1193-207.
Ruby, J. G., C. H. Jan, et al. (2007). "Intronic microRNA precursors that bypass Drosha processing." Nature **448**(7149): 83-U7.
Rybak, A., H. Fuchs, et al. (2008). "A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment." Nat Cell Biol.
Sanchez, R. and W. F. Marzluff (2002). "The stem-loop binding protein is required for efficient translation of histone mRNA in vivo and in vitro." Mol Cell Biol **22**(20): 7093-104.
Schmitter, D., J. Filkowski, et al. (2006). "Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells." Nucleic Acids Research **34**(17): 4801-4815.
Schratt, G. M., F. Tuebing, et al. (2006). "A brain-specific microRNA regulates dendritic spine development." Nature **439**(7074): 283-289.
Schubeler, D., D. Scalzo, et al. (2002). "Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing." Nat Genet **32**(3): 438-42.
Schwarz, D. S., G. Hutvagner, et al. (2003). "Asymmetry in the assembly of the RNAi enzyme complex." Cell **115**(2): 199-208.
Seggerson, K., L. Tang, et al. (2002). "Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation." Dev Biol **243**(2): 215-25.
Selbach, M., B. Schwanhausser, et al. (2008). "Widespread changes in protein synthesis induced by microRNAs." Nature.
Sen, G. L. and H. M. Blau (2005). "Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies." Nat Cell Biol **7**(6): 633-6.
Shao, Z. H., F. Raible, et al. (2000). "Regulation of chromatin structure by polycomb and by SWI/SNF." Faseb Journal **14**(8): A1582-A1582.
Shilatifard, A. (2008). "Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation." Curr Opin Cell Biol **20**(3): 341-8.
Si, M. L., S. Zhu, et al. (2007). "miR-21-mediated tumor growth." Oncogene **26**(19): 2799-803.
Smith, A. G. (2001). "Embryo-derived stem cells: of mice and men." Annu Rev Cell Dev Biol **17**: 435-62.
Sood, P., A. Krek, et al. (2006). "Cell-type-specific signatures of microRNAs on target mRNA expression." Proc Natl Acad Sci U S A **103**(8): 2746-51.
Stark, A., J. Brennecke, et al. (2005). "Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution." Cell **123**(6): 1133-46.

Steger, D. J., M. I. Lefterova, et al. (2008). "DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells." Mol Cell Biol **28**(8): 2825-39.
Su, A. I., M. P. Cooke, et al. (2002). "Large-scale analysis of the human and mouse transcriptomes." Proc Natl Acad Sci U S A **99**(7): 4465-70.
Suh, M. R., Y. Lee, et al. (2004). "Human embryonic stem cells express a unique set of microRNAs." Dev Biol **270**(2): 488-98.
Sylvestre, Y., V. De Guire, et al. (2007). "An E2F/miR-20a autoregulatory feedback loop." J Biol Chem **282**(4): 2135-43.
Takahashi, K. and S. Yamanaka (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." Cell **126**(4): 663-76.
Tam, O. H., A. A. Aravin, et al. (2008). "Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes." Nature **453**(7194): 534-8.
Tang, F., M. Kaneda, et al. (2007). "Maternal microRNAs are essential for mouse zygotic development." Genes Dev **21**(6): 644-8.
Tay, Y. M., W. L. Tam, et al. (2008). "MicroRNA-134 modulates the differentiation of mouse embryonic stem cells, where it causes post-transcriptional attenuation of Nanog and LRH1." Stem Cells **26**(1): 17-29.
Thermann, R. and M. W. Hentze (2007). "Drosophila miR2 induces pseudo-polysomes and inhibits translation initiation." Nature **447**(7146): 875-8.
Thompson, B. J. and S. M. Cohen (2006). "The Hippo pathway regulates the bantam microRNA to control cell proliferation and apoptosis in Drosophila." Cell **126**(4): 767-74.
Thomson, J. M., M. Newman, et al. (2006). "Extensive post-transcriptional regulation of microRNAs and its implications for cancer." Genes & Development **20**(16): 2202-2207.
Turner, B. M. (2002). "Cellular memory and the histone code." Cell **111**(3): 285-91.
Wagner, E. J., B. D. Burch, et al. (2007). "A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing." Mol Cell **28**(4): 692-9.
Wakiyama, M., K. Takimoto, et al. (2007). "Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system." Genes Dev **21**(15): 1857-62.
Walsh, C. P., J. R. Chaillet, et al. (1998). "Transcription of IAP endogenous retroviruses is constrained by cytosine methylation." Nat Genet **20**(2): 116-7.
Wang, B., T. M. Love, et al. (2006). "Recapitulation of short RNA-directed translational gene silencing in vitro." Mol Cell **22**(4): 553-60.
Wang, Q., Y. C. Li, et al. (2008). "miR-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor Rb2/p130." Proc Natl Acad Sci U S A **105**(8): 2889-94.
Wang, Y., R. Medvid, et al. (2007). "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal." Nat Genet **39**(3): 380-5.
Wassenegger, M., S. Heimes, et al. (1994). "RNA-directed de novo methylation of genomic sequences in plants." Cell **76**(3): 567-76.
Vasudevan, S. and J. A. Steitz (2007). "AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2." Cell **128**(6): 1105-18.

Vasudevan, S., Y. Tong, et al. (2007). "Switching from repression to activation: microRNAs can up-regulate translation." Science **318**(5858): 1931-4.
Watanabe, T., Y. Totoki, et al. (2008). "Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes." Nature **453**(7194): 539-43.
Watt, F. and P. L. Molloy (1988). "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter." Genes Dev **2**(9): 1136-43.
Vazquez, F. (2006). "Arabidopsis endogenous small RNAs: highways and byways." Trends Plant Sci **11**(9): 460-8.
Weber, M., J. J. Davies, et al. (2005). "Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells." Nat Genet **37**(8): 853-62.
Weber, M., I. Hellmann, et al. (2007). "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome." Nat Genet **39**(4): 457-66.
Ventura, A., A. G. Young, et al. (2008). "Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters." Cell **132**(5): 875-86.
Verdel, A., S. Jia, et al. (2004). "RNAi-mediated targeting of heterochromatin by the RITS complex." Science **303**(5658): 672-6.
Whitfield, M. L., L. X. Zheng, et al. (2000). "Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms." Mol Cell Biol **20**(12): 4188-98.
Wienholds, E., M. J. Koudijs, et al. (2003). "The microRNA-producing enzyme Dicer1 is essential for zebrafish development." Nat Genet **35**(3): 217-8.
Vigorito, E., K. L. Perks, et al. (2007). "microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells." Immunity **27**(6): 847-59.
Viswanathan, S. R., G. Q. Daley, et al. (2008). "Selective blockade of MicroRNA processing by Lin28." Science **320**(5872): 97-100.
Volpe, T. A., C. Kidner, et al. (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-7.
Voorhoeve, P. M., C. le Sage, et al. (2006). "A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors." Cell **124**(6): 1169-81.
Wu, L. and J. G. Belasco (2005). "Micro-RNA regulation of the mammalian lin-28 gene during neuronal differentiation of embryonal carcinoma cells." Mol Cell Biol **25**(21): 9198-208.
Wu, L., J. Fan, et al. (2006). "MicroRNAs direct rapid deadenylation of mRNA." Proc Natl Acad Sci U S A **103**(11): 4034-9.
Xie, X., J. Lu, et al. (2005). "Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals." Nature **434**(7031): 338-45.
Yang, L., D. Chen, et al. (2007). "Argonaute 1 regulates the fate of germline stem cells in Drosophila." Development **134**(23): 4265-72.
Yang, W. J., D. D. Yang, et al. (2005). "Dicer is required for embryonic angiogenesis during mouse development." J Biol Chem **280**(10): 9330-5.

Yekta, S., I. H. Shih, et al. (2004). "MicroRNA-directed cleavage of HOXB8 mRNA." Science **304**(5670): 594-596.
Yi, R., M. N. Poy, et al. (2008). "A skin microRNA promotes differentiation by repressing 'stemness'." Nature **452**(7184): 225-9.
Yi, R., Y. Qin, et al. (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." Genes & Development **17**(24): 3011-3016.
Yu, J., M. A. Vodyanik, et al. (2007). "Induced pluripotent stem cell lines derived from human somatic cells." Science **318**(5858): 1917-20.
Yuan, H., N. Corbi, et al. (1995). "Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3." Genes Dev **9**(21): 2635-45.
Zamore, P. D., T. Tuschl, et al. (2000). "RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals." Cell **101**(1): 25-33.
Zhao, Y., E. Samal, et al. (2005). "Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis." Nature **436**(7048): 214-20.
Zheng, L., R. G. Roeder, et al. (2003). "S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component." Cell **114**(2): 255-66.
Zhou, B., S. Wang, et al. (2007). "miR-150, a microRNA expressed in mature B and T cells, blocks early B cell development when expressed prematurely." Proc Natl Acad Sci U S A **104**(17): 7080-5.

3. Results and discussion

3.1 miRNAs control *de novo* DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells

Lasse Sinkkonen, Tabea Hugenschmidt, Philipp Berninger, Dimos Gaidatzis, Fabio Mohn, Caroline Artus-Revel, Mihaela Zavolan, Petr Svoboda and Witold Filipowicz

nature structural & molecular biology

MicroRNAs control *de novo* DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells

Lasse Sinkkonen[1], Tabea Hugenschmidt[1], Philipp Berninger[2], Dimos Gaidatzis[2], Fabio Mohn[1], Caroline G Artus-Revel[1], Mihaela Zavolan[2], Petr Svoboda[3] & Witold Filipowicz[1]

Loss of microRNA (miRNA) pathway components negatively affects differentiation of embryonic stem (ES) cells, but the underlying molecular mechanisms remain poorly defined. Here we characterize changes in mouse ES cells lacking Dicer (Dicer1). Transcriptome analysis of *Dicer*$^{-/-}$ cells indicates that the ES-specific miR-290 cluster has an important regulatory function in undifferentiated ES cells. Consistently, many of the defects in Dicer-deficient cells can be reversed by transfection with miR-290 family miRNAs. We demonstrate that *Oct4* (also known as *Pou5f1*) silencing in differentiating *Dicer*$^{-/-}$ ES cells is accompanied by accumulation of repressive histone marks but not by DNA methylation, which prevents the stable repression of *Oct4*. The methylation defect correlates with downregulation of *de novo* DNA methyltransferases (Dnmts). The downregulation is mediated by Rbl2 and possibly other transcriptional repressors, potential direct targets of miR-290 cluster miRNAs. The defective DNA methylation can be rescued by ectopic expression of *de novo* Dnmts or by transfection of the miR-290 cluster miRNAs, indicating that *de novo* DNA methylation in ES cells is controlled by miRNAs.

Short 20–25-nucleotide (nt) RNAs have emerged recently as important sequence-specific regulators of gene expression in eukaryotes[1–4]. Short RNAs are produced from long double-stranded RNA (dsRNA) and miRNA precursors, which are processed by the RNase III family enzymes Drosha and Dicer to yield mature effector molecules, small interfering RNAs (siRNAs) and miRNAs[1–5]. miRNAs are the dominant class of short RNAs in mammalian cells, from which several hundred different miRNAs have been identified and implicated in the regulation of many cellular processes[6,7]. Mammalian miRNAs typically base-pair imperfectly with the 3′ untranslated region (3′ UTR) of target mRNAs and induce their translational repression or degradation[8,9]. The eight 5′ terminal nucleotides form the critical miRNA region for target mRNA recognition. This region, generally referred to as the 'seed', hybridizes nearly perfectly with the target to nucleate the miRNA-mRNA interaction[10,11]. Most computational methods of miRNA target prediction incorporate this constraint[12].

ES cells are pluripotent cells derived from the inner cell mass of blastocysts. Depending on the culture conditions, ES cells can differentiate into various cell types[13]. The Oct4, Sox2 and Nanog transcription factors form a core circuit responsible for the transcriptional control of ES cell renewal and pluripotency[14,15]. Mouse ES cells contain numerous miRNAs, including a cluster of six miRNAs (miR-290 through miR-295) that share a 5′-proximal AAGUGC motif[16,17]. The cluster (for brevity referred to as the miR-290 cluster) is specific to ES cells[17]. Its expression increases during preimplantation development[18] and remains high in undifferentiated ES cells, but decreases after ES cell differentiation[17]. Genes and pathways regulated by the miR-290 cluster are unknown.

The loss of Dicer in mouse ES cells results in miRNA depletion[19,20] and causes differentiation defects *in vivo* and *in vitro*[19]. *Dicer*$^{-/-}$ cells make no contribution to chimeric mice and fail to generate teratomas *in vivo*. *In vitro*, *Dicer*$^{-/-}$ cells form embryoid body (EB)–like structures, but there is little morphological evidence of differentiation. Expression of *Oct4*, a characteristic marker of pluripotent ES cells, is only partially decreased in mutant EBs after day 5 of differentiation, and expression of endodermal and mesodermal markers is not detectable[19]. Similarly, the loss of Dgcr8, a protein required specifically for miRNA maturation, causes partial downregulation of pluripotency markers during retinoic acid (RA)–induced differentiation[21].

In this work, we investigated the molecular mechanisms underlying the inability of *Dicer*$^{-/-}$ ES cells to differentiate. We found that silencing of the *Oct4* pluripotency factor is properly initiated in differentiating *Dicer*$^{-/-}$ ES cells, but it is not followed by *de novo* DNA methylation of the promoter. Consistent with this, we observed that levels of *de novo* DNA methyltransferases are downregulated in *Dicer*$^{-/-}$ cells in an miR-290 cluster–dependent manner. Thus, our

[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland. [2]Division of Bioinformatics, Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland, and the Swiss Institute of Bioinformatics. [3]Institute of Molecular Genetics, Videnska 1083, Prague, Czech Republic. Correspondence should be addressed to W.F. (filipowi@fmi.ch) or P.S. (svobodap@img.cas.cz).

Received 22 July 2007; accepted 16 January 2008; published online 2 March 2008; doi:10.1038/nsmb.1391

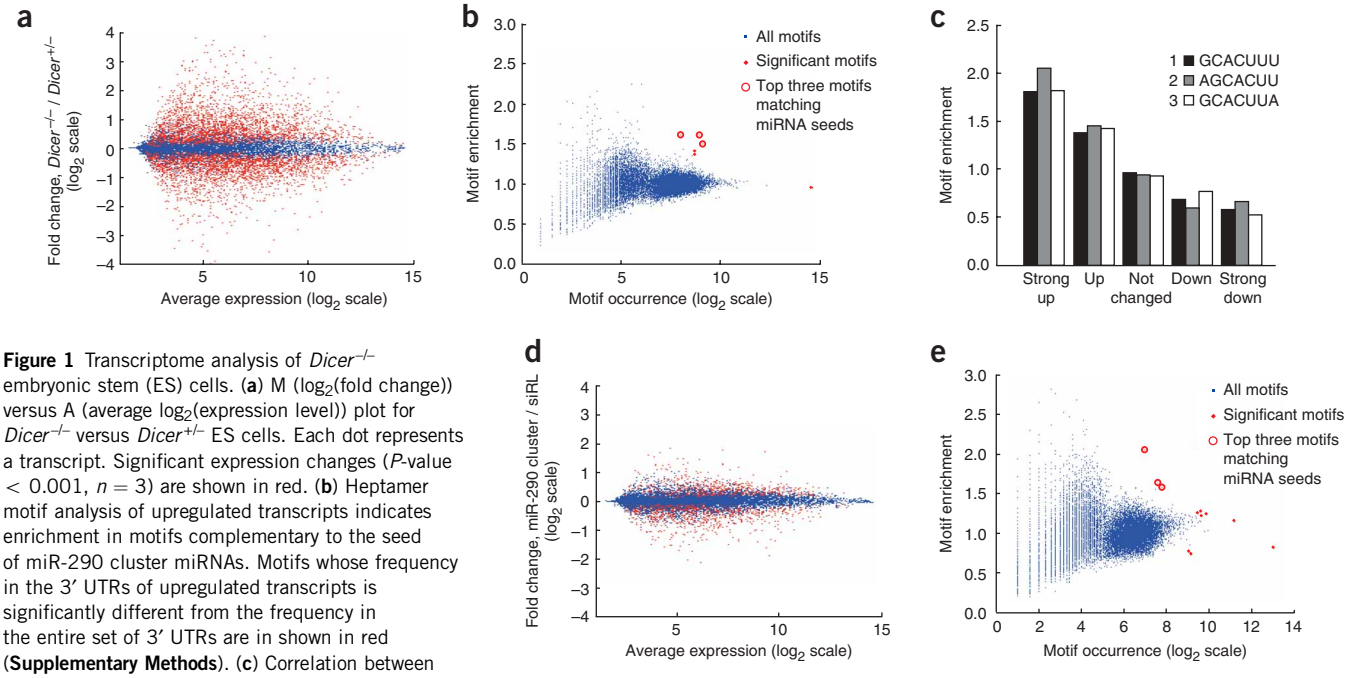


Figure 1 Transcriptome analysis of *Dicer*$^{-/-}$ embryonic stem (ES) cells. (**a**) M ($\log_2$(fold change)) versus A (average $\log_2$(expression level)) plot for *Dicer*$^{-/-}$ versus *Dicer*$^{+/-}$ ES cells. Each dot represents a transcript. Significant expression changes (*P*-value < 0.001, $n = 3$) are shown in red. (**b**) Heptamer motif analysis of upregulated transcripts indicates enrichment in motifs complementary to the seed of miR-290 cluster miRNAs. Motifs whose frequency in the 3′ UTRs of upregulated transcripts is significantly different from the frequency in the entire set of 3′ UTRs are in shown in red (**Supplementary Methods**). (**c**) Correlation between the occurrence of sequence motifs and the change in mRNA expression. Transcripts were divided into five sets on the basis of their change in expression in *Dicer*$^{-/-}$ compared with *Dicer*$^{+/-}$ cells as follows: strong down, more than 2-fold downregulation; down, 1.2-fold to 2-fold downregulation; not changed, 1.2-fold downregulation to 1.2-fold upregulation; up, 1.2-fold to 2-fold upregulation; strong up, more than 2-fold upregulation. (**d**) M versus A plot for *Dicer*$^{-/-}$ ES cells transfected with the miR-290 cluster versus *Dicer*$^{-/-}$ ES cells transfected with the small interfering RNA against *Renilla* luciferase mRNA (siRL), a nonspecific control. Significant expression changes (*P*-value < 0.001, $n = 3$) are shown in red. (**e**) Transcripts that were downregulated with a *P*-value < 0.001 in the miR-290 cluster–transfected *Dicer*$^{-/-}$ cells were extracted and analyzed as in **Figure 1b**. Many of the significantly enriched motifs are complementary to positions 2–7 of the miRNAs in the miR-290 cluster. The motifs complementary to the seed of siRL did not show any enrichment, indicating that there was a minimal off-target effect.

data indicate that the *de novo* DNA methylation in differentiating ES cells is regulated by ES-specific miRNAs from the miR-290 cluster.

RESULTS

Transcriptome analysis of *Dicer*$^{-/-}$ ES cells

To study the roles of miRNAs in gene regulation in ES cells, we profiled the transcriptomes of *Dicer*$^{-/-}$ and *Dicer*$^{+/-}$ ES cells using Affymetrix microarrays. We found a similar number of transcripts that were upregulated (2,551; *P*-value < 0.001) and downregulated (2,578; *P*-value < 0.001) upon the loss of Dicer (**Fig. 1a**). Analysis of core pluripotency regulators, as well as different differentiation markers, indicated that *Dicer*$^{-/-}$ cells retain characteristics of undifferentiated ES cells (**Supplementary Fig. 1** online).

The binding of miRNAs to the 3′ UTR of mRNAs commonly results in degradation of mRNA targets. Numerous studies have reported significant enrichment of sequences complementary to miRNA seeds in 3′ UTRs of mRNAs that are upregulated in miRNA knockdowns, or downregulated upon overexpression of miRNAs[22–24]. We searched for sequence motifs (heptamers) that are enriched in the 3′ UTRs of transcripts upregulated in the *Dicer*$^{-/-}$ cells and that could explain the mRNA expression changes (**Supplementary Methods** online). The three motifs that were most significantly enriched (**Fig. 1b**) were all complementary to the seed region of embryonic miRNAs[25]: miR-291a-3p, miR-291b-3p, miR-294 and miR-295 in the case of the first and second motifs (GCACUUU and AGCACUU), and miR-302 in the case of the third motif (GCACUUA). The seed region of miR-302 differs from that of miR-290 cluster members only in the first nucleotide (**Supplementary Fig. 2a** and **Supplementary Table 1** online). The enrichment of the GCACUUA motif may imply that miR-302 has an important role in regulating mRNA expression in ES cells. Alternatively, it may indicate that miRNAs prefer target sites with an A residue opposite the 5′-most nucleotide of the miRNA, as has been proposed before[11]. Because the same motif is also most significantly enriched in the 3′ UTRs of mRNAs that are downregulated upon transfection with miR-290 cluster miRNAs (see below), we favor the second explanation. We also note that the ubiquitously expressed oncogenic miRNAs of the miR-17/20/93/106 cluster share extensive similarity at their 5′ end with the embryonic miRNAs (**Supplementary Table 1**) and could also contribute to mRNA regulation in ES cells. As shown in **Figure 1c**, the frequency of the top three motifs decreased gradually from the mRNAs that are most strongly upregulated in *Dicer*$^{-/-}$ cells to the mRNAs that are strongly downregulated.

We examined expression of the miR-290 cluster primary transcript using available microarray data[26]. Quantification of the primary transcript indicated that expression of the cluster occurs zygotically and reaches the highest level in the blastocyst (**Supplementary Fig. 2b**). Notably, accumulation of the miR-290 cluster transcript was downregulated in *Dicer*$^{-/-}$ ES cells, indicating a possible feedback control of its expression by the cluster or other miRNAs (**Supplementary Fig. 2c**). Array analysis of miRNA levels in *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ ES cells using Exiqon arrays revealed that, as expected[17,27], miR-290 cluster miRNAs are abundantly expressed in ES cells, and miR-290 cluster and other miRNA levels are reduced in *Dicer*$^{-/-}$ cells (**Supplementary Fig. 2d** and **Supplementary Table 2** online).

Identification of primary miR-290 cluster targets

To increase the accuracy of the miRNA target prediction, we compared the transcriptome profile of *Dicer*$^{-/-}$ ES cells (transfected with a

nonspecific siRNA as a control) with that of *Dicer*$^{-/-}$ ES cells transfected with the siRNA-like form of miRNAs of the miR-290 cluster (**Fig. 1d**). Applying the same heptamer motif analysis used above, we found a few motifs enriched in transcripts that were downregulated after miR-290 cluster miRNA transfection. Among them are motifs complementary to seeds of miR-290 cluster miRNAs, identical to the top three motifs identified above (**Fig. 1e** and **Supplementary Table 1**). Analysis of both array experiments showed a good inverse correlation between transcript-level changes in *Dicer*$^{-/-}$ cells (compared to *Dicer*$^{+/-}$ cells) and *Dicer*$^{-/-}$ cells transfected with miR-290 cluster miRNAs (compared to control *Dicer*$^{-/-}$ cells) (**Supplementary Fig. 3a–c** online). The correlation holds for mRNAs that carry the miR-290 cluster seed-matching sequences in their 3′ UTR, as well as for those that do not (**Supplementary Fig. 3b,c**). The correlation for mRNAs lacking seed-matching sequences anywhere in the transcript was as good as that shown in **Supplementary Fig. 3c** (data not shown). These data suggest that not only primary miRNA effects, but also many secondary gene-expression changes controlled by miR-290 cluster miRNAs, are reversible in *Dicer*$^{-/-}$ ES cells.

To predict primary miR-290 cluster targets, we used data from both sets of microarray experiments. We intersected the lists of transcripts that showed a significant change (P-value < 0.001) in the expected direction in the *Dicer*$^{-/-}$ cells compared to *Dicer*$^{+/-}$ cells (upregulation) and in the miR-290 cluster–transfected *Dicer*$^{-/-}$ cells compared to control siRNA-transfected *Dicer*$^{-/-}$ cells (downregulation). The list was then filtered to keep only the transcripts whose 3′ UTRs had at least one match to the GCACUU hexamer, which is common to all significantly enriched heptamers. The resulting list of predicted targets contained 253 mRNAs (**Supplementary Table 3** online). However, it is likely that the number of targets is even larger, as not all expressed mRNAs are detectable by microarrays and some genes may be regulated at the protein rather than the transcript level.

Indirect control of *de novo* methyltransferases by miRNAs

Inspection of microarray data indicated that expression of *de novo* DNA methyltransferase genes *Dnmt3a*, *Dnmt3b* and *Dnmt3l* was significantly downregulated in undifferentiated *Dicer*$^{-/-}$ ES cells (**Fig. 2a**). Protein levels of Dnmt3a2, Dnmt3b1 and Dnmt3b6 were also lower in *Dicer*$^{-/-}$ cells, whereas the ubiquitously expressed isoform of Dnmt3a, Dnmt3a1 (ref. 28), remained unchanged (**Fig. 2b**). Notably, expression of *de novo* DNA methyltransferases could be rescued, at both mRNA and protein levels, upon transfection of all miR-290 cluster miRNAs or miR-291a-3p alone (**Fig. 2c–e**).

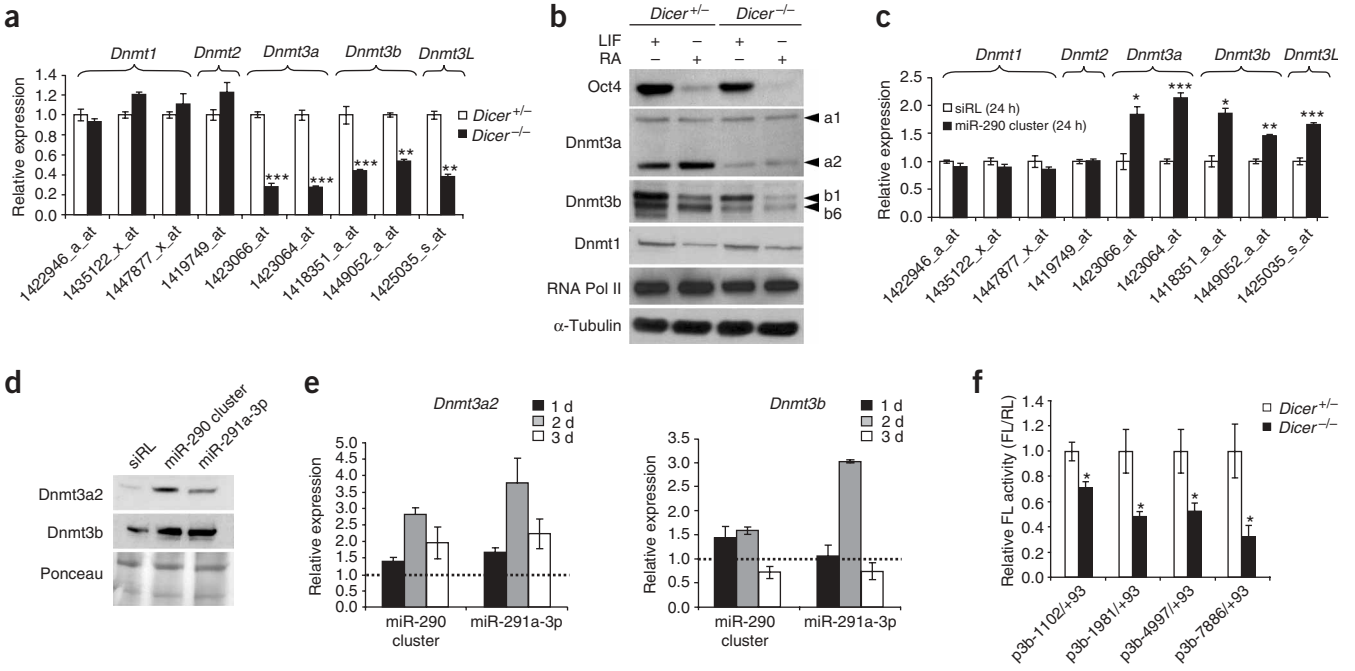


Figure 2 *De novo* DNA methyltransferases (Dnmts) are downregulated in *Dicer*$^{-/-}$ embryonic stem (ES) cells and their expression is rescued by miR-290 cluster miRNAs. (**a**) Expression of DNA methyltransferases in undifferentiated *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells as analyzed by Affymetrix microarrays. The probe sets detecting mRNAs encoding different DNA methyltransferases are indicated. Mean expression (±s.d.; $n = 3$) in *Dicer*$^{+/-}$ cells was set to one. Signals from probe sets detecting *Dnmt3a*, *Dnmt3b* and *Dnmt3l* were significantly downregulated in *Dicer*$^{-/-}$ cells (two-tailed *t*-test *P*-values, from left to right: 0.0001, 0.0006, 0.0093, 0.0022 and 0.0010). (**b**) Western blot analysis of Dnmt1, Dnmt3a and Dnmt3b levels in ES cells cultured in the presence of either leukemia inhibitory factor (LIF) or retinoic acid (RA) for 3 d. α-Tubulin was used as a loading control. Quantification of western blots shown in **b** and **d** and in **Figure 3f** by image densitometry revealed a 3.0-fold to 5.6-fold change in the level of Dnmt3a2 and a 2.0-fold to 4.4-fold change in the levels of Dnmt3b1/b6 between conditions of low and high expression of the proteins. (**c**) The miR-290 cluster miRNAs induce accumulation of mRNAs encoding Dnmt3a, Dnmt3b and Dnmt3l in *Dicer*$^{-/-}$ cells. Mean values (±s.d.; $n = 3$) observed for the siRL-transfected cells (a nonspecific control) were set to one. The *P*-values, from left to right, were: 0.0102, 0.0008, 0.0021, 0.0010 and 0.0009. (**d**) Dnmt3a2 and Dnmt3b expression 3 d after transfection with siRL, miR-290 cluster or miR-291a-3p. Ponceau staining served as a loading control. (**e**) Upregulation of *Dnmt3a2* and *Dnmt3b1/6* (quantified by RT-qPCR) in response to transfection of either all miR-290 cluster miRNAs or miR-291a-3p into *Dicer*$^{-/-}$ ES cells. Mean expression values (±s.d.; $n = 3$) were normalized to glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) and are shown relative to corresponding siRL samples, whose expression values were set to one (dashed line). (**f**) Dicer loss affects transcription from the *Dnmt3b* promoter. Firefly luciferase (FL) reporters containing *Dnmt3b* promoter fragments were co-transfected to *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells together with the pRL-TK control reporter. Mean FL activity values (±s.e.m., $n \geq 3$) in *Dicer*$^{+/-}$ cells were set to one. The *P*-values, from left to right, were: 0.0192, 0.0391, 0.0238 and 0.0230.

