# Analysis of small RNA-Argonaute complexes in different organisms

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) der Fakultät Biologie und Vorklinische Medizin der Universität Regensburg

vorgelegt von

Anne Dueck

aus

Bielefeld

Im Jahr 2014

Promotionsgesuch eingereicht am:
02.09.2014

Die Arbeit wurde angeleitet von:
Prof. Dr. Gunter Meister

Unterschrift:

_____

# TABLE OF CONTENTS

# SUMMARY/ABSTRACT

MicroRNAs (miRNAs) are small, non-coding RNAs of 20-24 nt in length. They are bound by members of the Argonaute protein family and play important roles in various processes including development, cellular homeostasis or response to external stimuli. Mis-regulations of miRNAs cause diseases such as cancers. In several chapters, this thesis describes miRNA profiles, miRNA and Argonaute protein characteristics as well as the *de novo* description of miRNAs and small RNAs in different cellular and organismal systems.

The four human Argonaute (Ago) proteins are highly similar proteins. One distinct difference is the enzymatic activity of Ago2 (Meister et al., 2004), however, detailed studies on the relationship between Ago proteins have been few. This work shows that Ago2 is the most abundant Ago protein in the human cell lines HEK 293T and HeLa (Petri et al., 2011). Ago1 and Ago3 are expressed only to a minor amount, whereas Ago4 expression is hardly detectable. Re-analysis of their enzymatic activity showed that Ago2 remains the only Ago protein capable of cleaving RNA targets (Hauptmann et al., 2013). Following this lead, it could also be demonstrated that not only do Ago1 and Ago3 have difficulties in replacing the passenger strand (the strand of a small RNA duplex that usually gets degraded), but that by increasing the stability of the duplex, on-target effects of an siRNA could be elevated and off-target effects could be alleviated (Petri et al., 2011).

To investigate the potential differences of the four Ago proteins with respect to their miRNA-binding properties, the distribution and the characteristics of miRNAs bound to endogenous Ago proteins were analyzed. By isolating Ago proteins using specific immunoprecipitation, cloning and sequencing of bound miRNAs, it could be shown that Ago proteins bind very similar sets of miRNAs. These miRNAs have in 10 to 20% of the cases non-templated additions of one adenosine. On the individual miRNA level, miRNAs can exist as 5' or 3' length variations that are called isomiRs. Investigating isomiRs, it could be shown that Ago2 binds to miRNAs with a broader isomiR spectrum than Ago1 and Ago3 (Dueck et al., 2012). The exceptional role of Ago2 is underscored in the processing of the non-canonical miR-451. Ago2 not only cleaves the precursor of miR-451 (Cheloufi et al., 2010; Cifuentes et al., 2010), but in my thesis it could also be shown that Ago2 exclusively binds miR-451 and does not exchange it with Ago1 or Ago3. Mutational studies on the miR-451 precursor highlight its unique and optimized structure, which renders its use as a backbone for Ago2-specific RNAi tools limited (Dueck et al., 2012).

Following up on the analysis of individual miRNAs and their distribution among Ago proteins, global miRNA profiles under the influence of extracellular stimuli in an immunological setting were examined. Wild-type as well as miR-155-deficient macrophages and dendritic cells (DCs) were stimulated or matured with LPS, LDL, oxidized LDL (oxLDL) or enzymatically lysed LDL (eLDL). MiRNAs were cloned and sequenced and the resulting libraries were compared. Depending on the stimulus, miRNA profiles changed mildly to strongly. The use of miR-155 deficient cells proved to be a powerful tool to establish that miR-155 strongly influences miRNA expression. Examination and comparison the

different global miRNA profiles highlighted that miR-155 is a master regulator of downstream miRNA expression.

The green alga *Volvox carteri* is a model organism for the development of multicellularity. Its proteome and transcriptome is predicted to be highly similar to its closest unicellular relative, *Chlamydomonas reinhardtii*. Up to date, no protein(s) or mechanism(s) could be identified that explain the big difference in development and morphology between the two organisms. The small RNA repertoire of *Volvox carteri* and the proteins involved in their biogenesis and exertion of their function was not described when the project was started. Here, we characterized an Argonaute protein of *V.carteri*, AGO3, and analyzed AGO3-bound small RNAs. A miRNA identification pipeline was constructed and we identified 288 miRNAs contained in 140 miRNA families. Several of the predicted miRNAs were shown to be expressed and modified at their 3' end, which is typical for plant miRNAs. Additionally, other small RNAs such as repeat-associated small RNAs or phased small RNAs were analyzed. Finally, no overlap of small RNA expression between *Volvox* and *Chlamydomonas* could be detected. The existence of a complete new set of small RNAs in *Volvox carteri* raises the question if the development of multicellularity and the division of labor between somatic and germ cells is at least partially regulated by miRNAs or small RNAs. Future studies, also on other members of the volvocine algae, need to elucidate the role of small RNAs in *Volvox carteri* further.

# ABBREVIATIONS

| | |
|---|---|
| **3'-UTR** | 3'-untranslated region |
| **5'-UTR** | 5'-untranslated region |
| **ADAR** | adenosine deaminase acting on RNA |
| **AEBSF** | 4-(2-Aminoethyl) benzenesulfonyl fluoride |
| **Ago** | Argonaute / Ago clade protein |
| **APS** | ammonium persulfate |
| ***A.thaliana*** | *Arabidopsis thaliana* |
| **ATP** | adenosine triphosphate |
| **bp** | base pair |
| **BSA** | bovine serum albumin |
| ***C.elegans*** | *Caenorhabditis elegans* |
| **cDNA** | complementary DNA |
| **CDS** | coding sequence |
| **Ci** | Curie |
| ***C.reinhardtii*** | *Chlamydomonas reinhardtii* |
| **d** | deoxy |
| **Da** | Dalton |
| **DC** | dendritic cell |
| **dd$H_2O$** | double-distilled water |
| **DGCR8** | DiGeorge syndrome critical region 8 |
| ***D.melanogaster*** | *Drosophila melanogaster* |
| **DNA** | deoxyribonucleic acid |
| **dNTP** | 2‟-deoxyribonucleoside triphosphate |
| **ds** | double-stranded |
| **dsRBD** | double-stranded RNA binding domain |
| **DTT** | 1,4-dithiothreitol |
| **DUF** | domain of unknown function |
| ***E.coli*** | *Escherichia coli* |
| **EDTA** | ethylenediaminetetraacetic acid |
| **FH** | Flag-HA |
| ***g*** | gravitational constant |
| **GFP** | green fluorescent protein |
| **GTP** | guanosine triphosphate |
| **h** | hour |
| **hc-RNA** | heterochromatic RNAs |
| **HEK** | human embryonic kidney |
| **HEPES** | 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid |
| ***H.sapiens*** | *Homo sapiens* |
| **HSC** | heat shock cognate |
| **HSP** | heat shock protein |
| **IgG** | immunoglobulin class G |
| **IP** | immunoprecipitation |
| **IRES** | internal ribosome entry site |
| **k** | kilo |
| **l** | liter |
| **LINE** | long interspersed nuclear elements |
| **LNA** | locked nucleic acid |
| **LTR** | long terminal repeat |
| **M** | molar |
| **MCS** | multiple cloning site |
| **miRNA** | microRNA |
| **miRNP** | micro-ribonucleoprotein |
| **mRNA** | messenger RNA |
| **NAT-siRNA** | natural antisense transcript-derived siRNA |
| **nt** | nucleotide(s) |
| **ORF** | open reading frame |
| **P** | phosphate |
| **P bodies** | processing bodies |
| **PAA** | polyacrylamide |
| **PABP** | poly(A)-binding protein |
| **PAGE** | polyacrylamide gel electrophoresis |
| **PAZ** | piwi-argonaute-zwille |
| **PBS** | phospate buffered saline |
| **PCR** | polymerase chain reaction |
| **piRNA** | Piwi-interacting RNA |
| **Pol** | polymerase |
| **pre-miRNA** | precursor miRNA |
| **pri-miRNA** | primary miRNA |
| **qRT-PCR** | quantitative real time polymerase chain reaction |
| **RdRP/ RDR** | RNA dependent RNA polymerase |
| **RISC** | RNA induced silencing complex |
| **RNA** | ribonucleic acid |
| **RNAi** | RNA interference |
| **RNP** | ribonucleoprotein |
| **rpm** | revolutions per minute |
| **RRM** | RNA recognition motif |
| **rRNA** | ribosomal RNA |
| **RT** | room temperature |
| **SD** | standard deviation |
| **SDS** | sodium dodecyl sulfate |
| **shRNA** | short hairpin RNA |
| **SINE** | short interspersed nuclear element |
| **siRNA** | small interfering RNA |
| **snoRNA** | small nucleolar RNA |
| **ss** | single-stranded |
| **SSC** | saline-sodium citrate buffer |
| **TE** | transposable element |
| **ta-siRNA** | trans-acting siRNA |
| **TEMED** | N,N,N‟,N‟-Tetramethylethylenediamine |
| **TNRC6** | trinucleotide repeat containing 6 |
| **Tris** | tris(hydroxymethyl)aminomethane |
| **tRNA** | transfer RNA |
| **UNA** | unlocked nucleic acid |
| **UTP** | uridine triphosphate |
| ***V.carteri*** | *Volvox carteri* |
| **v/v** | volume in volume |
| **w/v** | weight per volume |
| **wt** | wild type |

# CHAPTER 1 INTRODUCTION

Since the discovery of small interfering RNAs (siRNAs) and microRNAs (miRNAs) in the 1990s, many more different classes of small RNAs have been identified. Small RNAs are involved in the regulation of crucial cellular pathways such as post-transcriptional gene silencing, heterochromatin silencing or DNA elimination.

This introductory chapter will give an overview over the best-known small RNA classes, their biogenesis and function. Special emphasis will be given on miRNAs and their characteristics in animals as well as in plants. Another important aspect of miRNA function – namely their loading onto their effector proteins, the Argonaute protein family – will be described in more detail.

## 1.1 Biogenesis of miRNAs in animals and plants

### 1.1.1 Biogenesis in animals

#### *1.1.1.1 Canonical pathway*

MiRNAs are transcribed as primary miRNA transcripts (pri-miRNAs), which are processed to hairpin-structured miRNA precursors (pre-miRNAs) in the nucleus of eukaryotic cells (Figure 1-1). The microprocessor complex containing the RNase III Drosha and its co-factor DGCR8 catalyzes this initial processing step (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004a; Landthaler et al., 2004; Lee et al., 2003b). DGCR8 contains two double-stranded RNA-binding domains (dsRBD) and seems to be responsible for sensing the double-strand (ds)/single-strand (ss) junction of the pri-miRNA and positioning Drosha in a way that it is able to cleave ~11 nt away from the junction, thus releasing the pre-miRNA (Han et al., 2006). Exportin5 (XPO5) recognizes the 3' overhang of the pre-miRNA and exports it to the cytoplasm (Bohnsack et al., 2004; Lund et al., 2004; Yi et al., 2003), where the RNase III Dicer processes a short, ~21 nucleotides (nt) long double stranded RNA from the stem of the pre-miRNA (Grishok et al., 2001; Ketting et al., 2001). Both the 5' arm and the 3' arm of the pre-miRNA (referred to as miR-x-5p or -3p) can be selected as the mature single stranded miRNA that is subsequently incorporated into the RNA-induced silencing complex (RISC) (reviewed in (Kim et al., 2009; Seth et al., 2013). The other strand of the individual duplex is normally much less abundant. However, there are exceptions such as the *MIR-9* gene, which produces equally well-expressed 5p and 3p versions. The less abundant strand of the miRNA duplex is also often referred to as the miRNA* (Bartel, 2009). Within RISC, the mature miRNA strand binds to a member of the Argonaute protein family and guides it to partially complementary target sites located in 3' untranslated regions (UTR) of target mRNAs.