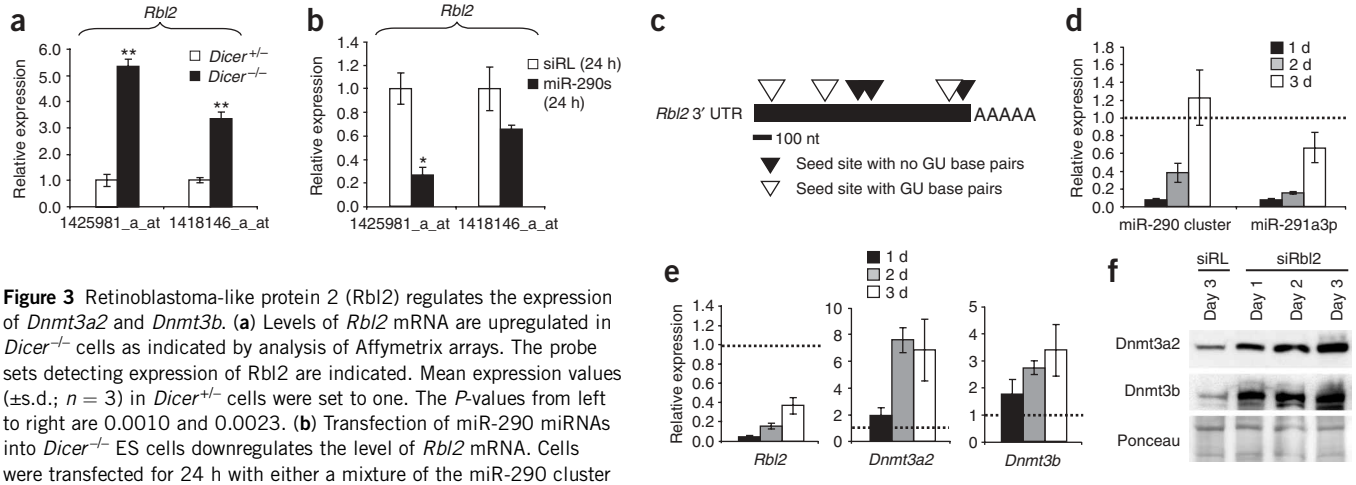


Figure 3 Retinoblastoma-like protein 2 (Rbl2) regulates the expression of *Dnmt3a2* and *Dnmt3b*. (**a**) Levels of *Rbl2* mRNA are upregulated in *Dicer*$^{-/-}$ cells as indicated by analysis of Affymetrix arrays. The probe sets detecting expression of Rbl2 are indicated. Mean expression values (±s.d.; $n = 3$) in *Dicer*$^{+/-}$ cells were set to one. The *P*-values from left to right are 0.0010 and 0.0023. (**b**) Transfection of miR-290 miRNAs into *Dicer*$^{-/-}$ ES cells downregulates the level of *Rbl2* mRNA. Cells were transfected for 24 h with either a mixture of the miR-290 cluster miRNAs or with siRL (small interfering RNA against *Renilla* luciferase mRNA). Mean expression values (±s.d.; $n = 3$) in siRL-transfected cells were set to one. The *P*-values from left to right were 0.0135 and 0.1082. (**c**) Schematic representation of the localization of predicted binding sites for AAGUGC seed–containing miRNAs in the 3′ UTR of *Rbl2* mRNA. Predicted binding sites that contain GU base pairs in the seed and those without GU base pairs in the seed are marked with white and black triangles, respectively. (**d**) Downregulation of *Rbl2* in response to transfection of *Dicer*$^{-/-}$ ES cells with either all miR-290 cluster miRNAs or miR-291a-3p. For other details see **Figure 2e**. (**e**) Effect of *Rbl2* knockdown on *Dnmt3a2* and *Dnmt3b* mRNA levels. Cells were transfected with siRNAs against *Rbl2* (siRbl2) or with siRL as a control. For other details see **Figure 2e**. (**f**) Western blot analysis of Dnmt3a2 and Dnmt3b expression 1 d, 2 d or 3 d after siRbl2 transfection. Expression after transfection of siRL (3 d) is shown as a control. Ponceau staining served as a loading control.

Similar downregulation of all three *Dnmt3* genes upon loss of Dicer and their upregulation in response to transfection of miR-290 cluster miRNAs indicated that miRNAs may regulate the expression of *Dnmt3* genes indirectly, possibly by controlling the activity of a common transcriptional repressor. This possibility is supported by the observations that *Dnmt3a2*, *Dnmt3b* and *Dnmt3l* contain similar TATA-less GC-rich promoters, are regulated by SP1-SP3 transcription factors, and are highly expressed in blastocysts and ES cells but are downregulated during differentiation into somatic lineages[28–31]. To corroborate the possibility of transcriptional regulation, we compared the activity of firefly luciferase (FL) reporters containing *Dnmt3b* promoter regions of different lengths. Activity of the reporters was significantly lower in *Dicer*$^{-/-}$ than in *Dicer*$^{+/-}$ ES cells (**Fig. 2f**), arguing that the *Dnmt3b* promoter is markedly repressed in cells lacking *Dicer* and suggesting that downregulation of *Dnmt3* genes in *Dicer*$^{-/-}$ ES cells may occur at the level of transcription.

Among the predicted primary targets of the miR-290 cluster (**Supplementary Table 3**), we identified several annotated[32] transcriptional repressors that are upregulated during embryonic differentiation after the blastocyst stage[31]. They include genes for the basic Kruppel-like factor *Klf3*, the nuclear receptor *Nr2f2*, the zinc-finger proteins *Zmynd11* and *Zbtb7*, and retinoblastoma-like 2 (*Rbl2*) (**Fig. 3a**,**b**). Several other observations make *Rbl2* a plausible candidate for the miR-290 cluster–regulated transcriptional repressor of *de novo* DNA methyltransferases. The *Rbl2* 3′ UTR contains conserved potential binding sites for miR-290 cluster miRNAs (**Fig. 3c**), and *Rbl2* mRNA is downregulated upon transfection of all miR-290 cluster miRNAs or miR-291a-3p alone into *Dicer*$^{-/-}$ ES cells (**Fig. 3b**,**d**). Rbl2 repressor was recently shown to associate with the *DNMT3B* promoter in human glioblastoma cells (ref. 33 and Discussion). In mouse ES cells, *Rbl2* is expressed at low levels, and during neuronal differentiation its expression correlates inversely with the expression of the miR-290 cluster and *de novo* DNA methyltransferases (F.M. and D. Schübeler, Friedrich Miescher Institute, unpublished results). We used RNA interference to obtain more direct evidence that *Rbl2* indeed regulates the expression of *de novo* DNA methyltransferases. Transfection of siRNAs against *Rbl2* resulted in a marked increase of *Dnmt3a2* and *Dnmt3b* expression at both mRNA and protein levels (**Fig. 3e**,**f**). Taken together, these data argue in favor of Rbl2 as a target of the miR-290 cluster that acts as a repressor, downregulating the expression of *de novo* DNA methyltransferases.

Defective DNA methylation of *Oct4* in *Dicer*$^{-/-}$ cells

To investigate in more detail the differentiation defects in *Dicer*$^{-/-}$ cells and the possible role of the miR-290 cluster, we examined expression of *Oct4*, the core pluripotency regulator of ES cells. When differentiation was induced with 100 nM RA in the absence of leukemia inhibitory factor (LIF), the mRNA and protein levels of *Oct4* decreased similarly in *Dicer*$^{-/-}$ and control cells at day 3 (**Fig. 4a**). The expression level of the orphan nuclear receptor gene *Gcnf*, an early repressor of *Oct4*, *Nanog* and other pluripotency markers[34], was upregulated to the same extent in *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells after 1 d of RA treatment (**Fig. 4b**). We could also detect accumulation of repressive histone marks at the *Oct4* promoter (**Fig. 4c**,**d**), indicating that the initiation of *Oct4* silencing was not strongly perturbed. However, repression of *Oct4* at day 6 of differentiation was clearly stronger in *Dicer*$^{+/-}$ ES cells (**Fig. 4e**). When RA was removed at day 6 and the cells were cultured in the presence of LIF for an extra 4 d, the *Oct4* mRNA levels in *Dicer*$^{-/-}$ cells increased to approximately 40% of the initial level, whereas *Oct4* expression remained repressed in *Dicer*$^{+/-}$ cells (**Fig. 4e**). A similar pattern of expression was observed for *Nanog* (**Supplementary Fig. 4** online).

Incomplete and reversible silencing of *Oct4* in RA-treated *Dicer*$^{-/-}$ ES cells is notably similar to findings demonstrating that the stable silencing of *Oct4* is dependent on a correct *de novo* methylation of DNA[35,36]. Therefore, we used bisulfite sequencing to analyze the methylation status of the *Oct4* promoter during the RA-induced differentiation. In *Dicer*$^{+/-}$ ES cells, DNA methylation was already detectable after 3 d of differentiation; it increased further at day 6 and remained high following the withdrawal of RA. In marked contrast,

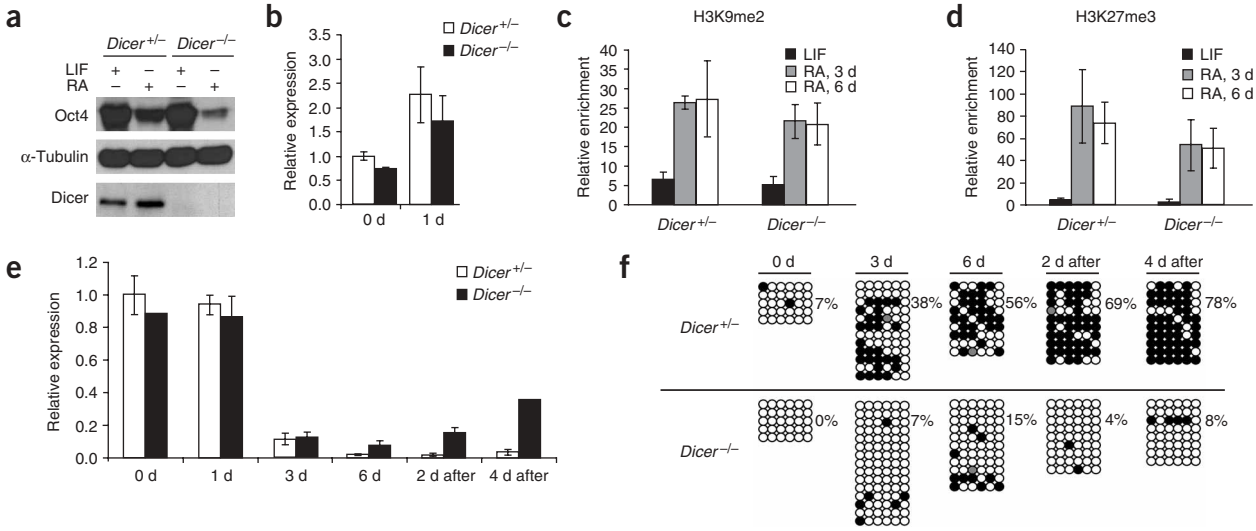


Figure 4 *Oct4* expression during differentiation of *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ embryonic stem (ES) cells. (**a**) Western blot analysis of Oct4 levels in *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells cultured in the presence of either leukemia inhibitory factor (LIF) or retinoic acid (RA) for 3 d. (**b**) Similar upregulation of the orphan nuclear receptor gene *Gcnf* expression in *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells in response to RA. Expression was estimated by RT-qPCR. The values, normalized to glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) expression, represent means (±s.e.m.; $n \geq 3$). Expression in control *Dicer*$^{+/-}$ cells at the 0 d time point was set as one. (**c**,**d**) Accumulation of repressive histone marks at the *Oct4* promoter. *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells, cultured in the presence of LIF, RA for 3 d (RA, 3 d) or RA for 6 d (RA, 6 d), were used for chromatin immunoprecipitation (ChIP) analysis using antibodies against dimethylated histone H3 lysine 9 (H3K9me2; **c**) and trimethylated histone H3 lysine 27 (H3K27me3; **d**). The enrichment values represent means (±s.e.m.; $n \geq 3$). (**e**) RT-qPCR analysis of *Oct4* expression during RA-induced differentiation at 0 d, 1 d, 3 d or 6 d, and after returning the cells to the LIF-containing medium for up to an additional 4 days (2 d after, 4 d after). *Oct4* expression was normalized to *Gapdh* as in **b** ($n \geq 3$). (**f**) Analysis of CpG methylation of the *Oct4* core promoter (positions –212 to –8) during differentiation, followed by 2 d or 4 d culture in the presence of LIF. Each row of dots represents CpGs in one sequenced clone. Black dots represent methylated CpGs and white dots represent unmethylated CpGs. Sites for which the methylation status was uncertain are in gray. The cells used were the same as those used for the experiment shown in **e**. Average percentages of the methylated CpG sites are indicated.

the *Oct4* promoter failed to undergo DNA methylation in differentiating *Dicer*$^{-/-}$ cells (**Fig. 4f**).

To address the possibility that impaired maintenance of DNA methylation is responsible for the observed methylation defect, we analyzed several typically hypermethylated sequences and found no loss of their methylation in undifferentiated or differentiated *Dicer*$^{-/-}$ ES cells (**Supplementary Fig. 5** online). Furthermore, expression of the maintenance DNA methyltransferase *Dnmt1* was not affected either by the loss of Dicer or upon transfection of miR-290 cluster miRNAs into *Dicer*$^{-/-}$ ES cells (**Fig. 2a–c**), suggesting that maintenance of DNA methylation is not impaired in *Dicer*$^{-/-}$ ES cells.

Rescue of *de novo* DNA methylation of *Oct4* by miRNAs

We tested whether ectopic expression of Dnmt3a2, Dnmt3b and Dnmt3l, or transfection with miR-290 cluster miRNAs, is sufficient to rescue the defective *Oct4* promoter methylation. Co-transfection of *Dicer*$^{-/-}$ ES cells with constructs expressing all three methyltransferases from a heterologous promoter restored the *de novo* DNA methylation in *Dicer*$^{-/-}$ cells treated with RA for 3 d (**Fig. 5a**). Transfection of *Dicer*$^{-/-}$ ES cells with miR-290 cluster miRNAs had a similar effect (**Fig. 5a**). These results indicate that the observed Oct4 promoter methylation defect is due to the repressed expression of *de novo* DNA methyltransferases in *Dicer*$^{-/-}$ ES cells.

To address whether the DNA methylation defect is more general, we analyzed the methylation status of two testis-specific genes, *Tsp50* and *Sox30*, which are silenced in ES cells and undergo *de novo* DNA methylation during differentiation (F.M. and D. Schübeler, unpublished results). *Dicer*$^{+/-}$ but not *Dicer*$^{-/-}$ ES cells showed limited DNA methylation at *Tsp50* and *Sox30* promoters, even in the undifferentiated state (**Supplementary Fig. 6** online). Differentiation of *Dicer*$^{+/-}$ but not *Dicer*$^{-/-}$ cells was accompanied by additional DNA methylation (**Supplementary Fig. 6**). Nevertheless, the DNA-methylation changes at *Tsp50* and *Sox30* promoters were less pronounced than those observed at the *Oct4* locus, and the *de novo* DNA methylation of *Tsp50* and *Sox30* promoters was not uniformly distributed along analyzed sequences (**Supplementary Fig. 6**). Ectopic expression of *de novo* DNA methyltransferases affected the accumulation of DNA methylation during differentiation, whereas transfection of miR-290 cluster miRNAs resulted in increased DNA methylation at the 3′ portion of the *Tsp50* sequence but had no appreciable effect at the *Sox30* promoter (**Supplementary Fig. 6**; see Discussion). Taken together, the data suggest that the defect in *de novo* methylation in *Dicer*$^{-/-}$ ES cells may be of more global character.

Dicer$^{-/-}$ ES cells grow substantially more slowly than *Dicer*$^{+/-}$ ES cells[20], and we found that transfection of miR-290 cluster miRNAs into *Dicer*$^{-/-}$ ES cells partially rescues the growth phenotype (**Fig. 5b**), possibly by regulating expression of p21, an established repressor of cell-cycle progression[37] (**Supplementary Fig. 7a–c** online; see Discussion). To eliminate the possibility that the observed changes of *Oct4* DNA methylation are a consequence of different proliferation rates rather than a specific miR-290 cluster–mediated regulation, we tested whether the proliferation rate of ES cells has an effect on the onset of *de novo* DNA methylation and the expression levels of the Dnmt3 enzymes.

To reduce proliferation of *Dicer*$^{+/-}$ ES cells to a rate similar to that of *Dicer*$^{-/-}$ ES cells (**Fig. 5b**), cells were treated with rapamycin, an inhibitor of mammalian target of rapamycin (TOR). Rapamycin reduces the proliferation of mouse ES cells without significantly

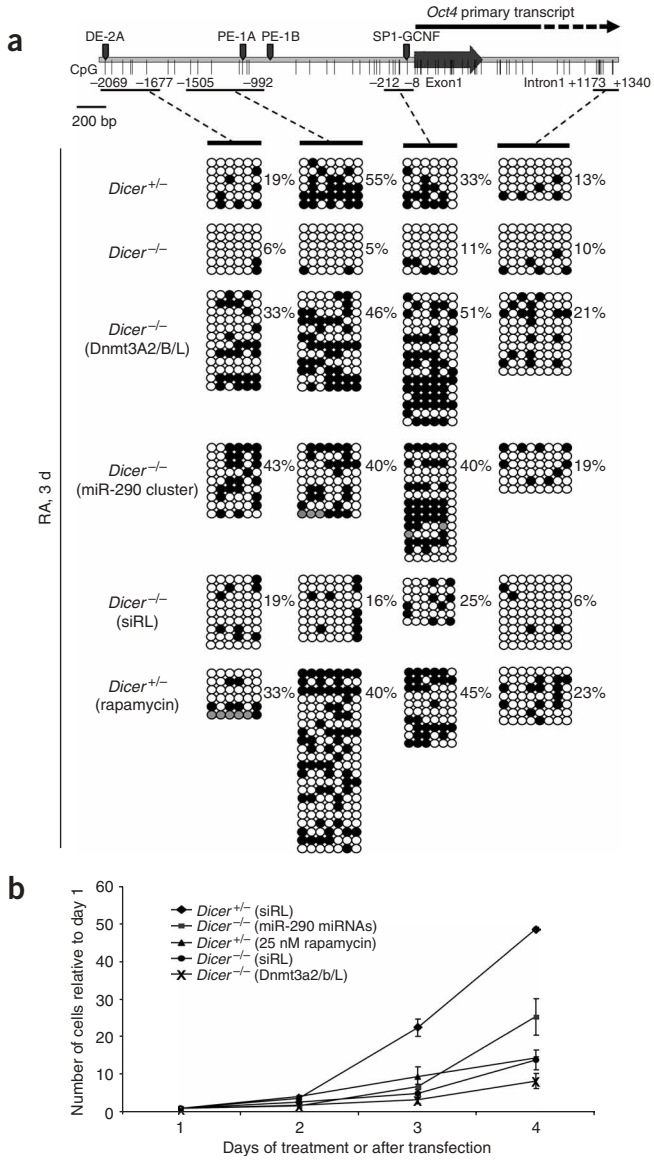


Figure 5 Deficient *de novo* DNA methylation of the *Oct4* promoter in *Dicer*$^{-/-}$ embryonic stem (ES) cells can be rescued by expression of *de novo* DNA methyltransferases (Dnmts) or by transfection of miR-290 cluster miRNAs. (**a**) Analysis of CpG methylation in four different *Oct4* regions. The scheme identifies positions of bisulfite-sequenced regions with respect to the *Oct4* transcription start site. SP1-GCNF depicts characterized transcription factor binding sites in the *Oct4* promoter. PE-1A and PE-1B show positions of previously characterized 1A and 1B sequences in the proximal enhancer and DE-2A is the position of 2A sequence in the distal enhancer (for the detailed *Oct4* promoter annotation, see ref. 54 and references therein). Represented from top to bottom: untransfected *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells; *Dicer*$^{-/-}$ cells co-transfected with plasmids expressing EGFP-Dnmt3a2, EGFP-Dnmt3b and EGFP-Dnmt3l; *Dicer*$^{-/-}$ cells transfected with miR-290 cluster mimics; *Dicer*$^{-/-}$ cells transfected with siRL (small interfering RNA against *Renilla* luciferase mRNA); and *Dicer*$^{+/-}$ cells treated with rapamycin. Both *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ ES cells were differentiated for 3 d with retinoic acid (RA) in the absence of leukemia inhibitory factor (LIF). For other details, see **Figure 4f**. The data originate from experiments independent of that shown in **Figure 4f**. (**b**) Effects of different treatments on proliferation of *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ ES cells. Equal numbers of undifferentiated *Dicer*$^{-/-}$ and *Dicer*$^{+/-}$ cells were transfected with miR-290 cluster miRNAs, siRL or a mix of plasmids expressing Dnmt3a2, Dnmt3b and Dnmt3l. Alternatively, cells were grown in the presence of rapamycin. Average number of cells is shown relative to the number of cells present at day 1 after transfection (±s.d.; $n = 3$).

affecting their cell-cycle profile[38], making the growth properties of rapamycin-treated *Dicer*$^{+/-}$ ES cells comparable to that of *Dicer*$^{-/-}$ cells[20]. In *Dicer*$^{+/-}$ cells grown in the presence of rapamycin, DNA methylation readily accumulated at the *Oct4* promoter after 3 d of RA treatment (**Fig. 5a**). Likewise, decreased proliferation had no significant effect on the expression of *Dnmt3a2* or *Dnmt3b1/6* (**Supplementary Fig. 7d**). Furthermore, restoration of *Oct4* promoter methylation by ectopic expression of *de novo* DNA methyltransferases occurred without an increase in the proliferation rate of *Dicer*$^{-/-}$ ES cells (**Fig. 5a**,**b**). Taken together, these data demonstrate that the *Oct4* promoter methylation defect is not caused by the slower proliferation of *Dicer*$^{-/-}$ ES cells but is dependent on the miR-290 cluster miRNAs.

DISCUSSION

Our data indicate that miRNAs bearing the AAGUGC seed, largely represented by the miR-290 cluster, are the functionally dominant miRNAs in mouse ES cells. In fact, the miR-290 cluster miRNAs were able to reverse many of the defects due to loss of Dicer when transfected into ES cells. We also found that *de novo* DNA methylation in differentiating ES cells is controlled by the miR-290 cluster and that this regulation is required for stable repression of *Oct4*. We propose that, in undifferentiated ES cells, the miR-290 cluster miRNAs suppress a transcriptional repressor that targets genes encoding *de novo* DNA methyltransferases. The predicted primary targets of the miR-290 cluster include several transcriptional repressors, and we identified *Rbl2* as a factor contributing to repression of *Dnmt3* genes.

The expression of approximately one-quarter of predicted primary miR-290 cluster targets in ES cells is high in the oocyte but reduced in the blastocyst and somatic cells (data not shown). This resembles the situation in zebrafish, where the zygotic AAGUGC seed–containing miR-430 miRNAs control the maternal mRNA degradation[39]. However, murine maternal mRNAs are largely degraded before zygotic genome activation[26], hence before the miR-290 cluster expression. Moreover, the transition between maternal and zygotic gene expression is much slower in mammals than in the zebrafish[40]. Thus, the miR-290 cluster and related miRNAs restrict embryonic expression of genes that are highly expressed in the oocyte rather than having an extensive role in the rapid elimination of maternal transcripts. However, miR-290 cluster miRNAs and miR-430 may share some conserved roles in development, as the mouse homologs of zebrafish *lft1* and *lft2*, important regulators of mesoderm formation and targets of miR-430 (ref. 41), are found among ~250 predicted primary targets of miR-290 cluster miRNAs (**Supplementary Table 3**).

The microarray analysis also identified several transcripts that showed inverse changes in the Dicer knockout and miR-290 cluster rescue microarray experiments, but contained no matches to the seed of miR-290 cluster miRNAs. These are probably secondary targets whose expression is regulated by the primary targets of the miRNAs. Notably, the microarray analysis indicated that many secondary effects, probably brought about by the primary targets, are reversible despite the fact that the Dicer$^{-/-}$ ES cell line was established a relatively long time ago.

Both primary and secondary targets probably contribute to the reduced proliferation rate of *Dicer*$^{-/-}$ ES cells, which can be partially rescued by transfecting miR-290 cluster miRNAs. Notably, one of the predicted primary targets of the miR-290 cluster is *p21* (also known as

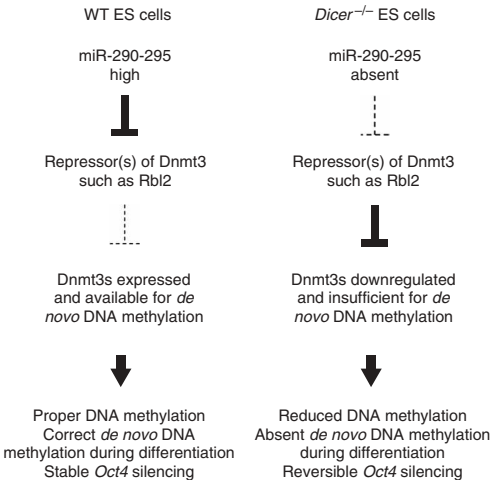


Figure 6 A model for a role of miRNAs in *de novo* DNA methylation in embryonic stem (ES) cells. Dnmt, DNA methyltransferase; Rbl2, retinoblastoma-like protein 2; WT, wild type.

Cdkn1a), a cyclin-dependent kinase inhibitor that has been shown to repress cell-cycle progression[42]. It is well established that control of *p21* expression is achieved through negative transcriptional regulators[37]. Our data argue for an additional layer of control of p21 expression by miRNAs carrying the AAGUGC seed sequence. *p21* mRNA has three GCACUU motifs in its 3′ UTR (**Supplementary Fig. 7a**), two of which are conserved across mammals. *p21* mRNA is upregulated more than three-fold in *Dicer*$^{-/-}$ ES cells, and this misregulation can be corrected by transfection of miR-290 cluster miRNAs (**Supplementary Fig. 7b,c**). Thus, upregulation of *p21* could be one of the mechanisms causing the slower-growth phenotype. Although in ES cells miRNAs carrying the AAGUGC seed sequence are primarily represented by miR-290 cluster miRNAs, other related miRNAs, such as the oncomirs of the miR-17/19/106 cluster[43], could regulate expression of *p21* in other tissues. Notably, the reverse complement of AAAGUGC (positions 2–8 in miR-17-5p) was one of the motifs that was highly enriched in 3′ UTRs of transcripts upregulated in human HEK293 cells depleted of Dicer or the argonaute protein AGO2 (ref. 24). At the same time, these cells grew more slowly, and the p21 transcript was upregulated. As miR-17/19/106 miRNAs are fairly ubiquitously expressed[27], they may provide another way to modulate expression of the *p21* tumor suppressor, with a predictable outcome for cellular growth.

The category of secondary targets includes *de novo* DNA methyltransferases, which are downregulated in *Dicer*$^{-/-}$ ES cells and upregulated upon miR-290 cluster miRNA transfection. Our data suggest that reduced expression of *Dnmt3* genes in *Dicer*$^{-/-}$ ES cells is the cause of *de novo* DNA-methylation defects observed during differentiation. Decreased expression of Dnmt3a2 and Dnmt3b, correlating with defective DNA methylation, has been described in mouse XX ES cells[44], arguing that even incomplete depletion of Dnmt3 enzymes may be limiting for proper *de novo* DNA methylation. Dnmt3a, Dnmt3b and possibly Dnmt3l may function as a complex[36]. Hence, even partial downregulation of each of them may strongly affect DNA methylation.