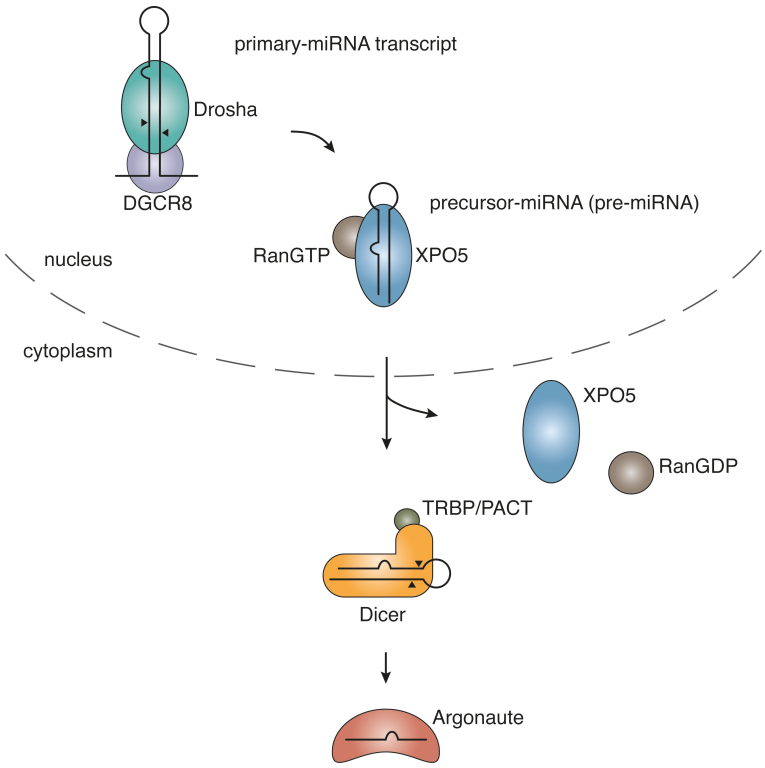



**Figure 1-1 miRNA biogenesis in mammals.** In mammals, miRNAs are transcribed as primary miRNAs (pri-miRNA) by Polymerase II from nuclear genes. The miRNA precursor (pre-miRNA) is excised by the Microprocessor complex, which consists of the RNase III enzyme Drosha and the ss/dsRNA junction sensing protein DGCR8. The pre-miRNA is then exported into the cytosol by Exportin5 in a RanGTP-dependent mechanism. Another RNase III enzyme, Dicer, associates with the pre-miRNA and cleaves ~22 nt away from the base of the hairpin. Dicer's cofactors TRBP and PACT, both double-stranded RNA-binding domain (dsRBD) proteins, facilitate both the processing of the pre-miRNA and the subsequent loading of the miRNA-duplex into one of the Argonaute (Ago) proteins (Dueck and Meister, 2014).

### *1.1.1.2 Non-canonical pathways*

MiRNAs that are independent of either Drosha or Dicer belong to the group of non-canonically processed miRNAs. The following paragraphs will briefly introduce the best-known pathways.

MiR-451 is transcribed and processed by Drosha/DGCR8 similar to a canonical miRNAs (Figure 1-2, left panel). Instead of being processed by Dicer, though, its unique stem-loop structure leads to the association with Ago2 (see also this work), which then cleaves the precursor to form a 30 nt intermediate (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010a). Recently, it has been shown that the exonuclease poly(A)-specific ribonuclease (PARN) trims miR-451 to its mature form of 22 nt in a $Mg^{2+}$-dependent manner (Yoda et al., 2013).

MiRtrons are miRNAs that are processed from short introns by splicing and are thus Drosha/DGCR8 independent. The splicing reaction generates a branched lariat structure, which is debranched to yield the miRNA precursor. There are several types of miRtrons: miRtrons without a tail, 5' tailed miRtrons and 3' tailed miRtrons (Westholm and Lai, 2011). Depending on the type, miRtrons are processed either by yet unknown nucleases (5' tail) or the exosome (3' tail) (Flynt et al., 2010) to generate an Exportin5 and subsequently a Dicer substrate. Conventional miRtrons, where the splicing reaction defines both ends of the precursor, are directly debranched and exported to the cytoplasm (Figure 1-2, second panel from the left).

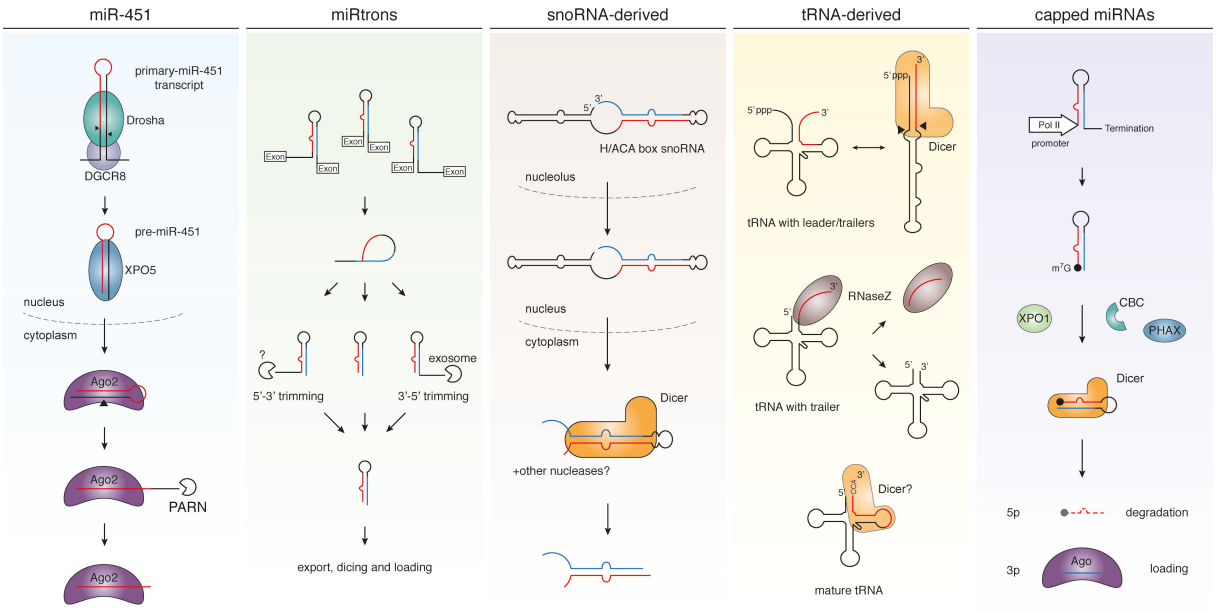



**Figure 1-2 Non-canonical biogenesis pathways of miRNAs.** The biogenesis of miR-451 and the processing of miRNAs from miRtrons, snoRNAs, tRNAs and capped transcripts are shown. Please refer to the text for detailed descriptions. Parts of the figure were adapted from Babiarz et al. 2008, Haussecker et al. 2010, Ladewig et al. 2012 and Xie et al. 2013.

MiRtrons have been described in many organisms, for example in *Drosophila* (Okamura et al., 2007; Ruby et al., 2007a), *C.elegans* (Chung et al., 2011; Ruby et al., 2007a) and vertebrates (Babiarz et al., 2011; Babiarz et al., 2008; Berezikov et al., 2007). Peculiarly, in *Drosophila*, the prevalent type of miRtrons is of the 3' tail class (Flynt et al., 2010), while vertebrate miRtrons are predominantly 5' tailed (Berezikov et al., 2007; Ladewig et al., 2012).

Another source for miRNAs are small nucleolar RNAs (snoRNAs). SnoRNAs are functioning in the maturation of other noncoding RNAs like rRNAs and small nuclear RNAs (snRNAs) (Matera et al., 2007). The structure of one class of snoRNAs, the H/ACA box snoRNAs, is similar to two connected miRNA precursor stem-loops. Indeed, it was shown that Dicer can generate a miRNA duplex from the snoRNA ACA45 (see Figure 1-2, middle panel) (Ender et al., 2008) and many more snoRNAs have been shown until now to give rise to bona fide miRNAs (Brameier et al., 2011; Taft et al., 2009).

The processing of tRNA-derived miRNAs has been shown from all major tRNA intermediates (Figure 1-2, second panel from the right). Blelloch and colleagues could identify a miRNA (mmu-miR-1983) originating from the precursor of a tRNA-Ile (Babiarz et al., 2008). While the tRNA still retains both so-called leader/trailer sequences, it probably can adopt an alternative fold that is more similar to a miRNA precursor and can subsequently be processed by Dicer. Some miRNAs are produced from the 3' trailer of a subset of tRNAs (Haussecker et al., 2010). In these cases, RNaseZ liberates a miRNA by maturing the tRNA (see Figure 1-2). The processing of a miRNA (designated CU1276) from a mature tRNA-Gly was reported by Dalla-Favera and colleagues (Maute et al., 2013). Since the first reports, other tRNAs have been identified that give rise to small RNAs (Babiarz et al., 2011; Cole et al., 2009;

Haussecker et al., 2010), but whether all fragments are bona fide miRNAs has to be elucidated in the future.

In a recent publication, Steitz and colleagues identify miRNAs that are transcribed by Polymerase II to yield short, capped transcripts that directly start and probably also end with the miRNA precursor sequence (see Figure 1-2, right panel) (Xie et al., 2013). They showed for human miR-320a and ~20 murine miRNAs that they are capped, exported into the cytoplasm by the cap-binding complex (CBC), PHAX and XPO1 (instead of XPO5) and then are further processed by Dicer. The 5p miRNA strand is only rarely incorporated into Ago, probably due to the weak binding affinity of Ago to the 5' cap-structure and is rapidly degraded. The 3p miRNA is readily incorporated into Ago.

Another pathway that produces miRNAs in a Microprocessor-independent manner was reported in early 2014. The authors could show that certain promoter-proximal RNA polymerase II (RNAPII) complexes produce hairpin RNA that are DGCR8/Drosha independent, but Dicer dependent (Zamudio et al., 2014). The hairpins are derived from RNAPII protein-coding gene promoters and are likely terminated and liberated by RNAPII pausing. The transcripts are probably de-capped and trimmed by 5'-3'-exonucleases before Dicer produces the mature miRNA duplex. Because of their origin, these miRNAs were named transcriptional start site miRNAs (TSSmiRNAs). Although their specific function is not elucidated yet, it is known that TSSmiRNAs are differentially expressed in mouse tissues and some seem to be conserved in human (Zamudio et al., 2014).

### 1.1.2 Biogenesis in plants

Plant miRNAs are usually 21-22 nt long, non-coding RNAs that are involved in post-transcriptional gene regulation. Opposed to the situation in animals, miRNAs in plants are not the most abundant small RNA species. Nevertheless, they are involved in important processes such as the targeting of transcription factors regulating organ polarity, cell division or hormonal control (reviewed in (Garcia, 2008)). The following paragraphs will describe the biogenesis of plant miRNAs (reviewed in (Czech and Hannon, 2011)), which differs in some points with respect to animal miRNAs (see 1.1), and the evolution and structure of their genes.

#### *1.1.2.1 Canonical pathway*

Plant miRNA genes are generally transcribed by RNA polymerase II (Pol II) into capped and polyadenylated primary-miRNAs (pri-miRNAs). The cap structure is recognized by the cap binding complex (CBC), whereas the junction of ssRNA and dsRNA at the base of the miRNA hairpin is bound and presumably stabilized by the RNA binding protein DAWDLE (DDL) (Figure 1-3) (Yu et al., 2008). Contrary to animals, all subsequent processing steps occur in the nucleus in so called D-bodies (Fang and Spector, 2007; Kurihara et al., 2006). There, a complex consisting of the RNase III protein DICER-

LIKE 1 (DCL1), the double-stranded RNA-binding protein HYPONASTIC LEAVES 1 (HYL1) and the Zinc-finger containing protein SERRATE (SE) processes the miRNA-harboring hairpin first to a stem-loop structure (precursor miRNA, pre-miRNA) and subsequently to a double-stranded duplex, containing the mature miRNA (guide strand) and its so-called star strand (passenger strand) (Dong et al., 2008; Grigg et al., 2005; Kurihara et al., 2006; Lobbes et al., 2006). In this processing complex, DCL1 fulfills both the function of Drosha and Dicer in animals (Henderson et al., 2006; Park et al., 2002; Reinhart et al., 2002), namely cleaving the hairpin, HYL1 presumably functions as a molecular ruler helping DCL1 to correctly process the pri-miRNA (Han et al., 2004b; Kurihara et al., 2006; Vazquez et al., 2004a; Yang et al., 2010b) and SE serves as docking platform for DCL1, HYL1 and CBC and is also thought to probably recognize structural features of miRNA precursors (Figure 1-3) (Machida et al., 2011).

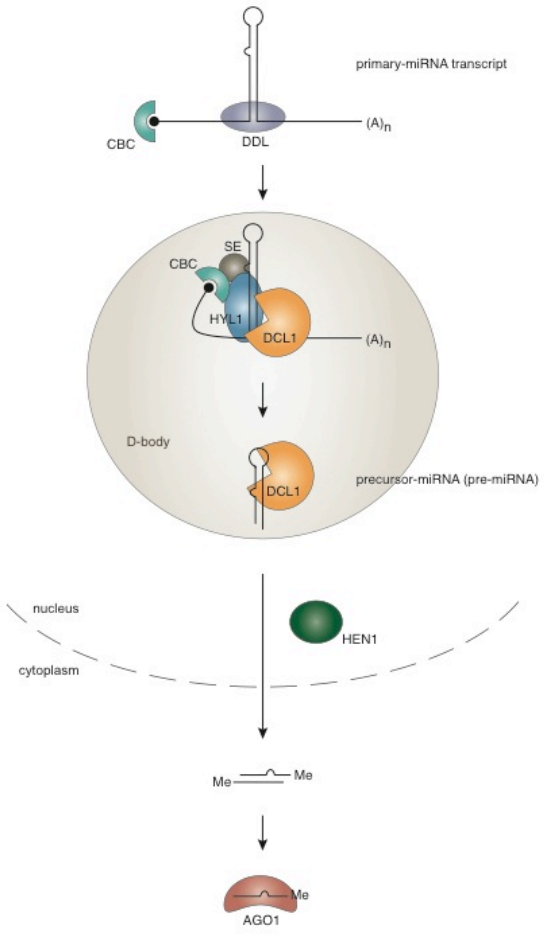



**Figure 1-3 MiRNA biogenesis in plants.** For a detailed description, please refer to the text. Figure was adapted from Voinnet, 2009 (Voinnet, 2009) (Dueck and Meister, 2014).

Before the miRNA duplex is exported to the cytoplasm, it is 2'-O-methylated by HUA ENHANCER 1 (HEN1) at each 3' end (Li et al., 2005; Yang et al., 2006; Yu et al., 2005). This modification serves as a protective measure against nucleases. Export of the methylated duplex is presumably mediated by

HASTY (HST), the plant homolog of Exportin5, and additional factors (Park et al., 2005). The guide strand is loaded primarily onto the Argonaute protein ARGONAUTE 1 (AGO1) (Baumberger and Baulcombe, 2005), while the passenger strand is degraded by the SMALL RNA DEGRADING NUCLEASE (SDN). The exact location of AGO1 loading is currently still under debate (Voinnet, 2009). However, one proposed model could be the maturation of the AGO1-miRNA complex in the nucleus and subsequent export to the cytoplasm (Eamens et al., 2009).

## 1.2 MiRNA genes, miRNA families and evolution of miRNA genes

Animal miRNAs originate from basically four different types of genetic settings (Figure 1-4) (reviewed in (Kim et al., 2009)). MiRNA genes can be located in introns of non-coding transcripts, in exons of non-coding transcripts and in introns or exons of coding transcripts. Very often (approximately ~50% in humans), canonical animal miRNAs are encoded in clusters containing several to many miRNAs like the miR-17~92 cluster in humans. They are transcribed from a single polycistronic transcription unit (Lee et al., 2002) and are then processed simultaneously or independently of each other (Chaulk et al., 2011).

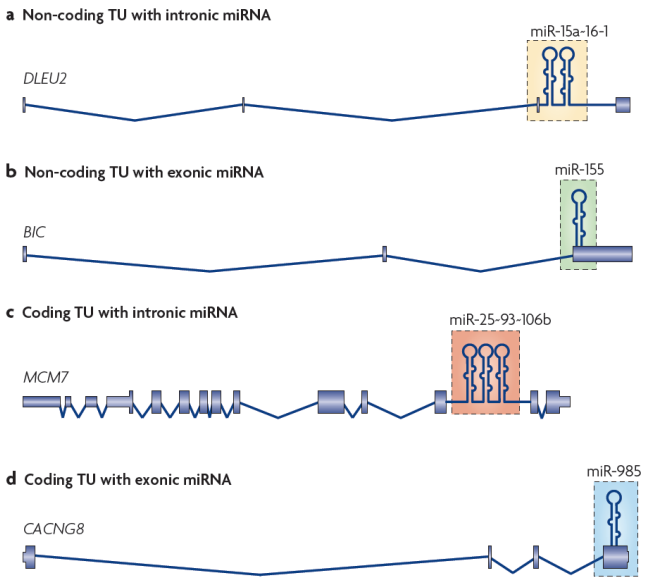



**Figure 1-4 Different types of genomic settings for animal miRNAs.** MiRNAs can be encoded in introns or exons of non-coding transcripts (A, B) or in introns or exons of coding transcripts (C, D). Examples are from the publications of (Calin et al., 2002; Tam, 2001). The figure is roughly to scale and taken from Kim et al. 2009.

Through the rise of high-throughput sequencing methods, also called “deep sequencing”, many animals from various lineages were analyzed on their miRNA content. MiRNAs seem to have been present from very early on in animal evolution. For example, miR-100 is found in all eumetazoans (animals with true tissues, i.e. all animals except sponges and placozoa) (Grimson et al., 2008). At several points of time, there seem to have been bursts of miRNA evolution, so-called miRNA

expansions (Christodoulou et al., 2010; Heimberg et al., 2008; Hertel et al., 2006; Peterson et al., 2009). Intriguingly, once miRNAs emerged in certain lineages, these shared miRNAs were rarely lost in descendant lineages (Heimberg et al., 2010; Heimberg et al., 2008; Peterson et al., 2009; Wheeler et al., 2009). Although there are conserved miRNAs across big evolutionary distances, one major part of the miRNA repertoire of a specific lineage has arisen only there. The emergence of new miRNA genes and the correlation between the amount of miRNA genes and the organismal complexity highlights the importance of miRNA innovation. For a detailed description of miRNA gene distribution across the animal kingdom see (Berezikov, 2011).

The acquirement of new miRNA genes seems to be a major driving force of evolution. The following paragraph will therefore highlight the most important mechanisms of miRNA gene innovation. A main source for novel miRNA genes is the duplication of an existing miRNA gene (Figure 1-5, A).

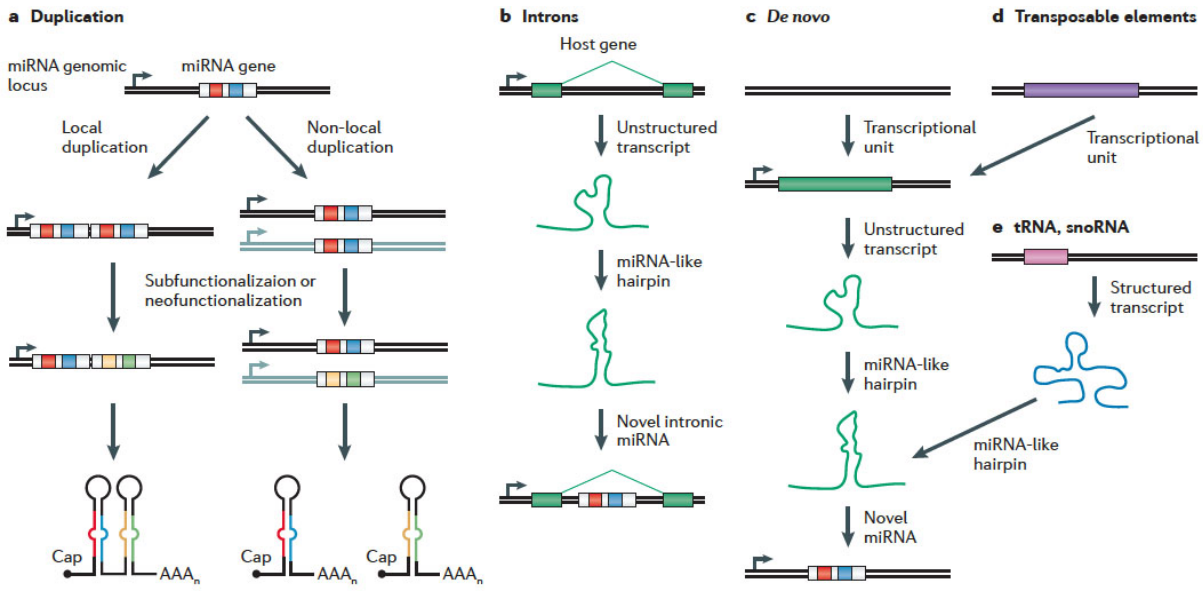



**Figure 1-5 Major sources for new miRNA genes in animals.** (A) Gene duplication. The miRNA gene can be duplicated either locally or non-locally (onto another chromosome). (B) MiRNAs can evolve from intronic sequences that acquire a miRNA-like hairpin structure. (C) *De novo* formation of a miRNA gene. First, a transcriptional unit needs to be established in the genome thus producing an RNA transcript. Later, this transcript can acquire a hairpin structure. (D) MiRNA genes can originate from transposable element genes or (E) tRNA and snoRNA genes. Figure was taken from (Berezikov, 2011).

Approximately one third of human miRNAs, for example, can be grouped into so-called families, namely having significant sequence homology to each other (Berezikov, 2011). Many miRNA genes are located in introns of protein-coding or long non-coding genes (Rodriguez et al., 2004). While some of them have their own transcriptional unit (Ramalingam et al., 2014), many utilize the transcription of their host gene. Since introns are naturally always transcribed when the gene itself is transcribed, this circumstance offers a good opportunity for miRNA rise and evolution (Figure 1-5, B). Indeed, almost 50% of human miRNAs are located in introns (Campo-Paysaa et al., 2011). The *de novo* formation of a miRNA gene requires first the acquirement of a promoter. This transcriptional unit can then evolve to define a miRNA hairpin. Transposable elements, pseudo genes, tRNAs and snoRNAs have all been

shown to give rise to miRNA genes (Figure 1-5, D and E). For a more detailed overview please see (Berezikov, 2011).