We investigated whether the proliferation rate itself affects *Dnmt3* expression and *de novo* DNA methylation. We found that *Dnmt3* expression and *de novo* DNA methylation are not impaired when the growth of control *Dicer*$^{+/-}$ ES cells is reduced by rapamycin. As the rapamycin-treated wild-type and *Dicer*$^{-/-}$ ES cells have comparable cell-cycle profiles and similarly slow proliferation rates[20,38], it is unlikely that the altered growth rate of *Dicer*$^{-/-}$ ES cells is responsible for decreased *Dnmt3* gene expression and the loss of *de novo* DNA methylation during differentiation. Furthermore, ectopic expression of *de novo* DNA methyltransferases rescued *de novo* DNA methylation without an apparent effect on proliferation of *Dicer*$^{-/-}$ cells. Because *de novo* DNA methylation proceeds normally in rapamycin-treated *Dicer*$^{+/-}$ ES cells, which show minimal proliferation during 3 d of RA-induced differentiation, it is unlikely that clonal effects in the cell culture would significantly distort the results of DNA-methylation analysis.

We propose that the transcription of *Dnmt3* genes is regulated in ES cells by a repressor protein whose mRNA is a target of miR-290 cluster miRNAs (**Fig. 6**). Loss of the miR-290 cluster miRNAs in *Dicer*$^{-/-}$ cells would cause the upregulation of the repressor, followed by the downregulation of *de novo* DNA methyltransferases. This type of *Dnmt3* regulation may be restricted to ES cells, as the levels of *Dnmt3* mRNAs are not affected in HEK293 cells with knockdown of Dicer or Argonaute proteins[24]. A suitable candidate for the repressor that targets *Dnmt3* genes is Rbl2, whose mRNA has all the features of a primary miR-290 cluster target. Consistent with our model, knockdown of *Rbl2* in *Dicer*$^{-/-}$ cells had a positive effect on *Dnmt3a2* and *Dnmt3b* expression. Rbl2 is a tumor suppressor that is capable of repressing E2f4 target genes as a part of the DREAM repressor complex[33]. Notably, the expression profile of human *Dnmt3b* during the cell cycle (low in G1 and G0 and upregulated in S phase[45]) is similar to that of the E2f4 target genes repressed by Rbl2 (ref. 33). RBL2 and the DREAM complex were recently shown to associate physically with the *Dnmt3b* promoter in human glioblastoma cells[33], suggesting that RBL2 can directly repress transcription of *Dnmt3* genes. Certainly, as the miR-290 cluster controls expression of a number of transcriptional repressors, *Rbl2* may not be the only regulator of *de novo* DNA methylation in ES cells. Fabbri *et al.*[46] have recently reported that the miR-29 family of miRNAs (miR-29s) can directly target *Dnmt3a* and *Dnmt3b* mRNAs and repress synthesis of *de novo* DNA methyltransferases in human lung cancer cells. miR-29 miRNAs are expressed in mouse ES cells and downregulated upon loss of Dicer, but our data argue against a major role of these miRNAs in controlling *Dnmt3a/b* mRNA or protein levels in mouse ES cells.

One of the functions of *de novo* DNA methylation during ES cell differentiation is the stable silencing of the pluripotency program. Our data indicate that, although the initial phase of transcriptional repression of *Oct4* seems to be undisturbed, the *de novo* DNA methylation of the *Oct4* promoter is severely impaired during differentiation of Dicer$^{-/-}$ cells. These results are consistent with the observation that stable silencing of *Oct4* is dependent on correct *de novo* methylation of DNA[35,36]. The defect in *de novo* DNA methylation may not be confined to *Oct4*, as *Nanog*, another core pluripotency factor, showed a similar expression profile (**Supplementary Fig. 4**). In addition, the promoters of *Tsp50* and *Sox30*, two testis-specific genes that are silent in ES cells and acquire *de novo* DNA methylation during differentiation, also failed to undergo DNA methylation in *Dicer*$^{-/-}$ cells. DNA-methylation data from these two loci are less conclusive, possibly resulting from slower kinetics of accumulation of methylation at these loci, exacerbated by a transient nature of the rescue with miR-290 cluster miRNAs. Nevertheless, accumulation of DNA methylation at these promoters is consistent with that of *Oct4*, suggesting a more general defect in *de novo* DNA methylation in *Dicer*$^{-/-}$ ES cells.

The defects in *de novo* DNA methylation in *Dicer*$^{-/-}$ ES cells may contribute decisively to the loss of the ability to differentiate *in vitro*

and *in vivo*. Notably, *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ double-mutant ES cells retain an undifferentiated morphology, and their late passages fail to form teratomas in nude mice[47]. The defects in *de novo* DNA methylation may also underlie the variable levels of centromeric DNA methylation reported for different *Dicer*$^{-/-}$ ES lines[19,20], because the loss of *de novo* DNA methyltransferases results in gradual DNA demethylation during prolonged culture[47].

In summary, our analysis of gene expression in mouse *Dicer*$^{-/-}$ ES cells indicates that many of the observed transcriptome changes that occur upon loss of Dicer can be attributed to miRNAs, particularly to those of the miR-290 cluster. We have identified ~250 candidate primary targets of the AAGUGC seed–containing miRNAs, and we also identified many genes that they regulate indirectly. Most notably, we demonstrated that *de novo* DNA methylation is defective in *Dicer*$^{-/-}$ ES cells, and that this is due to the indirect control of expression of the *de novo* DNA methyltransferases by the miR-290 cluster. The established link between miR-290 cluster miRNAs and *de novo* DNA methylation in ES cells indicates that miRNAs may contribute substantially to the epigenetic control of gene expression.

METHODS

Cell culture. The *Dicer* heterozygous (+/–; line D4) and *Dicer*-deficient (–/–; line 27H10) ES cells (referred to as *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$, respectively) were kindly provided by G. Hannon, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA[20]. They were maintained on gelatin-coated plates with DMEM supplemented with 15% (w/v) FCS, sodium pyruvate, β-mercaptoethanol, nonessential amino acids and LIF. Differentiation of ES cells was carried out in the absence of LIF and the presence of 100 nM RA. When indicated, cells were cultured for 4 d in the presence of 25 nM rapamycin (200 μM stock of rapamycin dissolved in ethanol). Control cells were grown in the presence of ethanol at equivalent concentration. For differentiation in the presence of rapamycin, the cells were cultured for 1 d with rapamycin and LIF followed by 3 d without LIF and with 100 nM RA and 25 nM rapamycin.

Plasmids. The control reporter constructs encoding firefly (FL) or *Renilla* (RL) luciferase (pGL3-FF and pRL-TK, respectively) were described earlier[24]. FL reporters under the control of *Dnmt3b* promoter fragments (p3b-1102/+93-FF, p3b-1981/+93-FF, p3b-4997/+93-FF and p3b-7886/+93-FF) and constructs encoding the EGFP-tagged *de novo* DNA methyltransferases (pCag-EGFP-Dnmt3a2, p-Cag-EGFP-Dnmt3b and p-Cag-EGFP-Dnmt3L) were kindly provided by K. Ura, Osaka University Graduate School of Medicine, Osaka, Japan[48,49].

Transfection of reporter constructs. At least three independent transfection experiments in triplicate were done in each case. For luciferase assays, *Dicer*$^{-/-}$ cells were transfected in six-well plates with 500 ng of indicated FL reporter constructs and 50 ng of pTK-RL as a control, using Lipofectamine 2000 reagent (Invitrogen). All luciferase assays were performed 24 h after transfection.

Other transfections. Other transfections were performed using the Mouse ES cell Nucleofection Kit (Amaxa Biosystems) and program A23 of Nucleofector I apparatus (Amaxa Biosystems). Approximately 3×10^6 *Dicer*$^{-/-}$ cells were used per transfection and the cells were plated immediately after electroporation. Transfections of siRNAs were performed according to the manufacturer's instructions, using 300 pmol of siRNA against RL mRNA (siRL) (Eurogentec), 50 pmol of siGENOME smartPOOL siRNAs against Rbl2 (Dharmacon), 50 pmol of each of the mmu-mir-290, mmu-mir-291a-3p, mmu-mir-292-3p, mmu-mir-293, mmu-mir-294 and mmu-mir-295 miRNA mimics (Dharmacon), or 300 pmol of mmu-mir-291a-3p, together with 2 μg of pCX-EGFP[50], which served as control for transfection efficiency. For rescue of *de novo* DNA methylation by a mixture of pCag-EGFP-Dnmt3a2, pCag-EGFP-Dnmt3b and pCag-EGFP-Dnmt3L plasmids, the *Dicer*$^{-/-}$ cells were co-transfected with 7 μg of each of these plasmids, using the Nucleofector I apparatus. The EGFP-expressing cells were collected using a MoFlow cell sorter (Dako Cytomation) after 3 d of culture in the presence of 100 nM RA and the absence of LIF.

Chromatin immunoprecipitation. Chromatin immunoprecipitations (ChIPs) were performed as described previously[51]. *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ ES cells, either undifferentiated or treated for 3 d with RA, were cross-linked by adding formaldehyde directly to the medium to a final concentration of 1% (w/v) at room temperature. The reaction was stopped after 8 min by adding glycine to a final concentration of 0.15 M. Cell lysates were sonicated to generate 300–1,500-bp DNA fragments. After preclearing the samples with Protein A Agarose (Upstate), the immunocomplexes were formed using anti-H3K9me2 or anti-H3K27me3 antibodies (Upstate). Immunocomplexes were collected with 30 μl of Protein A Agarose (Upstate). The purified DNA and a 1:100 dilution of the respective input DNA were used as templates for quantitative real-time PCR (RT-qPCR), using the ABI Prism 7000 Sequence Detection System (Applied Biosystems), Platinum SYBR Green qPCR SuperMix (Invitrogen) and primers specific for the glyceraldehyde-3-phosphate dehydrogenase (*Gapdh*) and *Oct4* promoters. Obtained values were first normalized to the respective input DNA and further to the enrichment at the *Gapdh* promoter where these modifications do not accumulate. Sequences of primers are listed in **Supplementary Table 4** online. Annealing of all primers was done at 55 °C.

Bisulfite sequencing. Bisulfite sequencing was performed using the Epitect Bisulfite sequencing kit (Qiagen) according to the manufacturer's conditions. Up to 2 μg of genomic DNA was used as a starting material. PCR amplification conditions were as described[52]. All primers are listed in **Supplementary Table 4**.

Statistical analysis. Analysis of microarray data and motifs, including statistical methods, is described in detail in the **Supplementary Methods**. All remaining statistical analysis used two-tailed *t*-tests.

Real-time quantitative RT-PCR (RT-qPCR). Total RNA from ES cells was extracted using the Absolutely RNA Miniprep Kit (Stratagene). A Thermoscript RT-PCR kit (Invitrogen) was used for the cDNA synthesis reaction with 1 μg template RNA and 250 pmol of oligo$(dT)_{20}$ primer, incubated for 1 h at 55 °C. Subsequently, cDNA was used as a template for RT-qPCR with the ABI Prism 7000 Sequence Detection System and Platinum SYBR Green qPCR SuperMix, using gene-specific primers. For Dnmt3 enzymes, splice-variant–specific primers were used. Sequences of primers are provided in **Supplementary Table 4**. Annealing of all primers was done at 55 °C. Relative expression levels were calculated using the formula $2^{-(\Delta Ct)}$, where ΔCt is $Ct_{(gene\ of\ interest)} - Ct_{(GAPDH)}$ and Ct is the cycle at which the threshold is crossed. For time course experiments, the expression level at day 0 in *Dicer*$^{+/-}$ ES cells or in siRL-transfected *Dicer*$^{-/-}$ cells was always set as 1 and expression levels at other time points were normalized to it.

Western blotting. Cells were lysed in lysis buffer (30 mM Tris-HCl, pH 7.5, containing 150 mM NaCl, 1 mM $MgCl_2$, 0.5% (v/v) Nonidet P-40, 1 mM DTT and protease inhibitors) and kept on ice for 10 min. Equal amounts of the lysed proteins were separated on polyacrylamide-SDS gels, blotted on polyvinylidene fluoride membrane and probed with the following primary antibodies: anti-Oct4 (Santa Cruz, dilution 1:2,000), anti–α-tubulin (5.2.1 Sigma, 1:10,000), anti-Dicer [D349 (ref. 53), 1:5,000], anti-Dnmt1 (Abcam, 1:500), anti-Dnmt3a (Imgenex, 1:250), anti-Dnmt3b (Imgenex, 1:250) and anti–RNA-polymerase II (Covance, 1:500). This was followed by incubation with secondary horseradish peroxidase–coupled antibodies. Detection was performed with ECL or ECL+ kits (Amersham).

Luciferase assays. Luciferase assays were performed using the Dual-Luciferase Reporter Assay kit (Promega) according to the manufacturer's instructions. FL activity was normalized to RL activity expressed from pRL-TK. Normalized FL activity in cells transfected with pGL3-FF was always set as one.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank G. Hannon and E. Murchison (Cold Spring Harbor Laboratory, New York, USA), for providing *Dicer*$^{-/-}$ ES cells, K. Ura (Osaka University, Japan), for providing DNMT3 plasmids, A. Peters for providing antibodies, D. Schubeler for helpful suggestions and comments (both Friedrich Miescher Institute, Basel, Switzerland). We also thank E. Oakeley, H. Angliker and M. Pietrzak for their contributions to array analysis and sequencing (Friedrich Miescher Institute).

P.S. is supported by the European Molecular Biology Organization (EMBO) SDIG program #1488, GAAV IAA501110701 and the Purkynje Fellowship. P.B. is supported by the Swiss National Science Foundation (SNF) grant #3100A0-114001 to M.Z., and D.G. is supported by the Swiss Institute of Bioinformatics. L.S. is partially supported by the EC FP6 STREP program LSHG-CT-2004. The Friedrich Miescher Institute is supported by the Novartis Research Foundation.

AUTHOR CONTRIBUTIONS

L.S., P.S. and W.F. designed the study; L.S., P.S. and M.Z. designed the computational part; L.S. carried out most of the experiments; T.H. contributed to some experiments with *Dicer*$^{-/-}$ ES cells and most of the western analyses; C.G.A.-R. contributed to *Rbl2* knockdown and western analyses; D.G. and P.B. performed computational analyses; P.S. carried out some of the bisulfite sequencing and initial analysis of microarray data; F.M. helped with Sox30 and Tsp50 methylation analysis; L.S., P.S., M.Z. and W.F. wrote the manuscript.

Published online at http://www.nature.com/nsmb
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

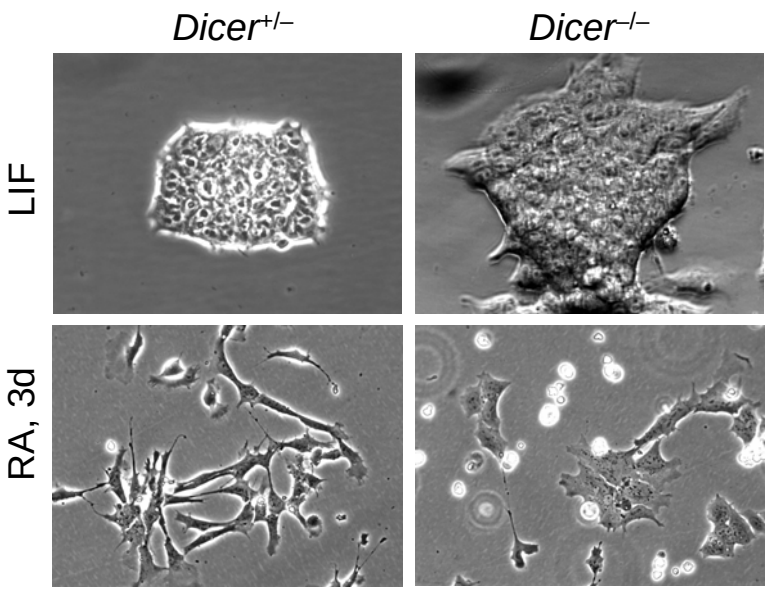
1. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
2. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 522–531 (2004).
3. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
4. Zamore, P.D. & Haley, B. Ribo-gnome: the big world of small RNAs. *Science* **309**, 1519–1524 (2005).
5. Kim, V.N. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
6. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
7. Bushati, N. & Cohen, S.M. MicroRNA functions. *Annu. Rev. Cell Dev. Biol.* **23**, 175–205 (2007).
8. Pillai, R.S., Bhattacharyya, S.N. & Filipowicz, W. Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.* **17**, 118–126 (2007).
9. Valencia-Sanchez, M.A., Liu, J., Hannon, G.J. & Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**, 515–524 (2006).
10. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
11. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
12. Rajewsky, N. MicroRNA target predictions in animals. *Nat. Genet.* **38** Suppl, S8–S13 (2006).
13. Smith, A.G. Embryo-derived stem cells: of mice and men. *Annu. Rev. Cell Dev. Biol.* **17**, 435–462 (2001).
14. Boyer, L.A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
15. Loh, Y.H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440 (2006).
16. Suh, M.R. *et al.* Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.* **270**, 488–498 (2004).
17. Houbaviy, H.B., Murray, M.F. & Sharp, P.A. Embryonic stem cell-specific MicroRNAs. *Dev. Cell* **5**, 351–358 (2003).
18. Tang, F. *et al.* Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* **21**, 644–648 (2007).
19. Kanellopoulou, C. *et al.* Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* **19**, 489–501 (2005).
20. Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S. & Hannon, G.J. Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **102**, 12135–12140 (2005).
21. Wang, Y., Medvid, R., Melton, C., Jaenisch, R. & Blelloch, R. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat. Genet.* **39**, 380–385 (2007).
22. Krutzfeldt, J. *et al.* Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**, 685–689 (2005).
23. Lim, L.P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
24. Schmitter, D. *et al.* Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res.* **34**, 4801–4815 (2006).
25. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144 (2006).
26. Zeng, F., Baldwin, D.A. & Schultz, R.M. Transcript profiling during preimplantation mouse development. *Dev. Biol.* **272**, 483–496 (2004).
27. Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
28. Chen, T., Ueda, Y., Xie, S. & Li, E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active *de novo* methylation. *J. Biol. Chem.* **277**, 38746–38754 (2002).
29. Aapola, U., Maenpaa, K., Kaipia, A. & Peterson, P. Epigenetic modifications affect Dnmt3L expression. *Biochem. J.* **380**, 705–713 (2004).
30. Jinawath, A., Miyake, S., Yanagisawa, Y., Akiyama, Y. & Yuasa, Y. Transcriptional regulation of the human DNA methyltransferase 3A and 3B genes by Sp3 and Sp1 zinc finger proteins. *Biochem. J.* **385**, 557–564 (2005).
31. Su, A.I. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA* **99**, 4465–4470 (2002).
32. Harris, M.A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D411–D417 (2004).
33. Litovchick, L. *et al.* Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol. Cell* **26**, 539–551 (2007).
34. Gu, P., Le Menuet, D., Chung, A.C. & Cooney, A.J. Differential recruitment of methylated CpG binding domains by the orphan receptor GCNF initiates the repression and silencing of Oct4 expression. *Mol. Cell. Biol.* **26**, 9471–9483 (2006).
35. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol.* **8**, 188–194 (2006).
36. Li, J.Y. *et al.* Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. *Mol. Cell Biol.* **27**, 8748–8759 (2007).
37. Gartel, A.L. & Radhakrishnan, S.K. Lost in transcription: p21 repression, mechanisms, and consequences. *Cancer Res.* **65**, 3980–3985 (2005).
38. Murakami, M. *et al.* mTOR is essential for growth and proliferation in early mouse embryos and embryonic stem cells. *Mol. Cell. Biol.* **24**, 6710–6718 (2004).
39. Giraldez, A.J. *et al.* Zebrafish miR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
40. Schultz, R.M. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Hum. Reprod. Update* **8**, 323–331 (2002).
41. Choi, W.Y., Giraldez, A.J. & Schier, A.F. Target protectors reveal dampening and balancing of nodal agonist and antagonist by miR-430. *Science* **318**, 271–274 (2007).
42. Dotto, G.P. p21(WAF1/Cip1): more than a break to the cell cycle? *Biochim. Biophys. Acta* **1471**, M43–M56 (2000).
43. He, L. *et al.* A microRNA polycistron as a potential human oncogene. *Nature* **435**, 828–833 (2005).
44. Zvetkova, I. *et al.* Global hypomethylation of the genome in XX embryonic stem cells. *Nat. Genet.* **37**, 1274–1279 (2005).
45. Robertson, K.D., Keyomarsi, K., Gonzales, F.A., Velicescu, M. & Jones, P.A. Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells. *Nucleic Acids Res.* **28**, 2108–2113 (2000).
46. Fabbri, M. *et al.* MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc. Natl. Acad. Sci. USA* **110**, 15805–15810 (2007).
47. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
48. Ishida, C. *et al.* Genomic organization and promoter analysis of the Dnmt3b gene. *Gene* **310**, 151–159 (2003).
49. Nimura, K. *et al.* Dnmt3a2 targets endogenous Dnmt3L to ES cell chromatin and induces regional DNA methylation. *Genes Cells* **11**, 1225–1237 (2006).
50. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
51. Sinkkonen, L., Malinen, M., Saavalainen, K., Vaisanen, S. & Carlberg, C. Regulation of the human cyclin C gene via multiple vitamin D3-responsive regions in its promoter. *Nucleic Acids Res.* **33**, 2440–2451 (2005).
52. Svoboda, P., Stein, P., Filipowicz, W. & Schultz, R.M. Lack of homologous sequence-specific DNA methylation in response to stable dsRNA expression in mouse oocytes. *Nucleic Acids Res.* **32**, 3601–3606 (2004).
53. Kotaja, N. *et al.* The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc. Natl. Acad. Sci. USA* **103**, 2647–2652 (2006).
54. Nordhoff, V. *et al.* Comparative analysis of human, bovine, and murine Oct-4 upstream promoter sequences. *Mamm. Genome* **12**, 309–317 (2001).

MicroRNAs control *de novo* DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells

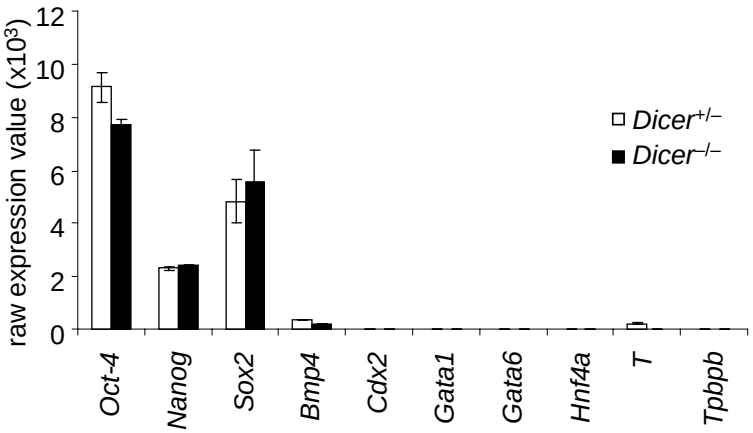
Lasse Sinkkonen, Tabea Hugenschmidt, Philipp Berninger, Dimos Gaidatzis, Fabio Mohn, Caroline G. Artus-Revel, Mihaela Zavolan, Petr Svoboda, and Witold Filipowicz

Sinkkonen et al. Supplementary Figure 1

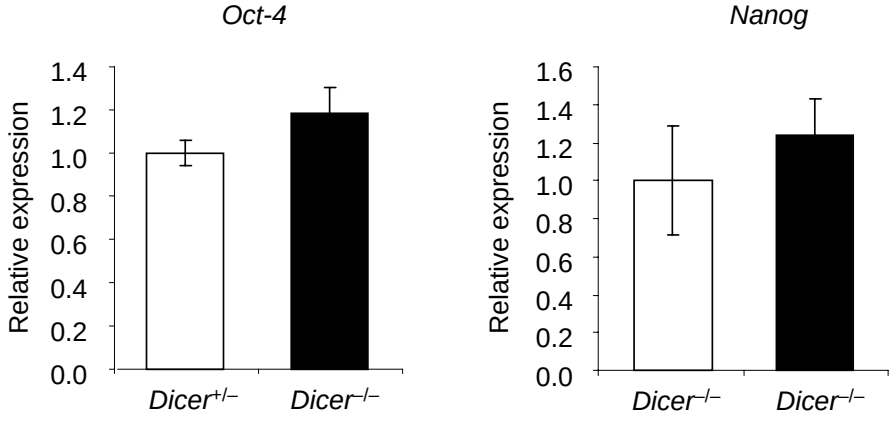
a



b



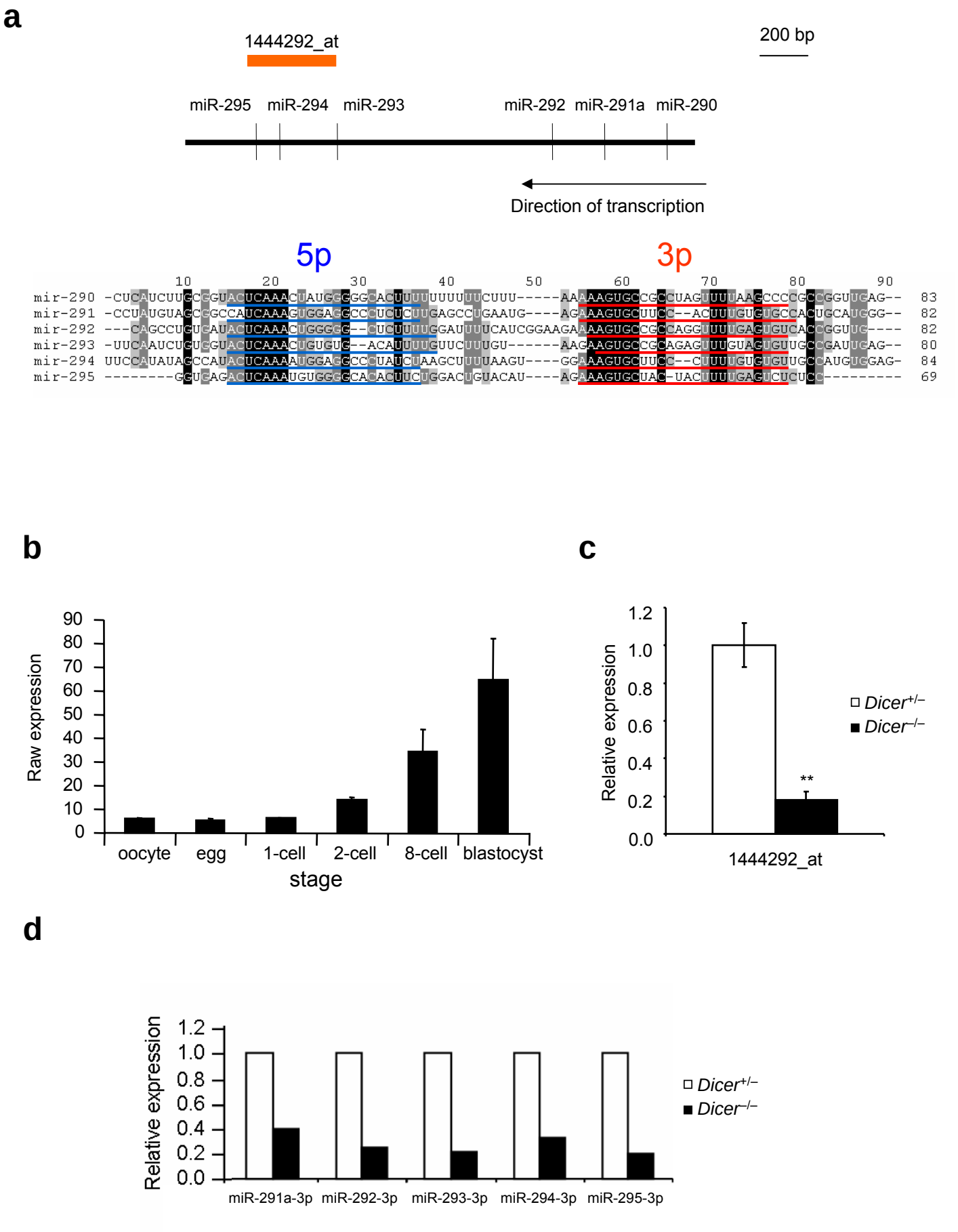
c



Supplementary Figure 1. Characterization of the differentiation status of *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ ES cells.

(a) Light microscopy images of undifferentiated cells (cells grown in the presence of LIF) and cells subjected to the differentiation treatment (grown for 3 days in the presence of 100 nM RA, in the absence of LIF; RA, 3 d). (b) Levels of mRNAs encoding core pluripotency and differentiation markers. Microarray analysis was performed with RNA isolated from undifferentiated *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ ES cells. Raw data were calculated as described in Materials and Methods. Median raw values (± s.d.) for each gene were taken from probe sets with the strongest hybridization signal. Other probe sets for the same genes also did not show significant differences in expression levels between *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells. Displayed differentiation markers were either used in a previous analysis of *Dicer*$^{-/-}$ ES cells [*T* (brachyury), *Hnf4a*, *Gata1*, *Bmp4*][7] or were culled from published articles (*Tpbpb*, *Cdx2*, *Gata6*)[8,9]. These markers are indicative of the presence of cells of trophectodermal (*Tpbpb*, *Cdx2*), extraembryonic endodermal (*Gata6*), embryonic mesodermal (brachyury, *Bmp4*, *Gata1*, *Gata6*), and embryonic endodermal (*Hnf4a*, *Gata6*) lineages. It is not known why *Dicer*$^{+/-}$ cells show a low microarray hybridization signal of brachyury. Possibly, a small fraction of cells spontaneously initiates differentiation. However, other mesodermal markers such as *Gata1* and *Gata6* remained absent. Detectable microarray hybridization signal for *Bmp4* has been previously reported for undifferentiated ES cells (GEO database, and[10]). (c) RT-qPCR analysis of *Oct-4* and *Nanog* mRNA levels confirms results of microarray analysis. Values, normalized to *Gapdh* expression, represent means (± s.e.m.) of at least 3 independent experiments.