In general, newly spawned, less conserved miRNAs are expressed at a low level (in contrast to broadly conserved miRNAs) (Berezikov et al., 2006; Lu et al., 2008), probably in order to avoid deleterious effects on mRNA expression or for not having any biological targets at all yet (Chen and Rajewsky, 2007; Lu et al., 2008).

Plant miRNA genes are usually intergenic and, unlike their animal counterparts, rarely arranged in tandem (Voinnet, 2009). Additionally, the foldback structure containing the mature miRNA differs from the structure seen in animals. While animal miRNA precursors follow quite strict rules in terms of secondary structure (Han et al., 2006), plant miRNA precursors exhibit far more diversity in total length, loop size or the position of the miRNA/miRNA* duplex (Figure 1-6).

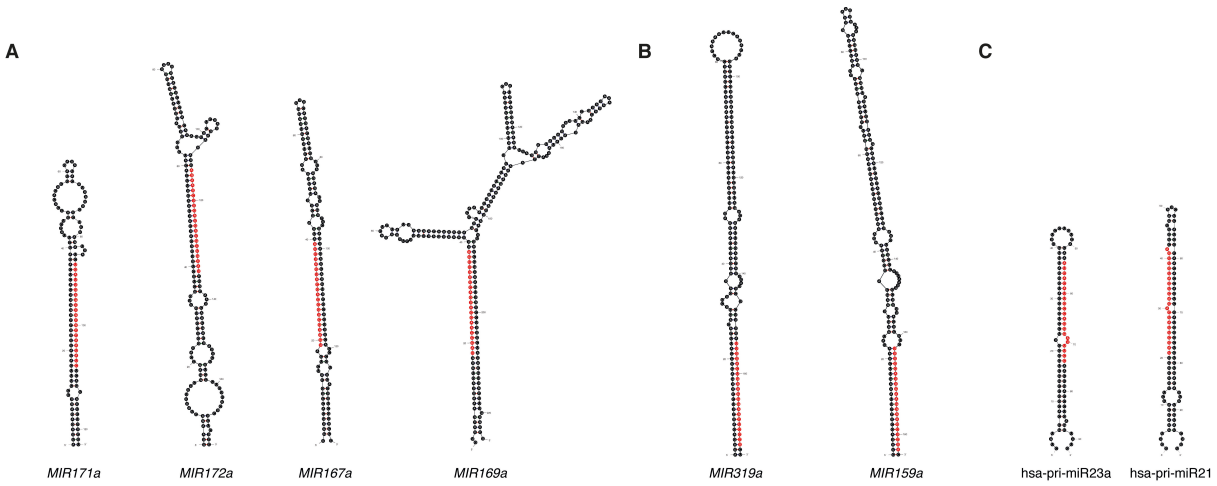



**Figure 1-6 Examples of structures of primary miRNAs in plants and humans.** (A) Canonically processed primary miRNAs from *A.thaliana*; DCL1 cleaves from bottom to top (B) Non-canonically processed miRNAs miR319a and miR159a from *A.thaliana*; DCL1 cleaves from top to bottom (C) Typical structures of human primary miRNAs (miR-23a and miR-21). Human primary miRNAs are first processed in the nucleus by Drosha and in a second step cleaved by Dicer in the cytoplasm. (A-C) Mature miRNA sequence is depicted in red color. *A.thaliana* hairpins were folded at 23°C, *H.sapiens* hairpins were folded at 37°C. Shown are only the foldback structures that contain the mature miRNA.

For example, some foldback structures contain a so-called proximal loop (*MIR172A*, Figure 1-6, A). Several groups have recently shown that DCL1 complexes initiate cleavage ~15nt from either such a structure or general ssRNA-dsRNA junctions (Mateos et al., 2010; Song et al., 2010; Werner et al., 2010). So far, two miRNA genes were shown to be processed from the top (terminal loop) to the bottom, where the mature miRNA is located (*MIRNA319A* and *MIRNA159A*, Figure 1-6, B) (Bologna et al., 2009), a mechanism yet unique to plants. In comparison, human miRNA genes are much smaller in size and contain smaller and less structured loops (Figure 1-6, C). The diverse properties of plant miRNAs described above make their prediction and *de novo* analysis a challenging task.

A detailed analysis of miRNAs from *Arabidopsis* and various plant species revealed possible mechanisms for miRNA evolution in plants. On the one hand, there exist 21 miRNA families that are expressed among diverse angiosperms (flowering plants) (Axtell and Bowman, 2008; Zhang et al., 2006). A subset of these can even be found in gymnosperms (e.g. conifers), lycopods (e.g. clubmosses) and bryophytes (e.g. mosses) and are hence even more ancient (Axtell and Bartel, 2005; Axtell et al., 2007; Floyd and Bowman, 2004). All of these so-called conserved miRNAs are in general highly expressed and since they presumably arose through multiple gene and large-scale genome duplications, they are represented by multiple copies in the genome (Maher et al., 2006). Younger miRNAs, on the other hand, are usually low abundant, are encoded on only one locus and are predicted to regulate a much broader target spectrum than their ancient counterparts (Zhang et al., 2006).

Several models attempt to explain the mechanism of the birth of plant miRNAs. The first model arose by the observation that young miRNAs bear a strong sequence similarity to their target genes, even beyond the stem-loop. Therefore, a new complementary foldback structure could emerge by an inverted gene duplication (Allen et al., 2004; Fahlgren et al., 2007). These new hairpins (named proto MIR) would mainly be perfectly complementary and therefore be substrates for the siRNA-generating enzymes DICER-LIKE 3 and 4 (DCL3 and DCL4) (Dunoyer et al., 2007). The sequence would acquire mutations through evolution, resulting in a shortening of the foldback structure and appearance of bulges and loops. At some point, processing of the hairpin would switch from DCL4 to DCL1 due to miRNA structure, which has been analyzed in depth in *Arabidopsis* (Rajagopalan et al., 2006; Vazquez et al., 2008). Only if the regulation of the young miRNA on the founder gene is favorable, the sequence is kept and continuously selected to finally result in an ancient *MIR* gene. This model relies on the selective influence of a miRNA on its target gene, but does not explain all cases of miRNA birth and evolution.

A second model, termed “spontaneous evolution”, suggests the rise of miRNA genes from the massive amount of small-to-medium sized hairpin structures throughout the genome of *Arabidopsis*. If one such structure would yield an advantageous target regulation and this incident would be stabilized by coevolution, this *MIR* gene may be contained in the genome (Felippes et al., 2008).

The third model was proposed on the observation that a certain type of transposable elements, called miniature inverted-repeat transposable elements (MITEs), can form partially complementary stem-loops that evolve into *MIR* genes (Felippes et al., 2008). Since transposable elements are also the source of 24 nt-long siRNAs, it is not clear if these sequences are bona fide miRNAs and therefore, this model is still debatable. However, this origin from transposable elements might explain why many conserved miRNAs in *Arabidopsis* are processed by both DCL1 and DCL3 into 21nt and 24nt long RNAs, respectively (Dunoyer et al., 2004; Vazquez et al., 2008).

In conclusion, although animal and plant miRNAs share some characteristics, e.g. birth of miRNA genes through transposable elements, their genetic structure and the modes of their evolution differ substantially.

## 1.3 Argonaute proteins

The Argonaute protein family is found in all kingdoms of life and can be divided into three subfamilies: The Ago clade, the Piwi clade and the worm-specific WAGO clade (Yigit et al., 2006). The number and identity of these proteins differ substantially between different organisms. The human genome, for example, encodes four Ago (Ago1-4) and four Piwi proteins (Hiwi, Hiwi2, Hili and Piwil3). For *Arabidopsis thaliana*, only Ago clade proteins are known (AGO1-10), whereas the *C.elegans* genome encodes 26 Argonaute family proteins, three of them are known to be Piwi proteins (reviewed in (Hutvagner and Simard, 2008; Peters and Meister, 2007; Tolia and Joshua-Tor, 2007)).

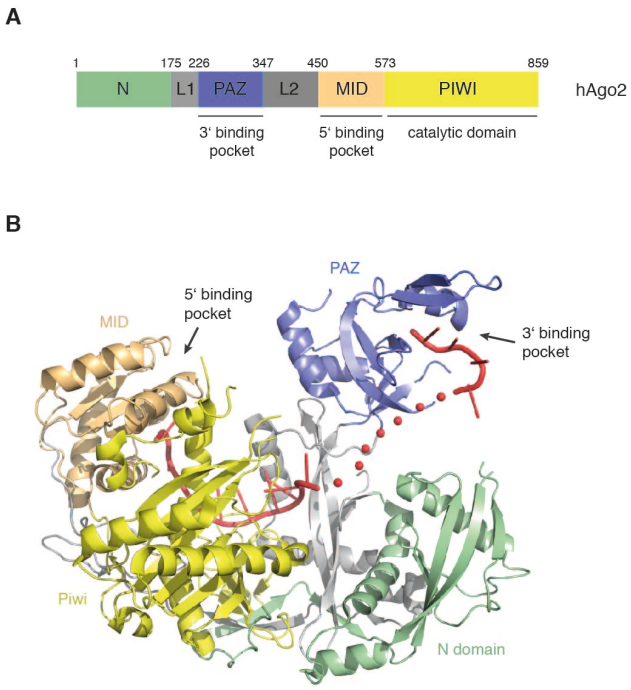



**Figure 1-7 Domain and crystal structure of human Ago2.** (A) Domain structure of Argonaute proteins using human Ago2 as representative example. Argonaute proteins have four conserved domains: The N-terminal domain, the PAZ domain (responsible for binding the 3' OH of the miRNA), the MID domain (anchors the 5' phosphate of the miRNA) and the PIWI domain (harbors the active cleavage site in hAgo2). The numbers above indicate the boundaries (in amino acids) for each domain. (B) Ribbon representation of the crystal structure of human Ago2 in complex with miR-20a (PDB_ID 4F3T; adapted from (Elkayam et al., 2012). The unresolved parts of miR-20a (red) are shown as a dotted line. The four domains are shown in color as depicted in (A) and the flexible linkers in grey (Dueck and Meister, 2014).

The domain structure of Argonaute proteins is well conserved. The PAZ domain recognizes the two nucleotide overhang of the small RNA duplex and binds to the 3' end of the mature small RNA (Figure 1-7, A/B). The MID domain specifically anchors the 5' phosphate group of the small RNA through several amino acid residues. The Piwi domain adopts an RNase H fold and is in some proteins able to cleave the opposing strand of the small RNA in an RNA-RNA duplex (Elkayam et al., 2012; Jinek and Doudna, 2009; Meister, 2013; Schirle and MacRae, 2012). The catalytic activity is dependent on the active site, which is comprised of four conserved residues, DEDH (Nakanishi et al., 2012; Song et al.,

2004; Yuan et al., 2005). This catalytic tetrad is necessary but not sufficient for cleavage, since human Ago2 is cleavage competent, whereas human Ago3 does not have the ability to cleave, although both have the same catalytic residues (Meister et al., 2004). Recent work has shown that two short unstructured loops within the N-terminal domain are involved in rendering an Argonaute protein cleavage competent (Faehnle et al., 2013; Hauptmann et al., 2013; Schurmann et al., 2013).