Sinkkonen et al. Supplementary Figure 2

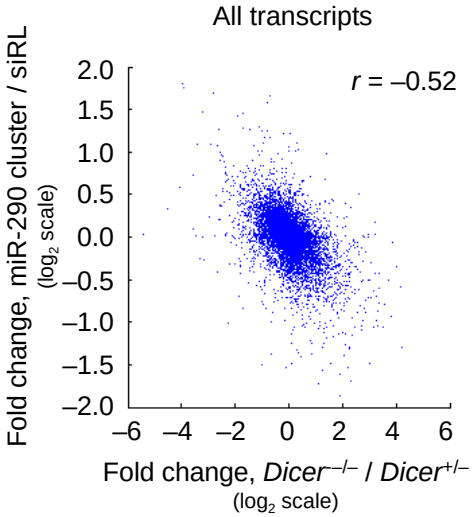


Supplementary Figure 2. Characterization of expression of the miR-290 cluster.

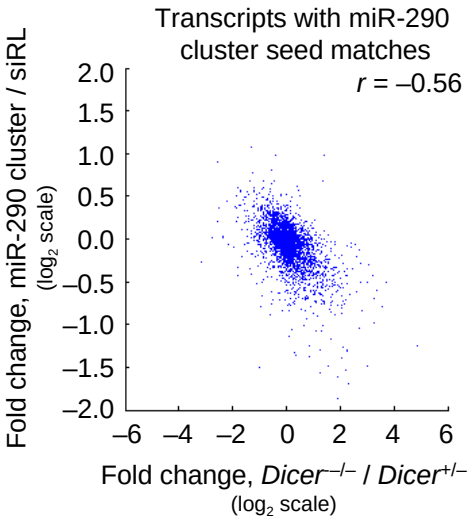
(a) Structure of the miR-290 locus and sequences of the miR-290 cluster miRNAs. The upper scheme depicts structure of the cluster. Shown is a 4 kb fragment from the chromosome 7 genomic contig NT_039413.7 (from 217,800 to 221,800) with relative positions of individual miRNAs and orientation of the pri-miRNA transcript. The orange line represents a region covered by oligonucleotide probes from the Affymetrix probe set 1444292_at, which detects pri-miRNA of the miR-290 cluster. Below the map is a ClustalW alignment of pre-miRNAs, which was downloaded from the miRbase[1] and edited using Genedoc alignment editor (http://www.psc.edu/biomed/genedoc). Blue lines represent mature miRNA sequences from the ascending (5p) strand and red lines from the descending (3p) strand of the pre-miRNA hairpin. Pre-miRNAs of the miR-290 cluster yield mature miRNAs primarily from descending (3p) hairpin strands. According to Landgraf et al.[6], the 5p strands also produce mature miRNAs but they represent only 13% of miRNAs generated from the cluster. Note that despite being expressed, not all mature miRNA sequences depicted here can be found in the current version of the miBase (9.2). (b) Quantification of the miR-290 cluster primary transcript, followed using the microarray probe 1444292_at, indicates that the cluster is expressed zygotically. The region of the primary transcript covered by the 1444292_at probe set is indicated in panel a. The analysis was performed using previously published microarray data[11]. Raw data were calculated as described previously[12]. Median raw values (± s.d.; n = 4) are shown. (c) The miR-290 cluster primary transcript is strongly down-regulated in *Dicer*$^{-/-}$ ES cells (two-tailed t-test: p = 0.0053). Expression of the primary trasncript was followed using the microarray probe 1444292_at as described in panel a. The mean expression value (± s.d.; n = 3) in *Dicer*$^{+/-}$ cells was set to one. (d) Levels of the mir-290 cluster miRNAs are down-regulated in *Dicer*$^{-/-}$ cells. Total RNA extracted from either *Dicer*$^{+/-}$ or *Dicer*$^{-/-}$ cells was analyzed on miRCURY^{TM} LNA arrays, following the manufacturer's protocol (www.exiqon.com). Values calculated for *Dicer*$^{+/-}$ cells were set as one. The figure shows expression levels of only 3p miRNAs of miR-290 cluster, which were reliably identified as present in the analyzed RNA samples. The 3p miRNAs represent 87% of all miRNAs generated from the cluster[6]. miR-290-3p is not shown since the Exiqon array does not contain a probe for its detection.

Sinkkonen et al. Supplementary Figure 3

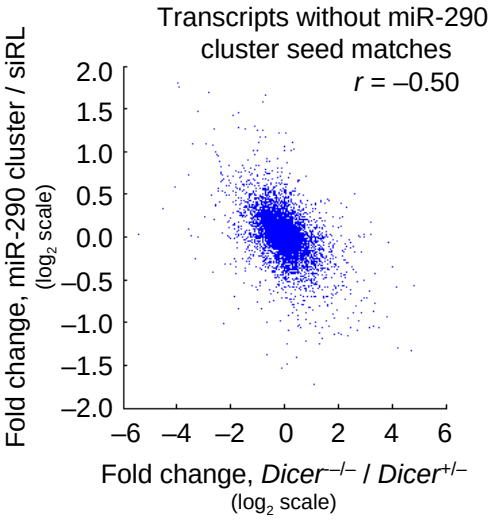
a



b

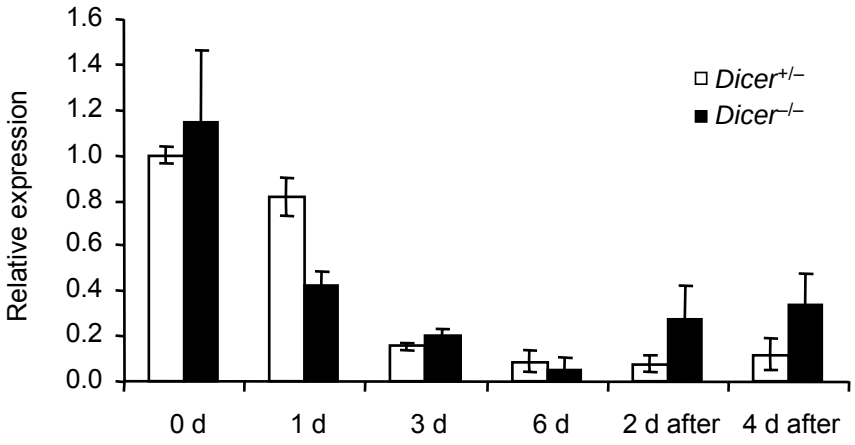


c



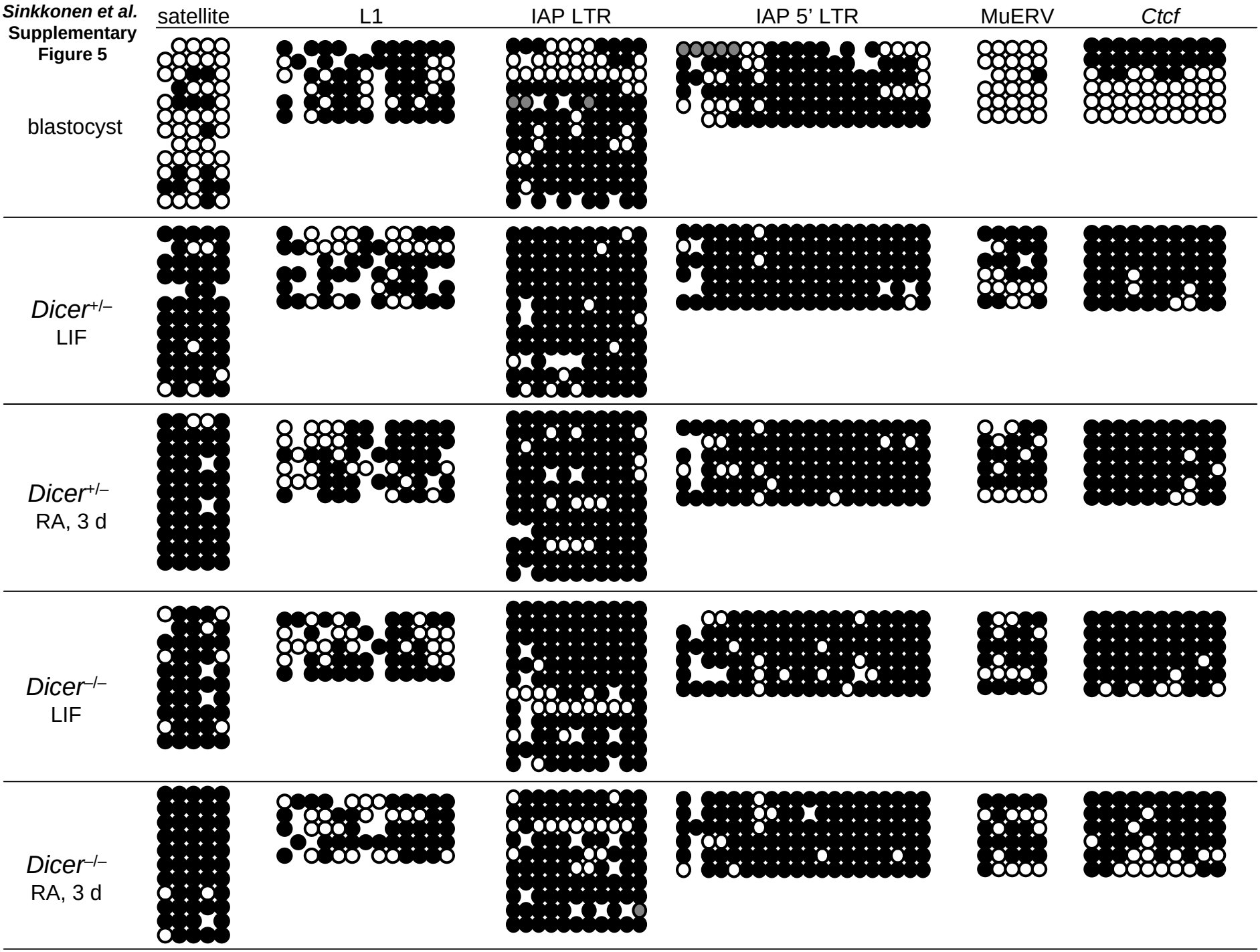
Supplementary Figure 3. Relationship between the expression changes in the *Dicer*$^{-/-}$ vs. *Dicer*$^{+/-}$ (x-axis) and miR-290s-transfected vs siRL-transfected *Dicer*$^{-/-}$ ES cells (y-axis). Each dot corresponds to a single transcript, and the panels represent: (a) all transcripts; (b) transcripts with at least one 7-mer match to one of the 1-8 positions of the miRNAs in the 290 cluster in their 3′-UTRs; and (c) transcripts with no 7-mer match to any of the miRNAs of the 290 cluster in their 3′-UTRs. The correlation coefficients are indicated in each panel.

Sinkkonen et al. Supplementary Figure 4



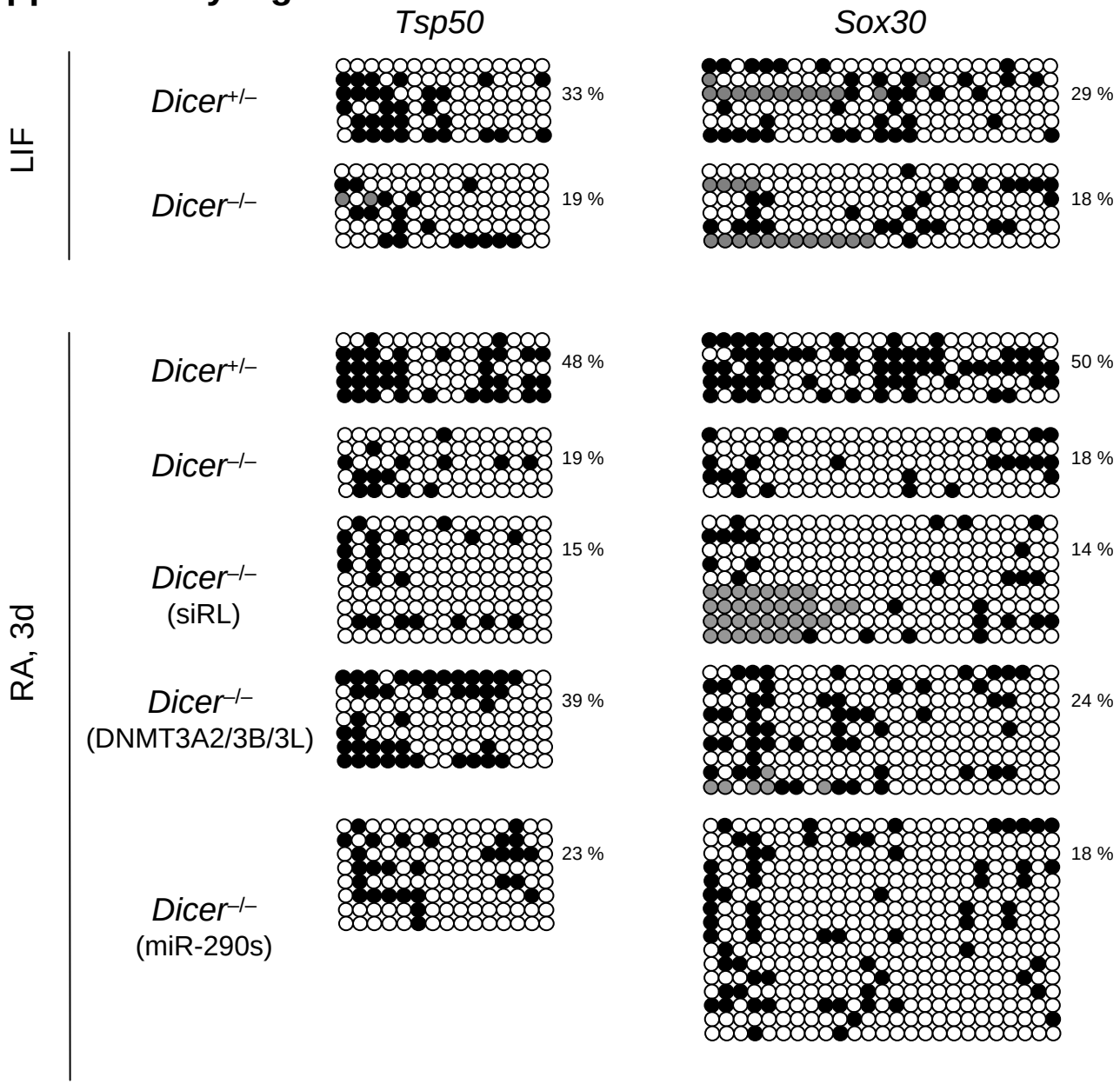
Supplementary Figure 4. RT-qPCR analysis of *Nanog* expression during the RA-induced differentiation for 0, 1, 3, or 6 days (0 d, 1 d, 3 d, 6 d), and after returning the cells to the LIF-containing medium devoid of RA for up to 4 additional days (2 d after, 4 d after). Values, normalized to GAPDH expression, represent means (± s.e.m.) of at least 3 independent experiments. Expression in control *Dicer*$^{+/-}$ cells at 0d time point was set as 1.

Sinkkonen et al.
Supplementary Figure 5



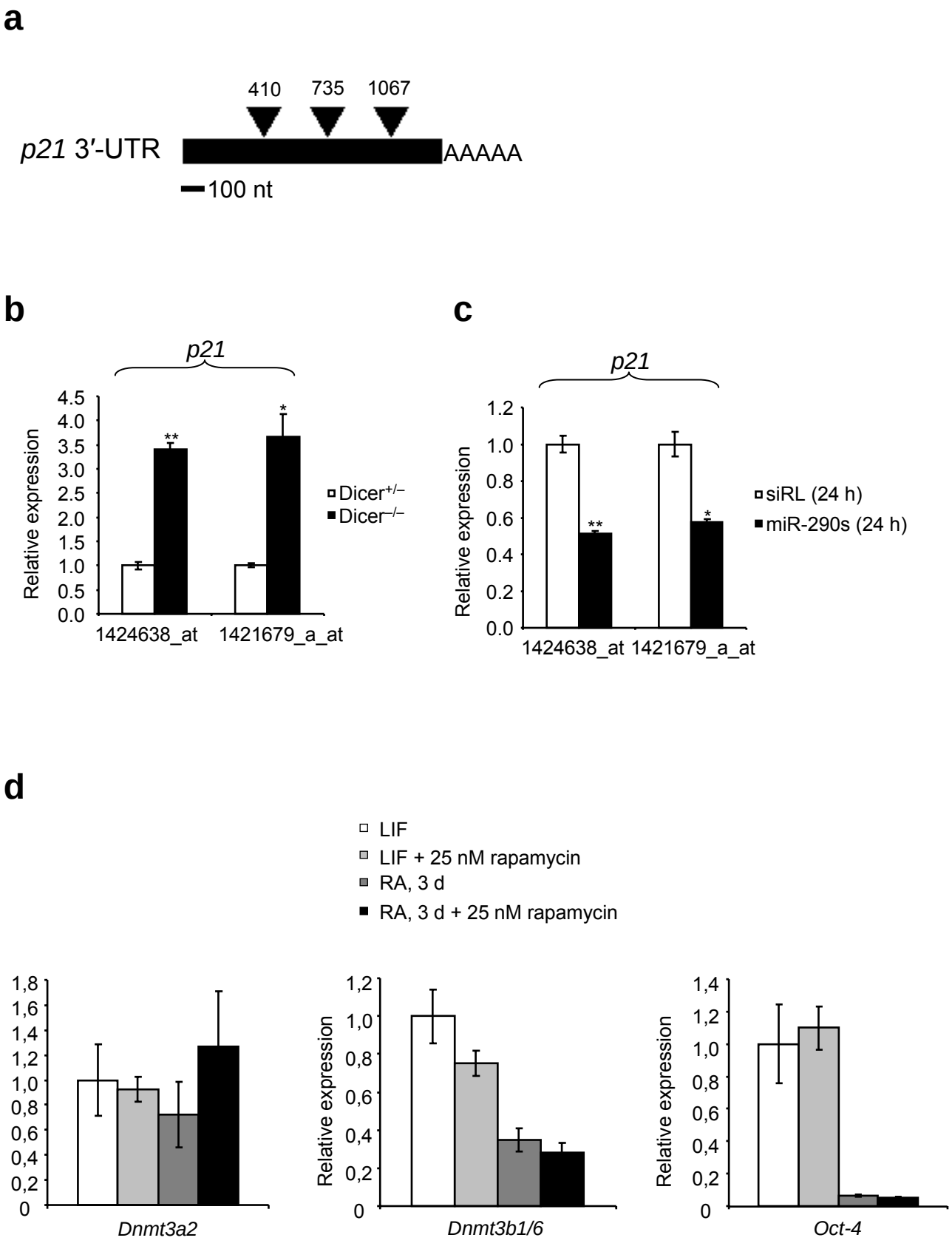
Supplementary Figure 5. Maintenance DNA methylation is not affected in *Dicer*$^{-/-}$ ES cells. Bisulfite analysis shows the same methylation pattern in undifferentiated and differentiated *Dicer*$^{-/-}$ and *Dicer*$^{+/-}$ ES cells for typical targets of DNA methylation: repetitive DNA and a body of a gene[13]. DNA methylation of tandemly arrayed repeat sequences (centromeric satellite), interspersed LTR retrotransposon sequences (IAP and MuERV-L), interspersed non-LTR transposon sequences (L1), and a single copy gene sequence (CTCF, exon 9) were analyzed. In the case of IAP, also its 5′ LTR sequence, which serves as a promoter for the retrotransposon, was analyzed. The exon 9 sequence of the CTCF gene is hypermethylated in numerous tissues[14]. The ES cell DNA samples used to obtain the data were the same as those described in Fig. 6. Blastocyst DNA samples served as a control providing the DNA methylation status prior to establishment of ES cells. The blastocyst DNA was obtained from a pool of 30 blastocysts from uteri of C57BL/6 female mice. An equivalent of three blastocysts was used for each PCR reaction. Black dots represent methylated and white dots non-methylated CpGs. Sites for which the methylation status was not certain are in grey. Average percentages of the methylated CpG sites for ES cell samples range from 89 to 97 % for satellite repeat, 62 to 70 % for L1, 84 to 92 % for IAP LTR, 90 to 97 % for IAP 5'LTR, 62 to 67 % for MuERV, and 75 to 92 % for *Ctcf*.

Sinkkonen et al. Supplementary Figure 6



Supplementary Figure 6. Analysis of *de novo* DNA methylation of promoters of testis-specific genes *Tsp50* and *Sox30*, which are transcriptionally silent in undifferentiated ES cells and acquire DNA methylation during differentiation in the presence of RA. The promoter regions, upstream of the *Tsp50* and *Sox30* transcription start sites, were analyzed in *Dicer*$^{+/-}$ ES cells, *Dicer*$^{-/-}$ ES cells, *Dicer*$^{-/-}$ ES cells co-transfected with plasmids expressing EGFP-DNMT3a2, EGFP-DNMT3b, and EGFP-DNMT3L, *Dicer*$^{-/-}$ ES cells transfected with miR-290 cluster miRNA mimics, and *Dicer*$^{-/-}$ ES cells transfected with siRL. Cells were cultured in presence of LIF and without RA (LIF) or differentiated for 3 days with RA in the absence of LIF (RA, 3 d). Each row of dots represents CpGs in one sequenced clone. Black dots represent methylated and white dots non-methylated CpGs. Sites for which the methylation status was not certain are in grey. Average percentages of the methylated CpG sites are indicated.

Sinkkonen et al. Supplementary Figure 7



Supplementary Figure 7. (a,b,c) *p21* is up-regulated in *Dicer*$^{-/-}$ ES cells and down-regulated in response to transfection of miR-290s. (a) Schematic representation of localization of predicted binding sites for AAGUGC-seed containing miRNAs in 3′-UTR of *p21* mRNA are marked with black triangles. (b) The level of *p21* mRNA is up-regulated in *Dicer*$^{-/-}$ cells. RNA isolated from *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells was analyzed by Affymetrix microarrays. The probe sets detecting expression of *p21* mRNA are indicated. Mean expression value (± s.d.; n = 3) in *Dicer*$^{+/-}$ cells was set to 1. The two-tailed t-test p-values were 0.0021 and 0.0120. (c) Transfection of miR-290s to *Dicer*$^{-/-}$ ES cells down-regulates the level of *p21* mRNA. Cells were transfected with either a mixture of the miR-290 cluster miRNAs or siRL used as a control. RNA was isolated 24 h after transfection. Mean expression value (± s.d; n = 3) in siRL transfected cells was set to 1. The two-tailed t-test p-values were 0.0012 and 0.0070. (d) Expression of *Dnmt3a2* and *Dnmt3b* is independent of the growth rate of the *Dicer*$^{+/-}$ ES cells. *Dicer*$^{+/-}$ ES cells were cultured with or without 25 nM rapamycin and differentiated for three days with RA in the absence of LIF (RA, 3 d). mRNA levels of *Dnmt3a2*, *Dnmt3b* and *Oct-4* were analyzed by RT-qPCR. Expression (± s.d.; n = 3) was normalized to that of *Gapdh* and is shown relative to corresponding samples cultured in the presence of LIF without rapamycin, whose expression values were set to 1.

Sinkkonen et al. **SUPPLEMENTARY TABLES**

Supplementary Table 1. Motifs most significantly enriched in 3′-UTRs of mRNAs up-regulated in *Dicer*$^{-/-}$ ES cells or down-regulated in *Dicer*$^{-/-}$ ES cells transfected with miRNAs of the miR-290 cluster.

Enrichments of the 7-mers were analyzed as described in Materials and Methods. Shown are the motifs most significantly (posterior probability > 0.99) enriched in 3′-UTRs of mRNAs up-regulated in *Dicer*$^{-/-}$ ES cells or down-regulated in *Dicer*$^{-/-}$ ES cells transfected with miRNAs of the miR-290 family. For each motif, the sequence of the motif, posterior probability of the enrichment, occurrence of the motif among up- or down-regulated 3′-UTRs, and the enrichment of the motif are shown. For motifs that are complementary to sequences within miRNAs, the names and sequences of the matching miRNAs are also shown. Sequences complementary to the enriched motif are in capitals. A number of top scoring 7-mer motifs enriched in 3′-UTRs of mRNAs down-regulated upon transfection of miR-290 cluster miRNAs contain seven or six U residues. The significance of these motifs and their enrichment is unknown.

Supplementary Table 1 - Significantly enriched motifs (posterior probability > 0.99)

Motifs enriched in transcripts upregulated in Dicer -/- cells						
Motif	**Posterior probability**	**Occurrence**	**Enrichment**	**Position**	**miRNA**	**Sequence of miRNA**
GCACUUU	0.9999999999	489	1.615			
				(2-8)	mmu-miR-17-5p	cAAAGUGCuuacagugcagguagu
				(1-7)	mmu-miR-291a-3p	AAAGUGCuuccacuuugugugcc
				(2-8)	mmu-miR-93	cAAAGUGCuguucgugcagguag
				(2-8)	mmu-miR-20b	cAAAGUGCucauagugcaggua
				(1-7)	mmu-miR-291b-3p	AAAGUGCauccauuuuguuuguc
				(2-8)	mmu-miR-106b	uAAAGUGCugacagugcagau
				(2-8)	mmu-miR-106a	cAAAGUGCuaacagugcaggua
				(2-8)	mmu-miR-20a	uAAAGUGCuuauagugcagguag
				(1-7)	mmu-miR-294-3p	AAAGUGCuucccuuuugugugu
				(1-7)	mmu-miR-295-3p	AAAGUGCuacuacuuuugagucu
AGCACUU	0.9999999994	545	1.495			
				(3-9)	mmu-miR-17-5p	caAAGUGCUuacagugcagguagu
				(1-7)	mmu-miR-302b	AAGUGCUuccauguuuuaguag
				(2-8)	mmu-miR-291a-3p	aAAGUGCUuccacuuugugugcc
				(3-9)	mmu-miR-93	caAAGUGCUguucgugcagguag
				(3-9)	mmu-miR-20b	caAAGUGCUcauagugcaggua
				(1-7)	mmu-miR-302d	AAGUGCUuccauguuugagugu
				(2-8)	mmu-miR-302c	cAAGUGCUuccauguuucagugg
				(3-9)	mmu-miR-106b	uaAAGUGCUgacagugcagau
				(3-9)	mmu-miR-106a	caAAGUGCUaacagugcaggua
				(3-9)	mmu-miR-20a	uaAAGUGCUuauagugcagguag
				(2-8)	mmu-miR-294-3p	aAAGUGCUucccuuuugugugu
				(2-8)	mmu-miR-295-3p	aAAGUGCUacuacuuuugagucu
				(2-8)	mmu-miR-302	uAAGUGCUuccauguuuugguga
GCACUUA	0.9999281990	255	1.614			
				(2-8)	mmu-miR-467*	gUAAGUGCcugcauguauaug
				(1-7)	mmu-miR-302	UAAGUGCuuccauguuuugguga
UGCACUU	0.9988487421	415	1.402			
				(2-8)	mmu-miR-291b-3p	aAAGUGCAuccauuuuguuuguc
AAGCACU	0.9967070829	406	1.392			

				(4-10)	mmu-miR-17-5p	caaAGUGCUUacagugcagguagu
				(2-8)	mmu-miR-302b	aAGUGCUUccauguuuuaguag
				(3-9)	mmu-miR-291a-3p	aaAGUGCUUccacuuugugugcc
				(11-17)	mmu-miR-471	uacguaguauAGUGCUUuucaca
				(2-8)	mmu-miR-302d	aAGUGCUUccauguuugagugu
				(3-9)	mmu-miR-302c	caAGUGCUUccauguuucagugg
				(4-10)	mmu-miR-20a	uaaAGUGCUUauagugcagguag
				(3-9)	mmu-miR-294-3p	aaAGUGCUUcccuuuugugugu
				(3-9)	mmu-miR-302	uaAGUGCUUccauguuuugguga

Motifs enriched in transcripts downregulated in Dicer -/- cells transfected with miR-290 cluster

Motif	Posterior probability	Occurrence	Enrichment	Position	miRNA	Sequence of miRNA
UUUUUUU	0.9999999542	2299	1.175			
GCACUUA	0.9999976429	126	2.047			
				(2-8)	mmu-miR-467*	gUAAGUGCcugcauguauaug
				(1-7)	mmu-miR-302	UAAGUGCuuccauguuuugguga
UUUGUUU	0.9999952321	931	1.265			
				(1-7)	mmu-miR-495	AAACAAAcauggugcacuucuu
UUUUGUU	0.999912871	776	1.272			
AGCACUU	0.999696131	224	1.576			
				(3-9)	mmu-miR-17-5p	caAAGUGCUuacagugcagguagu
				(1-7)	mmu-miR-302b	AAGUGCUuccauguuuuaguag
				(2-8)	mmu-miR-291a-3p	aAAGUGCUuccacuuugugugcc
				(3-9)	mmu-miR-93	caAAGUGCUguucgugcagguag
				(3-9)	mmu-miR-20b	caAAGUGCUcauagugcaggua
				(1-7)	mmu-miR-302d	AAGUGCUuccauguuugagugu
				(2-8)	mmu-miR-302c	cAAGUGCUuccauguuucagugg
				(3-9)	mmu-miR-106b	uaAAGUGCUgacagugcagau
				(3-9)	mmu-miR-106a	caAAGUGCUaacagugcaggua
				(3-9)	mmu-miR-20a	uaAAGUGCUuauagugcagguag
				(2-8)	mmu-miR-294-3p	aAAGUGCUucccuuuugugugu
				(2-8)	mmu-miR-295-3p	aAAGUGCUacuacuuuugagucu
				(2-8)	mmu-miR-302	uAAGUGCUuccauguuuugguga
GCACUUU	0.9996301511	193	1.635			
				(2-8)	mmu-miR-17-5p	cAAAGUGCuuacagugcagguagu

				(1-7)	mmu-miR-291a-3p	AAAGUGCuuccacuuugugugcc
				(2-8)	mmu-miR-93	cAAAGUGCuguucgugcagguag
				(2-8)	mmu-miR-20b	cAAAGUGCucauagugcaggua
				(1-7)	mmu-miR-291b-3p	AAAGUGCauccauuuuguuuguc
				(2-8)	mmu-miR-106b	uAAAGUGCugacagugcagau
				(2-8)	mmu-miR-106a	cAAAGUGCuaacagugcaggua
				(2-8)	mmu-miR-20a	uAAAGUGCuuauagugcagguag
				(1-7)	mmu-miR-294-3p	AAAGUGCuucccuuuugugugu
				(1-7)	mmu-miR-295-3p	AAAGUGCuacuacuuuugagucu
UUGUUUU	0.9989237718	802	1.245			
UAUUUUU	0.9983591123	691	1.264			

Supplementary Table 2. Profiling of miRNA levels in *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ ES cells using miRCURY^{TM} Exqon microarrays.