Argonaute proteins do not only anchor the 5' end by the Mid domain, this domain can also contain a so-called nucleotide-specificity loop that is responsible for the sensing of the 5' nucleotide (see also 1.6.2). The predominant 5' nucleotide that is bound by human Ago proteins is a U, followed by A. The structural study by Frank et al. shows that the nucleotide-specificity loop only interacts with U and A, therefore disfavoring C and G (Frank et al., 2010). In addition, this loop and its specificity were investigated for three AGO proteins of *Arabidopsis* as well (Frank et al., 2012), revealing different binding properties than for human Agos, yet a conserved mechanism. This nucleotide-specificity loop helps explaining why small RNAs in *Arabidopsis* are sorted into AGO proteins according to their 5' nucleotide. The term sorting refers to the distribution of small RNA subsets to different Ago proteins. Nucleotide binding preferences have been observed for many Argonaute proteins (Ago and Piwi clade), for a detailed overview see Supplement 2, Table in the Appendix.
Although most or all Argonaute proteins can sense the 5' nucleotide of small RNAs, this sensing is not necessarily a means to sort or distribute different small RNA population between different Argonaute proteins, but might also contribute to filter correctly processed small RNAs from the total pool of small RNAs.

### 1.3.1 Human Ago2 protein as a tool for exogenous RNA interference

For scientific purposes, the catalytic activity of Ago proteins is exploited for exogenous siRNA-mediated knock down of target mRNAs (RNA interference, RNAi). In humans, Ago2 is the only catalytically active Ago and therefore also called "slicer". Since Ago2 can only "slice" when the small RNA is fully complementary to its target mRNA, siRNAs are designed that way. However, siRNAs are initially small RNA duplexes with a 3' 2 nt overhang, similar to miRNA duplexes. The strand that is designed against the target mRNA and that is incorporated into Ago is called guide strand. The other strand, called passenger strand, is usually loaded to a much lower degree. Each of the siRNA strands also inadvertently contains a so-called seed sequence. The seed of a miRNA are nucleotides 2-7 or 2-8, which are mediating the binding to and targeting of an mRNA. The rest of the miRNA sequence seems not to play a major part in target recognition. The seed sequences of both siRNA strands could potentially lead to unwanted side effects, namely the inhibition of non-target mRNAs via the seed sequence (Wu et al., 2008). These unwanted effects are called "off-target" effects, in contrast to the desired "on-target" effects. The alleviation of off-target effects and the strengthening of on-target effects are of great importance for potential medical application and subject of many studies (see also this work).

## 1.4 Other classes of small RNAs

Besides miRNAs, many other types of small RNAs that bind to a member of the Argonaute protein family are known. This paragraph will give a broad overview over the many varieties of small RNAs and their functions in different organisms (see also Figure 1-8).

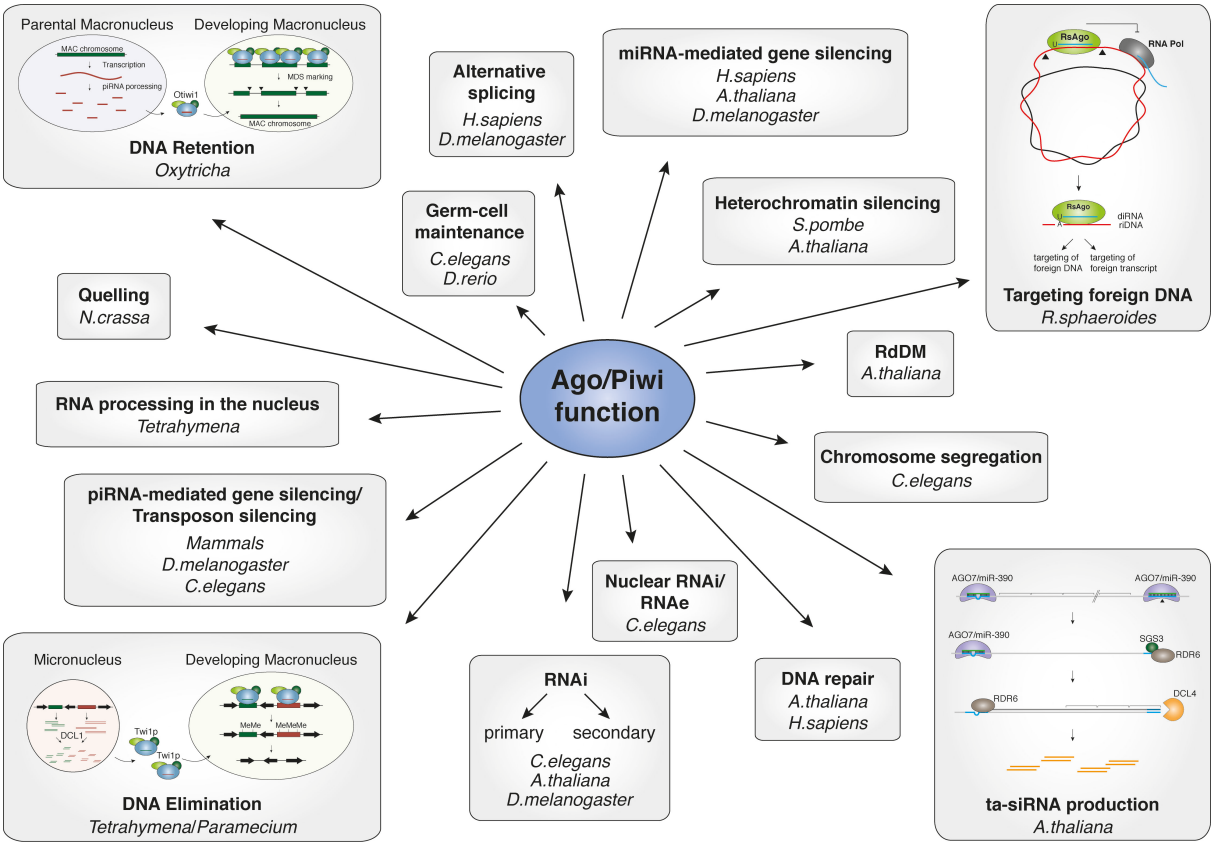



**Figure 1-8 Exemplary functions of Ago clade and Piwi clade proteins.** A few of the more unique mechanisms were depicted in detail. Please refer to the text for descriptions. Parts of the figure were adapted from (Chalker and Yao, 2011; Fang et al., 2012; Montgomery et al., 2008a; Olovnikov et al., 2013). Abbreviations: MAC = Macronucleus, MDS = macronuclear-destined sequences, Otiwi1/Twi1p/RsAgo/AGO7 = Ago/Piwi proteins, RNA Pol = RNA Polymerase, diRNA = DNA-interacting RNA, riDNA = RNA-interacting DNA, RdDM = RNA-dependent DNA methylation, SGS3 = SUPPRESSOR OF GENE SILENCING 3, RDR6 = RNA-dependent RNA polymerase 6, DCL1/DCL4 = DICER-LIKE 1/4, ta-siRNA = trans-acting siRNA, piRNA = Piwi-interacting RNA, Me = Methylation (Dueck and Meister, 2014).

Piwi-interacting RNAs (piRNAs) are germline-specific and have been found in many different organisms including *C.elegans*, *Drosophila* and mammals. PiRNAs derive from specific loci termed piRNA clusters, which typically contain mobile genetic elements. PiRNA clusters are transcribed as long single-stranded primary transcripts (Li et al., 2013) that are further processed to primary piRNAs (reviewed in Siomi et al. 2011). These primary piRNAs are typically 24-32 nt long and are incorporated into Piwi-clade Argonaute proteins. In a mechanism called ping-pong cycle, primary piRNAs are used as guides for the generation of secondary piRNAs. In most cells, mobile genetic elements are silenced epigenetically by heterochromatin induction. During spermatogenesis, where these epigenetic gene silencing marks such as DNA or histone methylation are erased, piRNAs repress the expression of

potentially harmful mobile genetic elements (reviewed in (Hartig et al., 2007; Siomi et al., 2011) (Dueck and Meister, 2014)).

*C.elegans* possesses 26 different Argonaute proteins and thus it is not surprising that a large variety of different small RNA pathways exist. Besides classical miRNA and siRNA functions, small RNA-Argonaute complexes have been implicated in nuclear RNAi, a worm-specific pathway that involves several different Argonaute proteins (WAGO proteins). Most interestingly, nuclear RNAi is required for a process termed RNA-induced epigenetic silencing (RNAe). This process involves chromatin modifications and can be stably inherited over many generations (Ashe et al., 2012; Buckley et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). Additional small RNA pathways in *C.elegans* affect chromosome segregation or germ cell maintenance, for example (Figure 1-8) (Dueck and Meister, 2014).

Another interesting organism with distinct small RNA pathways is *Tetrahymena thermophila*. In this organism, a macro- and a micronucleus exist. While the micronucleus is important for sexual reproduction, the macronucleus encodes all genes required for vegetative growth in culture. During macronucleus development, large parts of the DNA (internal eliminated sequences, IES) are eliminated and this process depends on small RNAs as well as *Tetrahymena* Argonaute proteins (Figure 1-8, lower left part). In addition, small RNA pathways are also involved in nuclear RNA processing (for references see Supplement 2, Table, Appendix).

In the protozoon *Oxytricha trifallax*, 95% of the DNA is removed (mainly transposons and other repetitive sequences) during somatic nuclear development (Fang et al., 2012). The piRNA pathway in this organism specifies genomic regions for retention. Interestingly, the piRNAs originate from the maternal nucleus, again demonstrating that small RNAs can function as powerful trans-generational carriers of epigenetic information (Figure 1-8, upper left panel) (Fang et al., 2012) (Dueck and Meister, 2014).

Argonaute proteins are not only found in eukaryotes but also in bacteria and archaea. Early structural work demonstrated that some archaeal Argonaute proteins bind small DNAs with high affinity, which function as guides for RNA cleavage (Yuan et al., 2005). However, their endogenous functions are still elusive. In the bacterium *Rhodobacter sphaeroides*, the Argonaute protein interacts with small RNAs and these small RNAs are complementary to DNA sequences mainly stemming from exogenous plasmids, transposons or phage DNA suggesting that Argonaute-small RNA complexes are important for the defense against invading foreign DNA in these bacteria (Figure 1-8, upper right part) (Olovnikov et al., 2013). Very recently, the *Thermus thermophilus* Argonaute protein has been reported to use DNA-guides to cleave DNA templates such as external plasmids (Sheng et al., 2014; Swarts et al., 2014).

Interesting new directions of Argonaute functions have been reported in human cells as well. Small RNAs have recently been implicated in DNA double-strand break repair (Francia et al., 2012; Michalik et al., 2012; Wei et al., 2012). Furthermore, it has been reported that nuclear Argonaute proteins function in alternative splicing (Allo et al., 2009; Ameyar-Zazoua et al., 2012; Taliaferro et al., 2013).

However, the precise mechanisms as well as the biological relevance of these findings remain to be further elucidated.

In plants, many different small RNA classes exist and therefore many functions have been reported. The most abundant group of small RNAs in plants are small interfering RNAs (siRNAs), which arise from dsRNA precursors. They can roughly be divided into three groups: secondary siRNAs, natural antisense-derived siRNAs (NAT-siRNAs) and heterochromatic RNAs (hc-RNAs) (reviewed in (Axtell, 2013)). For clarification, an overview with a classification of plant small RNAs is shown in Figure 1-9.