Microarray analysis was performed as described in Supplementary Materials and Methods. 190 miRNAs gave significant hybridization signals with both analyzed *Dicer*$^{+/-}$ RNA samples and thus were considered for further analysis. The miRNA expression levels in *Dicer*$^{-/-}$ ES cells were compared to the expression levels in *Dicer*$^{+/-}$ ES cells. The numbers of miRNAs down- or up-regulated, or not changing, with cut-offs of 1.2-fold, 1.5-fold and 2.0-fold are shown in the table. The miRNAs that were down-regulated more than 1.5-fold or 2.0-fold were considered as expressed in ES cells in a Dicer-dependent manner. These miRNAs are listed and their expression levels normalized to the expression levels in a control reference sample (a mixture of total RNA from 11 different tissues) are indicated to identify miRNAs expressed preferentially in ES cells.

Supplementary Table 2 - miRNA expression profiling in Dicer +/- and Dicer -/- ES cells

Changes in miRNA levels in Dicer -/- cells	Number of miRNAs	% of all detected miRNAs (n=190)
CUT-OFF 1.2-FOLD		
downregulated 1.2-fold	115	60,3
upregulated 1.2-fold	23	12,2
not changed above 1.2-fold	52	27,5
CUT-OFF 1.5-FOLD		
downregulated 1.5-fold	69	36,0
upregulated 1.5-fold	8	4,2
not changed above 1.5-fold	113	59,8
CUT-OFF 2.0-FOLD		
downregulated 2.0-fold	29	14,8
upregulated 2.0-fold	3	1,6
not changed above 2.0-fold	158	83,6

miRNAs DOWNREGULATED >2.0-fold		
Expressed > 1.5-fold higher relative to control reference mixture	**Expressed between 1.5- to 0.5-fold relative to control reference mixture**	**Expressed < 0.5-fold relative to control reference mixture**
mmu-let-7d*	mmu-miR-136	mmu-miR-101a
mmu-miR-291a-3p	mmu-miR-18	mmu-miR-103
mmu-miR-292-3p	mmu-miR-337	mmu-miR-106a
mmu-miR-293 (miR-293-3p)	mmu-miR-376a	mmu-miR-130a
mmu-mir294 (miR-294-3p)	mmu-miR-467*	mmu-miR-140*
mmu-miR-295 (miR-295-3p)	mmu-miR-541	mmu-miR-146b
		mmu-miR-16
		mmu-miR-17-5p
		mmu-miR-191
		mmu-miR-193
		mmu-miR-200a
		mmu-miR-20b
		mmu-miR-21
		mmu-miR-22
		mmu-miR-23a

		mmu-miR-335
		mmu-miR-98

miRNAs DOWNREGULATED >1.5-fold		
Expressed > 1.5-fold higher relative to control reference mixture	**Expressed between 1.5- to 0.5-fold relative to control reference mixture**	**Expressed < 0.5-fold relative to control reference mixture**
mmu-let-7d*	mmu-miR-136	mmu-let-7i
mmu-miR-291a-3p	mmu-miR-18	mmu-miR-101a
mmu-miR-292-3p	mmu-miR-337	mmu-miR-101b
mmu-miR-293 (miR-293-3p)	mmu-miR-341	mmu-miR-103
mmu-mir294 (miR-294-3p)	mmu-miR-376a	mmu-miR-106a
mmu-miR-295 (miR-295-3p)	mmu-miR-379	mmu-miR-107
mmu-miR-697	mmu-miR-467*	mmu-miR-10b
	mmu-miR-541	mmu-miR-122a
		mmu-miR-127
		mmu-miR-128a
		mmu-miR-130a
		mmu-miR-133a*
		mmu-miR-140
		mmu-miR-140*
		mmu-miR-142-5p
		mmu-miR-143
		mmu-miR-146
		mmu-miR-146b
		mmu-miR-148a
		mmu-miR-148b
		mmu-miR-15a
		mmu-miR-15b
		mmu-miR-16
		mmu-miR-17-5p
		mmu-miR-191
		mmu-miR-193
		mmu-miR-195
		mmu-miR-19b
		mmu-miR-200a

		mmu-miR-200b
		mmu-miR-20b
		mmu-miR-21
		mmu-miR-22
		mmu-miR-222
		mmu-miR-223
		mmu-miR-23a
		mmu-miR-23b
		mmu-miR-24
		mmu-miR-26a
		mmu-miR-27a
		mmu-miR-27b
		mmu-miR-29a
		mmu-miR-29c
		mmu-miR-30a-5p
		mmu-miR-30b
		mmu-miR-30c
		mmu-miR-30e
		mmu-miR-335
		mmu-miR-338
		mmu-miR-376b
		mmu-miR-449
		mmu-miR-451
		mmu-miR-9*
		mmu-miR-98

Supplementary Table 3 - Predicted primary targets of the miR-290 cluster miRNAs

RefSeq	Gene symbol	Description	Number of GCACUU motifs
NM_025326	0610011I04Rik	Mus musculus RIKEN cDNA 0610011I04 gene	1
XM_001003634	1110060D06Rik	PREDICTED: Mus musculus RIKEN cDNA 1110060D06 gene, transcript variant 2	1
XM_978179	1810013L24Rik	PREDICTED: Mus musculus RIKEN cDNA 1810013L24 gene, transcript variant 1	2
NM_026437	1810055E12Rik	Mus musculus RIKEN cDNA 1810055E12 gene	2
NM_175381	2700081O15Rik	Mus musculus RIKEN cDNA 2700081O15 gene	4
NM_172877	4732496O08Rik	Mus musculus RIKEN cDNA 4732496O08 gene	1
NM_029037	4930444A02Rik	Mus musculus RIKEN cDNA 4930444A02 gene	2
NM_175172	4930506M07Rik	Mus musculus RIKEN cDNA 4930506M07 gene	1
NM_173764	4932414K18Rik	Mus musculus RIKEN cDNA 4932414K18 gene	2
NM_175263	5730593N15Rik	Mus musculus RIKEN cDNA 5730593N15 gene	1
NM_146091	5730596K20Rik	Mus musculus RIKEN cDNA 5730596K20 gene	1
NM_025697	6330409N04Rik	Mus musculus RIKEN cDNA 6330409N04 gene	1
XM_991839	9030420J04Rik	PREDICTED: Mus musculus RIKEN cDNA 9030420J04 gene	1
NM_153117	9530068E07Rik	Mus musculus RIKEN cDNA 9530068E07 gene	4
NM_001007577	A630018P17Rik	Mus musculus RIKEN cDNA A630018P17 gene	1
NM_177118	A830073O21Rik	Mus musculus RIKEN cDNA A830073O21 gene	2
NM_015729	Acox1	Mus musculus acyl-Coenzyme A oxidase 1, palmitoyl	1
NM_007394	Acvr1	Mus musculus activin A receptor, type 1	1
NM_007404	Adam9	Mus musculus a disintegrin and metallopeptidase domain 9 (meltrin gamma)	2
NM_007408	Adfp	Mus musculus adipose differentiation related protein	1
NM_001005605	Aebp2	Mus musculus AE binding protein 2	2
NM_001033476	AI450948	Mus musculus expressed sequence AI450948	1
NM_177907	AI593442	Mus musculus expressed sequence AI593442	1
NM_145489	AI661453	Mus musculus expressed sequence AI661453	1
NM_011785	Akt3	Mus musculus thymoma viral proto-oncogene 3	1
NM_028270	Aldh1b1	Mus musculus aldehyde dehydrogenase 1 family, member B1	1
NM_019998	Alg2	Mus musculus asparagine-linked glycosylation 2 homolog (yeast, alpha-1,3-mannosyltransfe	1
NM_009667	Ampd3	Mus musculus AMP deaminase 3	1
NM_134071	Ankrd32	Mus musculus ankyrin repeat domain 32	1
XM_001000870	Ap1g1	PREDICTED: Mus musculus adaptor protein complex AP-1, gamma 1 subunit	1
NM_007457	Ap1s1	Mus musculus adaptor protein complex AP-1, sigma 1	1
NM_009686	Apbb2	Mus musculus amyloid beta (A4) precursor protein-binding, family B, member 2	1

NM_027144	Arhgef12	Mus musculus Rho guanine nucleotide exchange factor (GEF) 12	2
NM_001039515	Arl4a	Mus musculus ADP-ribosylation factor-like 4A	1
NM_007488	Arnt2	Mus musculus aryl hydrocarbon receptor nuclear translocator 2	1
NM_030711	Arts1	Mus musculus type 1 tumor necrosis factor receptor shedding aminopeptidase regulator	1
NM_025541	Asf1a	Mus musculus ASF1 anti-silencing function 1 homolog A (S. cerevisiae)	2
NM_138679	Ash1l	Mus musculus ash1 (absent, small, or homeotic)-like (Drosophila)	3
NM_007496	Atbf1	Mus musculus AT motif binding factor 1	2
NM_177632	BC022623	Mus musculus cDNA sequence BC022623	1
NM_153407	BC035295	Mus musculus cDNA sequence BC035295	1
XM_984947	BC053401	PREDICTED: Mus musculus cDNA sequence BC053401, transcript variant 2	1
NM_012060	Bcap31	Mus musculus B-cell receptor-associated protein 31	1
NM_080708	Bmp2k	Mus musculus BMP2 inducible kinase	1
NM_027430	Brp44	Mus musculus brain protein 44	1
NM_009790	Calm1	Mus musculus calmodulin 1	1
NM_177343	Camk1d	Mus musculus calcium/calmodulin-dependent protein kinase ID	4
XM_001000085	Camta1	PREDICTED: Mus musculus calmodulin binding transcription activator 1	1
XM_985577	Cand1	PREDICTED: Mus musculus cullin associated and neddylation disassociated 1	2
NM_007610	Casp2	Mus musculus caspase 2	4
NM_009817	Cast	Mus musculus calpastatin	1
NM_028763	Cbx6	Mus musculus chromobox homolog 6	1
NM_198164	Cdc2l6	Mus musculus cell division cycle 2-like 6 (CDK8-like)	2
NM_007669	Cdkn1a	Mus musculus cyclin-dependent kinase inhibitor 1A (P21)	3
NM_028760	Cep55	Mus musculus centrosomal protein 55	2
NM_019950	Chst5	Mus musculus carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 5	1
NM_013885	Clic4	Mus musculus chloride intracellular channel 4 (mitochondrial)	2
XM_921620	Cnot6	PREDICTED: Mus musculus CCR4-NOT transcription complex, subunit 6, transcript variant	4
NM_178854	Cnot6l	Mus musculus CCR4-NOT transcription complex, subunit 6-like	2
NM_013495	Cpt1a	Mus musculus carnitine palmitoyltransferase 1a, liver	1
NM_009963	Cry2	Mus musculus cryptochrome 2 (photolyase-like)	2
NM_026444	Cs	Mus musculus citrate synthase	1
NM_173185	Csnk1g1	Mus musculus casein kinase 1, gamma 1	4
NM_177662	Ctso	Mus musculus cathepsin O	2
NM_177640	D030056L22Rik	Mus musculus RIKEN cDNA D030056L22 gene	2
XM_984040	D630040G17Rik	PREDICTED: Mus musculus RIKEN cDNA D630040G17 gene	3
NM_011873	Dazap2	Mus musculus DAZ associated protein 2	3
NM_026302	Dctn4	Mus musculus dynactin 4	4

XM_001005781	Ddef2	PREDICTED: Mus musculus development and differentiation enhancing factor 2	1
NM_001039106	Ddhd1	Mus musculus DDHD domain containing 1	3
NM_007916	Ddx19a	Mus musculus DEAD (Asp-Glu-Ala-Asp) box polypeptide 19a	1
XM_898691	Dip2a	PREDICTED: Mus musculus DIP2 disco-interacting protein 2 homolog A (Drosophila), transc	1
XM_619261	Dock5	PREDICTED: Mus musculus dedicator of cytokinesis 5, transcript variant 1	1
NM_053090	Drctnnb1a	Mus musculus down-regulated by Ctnnb1, a	1
NM_001013371	Dtx3l	Mus musculus deltex 3-like (Drosophila)	3
NM_001013380	Dync1li2	Mus musculus dynein, cytoplasmic 1 light intermediate chain 2	2
NM_173386	E330016A19Rik	Mus musculus RIKEN cDNA E330016A19 gene	3
NM_001001932	Eea1	Mus musculus early endosome antigen 1	1
NM_007915	Ei24	Mus musculus etoposide induced 2.4 mRNA	1
NM_207685	Elavl2	Mus musculus ELAV (embryonic lethal, abnormal vision, Drosophila)-like 2 (Hu antigen B)	2
NM_007936	Epha4	Mus musculus Eph receptor A4	2
NM_007961	Etv6	Mus musculus ets variant gene 6 (TEL oncogene)	2
NM_153118	Fnbp1l	Mus musculus formin binding protein 1-like	1
NM_173182	Fndc3b	Mus musculus fibronectin type III domain containing 3B	1
NM_028194	Fryl	Mus musculus furry homolog-like (Drosophila)	1
XM_980423	Furin	PREDICTED: Mus musculus furin (paired basic amino acid cleaving enzyme), transcript varia	1
NM_008056	Fzd6	Mus musculus frizzled homolog 6 (Drosophila)	1
NM_019749	Gabarap	Mus musculus gamma-aminobutyric acid receptor associated protein	1
NM_013814	Galnt1	Mus musculus UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltran	1
NM_013529	Gfpt2	Mus musculus glutamine fructose-6-phosphate transaminase 2	1
NM_133236	Glcci1	Mus musculus glucocorticoid induced transcript 1	1
NM_025374	Glo1	Mus musculus glyoxalase 1	1
NM_021610	Gpa33	Mus musculus glycoprotein A33 (transmembrane)	1
NM_173747	Gpkow	Mus musculus G patch domain and KOW motifs	3
NM_019986	Habp4	Mus musculus hyaluronic acid binding protein 4	2
NM_026812	Hddc3	Mus musculus HD domain containing 3	1
NM_010437	Hivep2	Mus musculus human immunodeficiency virus type I enhancer binding protein 2	2
NM_008253	Hmgb3	Mus musculus high mobility group box 3	1
NM_008258	Hn1	Mus musculus hematological and neurological expressed sequence 1	2
NM_010470	Hp1bp3	Mus musculus heterochromatin protein 1, binding protein 3	2
XM_985333	Hs6st1	PREDICTED: Mus musculus heparan sulfate 6-O-sulfotransferase 1	2
NM_175185	Hsdl1	Mus musculus hydroxysteroid dehydrogenase like 1	2
NM_008331	Ifit1	Mus musculus interferon-induced protein with tetratricopeptide repeats 1	1
NM_019440	Iigp2	Mus musculus interferon inducible GTPase 2	1

NM_008371	Il7	Mus musculus interleukin 7	2
NM_172161	Irak2	Mus musculus interleukin-1 receptor-associated kinase 2	1
NM_008390	Irf1	Mus musculus interferon regulatory factor 1	1
XM_001002526	Irf2bp2	PREDICTED: Mus musculus interferon regulatory factor 2 binding protein 2	3
NM_008394	Isgf3g	Mus musculus interferon dependent positive acting transcription factor 3 gamma	2
NM_008402	Itgav	Mus musculus integrin alpha V	1
NM_010580	Itgb5	Mus musculus integrin beta 5	1
NM_008410	Itm2b	Mus musculus integral membrane protein 2B	1
NM_021310	Jmy	Mus musculus junction-mediating and regulatory protein	1
XM_978811	Kctd1	PREDICTED: Mus musculus potassium channel tetramerisation domain containing 1, transc	1
NM_207682	Kif1b	Mus musculus kinesin family member 1B	2
XM_994052	Klf3	PREDICTED: Mus musculus Kruppel-like factor 3 (basic)	2
NM_021284	Kras	Mus musculus v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	1
NM_145743	Lace1	Mus musculus lactation elevated 1	2
NM_010685	Lamp2	Mus musculus lysosomal membrane glycoprotein 2	3
NM_172153	Lcorl	Mus musculus ligand dependent nuclear receptor corepressor-like	2
NM_010094	Lefty1	Mus musculus left right determination factor 1	1
NM_177099	Lefty2	Mus musculus Left-right determination factor 2	1
NM_001029878	Lonrf2	Mus musculus LON peptidase N-terminal domain and ring finger 2	2
NM_172950	Lpin1	Mus musculus lipin 1	1
NM_080420	Lpo	Mus musculus lactoperoxidase	1
NM_172784	Lrp11	Mus musculus low density lipoprotein receptor-related protein 11	1
NM_173038	Lrrc35	Mus musculus leucine rich repeat containing 35	3
XM_984916	Lycat	PREDICTED: Mus musculus lysocardiolipin acyltransferase, transcript variant 3	1
NM_172865	Manea	Mus musculus mannosidase, endo-alpha	1
NM_027920	March8	Mus musculus membrane-associated ring finger (C3HC4) 8	3
NM_010773	Mbd2	Mus musculus methyl-CpG binding domain protein 2	1
NM_020007	Mbnl1	Mus musculus muscleblind-like 1 (Drosophila)	2
NM_175341	Mbnl2	Mus musculus muscleblind-like 2	1
NM_175088	Mdfic	Mus musculus MyoD family inhibitor domain containing	1
XM_976104	Mef2a	PREDICTED: Mus musculus myocyte enhancer factor 2A, transcript variant 3	2
XM_001002380	Mfn2	PREDICTED: Mus musculus mitofusin 2, transcript variant 4	1
XM_912670	Mgat5	PREDICTED: Mus musculus mannoside acetylglucosaminyltransferase 5, transcript variant 1	3
NM_008606	Mmp11	Mus musculus matrix metallopeptidase 11	1
NM_011985	Mmp23	Mus musculus matrix metallopeptidase 23	1
NM_008636	Mtf1	Mus musculus metal response element binding transcription factor 1	2

NM_001005864	Mtus1	Mus musculus mitochondrial tumor suppressor 1	1
NM_139063	Muted	Mus musculus muted	2
NM_008659	Myo1c	Mus musculus myosin IC	1
NM_019542	Nagk	Mus musculus N-acetylglucosamine kinase	1
XM_973478	Nck2	PREDICTED: Mus musculus non-catalytic region of tyrosine kinase adaptor protein 2, transc	1
NM_172495	Ncoa7	Mus musculus nuclear receptor coactivator 7	1
NM_008684	Neo1	Mus musculus neogenin	2
NM_008687	Nfib	Mus musculus nuclear factor I/B	1
NM_023526	Nkiras1	Mus musculus NFKB inhibitor interacting Ras-like protein 1	2
NM_009697	Nr2f2	Mus musculus nuclear receptor subfamily 2, group F, member 2	2
NM_172416	Ostm1	Mus musculus osteopetrosis associated transmembrane protein 1	1
NM_008775	Pafah1b2	Mus musculus platelet-activating factor acetylhydrolase, isoform 1b, alpha2 subunit	1
NM_011864	Papss2	Mus musculus 3'-phosphoadenosine 5'-phosphosulfate synthase 2	1
NM_028829	Paqr8	Mus musculus progestin and adipoQ receptor family member VIII	1
XM_992943	Pbx3	PREDICTED: Mus musculus pre B-cell leukemia transcription factor 3, transcript variant 6	1
XM_982935	Pcgf4	PREDICTED: Mus musculus polycomb group ring finger 4, transcript variant 4	1
NM_008786	Pcmt1	Mus musculus protein-L-isoaspartate (D-aspartate) O-methyltransferase 1	1
XM_912421	Pde3b	PREDICTED: Mus musculus phosphodiesterase 3B, cGMP-inhibited	1
XM_920266	Pdzd2	PREDICTED: Mus musculus PDZ domain containing 2, transcript variant 9	1
NM_019410	Pfn2	Mus musculus profilin 2	4
NM_171824	Pgbd5	Mus musculus piggyBac transposable element derived 5	2
XM_895539	Phip	PREDICTED: Mus musculus pleckstrin homology domain interacting protein, transcript varia	3
NM_201406	Pigs	Mus musculus phosphatidylinositol glycan anchor biosynthesis, class S	1
NM_181585	Pik3r3	Mus musculus phosphatidylinositol 3 kinase, regulatory subunit, polypeptide 3 (p55)	1
NM_001025309	Pja2	Mus musculus praja 2, RING-H2 motif containing	1
NM_018807	Plagl2	Mus musculus pleiomorphic adenoma gene-like 2	3
NM_031256	Plekha3	Mus musculus pleckstrin homology domain-containing, family A (phosphoinositide binding sp	1
NM_183034	Plekhm1	Mus musculus pleckstrin homology domain containing, family M (with RUN domain) member	2
NM_023564	Plscr3	Mus musculus phospholipid scramblase 3	1
NM_010127	Pou6f1	Mus musculus POU domain, class 6, transcription factor 1	2
NM_024209	Ppp6c	Mus musculus protein phosphatase 6, catalytic subunit	1
NM_178738	Prss35	Mus musculus protease, serine, 35	2
NM_008971	Ptk9	Mus musculus protein tyrosine kinase 9	1
NM_011877	Ptpn21	Mus musculus protein tyrosine phosphatase, non-receptor type 21	2
NM_027514	Pvr	Mus musculus poliovirus receptor	2
NM_024436	Rab22a	Mus musculus RAB22A, member RAS oncogene family	3

NM_026405	Rab32	Mus musculus RAB32, member RAS oncogene family	1
NM_144875	Rab7l1	Mus musculus RAB7, member RAS oncogene family-like 1	1
NM_001038621	Rabgap1l	Mus musculus RAB GTPase activating protein 1-like	1
XM_983626	Rapgef2	PREDICTED: Mus musculus Rap guanine nucleotide exchange factor (GEF) 2	1
NM_009826	Rb1cc1	Mus musculus RB1-inducible coiled-coil 1	2
NM_011250	Rbl2	Mus musculus retinoblastoma-like 2	3
NM_029777	Rhbdd1	Mus musculus rhomboid domain containing 1	1
NM_007484	Rhoc	Mus musculus ras homolog gene family, member C	1
NM_145491	Rhoq	Mus musculus ras homolog gene family, member Q	1
NM_023894	Rhox9	Mus musculus reproductive homeobox 9	1
NM_009068	Ripk1	Mus musculus receptor (TNFRSF)-interacting serine-threonine kinase 1	1
NM_023270	Rnf128	Mus musculus ring finger protein 128	1
NM_011277	Rnf2	Mus musculus ring finger protein 2	1
XM_903197	Rora	PREDICTED: Mus musculus RAR-related orphan receptor alpha, transcript variant 4	4
NM_009075	Rpia	Mus musculus ribose 5-phosphate isomerase A	1
NM_199476	Rrm2b	Mus musculus ribonucleotide reductase M2 B (TP53 inducible)	1
NM_030179	Rsnl2	Mus musculus restin-like 2	1
NM_030692	Sacm1l	Mus musculus SAC1 (suppressor of actin mutations 1, homolog)-like (S. cerevisiae)	1
NM_011452	Serpinb9b	Mus musculus serine (or cysteine) peptidase inhibitor, clade B, member 9b	2
NM_031179	Sf3b1	Mus musculus splicing factor 3b, subunit 1	2
XM_988661	Sh3glb1	PREDICTED: Mus musculus SH3-domain GRB2-like B1 (endophilin)	1
NM_177364	Sh3pxd2b	Mus musculus SH3 and PX domains 2B	3
NM_172966	Sh3rf2	Mus musculus SH3 domain containing ring finger 2	1
NM_134038	Slc16a6	Mus musculus solute carrier family 16 (monocarboxylic acid transporters), member 6	1
NM_025807	Slc16a9	Mus musculus solute carrier family 16 (monocarboxylic acid transporters), member 9	1
NM_172773	Slc17a5	Mus musculus solute carrier family 17 (anion/sugar transporter), member 5	2
NM_018861	Slc1a4	Mus musculus solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	1
NM_021435	Slc35b4	Mus musculus solute carrier family 35, member B4	2
NM_133741	Snrk	Mus musculus SNF related kinase	1
NM_029068	Snx16	Mus musculus sorting nexin 16	1
NM_028937	Sohlh2	Mus musculus spermatogenesis and oogenesis specific basic helix-loop-helix 2	1
NM_009274	Srpk2	Mus musculus serine/arginine-rich protein specific kinase 2	3
NM_138744	Ssx2ip	Mus musculus synovial sarcoma, X breakpoint 2 interacting protein	1
NM_011374	St8sia1	Mus musculus ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 1	1
NM_024239	Stambp	Mus musculus Stam binding protein	1
NM_133810	Stk17b	Mus musculus serine/threonine kinase 17b (apoptosis-inducing)	2

NM_134115	Stk38	Mus musculus serine/threonine kinase 38	1
NM_026343	Stx17	Mus musculus syntaxin 17	1
NM_025932	Syap1	Mus musculus synapse associated protein 1	2
XM_903544	Syde1	PREDICTED: Mus musculus synapse defective 1, Rho GTPase, homolog 1 (C. elegans)	1
XM_981719	Synj1	PREDICTED: Mus musculus synaptojanin 1	1
NM_145968	Tagap	Mus musculus T-cell activation Rho GTPase-activating protein	1
NM_198294	Tanc1	Mus musculus tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1	2
NM_023755	Tcfcp2l1	Mus musculus transcription factor CP2-like 1	2
NM_146142	Tdrd7	Mus musculus tudor domain containing 7	1
NM_009371	Tgfbr2	Mus musculus transforming growth factor, beta receptor II	3
NM_009373	Tgm2	Mus musculus transglutaminase 2, C polypeptide	1
NM_133352	Tm9sf3	Mus musculus transmembrane 9 superfamily member 3	1
XM_892747	Tmcc1	PREDICTED: Mus musculus transmembrane and coiled coil domains 1, transcript variant 2	2
NM_134020	Tmed4	Mus musculus transmembrane emp24 protein transport domain containing 4	1
NM_144792	Tmem23	Mus musculus transmembrane protein 23	1
NM_009395	Tnfaip1	Mus musculus tumor necrosis factor, alpha-induced protein 1 (endothelial)	2
XM_982893	Tnrc6a	PREDICTED: Mus musculus trinucleotide repeat containing 6a	1
NM_009277	Trim21	Mus musculus tripartite motif protein 21	2
NM_178872	Trim36	Mus musculus tripartite motif-containing 36	1
XM_987804	Trps1	PREDICTED: Mus musculus trichorhinophalangeal syndrome I (human)	1
NM_133681	Tspan1	Mus musculus tetraspanin 1	2
NM_026954	Tusc1	Mus musculus tumor suppressor candidate 1	1
NM_028339	Txndc1	Mus musculus thioredoxin domain containing 1	2
NM_153162	Txnrd3	Mus musculus thioredoxin reductase 3	1
NM_019586	Ube2j1	Mus musculus ubiquitin-conjugating enzyme E2, J1	1
NM_172300	Ube2z	Mus musculus ubiquitin-conjugating enzyme E2Z (putative)	1
NM_011670	Uchl1	Mus musculus ubiquitin carboxy-terminal hydrolase L1	2
NM_009466	Ugdh	Mus musculus UDP-glucose dehydrogenase	1
NM_177561	Usp46	Mus musculus ubiquitin specific peptidase 46	2
XM_977760	Wdr26	PREDICTED: Mus musculus WD repeat domain 26, transcript variant 7	1
NM_011701	Vim	Mus musculus vimentin	1
NM_172643	Zbtb41	Mus musculus zinc finger and BTB domain containing 41 homolog	4
NM_010731	Zbtb7a	Mus musculus zinc finger and BTB domain containing 7a	4
NM_028864	Zc3hav1	Mus musculus zinc finger CCCH type, antiviral 1	1
NM_011749	Zfp148	Mus musculus zinc finger protein 148	2
NM_175494	Zfp367	Mus musculus zinc finger protein 367	4

NM_009557	Zfp46	Mus musculus zinc finger protein 46	2
NM_133218	Zfp704	Mus musculus zinc finger protein 704	5
NM_133906	Zkscan1	Mus musculus zinc finger with KRAB and SCAN domains 1	5
NM_144516	Zmynd11	Mus musculus zinc finger, MYND domain containing 11	1
XM_893176	Znrf3	PREDICTED: Mus musculus zinc and ring finger 3, transcript variant 2	2

Supplementary Table 4. Primers used in the study.