Secondary siRNAs are defined as siRNAs produced downstream of small RNA triggered processes. The cleavage of the primary transcript leads to the recruitment of an RNA-dependent RNA polymerase (RdRP or RDR) that synthesizes the second strand. This dsRNA is further processed by a DICER-LIKE (DCL) enzyme to yield 21-22nt long secondary siRNAs (Allen et al., 2005; Yoshikawa et al., 2005). In most cases, RDR6 and DCL4 are involved in secondary siRNA biogenesis. This group of siRNAs cannot only be found in flowering plants, but also in distant phylogenies like bryophytes (mosses) (Axtell et al., 2006; Talmor-Neiman et al., 2006). Secondary siRNAs can be further subdivided into *trans*-acting siRNAs (ta-siRNAs) and phased siRNAs (Figure 1-9).

Ta-siRNAs, as their name implies, act in *trans*. Their function has been investigated in detail only for a few cases in *Arabidopsis*, but the mechanism, its cellular impact and its conservation will be elucidated further in the upcoming years. The best studied example for ta-siRNAs is the TAS3 family of secondary siRNAs. Here, the ta-siRNA-specific Ago protein AGO7 is guided by miR390 to the *TAS3* transcript and mediates its cleavage (Figure 1-8, bottom right) (Allen et al., 2005; Fahlgren et al., 2006; Montgomery et al., 2008a). Through this sequence-specific cleavage, the RNA-dependent polymerase RDR6 and DCL4 are recruited to *TAS3*, thus generating secondary siRNAs. These ta-siRNAs have been shown to target *Auxin Response Factor 3* (*ARF3*) and *ARF4*, which are involved in plant flowering. This mechanism is widely conserved and plays an important role in organ polarity, meristem identity and developmental timing (reviewed in Chitwood et al., 2007 (Chitwood et al., 2007)).

Phased siRNAs are generated from certain genomic loci and are defined by being uniformly processed into always the same or a very similar set of siRNAs. This nomenclature has arisen through their specific biogenesis pathway. However, if their function is to act in *trans*, they can also be called ta-siRNAs. The advantage of secondary siRNAs is that an siRNA response is amplified and they are a powerful tool to simultaneously regulate a network of mRNAs highly similar in sequence and function.

Natural antisense-derived siRNAs are produced from independently transcribed RNAs that can hybridize and form a perfectly complementary dsRNA molecule (Figure 1-9). The transcribed RNAs can either be stemming from the same locus (overlapping genes, *cis*-NAT-siRNAs) or from independent loci, but still with complementarity to each other (*trans*-NAT-siRNAs) (Axtell, 2013). Examples for *trans*-NAT-siRNAs have not been found so far and therefore remain a hypothetical class. In *Arabidopsis*, there are three functionally characterized *cis*-NAT-siRNAs so far, induced or accumulating under diverse circumstances such as salt-stress, infection with avirulent bacteria or pollen development (Borsani et al., 2005; Katiyar-Agarwal et al., 2006; Ron et al., 2010).

Heterochromatic RNAs originate from intergenic loci or from repetitive regions. They are 23-24nt long and are usually transcribed by Pol IV (Mosher et al., 2008). They require the processing by RDR2 (RNA-dependent RNA polymerase-2) and DCL3 (Kasschau et al., 2007; Lu et al., 2006) and are subsequently loaded onto an AGO protein of the AGO4 subfamily, which includes AGO4, AGO6 and AGO9 in *Arabidopsis* (Havecker et al., 2010). This complex is necessary for RNA-mediated DNA methylation and therefore heterochromatin silencing (reviewed by Law and Jacobson, 2010 (Law and Jacobsen, 2010)) (Chan et al., 2004; Herr et al., 2005; Onodera et al., 2005; Pontier et al., 2005).

The exploding literature on novel functions of Argonaute-small RNA complexes of the last few years underscores the importance of small RNA-guided processes in a large variety of different organisms. It is very likely that not only the abovementioned pathways will be further characterized but also novel, so far unrecognized functions will be identified and reported in the future.

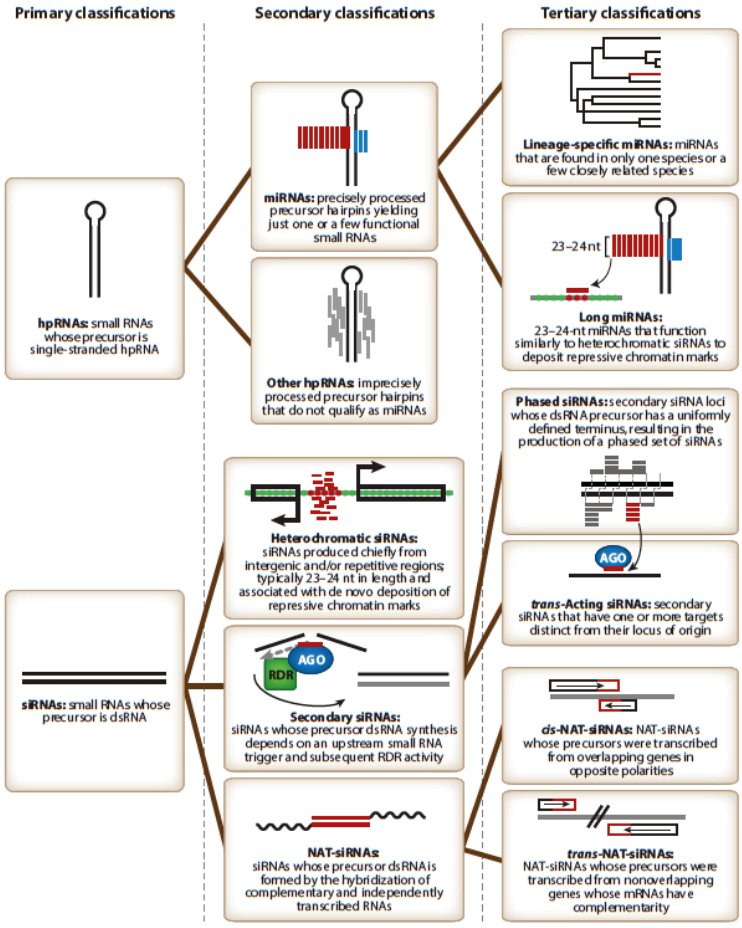



**Figure 1-9 Classification of plant small RNAs.** According to their origin from either a hairpin or a ds RNA molecule, small RNA can be divided on a first level (primary classification). After this step, small RNAs can be subdivided into secondary and tertiary groups. Figure was taken from (Axtell, 2013).

## 1.5 Function of miRNAs

The next paragraphs will describe the function of miRNAs in animals as well as in plants. Plant and animal pathways differ in properties like Argonaute-binding specificity, mRNA-target recognition or several mechanistic aspects.

### 1.5.1 Function of miRNAs in animals

After Argonaute is loaded with a miRNA strand, the complex binds to an mRNA target. The most relevant part of a miRNA for binding its target-mRNA is the so-called seed-region (nucleotides 2-7 or 2-8) of the miRNA. The different types of miRNA target sites have been reviewed by David Bartel (Bartel, 2009), for example.

Upon target recognition, the Argonaute protein recruits a member of the GW182 protein family, which coordinates all downstream silencing events (see Figure 1-10) (Jakymiw et al., 2005; Liu et al., 2005a; Liu et al., 2005b; Meister et al., 2005). In mammals, the GW182 protein family is composed of three members referred to as TNRC6A-C. While the N-terminus of the GW182 protein binds to Argonaute, its M2 and PAM2 domains interact with the Poly(A)-binding protein (PABP) that is present on the poly(A)-tail of the mRNA (Pfaff and Meister, 2013). This association leads to the disruption of 5' cap – 3' tail interactions, thus opening the closed loop conformation of the mRNA necessary for efficient translation. In further steps, the GW182 protein recruits the deadenylation machinery such as the CCR4/NOT complex (Chekulaeva et al., 2011; Chen et al., 2014; Fabian et al., 2011; Kuzuoglu-Ozturk et al., 2012; Mathys et al., 2014), which leads to the degradation of the poly(A)-tail of the uncapped mRNA. The mRNA is now silenced and potentially stored in P-bodies (cytoplasmic foci, where RNA metabolizing enzymes are enriched) or further degraded. The decay process is initiated by the decapping complex (consisting of DCP2, DCP1, DDX6 and EDC4) (Behm-Ansmant et al., 2006; Rehwinkel et al., 2005) and the decapped mRNA is degraded by 5'-3' (XRN1) and/or 3'-5' exonucleases (reviewed in (Huntzinger and Izaurralde, 2011; Krol et al., 2010). At early time points after miRNA-guided silencing, repression occurs predominantly on the level of translation, while mRNA degradation dominates at later time points (Huntzinger et al., 2012).

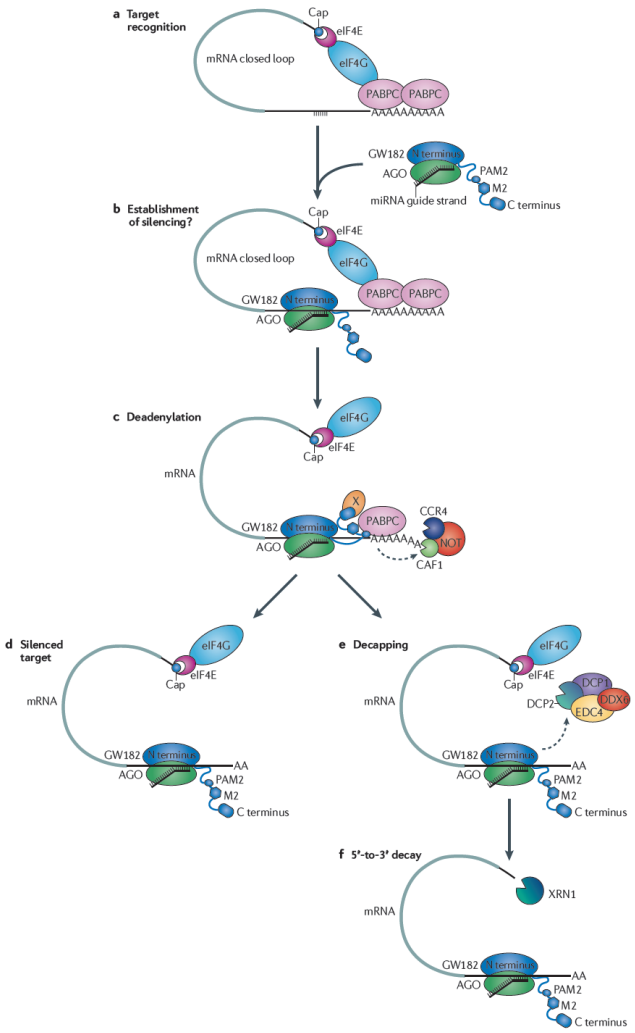



**Figure 1-10 Mechanisms of miRNA-mediated mRNA silencing.** (A) Initially, an Ago-miRNA complex binds a partially complementary target site on an mRNA which is typically in a closed loop conformation. (B, C) Ago recruits a member of the GW182 family which in turn interacts with the cytoplasmic Poly(A)-binding protein (PABPC) and thus breaks the closed loop configuration and leads to the establishment of silencing. GW182 also interacts with the CCR4/NOT complex which contains the deadenylases CCR4 and CAF1 that degrade the poly(A)-tail. (E, F) Subsequently, additional factors are recruited that lead to the decapping and then complete degradation of the mRNA. (D) An mRNA without a poly(A)-tail and an open conformation is silenced. In certain tissues or circumstances, the translationally repressed mRNA can be stored. Figure was taken from (Huntzinger and Izaurralde, 2011).