Primer pair		Primer sequences (5´-3´)	Reference
Primers for chromatin immunoprecipitation			
GAPDH	forward	TCCCCTCCCCCTATCAGTTC	
	reverse	TTGGACCCGCCTCATTTTT	
Oct-4	forward	TGGGCTGAAATACTGGGTTC	
	reverse	TTGAATGTTCGTGTGCCAAT	
Primers for real-time quantitative RT-PCR			
GAPDH	forward	CCATCACCATCTTCCAGG	
	reverse	CCTGCTTCACCACCTTCTTG	
Oct-4	forward	GGCGTTCGCTTTGGAAAGGTGTTC	
	reverse	CTCGAACCACATCCTTCTCT	
Nanog	forward	TGATTCAGAAGGGCTCAGCAC	
	reverse	GGGATAGCTGCAATGGATGC	
GCNF	forward	TGAATTGGCAGAGCTTGATCC	
	reverse	CGATCATCTGGGACGGAAAC	
Dnmt3a2	forward	AGGGGCTGCACCTGGCCTT	[1]
	reverse	TCCCCCACACCAGCTCTCC	[1]
Dnmt3b1/b6	forward	TGGGATCGAGGGCCTCAAAC	[1]
	reverse	TTCCACAGGACAAACAGCGG	[1]
Rbl2	forward	CCCGGAGCCAGGTGTACA	
	reverse	CCTCATCACTGGGCTGGAAT	
Primers for bisulfite sequencing			
Oct-3/4 [-2069]	forward	GGGAGGAATTGGGTGTGGGGAGGTT	[2]
Oct-3/4 [-1677]	reverse	AAAAATCCCCTCCTTCTACCACAT	[2]
Oct-3/4 [-1505]	forward	TGATGAAGATTATTATTAAGAGAT	[2]
Oct-3/4 [-992]	reverse	CCCCAATCCCCTCACACAAAACTT	[2]
Oct-3/4 [-212]	forward	AGGATTTTGAAGGTTGAAAATGAAGG	[2]
Oct-3/4 [-8]	reverse	TCCCTCCCCAATCCCACCCTC	[2]
Oct-3/4 [+1173]	forward	GTAATTAGTTTTAAGAATAAGGTG	[2]
Oct-3/4 [+1340]	reverse	AAATAAACTATTAATACCTTCCTA	[2]
CTCF.bis.9.fwd	forward	GTTAATTTTTAAGGATGATAGTTTTGTGATT	
CTCF.bis.9.rev	reverse	AAAACCATAACAAAAACCTAAACCTTAC	
IAP.bis.fwd	forward	TTGATAGTTGTGTTTTAAGTGGTAAATAAA	[3]
IAP.bis.rev	reverse	AAAACACCACAAACCAAAATCTTCTAC	[3]
IAP 5'.bis.fwd	forward	ATGGGTTGTAGTTAATTAGGGAGTGATA	
IAP 5'.bis.rev	reverse	CATACAATTAAATCCTTCTTAACAATCTAC	
LINE1.bis.fwd	forward	TAGGAAATTAGTTTGAATAGGTGAGAGGT	
LINE1.bis.rev	reverse	TCAAACACTATATTACTTTAACAATTCCCA	
satellite.bis.fwd	forward	ATACACACTTTAAAACATAAAATATAA	[5]
satellite.bis.rev	reverse	TTYGTTATATTTTAGGTTTTTTAGA	[5]
MuERV.bis.fwd	forward	GTTATTATGTGATTTGAATTA	[3]
MuERV.bis.rev	reverse	ACATACAAAACCATCAATAAAC	[3]

Sox30.bis.fwd	forward	AGGTGTTTTTATATTTGAGAATGATTAGAA	[4]
Sox30.bis.rev	reverse	ATTAAAACCCTTCCAAAACCTTAACTA	[4]
Tsp50.bis.fwd	forward	TAAAAATTGTTATTGAAGTTAAGTTTGG	[4]
Tsp50.bis.rev	reverse	CTAAACCCTTTCTCTAAATCCCTATAC	[4]

References for primer sequences

1. Chen, T., Ueda, Y., Xie, S. & Li, E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J Biol Chem* **277**, 38746-54 (2002).
2. Sato, N., Kondo, M. & Arai, K. The orphan nuclear receptor GCNF recruits DNA methyltransferase for Oct-3/4 silencing. *Biochem Biophys Res Commun* **344**, 845-51 (2006).
3. Ramirez, M.A. et al. Transcriptional and post-transcriptional regulation of retrotransposons IAP and MuERV-L affect pluripotency of mice ES cells. *Reprod Biol Endocrinol* **4**, 55 (2006).
4. Mohn, F. et al. Concerted reprogramming of DNA methylation and Polycomb targets defines stem cell commitment and terminal neuronal differentiation. **submitted**(2007).
5. Kim, S.H. et al. Differential DNA methylation reprogramming of various repetitive sequences in mouse preimplantation embryos. *Biochem Biophys Res Commun* **324**, 58-63 (2004).

Sinkkonen et al. SUPPLEMENTARY MATERIALS AND METHODS

mRNA microarray analysis

Undifferentiated *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells were grown in the presence of LIF as described in the main text. Prior to harvesting, *Dicer*$^{+/-}$ and *Dicer*$^{-/-}$ cells were grown in triplicates for more than one week. For the rescue with the miR-290s mimics, three independently cultured *Dicer*$^{-/-}$ cell samples were transfected separately with either miR-290 cluster miRNA mimics or siRL, as described in Materials and Methods (main text), and harvested 24 h later. We note that at the time of study the miR-290 annotation in miRBase[1] was for the miR-290-5p and not miR-290-3p. The miR-290-5p mimic was therefore included in the transfection together with miR-291a-3p through miR-295-3p miRNAs, which represent the main products of their respective hairpins. The complement of miR-290-5p seed was not found to be significantly enriched in any seed motif analysis, arguing that miR-290-5p does not play a major role in ES cells.

Total RNA was isolated using Absolutely RNA Miniprep Kit (Stratagene). 5 µg of total RNA from each triplicate culture was reverse transcribed with the Affymetrix cDNA synthesis kit and cRNA was produced by *in vitro* transcription (IVT) by T7 RNA polymerase, using the Affymetrix IVT kit as per manufacturer's instructions. 20 µg of biotinylated cRNA was fragmented by heating in the presence of Mg^{2+} (as per Affymetrix's instructions) and 15 µg of fragmented cRNA from each triplicate was hybridized to Mouse MOE430 v2.0 GeneChips™. All arrays yielded hybridization signals of comparable intensity and quality. BioConductor[2] Affymetrix package of the R software was used to import the CEL files from the Affymetrix Mouse Genome 430 2.0 Array. Probe set intensities were then background-corrected, adjusted for non-specific binding and quantile normalized with the GCRMA algorithm[3]. GCRMA-normalized microarray data were deposited in the GEO database (GSE7141 and GSE8503).

Data analysis

To extract a non-redundant set of transcripts for subsequent analyses of 3′-UTR sequences, probe sets with _s or _x tags, which map to multiple transcripts from different genes, were discarded. Then, the Affymetrix annotation from December 2006 was used

to obtain the corresponding reference sequence (RefSeq[4]) for each probe set. When the Affymetrix array contained probe sets for alternative RefSeq transcripts for the same gene, we only used the RefSeq transcript with the median length 3′-UTR. Through this procedure, we obtained an n-to-1 probe set to RefSeq transcript mapping. For transcripts that had multiple probe sets, we discarded those that were deficient, as indicated by their very low variance across a set of unrelated experiments performed with different cell types using the same platform (Affymetrix Mouse Genome 430 2.0). Finally, the log2 intensities of the probe sets corresponding to a given transcript were averaged to obtain a transcript level measurement. We used Limma[5] to estimate the fold change and the corresponding p-value in the three replicate experiments for each condition.

To identify those motifs whose frequency in up-regulated (in *Dicer*$^{-/-}$) or down-regulated (in *Dicer*$^{-/-}$ ES cells transfected with miRNA mimics of the miR-290 family) 3′-UTRs is significantly different relative to the frequency in the entire set of 3′-UTRs, we extracted the set of transcripts up-regulated in the Dicer-/- cells (p-value < 0.001) and computed the relative frequency of all 7-mers in the 3′-UTRs of these transcripts compared with the entire set of 3′-UTRs represented on the microarray. For each 7-mer, we then plotted the log2(number of occurrences in up-regulated 3′-UTRs) on the x-axis, and the enrichment in up-regulated 3′-UTRs compared to the entire set of 3′-UTRs on the y-axis (Fig. 1b and 1e). We then used a Bayesian model that we previously introduced for comparing miRNA frequencies between samples[6]. Briefly, we estimate the posterior probabilities of the model that assumes that the frequency of a given motif is different between two sets of transcripts (call this "different" model), and the model that assumes that the frequency is the same (call this "same" model), given the observed counts m and n of the motif among M and N total motifs in the two samples. We selected as significant those motifs that were enriched in the up-regulated or down-regulated set, respectively, with a posterior probability of the "different" model > 0.99

miRNA microarray analysis

Total RNA from two independent cultures of *Dicer*$^{+/-}$ cells and single culture of *Dicer*$^{-/-}$ cells was extracted using MirVana miRNA Isolation Kit (Ambion). 5 µg of each RNA preparation was used for miRNA miRCURY^{TM} microarray analysis as a service by

Exiqon (Vedbäck, Denmark). As a control, a mixture of 5 μg of total RNA originating from 10 mouse tissues (Ambion) supplemented with 500 ng of total RNA from *Dicer*$^{+/-}$ cells was labeled with Hy5 (spectrally equivalent to Cy5) and co-hybridized with either the *Dicer*$^{+/-}$ or the *Dicer*$^{-/-}$ RNA samples, which were labeled with Hy3 (spectrally equivalent to Cy3). The expression level of reliably detected miRNAs was calculated relative to the levels in *Dicer*$^{+/-}$ sample as well as relative to the level in the control mixture of total RNAs (reference sample). Most miRNA probes exhibited hybridization signal also with *Dicer*$^{-/-}$ samples, suggesting that the arrays also detect precursors of miRNAs or cross-hybridize to unrelated RNAs. The original data are available upon request.

REFERENCES

1. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-4 (2006).

2. Gentleman, R.C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).

3. Wu, Z., Irizarryl, A.R., Gentleman, R., Martinez-Murillo, F. & Spencer, F.M. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *JASA* **99**, 909-17 (2004).

4. Pruitt, K.D., Katz, K.S., Sicotte, H. & Maglott, D.R. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet* **16**, 44-7 (2000).

5. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).

6. Landgraf, P. et al. A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* **129**, 1401-14 (2007).

7. Kanellopoulou, C. et al. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev* **19**, 489-501 (2005).

8. Deussing, J. et al. Identification and characterization of a dense cluster of placenta-specific cysteine peptidase genes and related genes on mouse chromosome 13. *Genomics* **79**, 225-40 (2002).

9. Hough, S.R., Clements, I., Welch, P.J. & Wiederholt, K.A. Differentiation of mouse embryonic stem cells after RNA interference-mediated silencing of OCT4 and Nanog. *Stem Cells* **24**, 1467-75 (2006).

10. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-40 (2006).

11. Zeng, F., Baldwin, D.A. & Schultz, R.M. Transcript profiling during preimplantation mouse development. *Dev Biol* **272**, 483-96 (2004).

12. Schmitter, D. et al. Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res* **34**, 4801-15 (2006).

13. Weber, M. & Schubeler, D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol* **19**, 273-80 (2007).

14. Svoboda, P., Stein, P., Filipowicz, W. & Schultz, R.M. Lack of homologous sequence-specific DNA methylation in response to stable dsRNA expression in mouse oocytes. *Nucleic Acids Res* **32**, 3601-6 (2004).

3.1.3 The silencing of pri-miR-290 locus by *de novo* DNA methylation during neuronal differentiation enables upregulation of neuronal genes

Lasse Sinkkonen, Fabio Mohn, Dirk Schübeler and Witold Filipowicz

3.1.3.1 Aim of the project

Mature miRNAs of the miR-290 cluster are known to become downregulated upon differentiation of ESCs but the mechanism of this downregulation is unknown (Houbaviy et al. 2003). Here we aimed to analyze the downregulation of these miRNAs and to determine whether transcriptional silencing of the miRNA locus is contributing to this downregulation. Since we have previously discovered that miR-290 miRNAs are necessary for proper *de novo* DNA methylation in mouse ESCs (see chapter 3.1.1), we were interested to find out whether the silencing of these miRNAs would involve *de novo* DNA methylation, thus creating a potential autoregulatory loop.

3.1.3.2 Results and discussion

In order to decipher whether the repression of miR-290 miRNAs takes place at the transcriptional level, we focused on the pri-miR-290 transcript (Houbaviy et al. 2005). To analyze pri-miR-290 in a physiologically relevant system, we took advantage of a recently established neuronal differentiation protocol (Bibel et al. 2004; Mohn et al. 2008). In this protocol the mouse ESCs are differentiated in a course of eight days into neuronal precursors (NPs). These precursors are then differentiated additional ten days into specific neuronal subtype of radial glial cells (terminal neurons or TNs). First we made use of the previously published microarray data (Mohn et al. 2008) to analyze the transcriptional changes at these three different developmental stages (ESC, NP and TN). Figure 5A shows a schematic representation of the structure of pri-miR-290 locus and the location of array probes and PCR primers used to analyze the locus. We analyzed the level of pri-miR-290 based on the Affymetrix array probe (1444292_at) recognizing the pri-miR-290 as well as the mRNA levels of primary and secondary miR-290 cluster targets Rbl2 and Dnmt3s, respectively (Figure 5B). As expected from a primary miR-290

cluster target, Rbl2 mRNA level was strongly induced upon differentiation and this upregulation could be confirmed also by real-time quantitative reverse transcription-PCR (RT-qPCR) (data not shown). At the same time the mRNAs for targets of RBL2 repression, Dnmt3a2, Dnmt3b and Dnmt3L, were downregulated. It is important to note that some probe sets not depicted here, especially for Dnmt3a, showed a different pattern of expression. This is most likely due to crosshybridization to alternative transcription or splicing variants such as Dnmt3a1, which is known to have different expression pattern from Dnmt3a2 (Chen et al. 2002). Interestingly, also pri-miR-290 showed a very strong downregulation upon neuronal differentiation (up to 30-fold). Moreover, similar extent of repression could be detected by RT-qPCR when the expression of pri-miR-290 was normalized to that of glyceraldehyde-3-phosphate dehydrogenase (Gapdh) (Figure 5C). This analysis revealed that in ESCs pri-miR-290 is expressed at very high levels at close to 50% of the expression of Gapdh. The slight increase observed in pri-miR-290 expression from NP to TN stage transition is due to a modest decrease in the levels of Gapdh and not because of re-expression of pri-miR-290.

The silencing of pri-miR-290 in neurons was robust and very reminiscent of the irreversible silencing of pluripotency genes such as *Oct-4* and *Nanog*. In addition, pri-miR-290 has similar ESC specific expression pattern as *Oct-4* and *Nanog*. For these reasons we hypothesized that silencing of pri-miR-290 might be accompanied by similar changes in its chromatin structure as the silencing of these pluripotency genes (see for example Mohn et al. 2008). There are only very few pri-miRNAs for which the epigenetic regulation at their promoter regions has been described. Thus, we performed ChIP analysis using antibodies against RNA Pol II, H3K4me2 and H3K27me3. ChIP was perfomed at ESC, NP and TN stages and analyzed by primers detecting the TSS (ChIP proximal) or promoter region (ChIP distal) of pri-miR-290. As a control we monitored the promoter of the highly expressed Gapdh gene. Consistently with high expression in ESCs, RNA Pol II was found highly enriched at the TSS of pri-miR-290 as well as at the Gapdh promoter but not in the more distal region of the pri-miR-290 promoter (Figure 6A). Transcriptional silencing of pri-miR-290 in NPs and TNs was accompanied by complete loss of RNA Pol II while it remained present at the active Gapdh locus.

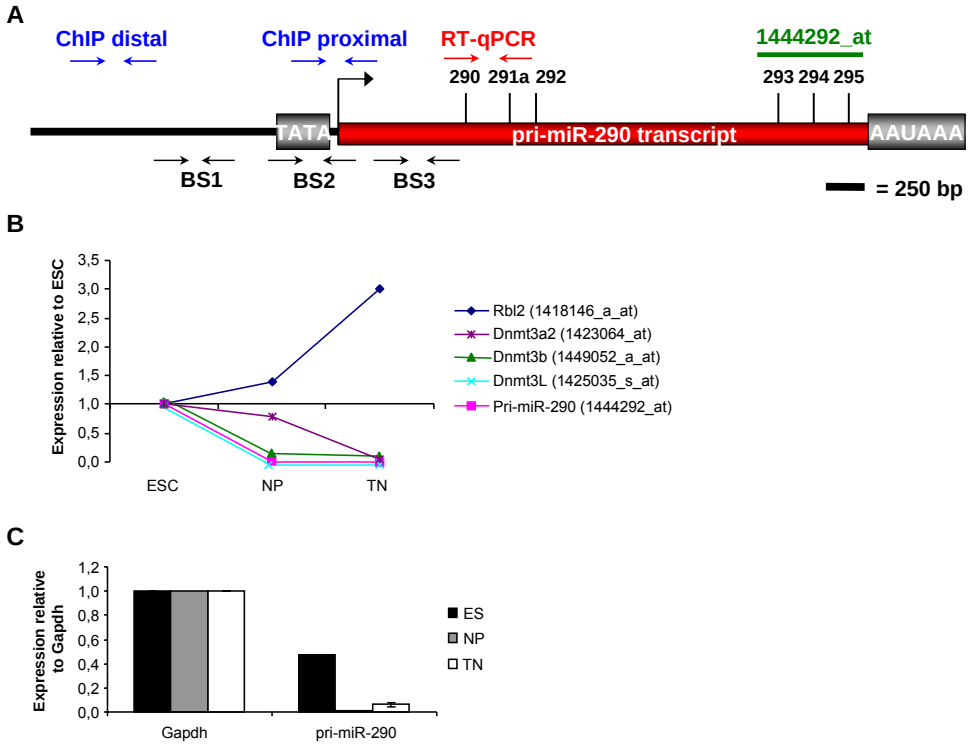


Figure 5. Pri-miR-290 is silenced during neuronal differentiation of mouse ESCs.

(A) Schematic structure of the pri-miR-290 locus depicting the location of the conserved TATA-box, mature miRNA hairpins, the classical polyadenylation signal and the probes/primers used in the experiments. ChIP, chromatin immunoprecipitation; BS, bisulphite sequencing. (B) Expression of the indicated transcripts during neuronal differentiation was analyzed by Affymetrix microarrays and the indicated probe sets. Relative expression levels are shown and expression at the ESC stage was set to 1. The values represent the mean of two independent microarray experiments. (C) The silencing of pri-miR-290 was confirmed by RT-qPCR analysis. The expression levels were normalized to the respective expression of Gapdh. Values represent the mean of two independent experiments. Error bars show standard deviation (SD).

Furthermore, the chromatin modification associated with high transcriptional activity, H3K4me2, was well enriched throughout the pri-miR-290 promoter in ESCs and was fully removed in the neurons (Figure 6B). H3K27me3, which is associated with Polycomb-mediated repression, was not found to be present at Gapdh promoter at any stage (Figure 6C). Also, the pri-miR-290 locus was free of this modification in the pluripotent ESCs but upon differentiation high levels accumulated at the TSS, suggesting that pri-miR-290 might be silenced via activity of PRC2. Interestingly, the enrichment of

H3K27me3 was much weaker at the distal promoter region, indicating that the modification is present specifically at the TSS.

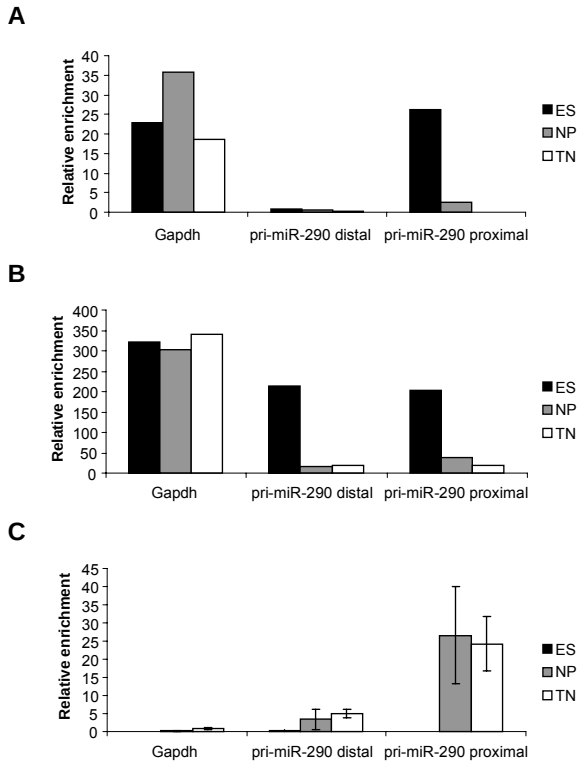


Figure 6. Chromatin changes in transcriptional silencing of pri-miR-290 during differentiation.
ChIP analysis of Gapdh promoter and distal and proximal promoter regions of pri-miR-290 at ESC, NP and TN stages of neuronal differentiation by antibodies against (A) RNA Pol II, (B) H3K4me2 and (C) H3K27me3. Relative enrichments after normalization to the respective input DNA are shown. For H3K27me3 the enrichment was additionally normalized to the enrichment at an unrelated intergenic regions not accumulating H3K27me3. For H3K27me3 the values represent means (± SD) of two independent experiments. For RNA Pol II and H3K4me2, the values come from single experiments.

Complete silencing and heterochromatinization of many pluripotency genes by the Polycomb group proteins is often accompanied by DNA methylation of the locus. Since miR-290 miRNAs are contributing to high expression of *de novo* DNMTs in ESCs, it is possible that *de novo* DNA methylation also contributes to silencing of pri-miR-290, creating an autoregulatory circuit. For this reason, we analyzed the DNA methylation in ESCs, NPs and TNs at three adjacent regions of the pri-miR-290 locus by bisulphite sequencing (Figure 7). Together, the studied regions contain 24 CpGs flanking the TSS

of pri-miR-290. As expected, DNA from ESCs contained only low level of CpG-methylation. But, consistently with the kinetics of *de novo* DNA methylation of other pluripotency genes, pri-miR-290 promoter had become highly methylated in the NPs and maintained this methylation also in TNs. Interestingly, the CpG-dinucleotide most resistant to methylation (the CpG depicted by the fourth circle from the right in region BS2 of Figure 7) is located immediately upstream of the pri-miR-290 TSS, at the position -9. None of the sequenced clones from NPs was methylated at this position. And even in

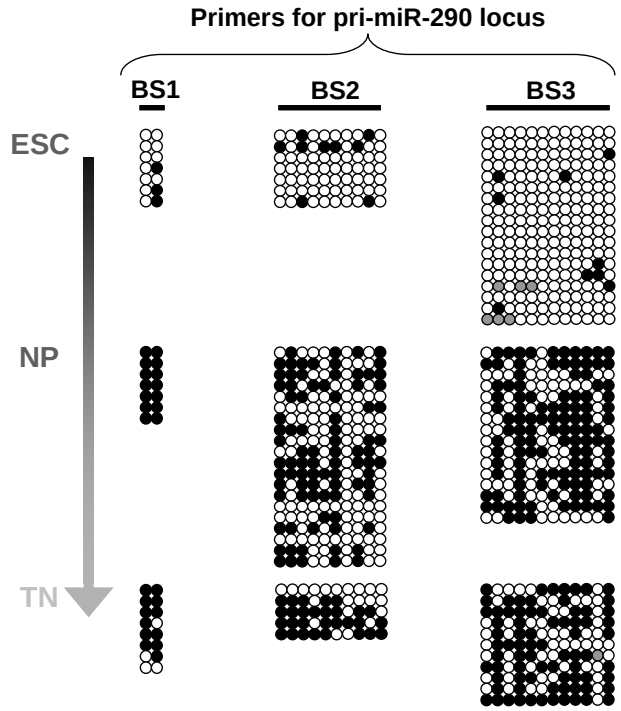


Figure 7. DNA methylation of pri-miR-290 locus during neuronal differentiation.
DNA was extracted from the ESC, NP and TN stages of neuronal differentiation and analyzed for CpG methylation by bisulphite sequencing. The location of the three examined regions (BS1, BS2 and BS3) is depicted in Figure 5A. Each row of dots represents CpGs in one sequenced clone. Black dots represent methylated CpGs and white dots represent unmethylated CpGs. Sites for which the methylation status was uncertain are in gray.

TNs only one out of the five sequences had methylation at this CpG. The high gain of DNA methylation at the pri-miR-290 locus, together with the intermediate CpG density of this locus imply that complete silencing of pri-miR-290 and its miRNA products might depend on *de novo* DNA methylation. It will be interesting to examine silencing of the

locus in DNMT3A/DNMT3B as well as PRC2 depleted ESCs and to estimate the contribution of pri-miR-290 silencing for successful differentiation of ESCs.

Since expression of miR-290 miRNAs becomes fully silenced in NPs and TNs, we asked whether the targets of miR-290 miRNAs (listed in chapter 3.1.3) are enriched among transcripts highly expressed in neurons. For this purpose we inspected the microarray data for the transcriptional changes between ESCs and NP or TN stage neurons. The arrays contained altogether 20,872 probe sets that were reliably detected at all three stages. We analyzed what fraction of these probe sets was strongly upregulated (> 3-fold), upregulated (> 1.5-fold), did not change (<1.5-fold), was downregulated (>1.5-fold) or strongly downregulated (> 3-fold) either between ESCs and NPs or between ESCs and TNs (Figure 8A or B, respectively). Then we did the same analysis for the 400 reliably detectable probe sets monitoring the expression of miR-290 targets. Comparison of distribution of individual fractions in the set of all transcripts and in the set of miR-290 target transcripts revealed clear differences between these two groups.

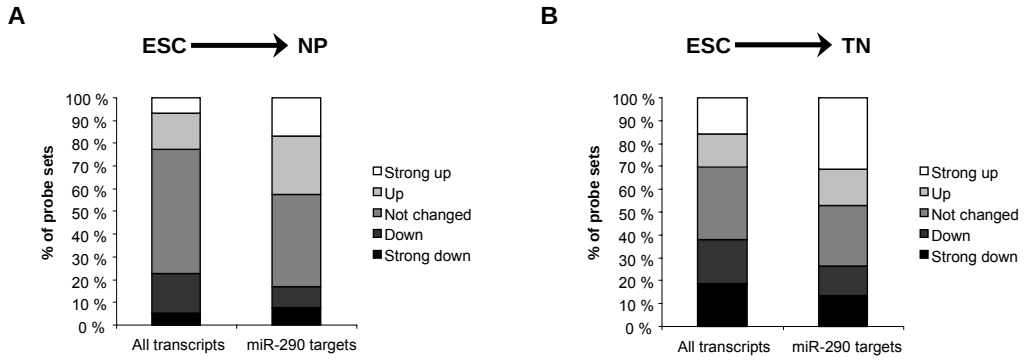


Figure 8. The primary targets of miR-290 cluster miRNAs are enriched for neuronal transcripts.
The reliably detectable probe sets from the microarray analysis of three neuronal differentiation stages (ESC, NP and TN) were divided into five different subgroups based on the change in their signal between (A) ESCs and NPs or (B) ESCs and TNs. The divide into the subgroups for all transcripts and for miR-290 target transcripts was compared. Strong up, > 3-fold upregulation; up, > 1.5-fold upregulation; not changed, < 1.5-fold change to either direction; down, > 1.5-fold downregulation; and strong down, > 3-fold downregulation.

While only 7% and 16% of all probe sets are either strongly upregulated or upregulated between ESCs and NPs, respectively, as many as 17% and 26% of the probe sets detecting miR-290 targets showed similar upregulation (Figure 8A). Also in the comparison of ESCs and TNs, the fraction of probe sets with strong upregulation was twice as big for miR-290 targets as the one for all probe sets (from 16% to 31%) (Figure 8B). Curiously, the fractions of mildly upregulated probes sets were similar. In all cases, the increased fraction of upregulated probe sets in the miR-290 targets was compensated by smaller fractions of mildly downregulated and not-changed probe sets. Thus, consistent with miR-290 mediated repression in undifferentiated ESCs, putative miRNA targets appear to be enriched in transcripts whose expression increases during neuronal differentiation.

3.1.3.3 Conclusions

Taken together, we have shown that the downregulation of miR-290 miRNAs during neuronal differentiation is mediated at the transcriptional level and leads to complete silencing of the pri-miR-290 expression. The silencing is characterized by complete loss of RNA Pol II and H3K4me2 from the locus and accompanied by strong increase in H3K27me3. This leaves open the possibility that pri-miR-290 might be a target of Polycomb mediated silencing in neuronal differentiation. Like many Polycomb target genes, pri-miR-290 promoter accumulated DNA methylation during differentiation, suggesting that *de novo* DNA methylation by DNMT3 enzymes might be necessary for irreversible silencing of expression of miR-290 miRNAs. All the features of pri-miR-290 silencing resemble the silencing of many pluripotency genes like *Oct-4*. This suggests that in order for normal development to take place, like expression of *Oct-4*, the expression of miR-290 miRNAs has to be restricted to early embryonic development. Indeed, the targets of miR-290 miRNAs are enriched among the transcripts upregulated during neuronal differentiation. This further argues that miR-290 miRNAs are important for maintaining the pluripotency of ESCs.

Since pri-miR-290 locus is a target of DNMT3 enzymes while miR-290 miRNAs regulate the expression of DNMT3 enzymes via targeting RBL2, it is concievable that an autoregulatory loop exists between these factors (Figure 9). In this regulation, high

expression of pri-miR-290 would allow high numbers of mature miR-290 miRNAs in ESCs. This in turn would lead to strong downregulation of their primary target RBL2, a transcriptional repressor of *Dnmt3a2/Dnmt3b*, and possibly *Dnmt3L*, expression, thus allowing high expression levels of these enzymes. In this manner, the cells would express sufficient numbers of *de novo* DNMTs to succesfully methylate DNA at their target promoters during initiation of differentiation, and allow complete silencing of many pluripotency genes, including pri-miR-290. This silencing would then eventually lead to upregulation of RBL2 and to RBL2-mediated repression of DNMT3 enzymes, which can be observed during neuronal differentiation (Figure 5B). Similar autoregulatory loops have already been desribed to exist between miRNAs and transcriptional regulators, for example between miR-17-92 cluster and E2F family transcription factors (Sylvestre et al. 2007).

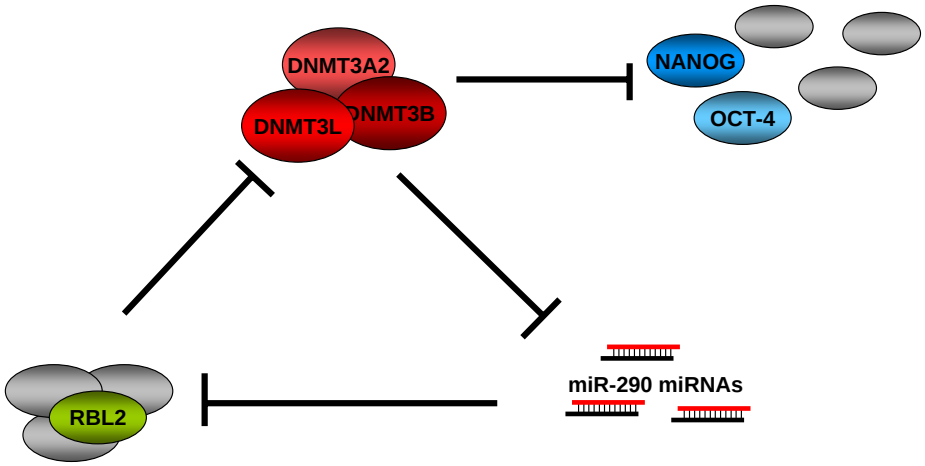


Figure 9. Model for autoregulation between miR-290 miRNAs and DNMT3 enzymes in mouse ESC differentiation.

High expression of miR-290 miRNAs leads to downregulation of RBL2, allowing high expression of DNMT3 enzymes. In this way, sufficient number of DNMT3A/3B/3L complexes are available upon initiation of differentiation to mediate irreversible silencing of pluripotency genes such as *Nanog* and *Oct-4* as well as pri-miR-290. Silencing of pri-miR-290 during differentiation leads to increased expression of RBL2 and, in turn, causes downregulation of DNMT3s.

3.1.3.4 Methods

ESC differentiation

The differentiation was performed as previously described (Bibel et al. 2004; Mohn et al. 2008). In short, ESCs were deprived of feeder cells during 3-4 passages and this was followed by formation of cellular aggregates by 4 x 10^6 cells. The aggregates were then cultivated in non-adherent dishes for 8 days. At day 4 retinoic acid (5 μM) was added and left for the 4 remaining days. Subsequently, the aggregates were dissociated with trypsin and plated (2 x 10^5 cells per cm^2) on cationic substrate coated with laminin. After plating a medium enriched with supplements was added for 10 days of terminal neuronal maturation.

RT-qPCR

Total RNA was extracted with Trizol (Invitrogen) and purified using RNAeasy columns (Qiagen). Thermoscript RT-PCR kit (Invitrogen) was used for the cDNA synthesis reaction with 1 μg template RNA and 250 pmol of oligo(dT)20 primer, incubated for 1 h at 55 °C. Subsequently, cDNA was used as a template for RT-qPCR with the ABI Prism 7000 Sequence Detection System and Platinum SYBR Green qPCR SuperMix, using primers specific for Gapdh and pri-miR-290. Sequences of primers are provided in Table 1. Annealing of all primers was done at 55 °C. Relative expression levels were calculated using the formula 2–(ΔCt), where ΔCt is Ct(gene of interest)–Ct(Gapdh) and Ct is the cycle at which the threshold is crossed.

ChIP

ChIP was performed mainly as previously described (Weber and Schubeler 2007). The ESCs, NPs or TNs were cross-linked in medium containing 1% formaldehyde for 10 min at room temperature, scraped off and rinsed with 10 ml of 1xPBS. Pellets were resuspended in 15 ml of buffer 1 (10 mM Tris (pH 8.0), 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) and twice in 15 ml of buffer 2 (10 mM Tris (pH 8.0), 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl). Following the washes the cells were lysed in 1 ml of lysis buffer (50 mM HEPES/KOH (pH 7.5), 500 mM NaCl, 1 mM EDTA, 1% Triton X-

100, 0.1% DOC, 0.1% SDS, protease inhibitors) and sonicated three times for 15 s (using a Branson sonicator, amplitude 70%). 70 µg of chromatin was incubated overnight at 4 °C with 5 µg of the following antibodies: anti-trimethyl-H3K27 (Upstate, #07-449), anti-dimethyl-H3K4 (Upstate, #07-030), anti-RNA Pol II (Santa Cruz Biotechnology, #SC899). The formed immunocomplexes were incubated for 3 h at 4 °C with 30 µl protein A-Sepharose beads preblocked with tRNA. Beads were washed twice with 1 ml lysis buffer and once with 1 ml DOC buffer (10 mM Tris (pH 8.0), 0.25 M LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA), and bound chromatin was eluted in 1% SDS/0.1 M NaHCO3. After RNase A treatment, cross-linking was reversed by overnight incubation at 65 °C followed by proteinase K digestion. DNA was isolated by phenol/chloroform extraction followed by ethanol precipitation and resuspension in 50 µl TE. A sample of the input chromatin was treated in the same way to generate total input DNA. The purified DNA and the respective input DNA were used as templates for quantitative real-time PCR, using the ABI Prism 7000 Sequence Detection System (Applied Biosystems), Platinum SYBR Green qPCR SuperMix (Invitrogen) and primers specific for the Gapdh and pri-miR-290 promoters. Obtained values were first normalized to the respective input DNA and further to the enrichment of an unrelated intergenic region in the case of H3K27me3. Sequences of primers are listed in Table 1. Annealing of all primers was done at 55 °C.

Bisulfite sequencing

1 µg of genomic DNA extracted from ESCs, NPs, and TNs was bisulfite converted using the Epitect Bisulfite Kit (Qiagen). Three different regions (BS1, BS2, and BS3) of pri-miR-290 locus were amplified by PCR, the PCR products were gel purified, cloned by TOPO-TA cloning (Invitrogen) and sequenced using SP6 reverse sequencing primer. The sequences of primers to amplify converted DNA are listed in Table 1.

Table 1. Primers used for analysis of pri-miR-290 during neuronal differentiation.

Primer pair		Primer sequences (5' to 3')
Primers for RT-qPCR		
Gapdh	forward	CCATCACCATCTTCCAGG
	reverse	CCTGCTTCACCACCTTCTTG
pri-miR-290	forward	CCACGTGTTCCGGGTTAACT
	reverse	ACCGTCTACTGGGCAGGATG
Primers for ChIP		
Gapdh promoter	forward	TCCCCTCCCCCTATCAGTTC
	reverse	TTGGACCCGCCTCATTTTT
pri-miR-290 proximal	forward	AGCAGCCCAGTTTGACCATC
	reverse	CTGGAGCAGAGGCTATCCCA
pri-miR-290 distal	forward	AGTGAAGGTCACTCTGCCGC
	reverse	AGAGACCAGCATTCCCGATG
Primers for bisulphite sequencing		
pri-miR-290 BS1	forward	AACAAAAGAAAAACAGCCGGGCATGGTG
	reverse	TCAAATCCTCCCTCTTTTTTACCTT
pri-miR-290 BS2	forward	CCTAGTCACCATAGTAGACCAAGCTGGC
	reverse	CTGGAGGCAGAGAGGCAGGCAGAAA
pri-miR-290 BS3	forward	CCAGGCTGGCCTTTAACTCTCATTTCTGC
	reverse	CTCCAACCTGAAGGAAACCTGGATG

3.1.3.5 References

Bibel, M., J. Richter, et al. (2004). "Differentiation of mouse embryonic stem cells into a defined neuronal lineage." Nat Neurosci **7**(9): 1003-9.

Chen, T., Y. Ueda, et al. (2002). "A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation." J Biol Chem 277(41): 38746-54.

Houbaviy, H. B., L. Dennis, et al. (2005). "Characterization of a highly variable eutherian microRNA gene." Rna **11**(8): 1245-57.

Houbaviy, H. B., M. F. Murray, et al. (2003). "Embryonic stem cell-specific MicroRNAs." Dev Cell **5**(2): 351-8.

Mohn, F., M. Weber, et al. (2008). "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." Mol Cell **30**(6): 755-66.

Sylvestre, Y., V. De Guire, et al. (2007). "An E2F/miR-20a autoregulatory feedback loop." J Biol Chem **282**(4): 2135-43

Weber, M. and D. Schubeler (2007). "Genomic patterns of DNA methylation: targets and function of an epigenetic mark." Curr Opin Cell Biol **19**(3): 273-80.

3.2. Intact RNA silencing machinery is necessary for proper 3' end processing of replication-dependent histone mRNAs.

Lasse Sinkkonen, Caroline Artus-Revel, Petr Svoboda and Witold Filipowicz

3.2.1 Aim of the project

Accumulating evidence suggests that miRNAs are involved in regulation of number of pathways and processes in mammalian cells. Characterization of these regulatory pathways is of great interest. Here we aimed to identify novel processes under miRNA-mediated regulation by taking advantage of depletion of miRNAs from human HEK293 cells. miRNAs were depleted by inducible knock-down (KD) of different RNA silencing pathway components, namely Dicer and the four human Argonaute proteins. miRNAs are known to exhibit much of their regulation via mRNA degradation in HEK293 cells, more so than for example in HeLa cells (Schmitter et al. 2006). Thus, HEK293 cells are a good model system to identify miRNA targets through analysis of transcriptome changes upon depletion of miRNAs. By detailed analysis of the genome-wide transcriptome changes after miRNA depletion we were able to identify 3' end processing of histone mRNAs as a process under miRNA-mediated regulation.

3.2.2 Results and discussion

Previously, stable human HEK293 cell lines suitable for inducible depletion of Dicer or any of the four Argonautes were generated in our laboratory (Schmitter et al. 2006). Plasmids expressing short hairpin RNAs (shRNAs) were cloned and stably integrated into the genome of 293T-REx cell line, which expresses a Tet-repressor. Tet-repressor binds to the promoter of the shRNA and represses its expression. Treatment of these cells with tetracyclin or its analog doxycyclin leads to relief of repression by Tet-repressor and allows expression of the shRNA. The shRNA enters the RNAi pathway and induces KD of the targeted gene. RNA isolated from cell lines with depletion of either Dicer or one of the Argonaute proteins was used for microarray analysis in order to identify transcripts that are regulated by these components of the RNA silencing pathway. The results from these experiments suggest that most transcriptomic changes upon loss of Dicer and AGO2 are due to depletion of miRNAs (Schmitter et al. 2006).

A detailed analysis of the microarray data revealed that many of the human replication-dependent histone genes are upregulated upon loss of Dicer in 293T-REx cells (Figure 10). There are 61 probe sets on Affymetrix Human U133 2.0 Plus arrays

that monitor the levels of 54 different histone genes (Table 2). 17 of these probe sets (representing the expression levels of 14 histone genes), showed a more than 1.5-fold increase in hybridization signal in both of the tested Dicer-KD cell lines (2-2 and 2b2) after 6 days of shRNA induction (Figure 10). For many of the probe sets the increase could be observed already 2 days after the Dicer-KD. Likewise, many of the 17 probe sets showed an increased signal after 2 days of AGO2 or AGO3 KD when compared to the signal in similarly treated control cell lines. The control 293T-REx cell lines express either a scrambled hairpin RNA (293T-REx controls 2&3; Figure 10) or have an integration of an empty plasmid (293T-REx controls 1&4; Figure 10). In contrast, loss of AGO1 or AGO4 seemed to have no effect on the expression of histone genes.

The apparent upregulation of histone genes could have several causes. Since expression of histone genes is fluctuating during cell cycle, the upregulation could be a result of a prolonged S-phase (Harris et al. 1991). Hence, we analyzed the cell cycle profile of the cells after 6 days of Dicer-KD. As shown in Figure 11A, after 6 days of tetracycline treatment the fraction of the cells in S-phase was around 13% in both

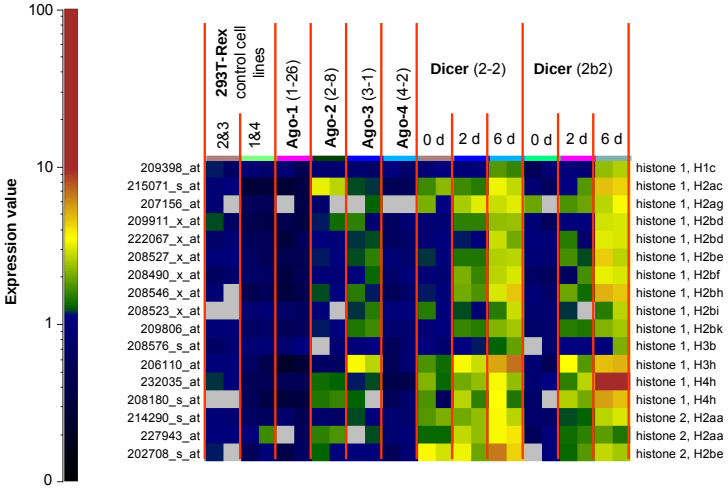


Figure 10. Microarray analysis of replication-dependent histone genes after depletion of Dicer or Argonaute proteins in 293-T-REx cells.

The raw expression values of the 17 probe sets monitoring a more than 1.5-fold upregulation of histone mRNAs in two different Dicer-KD cell lines after 6 days of tetracycline induction are illustrated as a heat map. The numerical values corresponding to the different colours are indicated by the bar on the left. Grey color indicates that the microarray probe signal was evaluated as “absent”. The names of the probes sets, the respective genes as well as tested cell lines are shown. Dicer-KD cell lines were induced for the indicated time points. All other cell lines were induced for 2 days.

control and Dicer-KD cells. This result is consistent with the fact that only a subpopulation of histone genes is upregulated. In case the upregulation would be due to prolonged S-phase all rather than only some of histone genes would be expected to be upregulated. Hence, the observed upregulation is likely to be caused by a more specific type of misregulation.

In order to confirm the microarray results, we performed RT-qPCR analysis of RNA extracted from Dicer-KD cells over a time course of 1 to 9 days of induction (Figure 11B). We tested five histone genes upregulated on microarrays (one for each histone class: H1, H2A, H2B, H3 and H4), using gene specific primer pairs. In addition, we tested the expression levels of HIST1H2AB and HIST1H4I, two histone genes not affected in the microarray experiments. The levels of upregulated histone mRNAs started to increase already a few days after KD induction and continued to accumulate to as high as 10-fold increase after 9 days. Interestingly, also HIST1H4I started to show some accumulation by 9 days of Dicer-KD, suggesting that most histone mRNAs might be affected by the loss of Dicer, although many of them to a lesser extent. Still, HIST1H2AB remained unchanged. Importantly, transfection of the Dicer-KD cells after 7 days of induction with a Dicer construct modified to escape the repression by Dicer-specific shRNA was sufficient to rescue the effect of the KD on increase in histone mRNAs by 9 days. This suggests that the effect is specific for loss of Dicer and is not caused simply by expression of the shRNA or its possible off-target effect.

Similar RT-qPCR analysis of the AGO2-KD cells was able to confirm the upregulation of the histone mRNAs also in the absence of AGO2 (Figure 11C). Consistent with AGO2 acting downstream of Dicer in the RNA silencing pathway, the effect of loss of AGO2 on histone mRNAs was more robust and rapid that in the case of Dicer-KD. By 2 days the upregulation could be clearly observed and by 4 days expression of HIST1H2BK mRNA had already increased over 10-fold. Curiously, also HIST1H2AB became upregulated upon loss of AGO2.

Tetracyclin inducible KD of both Dicer and AGO2 leads to upregulation of many of the replication-dependent histone genes in HEK293 cells. To confirm, that this increase in histone mRNAs takes place also in other cell types and upon tetracyclin-

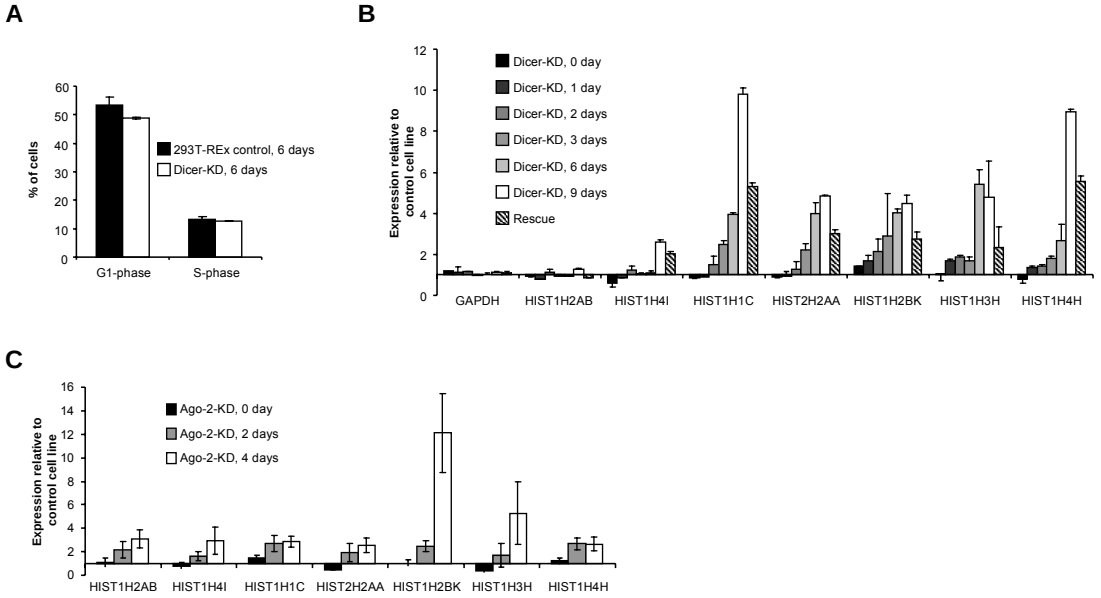


Figure 11. Histone mRNAs accumulate after loss of Dicer and AGO2 without an effect on the cell cycle.

(A) Fluoresence-assisted cell sorting (FACS) analysis for percentage of cells in G1- or S-phase of the cell cycle after 6 days tetracycline treatment of 293T-REx control cells and Dicer-KD cells. Mean values (±SD) from two independent FACS analysis are shown. (B and C) Expression of indicated transcripts in (B) Dicer-KD cells or (C) AGO2-KD cells relative to the 293T-REx control cells during a time course of tetracycline treatment. For the rescue experiment in panel B the Dicer-KD cells were induced for 7 days, transfected by a modified Dicer expression construct and induced for further 2 days. The values were normalized to GAPDH and represent mean (± standard error of mean, SEM) of two to five independent experiments.

independent Dicer-KD, we transiently transfected HeLa cells with the shDicer-expressing plasmid. Already 1 day after transfection the mRNA and protein levels of Dicer were reduced and by three days no Dicer protein could be detected (Figure 12A and B). Although at day 1 there was no significant misregulation of histone mRNAs, already by 3 days the expression of all five histones genes of interest was strongly upregulated (Figure 12C). In fact, the upregulation was stronger than in the stable Dicer-KD cells, possibly due to more efficient KD of Dicer by several copies of shDicer plasmid, enhanced in addition by lower starting levels of Dicer in HeLa than in 293T-REx cells (Su et al. 2002 and our unpublished results). Unexpectedly, upregulation of HIST1H3H was somewhat

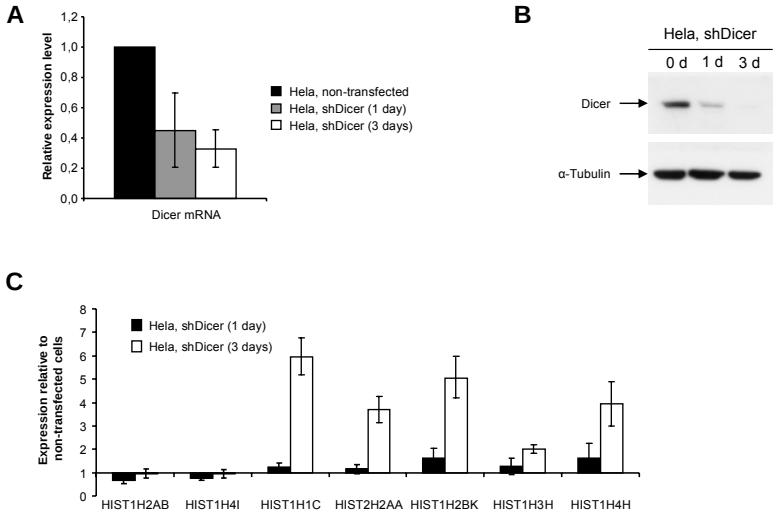


Figure 12. Histone mRNAs are upregulated upon transient KD of Dicer in HeLa cells.

(A) The Dicer mRNA level in 1 and 3 days shDicer transfected HeLa cells relative to non-transfected cells. (B) Western blot analysis of Dicer levels in HeLa cells before and after transfection of shDicer construct. α-Tubulin was used as a loading control. (C) Expression of indicated transcripts in 1 and 3 days shDicer transfected HeLa cells relative to non-transfected cells. In panels A and C the values were normalized to GAPDH and represent mean (±SEM) of three to four independent experiments.

weaker in HeLa than in the stably-transformed Dicer-KD cells when compared to the upregulation of other histone mRNAs. Again, HIST1H2AB and HIST1H4I were not affected. Taken together, these results further support the idea that the observed upregulation of the replication-dependent histone genes is specific for loss of Dicer and takes place independently of the used cell type.

The microarray experiments as well as the abovementioned RT-qPCR experiments were performed using cDNA, synthesis of which was primed by oligo-dT. Thus, the results reflect the changes of polyadenylated transcripts and not necessarily of the levels of histone mRNAs physiologically 3' processed at the 3'-terminal stem-loop. Since replication-dependent histone genes are usually following this mode of the 3' end processing, our results might not apply to the total population of the transcribed histone mRNAs. In order to test this, we performed RT-qPCR analysis of some of the histone mRNAs using random hexamer priming during the cDNA synthesis. Oligo-dT and random hexamer primed RT-qPCR analyses of Dicer-KD cells induced for 6 days revealed that the upregulation of HIST1H1C and HIST1H3H is clearly more robust in the

oligo-dT primed samples (Figure 13A). This result indicates that, instead of an increase in total populations of histone mRNAs, loss of Dicer is causing enhanced polyadenylation of the mRNAs, possibly due to affected 3' end processing at the histone stem-loop. A notable exception in Figure 13A is HIST1H2BK, which is robustly upregulated in both conditions. This might reflect a normally poor processing of HIST1H2BK by U7 and SLBP, leading to a low stability of the S-phase specific HIST1H2BK mRNA. And as the polyadenylation would become a more favored processing mode, the total level of HIST1H2BK mRNA would increase. This is probable since *HIST1H2BK* is one of the only three human replication-depent histone genes that have introns, implying that it might be processed by polyadenylation rather than at the histone stem-loop structure.

To confirm that the shift from production of non-polyadenylated to polyadenylated histone mRNAs can indeed be detected as an upregulation in oligo-dT primed RT-qPCR experiments, we disrupted the normal histone 3' end processing in HeLa cells by depleting SLBP. This was done by transfecting a pool of siRNAs against SLBP into HeLa cells and collecting RNA from these cells 1 and 3 days after the original transfection. The siRNA transfection led to a robust downregulation of SLBP mRNA, which was accompanied by extremely strong increase in polyadenylated histone transcripts (Figure 13B and C). The robustness of the increased polyadenylation upon SLBP-KD is likely to be due to complete failure of these cells to process new histone mRNAs in a canonical way at the histone stem-loop structure. This may lead to inadequate production of histones, what is supported by the observation that all SLBP-KD cells die already soon after 3 days of KD. Interestingly, also upon loss of SLBP, HIST1H3H is affected to a lesser extent than the other tested histone mRNAs in HeLa cells.

Consistent with our observations, recent results from Narita et al. show that disruption of normal histone 3' end processing by KD of NELF-E or CBP80, and the subsequent enhancement of histone mRNA polyadenylation, can be observed bymicroarrays as a more intense signal from probe sets monitoring histone mRNAs (Narita et al. 2007). Comparison of their microarray data to ours reveals that 11 of 12

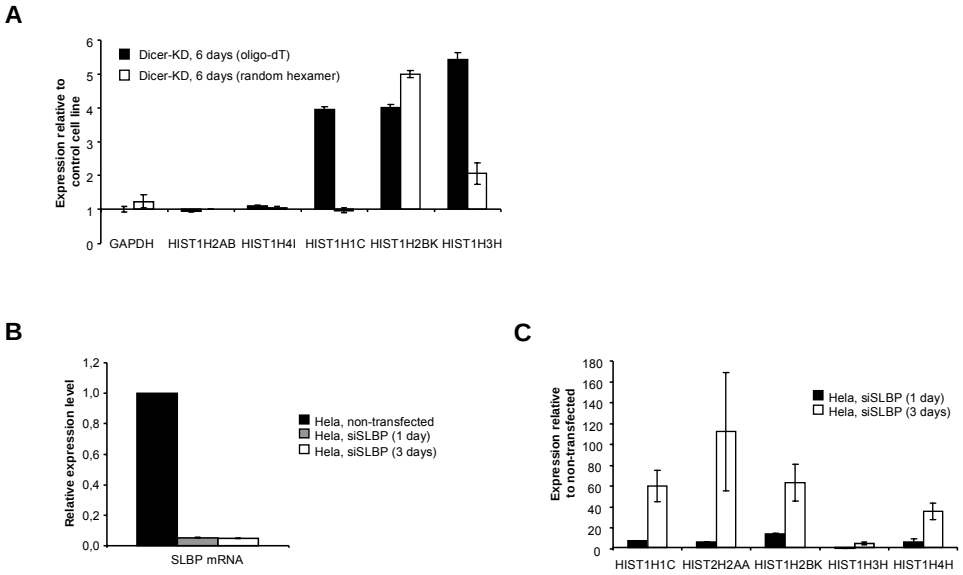


Figure 13. The KD of Dicer or SLBP leads to increase in polyadenylated histone mRNAs.

(A) The observed upregulation of histone mRNAs upon loss of Dicer is more robust after oligo-dT priming in cDNA synthesis. Expression of indicated transcripts in Dicer-KD cells relative to the 293T-REx control cells using either olido-dT or random hexamer primed cDNA after 6 days of tetracycline treatment. (B) The SLBP mRNA level in 1 and 3 days siSLBP transfected HeLa cells relative to non-transfected cells. (C) Expression of indicated transcripts in 1 and 3 days siSLBP transfected HeLa cells relative to non-transfected cells. In all panels the values were normalized to GAPDH and represent mean (±SEM) of two independent experiments.

histone mRNAs "upregulated" after KD of both NELF-E and CBP80 are also "upregulated" upon loss of Dicer (data not shown).

To find out how general the polyadenylation of replication-dependent histone mRNAs is, we studied the publicly available expression data from Genomics Institute of Novartis Research Foundation (GNF) (http://symatlas.gnf.org/SymAtlas/) (Su et al. 2002). We analyzed which of the 61 GNF probe sets detecting human histone mRNAs show reasonably high expression values (raw expression value of more than 100) across a panel of 9 tested human cell lines (Table 2). Remarkably, all but one of the probe sets reporting increased levels after Dicer-KD also show high basal expression level in these cell lines. Most other histone probe sets reported low signals. This suggests that transcription of these genes also results in formation of polyadenylated transcripts, in addition to the transcripts normally processed at the histone stem-loop.

Table 2. Human replication-dependent histones detected by Affymetrix Human U133 Plus 2.0 arrays.