### 1.5.2 Function of miRNAs in plants

MiRNAs in plants are in majority 21-22nt in length and bear predominantly a U at the 5' end. They are specifically processed by DCL1 and sorted into AGO1. The AGO1-miRNA binds to its target (mostly to the open reading frame (ORF), but also to the 5' or 3' UTR) near perfectly complementary and this leads in most cases to its cleavage and degradation (Kasschau et al., 2003; Llave et al., 2002; Qi et al., 2005; Rhoades et al., 2002), much like RNAi-mediated knockdown of mRNAs in animals. AGO1 cannot only slice its target (Baumberger and Baulcombe, 2005), but can also inhibit its translation (Brodersen et al., 2008). Since miRNAs in plants bind their targets with near perfect complementarity, their target spectrum is much smaller than that of animal miRNAs. Additionally, miRNA and target

usually are evolutionary related. An overwhelming majority of the ancient miRNAs target developmentally important transcription factors (Jones-Rhoades and Bartel, 2004; Jones-Rhoades et al., 2006; Rhoades et al., 2002), whereas younger miRNAs have a wider range of targets (Axtell and Bowman, 2008).

## 1.6 Loading of Ago proteins and miRNA diversity

Loading efficiency of Argonaute proteins and the identity and characteristics of the miRNA strand that is incorporated are influenced by several factors. The most crucial factors will be described in the next paragraphs: In the last few years, the importance of chaperones in Argonaute loading has become apparent. Furthermore, the features of the miRNA itself (both as a duplex and mature strand) decide how it is sensed and treated by Dicer, its co-factors and finally, by Argonaute.

### 1.6.1 Chaperones and co-chaperones in RISC loading

Argonaute proteins alone bind to double-stranded RNA duplexes only with low affinity. Although they are capable of loading small RNAs from duplexes *in vitro* (Hauptmann et al., 2013; Noland and Doudna, 2013), they prefer a single-stranded substrate, when left without any auxiliary proteins (Hauptmann et al., 2013; Lima et al., 2009; Meister et al., 2005). This is probably due to the need for a conformational change within the protein to allow for the accommodation of the RNA duplex. In the last couple of years, several papers have elucidated the mechanism by which Argonaute proteins switch from an unloaded to a loaded state and a recurring theme of Argonaute loading competence involves the heat shock protein 90 (HSP90) protein family.

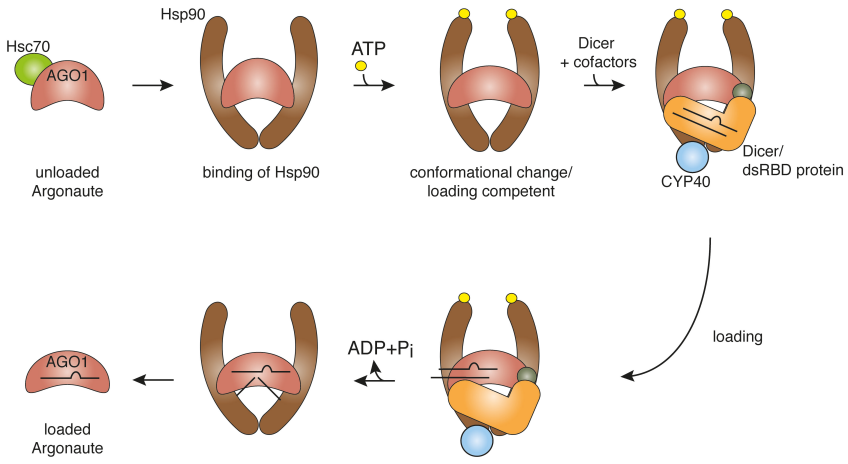



**Figure 1-11 HSP90-mediated loading of miRNAs into Argonaute proteins.** In plants, nascent AGO1 is first bound by the chaperone Hsc70 and then transferred to an HSP90 dimer. The binding of ATP to HSP90 causes a conformational change of HSP90 and consequently a conformational change of AGO1. In this complex, AGO1 is stable and also loading competent. The RISC loading complex is completed by the association of Dicer and the HSP90 co-chaperone CYP40, which facilitates loading of the miRNA duplex. Different to other organisms, the removal of the passenger strand is triggered by the hydrolysis of ATP and the resulting conformational changes. (Dueck and Meister, 2014)

HSP90 proteins constitute a large class of molecular chaperones that utilize ATP hydrolysis to support and control proper protein folding. HSP90 primarily binds to folded proteins that need to undergo conformational changes in order to reach their active states (Rohl et al., 2013). In several biochemical purifications of Argonaute protein complexes, chaperones such as HSP90 have been identified (Frohn et al., 2012; Hock et al., 2007; Liu et al., 2004; Maniataki and Mourelatos, 2005; Tahbaz et al., 2001; Weinmann et al., 2009). In plants and *Drosophila*, it was demonstrated that the chaperone HSP90 is necessary for the loading of small RNA duplexes into Argonaute proteins (Iki et al., 2010; Iwasaki et al., 2010; Miyoshi et al., 2010b). In a proposed model based on these findings, another heat shock protein, Heat shock cognate 70 (Hsc70), binds and folds the newly synthesized Argonaute protein during and after translation. It is then handed over to HSP90, which induces a conformational change to the newly synthesized Argonaute protein by binding ATP (Figure 1-11). This Argonaute state is stabilized until the small RNA is loaded. Now the Argonaute protein is loading competent and receives its small RNA duplex from a complex of Dicer and its associated factors (e.g. TRBP or PACT in mammals or R2D2 or Loqs in flies). In *Drosophila*, the following steps of RISC formation are independent of HSP90, whereas in plants, the removal of the passenger strand by AGO1 is triggered by ATP hydrolysis of HSP90 (Figure 1-11) (Iki et al., 2010; Iwasaki et al., 2010).

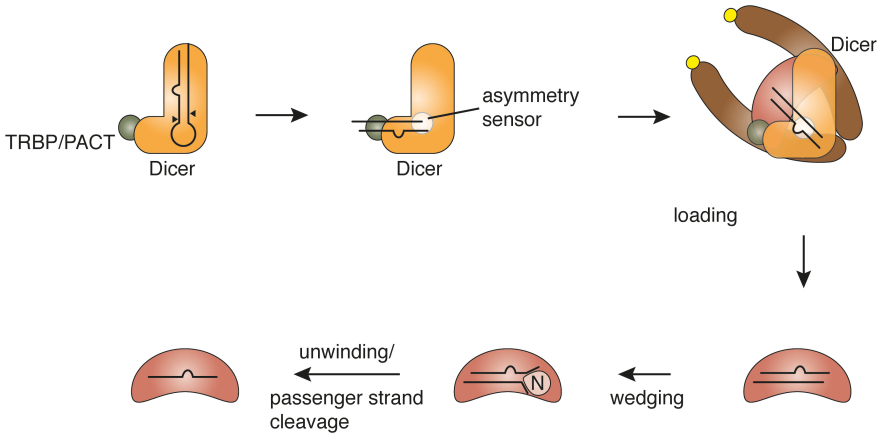



**Figure 1-12 Asymmetry sensing of the small RNA duplex and loading into Ago.** Dicer binds the miRNA precursor with its PAZ domain and cleaves off the terminal loop. The miRNA duplex shifts towards the helicase domain and a dsRBD protein (TRBP or PACT) that is bound by Dicer. Dicer, together with TRBP/PACT, senses the asymmetry of the duplex and aligns the duplex for loading it into Argonaute. Dicer and Argonaute interact via the RNase III domain and the Piwi domain, respectively (Tahbaz et al., 2004). Additionally, Argonaute is in complex with HSP90 that maintains Argonaute's open state (see also Figure 1-11). Presumably, the duplex is loaded into Argonaute, followed by a wedging mechanism by the N domain of Argonaute, thus unwinding the strand (Kwak and Tomari, 2012). Alternatively, cleavage-competent Argonautes can cleave the passenger strand in order to facilitate passenger strand removal (Diederichs and Haber, 2007). For simplification, HSP90 and Dicer have been omitted after the third step (Dueck and Meister, 2014).

After the transfer of the small RNA duplex to the Argonaute protein, catalytic Argonaute proteins can remove the passenger strand or the miRNA* sequence much faster than non-catalytic Argonaute proteins (see also this work) (Gu et al., 2011; Petri et al., 2011). Catalytic Argonaute proteins treat the passenger strand as complementary target RNA and cleave it. The two ~10 nt long fragments dissociate from the duplex more efficiently than the full length passenger strand (Leuschner et al.,

2006; Matranga et al., 2005; Rand et al., 2005). It is still not fully clear whether or not specific helicases are involved in duplex unwinding and passenger strand removal. However, it is becoming apparent that Argonaute proteins themselves might play a more active role in this process (Figure 1-12) (Kwak and Tomari, 2012; Wang et al., 2009).

In human and *Drosophila* cells, Argonaute proteins are unstable when not loaded or when HSP90 is inhibited (Johnston et al., 2010; Martinez et al., 2013; Pare et al., 2009; Smibert et al., 2013; Tahbaz et al., 2001). In addition, HSP90 is important for the interaction between Argonaute and Dicer via the Piwi and the RNase III domain, respectively (Tahbaz et al., 2004). In humans, the dependence of ATP and HSP90 on the actual loading step is under debate. On the one hand, several groups used immunopurified or recombinant Ago2 complexes to show the independence of Argonaute loading from ATP (Gregory et al., 2005; MacRae et al., 2008; Maniataki and Mourelatos, 2005), whereas on the other hand, Yoda and colleagues used *in vivo* and *in vitro* systems to show that there is a difference in miRNA incorporation and target cleavage only when using a cell lysate, not purified or recombinant Ago2 (Yoda et al., 2010). In such a pathway there likely exist bypasses for treating exceptions. Clearly, more evidence is needed to decide, which main route Argonaute proteins are taking to be loaded *in vivo*.

Recently, co-chaperones of HSP90 have been implicated in regulating the loading and consequently the level of miRNAs. These co-factors confer specificity to the HSP90 machinery by recruiting distinct client proteins to HSP90. Many of the involved co-chaperones belong to the family of immunophilins and contain a peptidyl prolyl cis/trans isomerase (PPI) domain as well as a tetratricopeptide repeat (TPR). They bind to the MEEVD motif of HSP90 via their TPR domain and also compete for this motif (Barent et al., 1998; Carrello et al., 1999). Although their functions have not been elucidated in detail so far, it is becoming clear that co-chaperones can facilitate the recognition of client proteins, display chaperone activity themselves or influence the ATPase activity of HSP90 (Rohl et al., 2013). In 2009, Poethig and colleagues showed that AGO1 probably is a major client of SQUINT, the *Arabidopsis* homolog of cyclophilin 40 (CYP40) (Smith et al., 2009). Studies in tobacco showed that CYP40 associates with AGO1 transiently and facilitates the binding of miRNA duplexes to AGO1. This process is HSP90 dependent (Iki et al., 2010). For humans, an association of Ago2 with the FK506-binding immunophilins FKBP4 and FKBP5 was detected (Frohn et al., 2012) and either FKBP4 together with p23, another co-chaperone of HSP90, or FKBP4/5 were implicated in the activation of RNAi (Martinez et al., 2013; Pare et al., 2013). Taken together, co-chaperones of HSP90 seem to play important roles in facilitating the assembly of RISC. It will be interesting to see whether different Argonaute proteins utilize different co-chaperone systems for small RNA loading and function.