Gene name	Probe set	High signal†	≥1,5x up in Dicer-KD	poly-A cloned*	Gene name	Probe set	High signal†	≥1,5x up in Dicer-KD	poly-A cloned*
HIST1H1A	208484_at	N	N	N	HIST1H3B	208576_s_at	N	Y	N
HIST1H1B	214534_at	N	N	N	HIST1H3C	208577_at	N	N	N
HIST1H1C	209398_at	Y	Y	Y	HIST1H3D	214472_at	Y	N	Y
HIST1H1D	214537_at	N	N	N		214522_x_at	N	N	Y
HIST1H1E	208553_at	N	N	N		214472_at	Y	N	Y
HIST1H1T	207982_at	N	N	N		214522_x_at	N	N	N
HIST1H2AB	208569_at	N	N	N	HIST1H3E	214616_at	N	N	N
HIST1H2AC	215071_s_at	Y	Y	Y	HIST1H3F	208506_at	N	N	N
HIST1H2AE	214469_at	N	N	N		208506_at	N	N	N
HIST1H2AG	207156_at	Y	Y	Y	HIST1H3G	208496_x_at	Y	N	Y
HIST1H2AI	214542_x_at	N	N	N	HIST1H3I	214509_at	N	N	N
	206110_at	N	N	N	HIST1H3J	214646_at	N	N	N
	206110_at	N	N	N	HIST1H4A	208046_at	N	N	N
HIST1H2AJ	208583_x_at	N	N	N	HIST1H4B	214516_at	N	N	N
HIST1H2AK	214644_at	N	N	N	HIST1H4C	205967_at	Y	N	Y
	214644_at	N	N	N	HIST1H4D	208076_at	N	N	N
HIST1H2AL	214554_at	N	N	N	HIST1H4E	206951_at	N	N	Y
HIST1H2AM	214481_at	Y	N	Y	HIST1H4F	208026_at	N	N	N
HIST1H2BB	208547_at	N	N	N	HIST1H4G	208551_at	N	N	N
HIST1H2BC	214455_at	N	N	N	HIST1H4H	208180_s_at	Y	Y	Y
HIST1H2BD	209911_x_at	Y	Y	Y		208181_at	N	Y	Y
	222067_x_at	Y	Y	Y	HIST1H4I	214634_at	N	N	Y
	222067_x_at	Y	Y	Y	HIST1H4J	214463_x_at	N	N	N
HIST1H2BE	208527_x_at	Y	Y	N		208580_x_at	N	N	N
HIST1H2BF	208490_x_at	Y	Y	N		214463_x_at	N	N	N
	208490_x_at	Y	Y	N		208580_x_at	N	N	N
HIST1H2BG	215779_s_at	Y	N	Y		214463_x_at	N	N	N
	210387_at	Y	N	Y		208580_x_at	N	N	N
HIST1H2BH	208546_x_at	Y	Y	N	HIST1H4L	214562_at	N	N	N
HIST1H2BI	208523_x_at	Y	Y	N	HIST2H2AA	214290_s_at	Y	Y	N/A
HIST1H2BJ	214502_at	N	N	Y		218279_s_at	N	N	N/A
	214502_at	N	N	Y		218280_x_at	Y	N	N/A
HIST1H2BK	209806_at	Y	Y	Y	HIST2H2BE	202708_s_at	Y	Y	Y
HIST1H2BL	207611_at	N	N	N	HIST2H4	207046_at	N	N	N/A
HIST1H2BM	208515_at	N	N	N	HIST3H2A	221582_at	Y	N	Y
HIST1H2BN	207226_at	N	N	Y	HIST3H3	208572_at	N	N	N
HIST1H2BO	214540_at	N	N	N					
HIST1H3A	208575_at	N	N	N					

† = whether the respective probe set shows in average a raw expression signal of above 100 across a panel of 9 tested cell lines (HEK293, HEK293T, 293T-REx, HeLa, HepG2, Huh-7, Jurkat, K562 and MCF-7) (Su et al. 2002).

* = whether a longer, polyadenylated variant of the gene has been cloned by the Mammalian Gene Collection (MGC) of National Institute of Health (NIH) (Strausberg et al. 2002).

Y = yes

N = no

N/A = not available

To further address this possibility, we studied whether any longer and polyadenylated transcripts of the human histone genes had been cloned and sequenced by the Mammalian Gene Collection (MGC) consortium of the National Institute of Health (NIH) (http://mgc.nci.nih.gov/) (Strausberg et al. 2002). As shown in Table 2, for 7 out of the 14 upregulated histone mRNAs, a polyadenylated variant has indeed been identified. For all detectable histone mRNAs the same numbers are 18 out of 51. Taken together, these observations indicate that under normal conditions a subpopulation of replication-dependent histone genes is giving rise to low levels of polyadenylated transcripts, in addition to their normal mature mRNAs, processed at the 3'-terminal stem-loop. Disruption of the normal histone mRNA processing, for example by loss of SLBP, leads to increased production of mainly longer, polyadenylated transcripts. This is manifested in increased mRNA levels in assays relying on oligo-dT priming.

To investigate whether the normal processing of histone mRNAs is in fact disrupted by the loss of intact RNA silencing pathway, we focused our analysis on one of the candidate histones, HIST1H3H. Under normal conditions *HIST1H3H* is transcribed into 473-nt long mature mRNA that is cleaved 5 nts after the stem-loop (Figure 14A). In addition, a 1253-nt long polyadenylated transcript arising from the same gene has been cloned. To test whether the proportion of the longer HIST1H3H transcripts increases upon Dicer-KD, we designed several primer pairs to monitor different regions of the gene. The first primer pair (H3H-1) detects all HIST1H3H transcripts due to its location in the CDS; the second pair (H3H-2) is flanking the normal processing site and detects only the misprocessed, polyadenylated transcripts; the third pair (H3H-3) is located several hundred nts downstream of the mature mRNA but still within the longer poly-A+ transcript; while the fourth (H3H-4) is located beyond the polyadenylation signals of the longer transcript. These primer pairs were used in random hexamer primed RT-qPCR experiments to detect the relative expression levels of the different length transcripts following 6 or 9 day KD of Dicer (Figure 14B). As measured by the primer pairs H3H-2 and H3H-3, in 293T-Rex control cells as well as non-induced Dicer-KD cells the longer transcripts amount to about 10% of the total HIST1H3H mRNA population (level of which is measured by H3H-1). Consistent with being located downstream of either mature mRNAs, the H3H-4 primer pair showed a 100-fold lower signal than H3H-1.

Still, H3H-4 was amplifying a specific product, likely representing the occasional read-through products of RNA Pol II. Upon loss of Dicer, after 6 or 9 days of tetracycline treatment, the proportion of longer HIST1H3H mRNA had increased to more than 60% of total HIST1H3H. Thus, a significant shift in the predominant 3' end formation mechanism had taken place. Interestingly, also amount of transcript detected by H3H-4 primer pair increased upon loss of Dicer. This is likely due to the increased RNA Pol II read-through upon repressed production at the histone stem-loop.

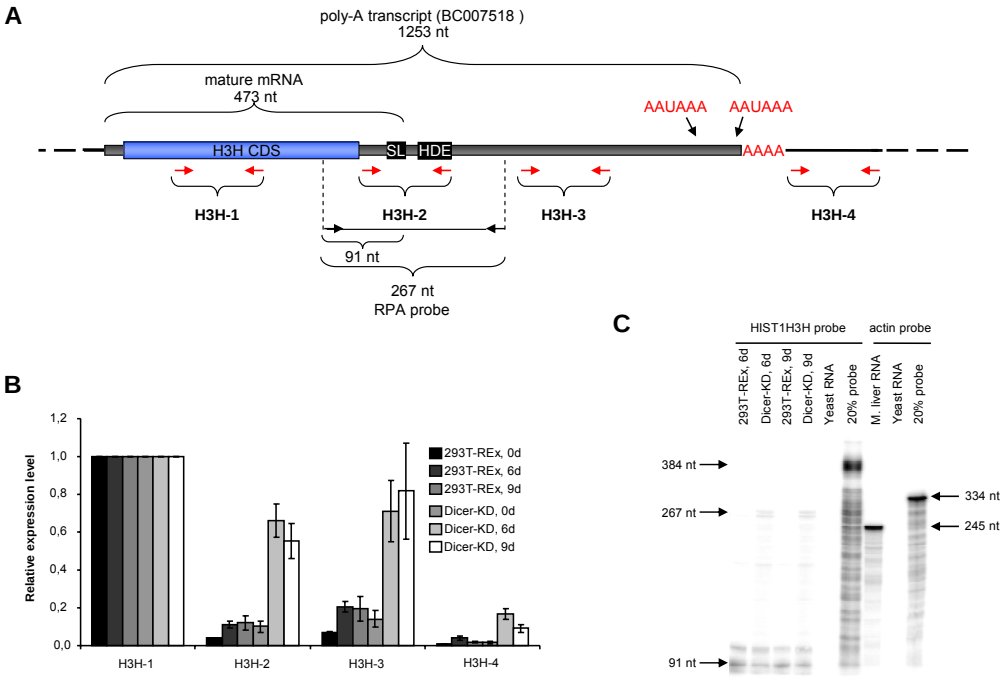


Figure 14. Normal 3' end processing of HIST1H3H disrupted upon loss of Dicer.

(A) Schematic structure of the *HIST1H3H* locus depicting the relative location of HIST1H3H CDS, the stem-loop structure (SL), the histone downstream element (HDE) and polyadenylation signals (AAUAAA) downstream of the mature mRNA. Also location of primer pairs and the RPA probe used are indicated. (B) Expression of different length HIST1H3H transcripts relative to the total population of HIST1H3H transcripts in 293T-Rex control cells and in Dicer-KD cells at different time points as measured by random hexamer primed RT-qPCR. The values were normalized to GAPDH and represent mean (±SEM) of two to three independent experiments. (C) RPA analysis of HIST1H3H processing 6 and 9 days after tetracycline treatment of 293T-REx control cells and Dicer-KD cells. The used probes and the source of RNA hybridized with them are indicated. No intact probe can be detected in the negative control lane where the probe was incubated with yeast RNA. 20% of the non-RNase treated free probe was loaded. As a positive control, β-actin probe together with mouse total liver RNA was used.

Intriguingly, analysis of inducible TRBP-KD cell lines (TRBP is a co-factor of Dicer required for pre-miRNA processing) demonstrated a similar shift towards longer HIST1H3H transcripts as observed for Dicer-KD cell lines (data not shown; Haase et al 2005). Thus, TRBP seems to contribute to the function of Dicer in regulation of histone 3' end processing.

As an alternative approach to verify the increase in longer poly-A+ transcripts we used RNase protection assay (RPA). A 384-nt radioactively labeled probe should protect a 267-nt long fragment diagnostic of poly-A+ HIST1H3H mRNA (Figure 14A) and 91 nt fragment diagnostic of normal histone mRNA. A probe detecting the mouse β-actin mRNA was used as a positive control together with mouse liver RNA. When the RPA was performed with the HIST1H3H probe and RNA from 6 and 9 day tetracycline induced 293T-Rex control cells, only short, 90-nt fragments could be detected (Figure 14C). This result further argues that most of HIST1H3H mRNA is normally processed at the stem-loop structure in these cells. When the same probe was incubated with RNA from 6 and 9 day induced Dicer-KD cells, the 90-nt fragments were also detectable. But, in addition, longer, ~267-nt fragment was also detectable, demonstrating that a significant misregulation of the HIST1H3H processing was taking place upon loss of Dicer. Moreover, lower levels of intermediate size RNA fragments were visible in the Dicer-KD lanes, likely reflecting hybridization of degraded fragments of the probe to the longer HIST1H3H transcript. This data further confirms that KD of Dicer leads to disruption of normal histone 3' end processing and production of longer HIST1H3H transcripts.

Finally, to see whether regulation of histone mRNA processing by the RNA silencing pathway is conserved to other species, we analyzed expression of histone mRNAs in mouse *Dicer*$^{-/-}$ embryonic stem cells (ESCs). First, we analyzed the expression profiles of histone mRNAs in *Dicer*$^{+/-}$ and *Dicer* $^{-/-}$ ESCs, based on the microarray data described in chapter 3.1. On Affymetrix Mouse MOE430 v2.0 GeneChips, there are 34 probe sets monitoring expression of only 20 different replication-dependent histone mRNAs. Of these, 5 probe sets, representing 3 histone mRNAs, showed clear upregulation of more than 1.5-fold (Figure 15A). The upregulation of all three mRNAs, Hist1h1c, Hist1h2bc and Hist1h2bp, could be confirmed by RT-qPCR after oligo-dT primed cDNA synthesis (Figure 15B). Thus, RNA silencing

machinery appears to be necessary for histone 3' end processing in both human and mouse cells.

Still, the number of histone genes affected, as well as extent of upregulation, are smaller in mouse ESCs than in the tested human cell lines. There could be several reasons for this. First, although conserved, the 3' end processing of histone mRNAs in mouse and human cells might differ in some aspects. Second, the analyzed mouse cells are ESCs that are special in many ways e.g. they have an altered cell cycle profile and practically move from mitosis directly to a new S-phase, omitting most of the G1-phase (White and Dalton 2005). Third, while in human cells the depletion of Dicer takes place in a rapid, inducible manner, the *Dicer*$^{-/-}$ ESCs have been cultured for many passages since the removal of Dicer and have possibly acquired additional mutations to compensate for their initial defects.

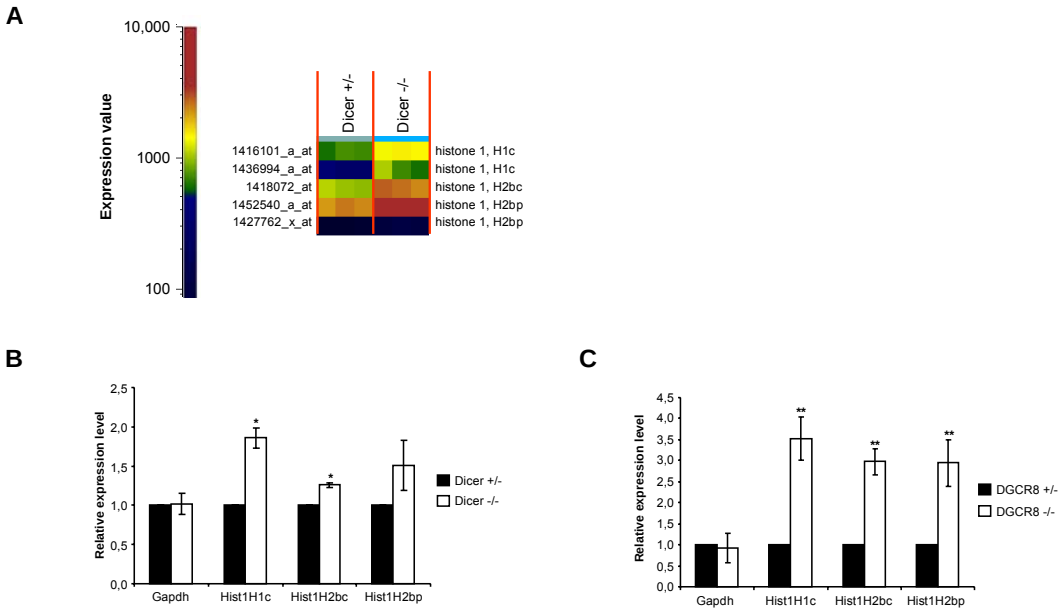


Figure 15. Histone mRNAs are upregulated in miRNA deficient mouse ESCs.

(A) The raw expression values of the 5 probe sets monitoring a more than 1.5-fold upregulation of histone mRNAs in mouse *Dicer*$^{-/-}$ ESCs are illustrated as a heat map. The numerical values corresponding to the different colours are indicated by the bar on the left. The names of the probes sets, the respective genes as well as tested cells are shown. (B and C) Expression of indicated transcripts in (B) *Dicer*$^{-/-}$ ESCs and (C) *Dgcr8*$^{-/-}$ ESCs relative to the respective heterozygous ESCs as measured by RT-qPCR. The values were normalized to Oct-4 and represent mean (±SEM) of three to five independent experiments. Expression value in heterozygous cells was set to 1. p-values calculated by two-tailed Student's t-test are: * ≤ 0.05; ** ≤ 0.01.

Based on our results, Dicer, AGO2, AGO3, and TRBP appear to contribute to the correct balance between different 3' end processing pathways of histone mRNAs and, at least for Dicer, this regulation is conserved from human to mouse. Yet, it is not clear how this regulation takes place. One possibility is that these proteins, or other proteins functioning in a complex with them, could play a direct role in histone mRNA processing. Another, perhaps more likely explanation, is that small RNAs generated by Dicer with the help of TRBP are loaded on AGO2 and AGO3, and are capable of regulating factors involved in histone 3' end formation (for example SLBP or U7). So far we were unable to identify any genes involved in the histone mRNA processing to be misregulated at the mRNA level in Dicer-depleted cells (data not shown). But, we have not yet ruled out their possible misregulation at the protein level.

To address whether the increased histone mRNA polyadenylation depends on miRNAs or some other Dicer products such as endogenous siRNAs, we analyzed the expression of polyadenylated histone mRNAs in mouse *Dgcr8*$^{-/-}$ ESCs (Wang et al. 2007). DGCR8 is a co-factor Drosha and is needed for processing of pri-miRNAs, and thus, for biogenesis of mature miRNAs. The RNA from *Dgcr8*$^{+/-}$ and *Dgcr8*$^{-/-}$ ESCs (kindly provided by Dr. R. Blelloch) was analyzed by RT-qPCR after the oligo-dT primed cDNA synthesis for expression levels of Hist1h1c, Hist1h2bc and Hist1h2bp (Figure 15C). As shown in Figure 15C, all three mRNAs were significantly upregulated upon loss of DGCR8, indicating that the proper histone 3' end formation depends on expression of mature miRNAs.

3.2.3 Conclusions

Taken together, these results demonstrate that the presence of the RNA silencing machinery is required for normal 3' end processing of many replication-dependent histone genes in mouse and human cells. Most likely, this regulation is mediated by conserved miRNAs that are present in all proliferating cells. Possibly, these miRNAs regulate one or several components of the 3' end formation complexes and, in this way, contribute to the decision between different 3' end processing pathways. Identification of

these miRNAs and their downstream targets would be of great interest and might reveal so far unknown aspects of the molecular decision making that takes place during transcription termination.

3.2.4 Methods

Cell culture and transfections of plasmids and siRNAs

293T-REx and HeLa cell lines were maintained in DMEM supplemented with 10% (w/v) FCS (fetal calf serum). In addition, to maintain the expression of Tet-repressor, 293T-REx cells were cultured in the presence of 5 μg/ml blasticidin. Stable KD cell lines were cultured in presence of 200 μg/ml zeocin to maintain shRNA construct. For KD induction the cells were treated with 10 μg/ml tetracycline. The Dicer heterozygous (+/–; line D4) and Dicer-deficient (–/–; line 27H10) ESCs (kindly provided by G. Hannon, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA) were maintained on gelatin-coated plates in DMEM supplemented with 15% (w/v) FCS, sodium pyruvate, β-mercaptoethanol, nonessential amino acids and LIF. Transfections of 293T-REx and HeLa cells were performed using Lipofectamine 2000 (Invitrogen). shDicer construct was generated by Kaifu Tang (Schmitter et al. 2006). Point mutated Dicer expression construct able to avoid shDicer was generated by Astrid Haase (Haase et al. 2005). SLBP-KD was performed by transcfecting HeLa cells with 10 nM siGENOME smartPOOL siRNAs against SLBP (Dharmacon).

FACS analysis

For 30 min prior to FACS analysis, the control and KD cells were cultured in the presence of 10 μg/ml Hoechst33342 dye. Subsequenctly the cells were collected by trypsinization to PBS with 3% FCS, filtered (0.22 μm) and analyzed using MoFlow cell sorter (Dako Cytomation).

RT-qPCR

Total RNA from different cell lines was extracted using the Absolutely RNA Miniprep Kit (Stratagene). Total RNA from *Dgcr8*$^{+/-}$ and *Dgcr8*$^{-/-}$ ESCs was kindly provided by R. Blelloch (University of California San Franciso, San Franciso, USA). AThermoscript RT-PCR kit (Invitrogen) was used for the cDNA synthesis reaction with 1 μg template RNA and 250 pmol of oligo(dT)20 primer or 50 ng of random hexamers, incubated for 1 h at 55 °C. Subsequently, cDNA was used as a template for RT-qPCR with the ABI Prism 7000 Sequence Detection System and Platinum SYBR Green qPCR SuperMix,

using gene-specific primers. Sequences of primers are provided in Table 3. Annealing of all primers was done at 55 °C. Relative expression levels were calculated using the formula 2–(ΔCt), where ΔCt is Ct(gene of interest)–Ct(control gene) and Ct is the cycle at which the threshold is crossed.

Western blotting

Cells were lysed in lysis buffer (30 mM Tris-HCl, pH 7.5, containing 150 mM NaCl, 1 mM MgCl2, 0.5% (v/v) Nonidet P-40, 1 mM DTT and protease inhibitors) and kept on ice for 10 min. Equal amounts of the lysed proteins were separated on polyacrylamide-SDS gels, blotted on polyvinylidene fluoride membrane and probed with primary antibodies against α-Tubulin (5.2.1 Sigma, 1:10,000) and Dicer (D349, 1:5,000). This was followed by incubation with secondary horseradish peroxidase–coupled antibodies. Detection was performed with ECL kit (Amersham).

RPA

The sequence for HIST1H3H probe was amplified from genomic DNA of 293T-REx cells using specific primers. The primer sequences are provided in Table 3. The amplified PCR product was gel purified and cloned into pCRII-TOPO vector downstream of a SP6 promoter (Invitrogen). The cloned plasmid was used as a template for *in vitro* transcription of the radiolabeled probe together with 10 mCi/ml of [α-^{32}P]UTP. Transcription was done using SP6 MAXIscript Kit for 1 h at 37 °C (Ambion). Full length radiolabeled probe was gel purified and 40,000 cpm of the probe was used for hybridization overnight at 42 °C together with 10 μg of total RNA from 293T-Rex or Dicer-KD cells using reagents from and according to RPAIII Kit (Ambion). As a negative control yeast RNA was used. In parallel a β-actin probe provided by the kit was transcribed and used as a positive control in hybridization to mouse liver RNA (also provided). After the hybridization, non-hybridized, single-stranded RNA was digested with RNaseA/RNase T1 Mix for 30 min at 37 °C. Remaining RNA was ethanol precipitated and run on a 6% denaturating polyacrylamide gel together with 20% of non-digested free probe. Detection was done using Storm 860 Phosphoimager (Fuji) after overnight exposure.

Table 3. Primers used in the study.

Primer pair		Primer sequences (5' to 3')
Primers for human genes		
GAPDH-1	forward	CGCTCTCTGCTCCTCCTGTT
	reverse	CCATGGTGTCTGAGCGATGT
GAPDH-2	forward	TCAAGAAGGTGGTGAAGCAGG
	reverse	TTCCTCTTGTGCTCTTGCTGG
HIST1H2AB	forward	ACTCGGTCTTCTCGTGCAGG
	reverse	CGCTCGGAGTAGTTGCCTTT
HIST1H4I	forward	GGTATCACCAAGCCAGCCAT
	reverse	ACACCTTCAACACTCCGCG
HIST1H1C	forward	GCCGGCTATGATGTGGAGAA
	reverse	GCCTTTCGTTTGCACCAGAG
HIST2H2AA	forward	GTCTTGCCTAACATCCAGGCC
	reverse	TTGAGTTCACAGGTGCCCCT
HIST1H2BK	forward	CCAAGGCCGTCACCAAGTAC
	reverse	CCGAAGGCAATTGTGCTTCT
HIST1H3H	forward	CGCTATCGGCCTGGTACAGT
	reverse	GCGCAAGTCGGTCTTGAAGT
HIST1H4H	forward	TGTCTGGCCGTGGTAAAGGT
	reverse	GGATGTTATCGCGCAAAACC
Dicer	forward	AATTGTCCATCATGTCCTCGC
	reverse	CACATGGCTGAGAAGTATACCTGTCT
SLBP	forward	GACACCTTCGACAACCTGGC
	reverse	GCTCGGAGCTGCTTTCTGC
Primers for mouse genes		
Gapdh	forward	CCATCACCATCTTCCAGG
	reverse	CCTGCTTCACCACCTTCTTG
Oct-4	forward	GGCGTTCGCTTTGGAAAGGTGTTC
	reverse	CTCGAACCACATCCTTCTCT
Hist1H1c	forward	CGGCGCTAGTTCGTATTGGAC
	reverse	GCGCCCAATCAAACGAAGA
Hist1H2bc	forward	CTGTGTTTGGAAATCCGAAGATG
	reverse	GCGCTTCTTGCCGTCCTT
Hist1H2bp	forward	GTAGAGTTCCTGACCTAACATGCCTG
	reverse	GCGCTTCTTGCCATCCTTC
Primers for amplifzing RPA probe		
HIST1H3H	forward	ATCCAGCTCGCACGTCGTAT
	reverse	CGGAAAAATGCCGGACAT T

3.2.5 References

Haase, A. D., L. Jaskiewicz, et al. (2005). "TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing." EMBO Rep 6(10): 961-7.

Harris, M. E., R. Bohni, et al. (1991). "Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps." Mol Cell Biol **11**(5): 2416-24.

Narita, T., T. M. Yung, et al. (2007). "NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs." Mol Cell **26**(3): 349-65.

Schmitter, D., J. Filkowski, et al. (2006). "Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells." Nucleic Acids Res **34**(17): 4801-15.

Strausberg, R. L., E. A. Feingold, et al. (2002). "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences." Proc Natl Acad Sci U S A **99**(26): 16899-903.

Su, A. I., M. P. Cooke, et al. (2002). "Large-scale analysis of the human and mouse transcriptomes." Proc Natl Acad Sci U S A **99**(7): 4465-70.

Wang, Y., R. Medvid, et al. (2007). "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal." Nat Genet **39**(3): 380-5.

White, J. and S. Dalton (2005). "Cell cycle control of embryonic stem cells." Stem Cell Rev **1**(2): 131-8.

4. Curriculum vitae

CURRICULUM VITAE

Sinkkonen, Lasse Tapani
Mattenstrasse 31
CH-4058 Basel
Tel. +4178 885 0736
E-mail: Lasse.Sinkkonen@fmi.ch

DATE OF BIRTH March 10, 1980

CITIZENSHIP Finnish

MARITAL STATUS engaged

EDUCATION

2004-date	PhD Studies at University of Basel and Friedrich Miescher Institute, Basel, Switzerland
2000-2004	M.Sc. Degree in Biochemistry, University of Kuopio, Kuopio, Finland
1996-1999	High school, Vuoksenniskan Yhteislukio, Imatra, Finland
June-Dec 1999	Military service in the Finnish Army

CURRENT POSITION

PhD student, Friedrich Miescher Institute for Biomedical Research, Novartis Research Foundation

LANGUAGE SKILLS

Native Finnish, excellent English, passable Swedish and German

WORK EXPERIENCE

November 2004-date	Ph.D. Student, Friedrich Miescher Institute for Biomedical Research
June 2002-November 2004	Research assistant, University of Kuopio, Department of Biochemistry
March-April 2002	Part-time assistant on an organic chemistry lab course, University of Kuopio, Department of Chemistry
January-May 2000	Substitute teacher in mathematics, physics and chemistry in Vuoksenniska Junior High School

PUBLICATIONS

Sinkkonen L., Malinen M., Saavalainen K., Väisänen S., Carlberg C. Regulation of the human *cyclin C* gene via multiple vitamin D3-responsive regions in its promoter. *Nucleic Acids Research*, 2005 Apr 29;33(8):2440-51.

Väisänen S., Dunlop T.W., Sinkkonen L., Frank C., Carlberg C. Spatio-temporal activation of chromatin on the human *CYP24* gene promoter in the presence of 1a,25-dihydroxyvitamin D3. *Journal of Molecular Biology*, 2005 Jul 1;350(1):65-77.

Dunlop T.W., Väisänen S., Frank C., Molnar F., Sinkkonen L., Carlberg C. The human peroxisome proliferator-activated receptor d gene is a primary target of 1a,25-dihydroxyvitamin D3 and its nuclear receptor. *Journal of Molecular Biology*, 2005 Jun 3;349(2):248-60.

Carlberg C., Dunlop T.W., Saramäki A., Sinkkonen L., Matilainen M., Väisänen S. Controlling the chromatin organization of vitamin D target genes by multiple vitamin D receptor binding sites. *The Journal of Steroid Biochemistry and Molecular Biology*, 2007 Mar;103(3-5):338-43.

Sinkkonen L., Hugenschmidt T., Berninger P., Gaidatzis D., Mohn F., Artus-Revel C., Zavolan M., Svoboda P., Filipowicz W. miRNAs control *de novo* DNA methylation in mouse embryonic stem cells. *Nature Structural and Molecular Biology*, 2008 Mar; 15: 259-267.

TALKS GIVEN IN SCIENTIFIC MEETINGS

Microsymposium on small RNAs, Vienna, Austria, 2008

3rd Novartis meeting on epigenetics, Cape Cod, MA, USA 2007

Annual meeting of Friedrich Miescher Institute, Grindewald, Switzerland 2007

SWISS RNA meeting, Bern, Switzerland 2006

Annual meeting of Network of Excellence in Epigenetics, Naples, Italy 2006

2nd Novartis meeting on epigenetics, Les Diablerets, Switzerland 2006