### 1.6.2 Features of small RNAs that are important for loading

After the first small RNA cloning and sequencing studies it became clear that miRNAs predominantly start with a U at their 5' end. The crystallization of an isolated MID domain of human Ago2 identified a specific loop that discriminates between different 5' nucleotides and is therefore responsible for the

observed U bias (Frank et al., 2010). In many organisms, various different small RNA pathways exist and many of these small RNA classes contain specific 5' nucleotide biases (Czech and Hannon, 2011). It has been demonstrated in plants that the 5' nucleotide is sensed by the MID domain of the specific Argonaute protein indicating that the 5' nucleotide is an important feature of small RNAs for efficient loading into Argonaute proteins (see also Supplement 2, Table, Appendix) (Frank et al., 2012).

After Dicer processing, the small RNA duplex is transferred to the Argonaute protein and one strand is selected as the guide strand or the predominant miRNA arm. How is this selection achieved? SiRNA as well as miRNA duplexes are characterized by thermodynamic asymmetry, i.e. there are stronger and weaker paired ends. It has been shown that the strand with the less stably paired 5' end is preferentially selected and loaded into the Argonaute protein (Khvorova et al., 2003; Schwarz et al., 2003).

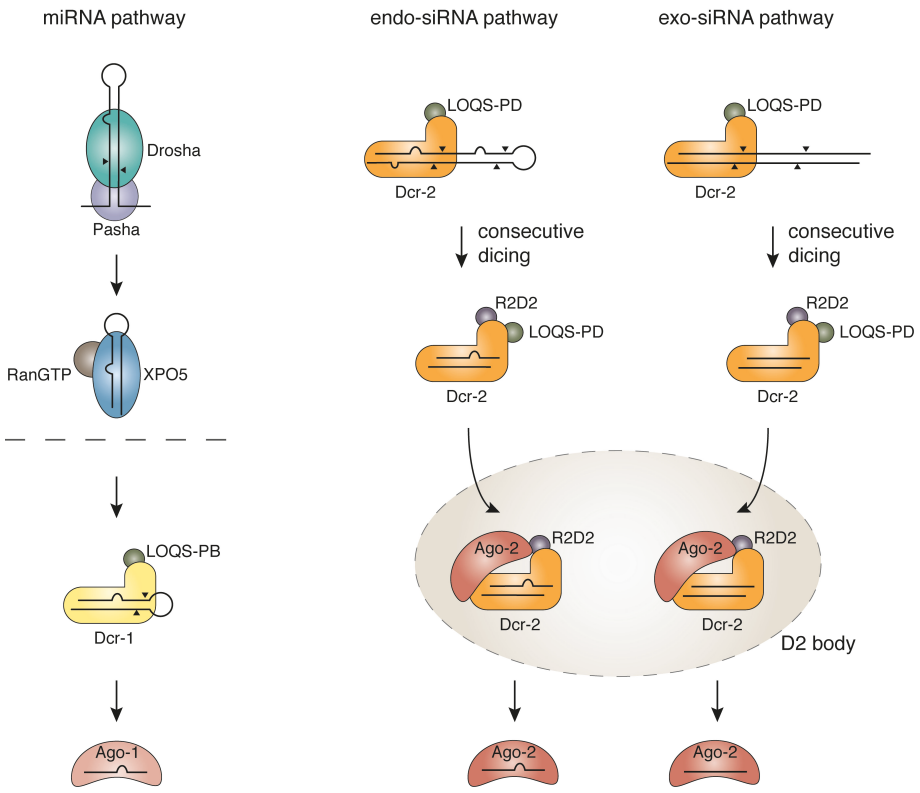



**Figure 1-13 Small RNA biogenesis pathways in *D.melanogaster*.** In somatic cells of *Drosophila melanogaster*, three major pathways of small RNA biogenesis are known. The miRNA biogenesis pathway is very similar to the mammalian pathway (see Figure 1-1). MiRNAs are transcribed in long transcripts in the nucleus and are processed by the Microprocessor complex (RNase III enzyme Drosha, in complex with the RNA-sensing protein Pasha). The resulting miRNA precursor is exported into the cytosol via XPO5 in a RanGTP dependent manner where it is bound by Dcr-1 in complex with the dsRBD protein LOQS-PB and processed into an RNA duplex. The mature miRNA strand is loaded into Ago-1 (see also text, Figure 1-11 and Figure 1-12). Endo-siRNA precursors are very long foldback structures with only very few mismatches. They are consecutively processed by Dcr-2 in association with the dsRBD protein LOQS-PD. When another dsRBD protein, R2D2, binds the Dcr-2/LOQS-PD/miRNA-duplex complex, it promotes the re-localization to so-called D2 bodies, where Ago-2 is loaded with the endo-siRNA. Similarly, Dcr-2 and LOQS-PD process long dsRNA molecules into exo-siRNAs. The complex re-localizes together with R2D2 to D2 bodies and Ago-2 is loaded with exo-siRNAs (Dueck and Meister, 2014).

In *Drosophila*, the dsRBD protein R2D2 is important for the asymmetry recognition (Tomari et al., 2004). R2D2 has two dsRBDs and tightly associates with Dcr-2 via its C-terminal part (Liu et al., 2003; Miyoshi et al., 2010a). It is able to sense the thermodynamic properties of the siRNA duplex and selects the strand with a less stable 5' end, whereas the other strand is usually degraded. Thus, thermodynamic asymmetry is a second important structural feature of the siRNA duplex that is important for loading.

Flies have a very efficient system of separating the miRNA from the siRNA pathway. In *Drosophila*, miRNAs are primarily associated with AGO1, while endo- and exo-siRNAs are associated with AGO2. How is this relatively clear separation managed? The answer to this question seems to lie both in the nature of the RNA molecule as well as the dsRBD cofactors of the Dicer proteins that sense some of these molecular differences. MiRNA biogenesis relies on Dicer-1 (Dcr-1) with its cofactor Loquacious (Loqs) (Figure 1-13) (Forstemann et al., 2005). The Loqs gene is expressed as 4 different isoforms, Loqs-PA, Los-PB, Loqs-PC and Loqs-PD. Loqs-PA and Loqs-PB contain 3 dsRBDs, while Loqs-PD lacks the third dsRBD and exhibits a unique C-terminal sequence (Hartig et al., 2009; Zhou et al., 2009). Loqs-PA and Loqs-PB both bind exclusively to Dcr-1 via their third dsRBD (Forstemann et al., 2005; Ye et al., 2007). Although they have partially redundant roles, it is Loqs-PB that is necessary and sufficient for processing of a subset of miRNAs and maintenance of female germ stem cells (Fukunaga et al., 2012; Park et al., 2007). Loqs-PB increases the affinity and turnover of Dcr-1 for some precursor miRNAs and can also influence the length of a certain subset of miRNAs (Fukunaga et al., 2012). The unique C-terminus of Loqs-PD mediates the binding to Dcr-2 rather than Dcr-1 (Hartig et al., 2009; Hartig and Forstemann, 2011; Zhou et al., 2009). Mutational studies *in vivo* already pointed to a direct role of Loqs-PD in endo-siRNA processing (Hartig et al., 2009; Hartig and Forstemann, 2011; Marques et al., 2010; Miyoshi et al., 2010a; Zhou et al., 2009). By generating flies expressing only one of the three Loqs isoforms, Zamore and colleagues showed that Loqs-PD indeed promotes the processing of both endo- and exo-siRNAs by Dcr-2 (Figure 1-13) (Fukunaga et al., 2012).

MiRNA precursors and endo-siRNAs often contain intra-molecular bulges. Exo-siRNA-precursors, however, are perfectly complementary dsRNA molecules derived from viruses, for example. Usually, bulged double stranded precursors are fueled into Ago1 while perfectly paired precursors are loaded into Ago2. Thus, a third crucial feature of small RNAs for loading in *Drosophila* is the structure and the double stranded nature of the small RNA precursor. In the case of endo-siRNAs, however, a mechanism seems to exist that results in the exclusive loading of Ago2 with endo-siRNAs, but not Ago1. Recent studies have uncovered that depletion of R2D2 leads to the mis-direction of endo-siRNAs into Ago1, although a detailed mechanism is still elusive (Ameres et al., 2011; Marques et al., 2010; Okamura et al., 2011).

In mammals, only one Dicer exits and the miRNA pathway is not as clearly separated from the endo-siRNA pathway with respect to their biogenesis factors as observed for *Drosophila*. The dsRBD proteins TRBP and PACT have been identified as critical co-factors for Dicer processing

(Chendrimada et al., 2005; Gregory et al., 2005; Haase et al., 2005; Lee et al., 2006). However, whether these proteins function in a similar way as their fly relatives, is currently unclear.

In conclusion, several factors influence the processing, the loading and choice of Ago protein. Although a lot of effort has been put into analyzing these characteristics, many aspects of small RNA biogenesis and loading need to be elucidated in the future.

## 1.7 Aim of this work

The aim of this work was to unravel principles and mechanisms underlying the specific association of small RNAs with different Ago proteins, their distinct expression patterns and their functional impact in different cell types and organisms.

Beside their small RNA content, the four human Argonaute proteins can be discriminated by the presence or absence of catalytic activity (Meister et al., 2004). Additional discriminators are their abundance or potentially their ability to eject the passenger strand of a small RNA duplex. Therefore, the experiments in Chapter 2 serve to analyze in more detail the abundance and cleavage activity of human Ago proteins. Further, the impact of the thermodynamic stability of the small RNA duplex on Ago loading and targeting were dissected. The results of this chapter have been published in *RNA* (Petri et al., 2011) and *Nature Structural and Molecular Biology* (Hauptmann et al., 2013).

Human Ago proteins are almost identical and it is not clear what their specific functions are. To understand this better, we have analyzed small RNAs specifically associated with human Agos in Chapter 3. I established a cloning protocol in the lab and analyzed the small RNAs bound to all four human Ago proteins in HeLa S3 cells by high throughput sequencing and subsequent validation by Northern blotting. MiR-451 is unique with respect to its processing (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010a). Since it was already known that miR-451 is processed in a Dicer-independent manner by the cleavage-active Ago2, it was tempting to speculate that mature miR-451 is bound by Ago2 only. Therefore, the aim of this analysis was to also characterize miR-451 association with Ago proteins and the determinants for its efficient processing by Ago2. The results of this project have been published in *Nucleic Acids Research* (Dueck et al., 2012).

In addition to the specific Argonaute characteristics mentioned above, miRNA levels are important regulators of cellular function. It is well known that miRNAs are induced and regulated under different conditions like differentiation processes, stress or cancer. However, the impact and influence of one miRNA on putative downstream miRNAs has not been analyzed in detail so far. MiR-155 is a highly important miRNA in immune cells, it has been characterized in different cell types and is known to be dramatically induced upon immune cell maturation, for example. For miR-155, a mouse model deficient in miR-155 exists. In Chapter 4, I therefore chose an immunological cell type setting to measure miRNA levels in wild type and miR-155 deficient cells under different conditions and to

analyze potential miRNA hierarchies. The findings of this chapter have been published in *FEBS Letters* (Dueck et al., 2014).

The knowledge of miRNAs in animals has been expanding greatly. In plants, many small RNA classes and their mechanisms seem to be very similar, however, there are distinct differences. *Volvox carteri* is a simple green alga and its status as a model system for the evolution of multicellularity and its close relationship to the unicellular green alga *Chlamydomonas reinhardtii* renders it an intriguing organism to study small RNA function. It was the aim of this part (Chapter 5) to identify genes involved in small RNA biogenesis and function and to characterize the small RNAs associated with an Argonaute protein of *Volvox carteri*. This approach yields to improve our knowledge of small RNA expression and function in a simple yet specialized organism such as *Volvox carteri* and understand its implications for the evolution of small RNAs and their functions.

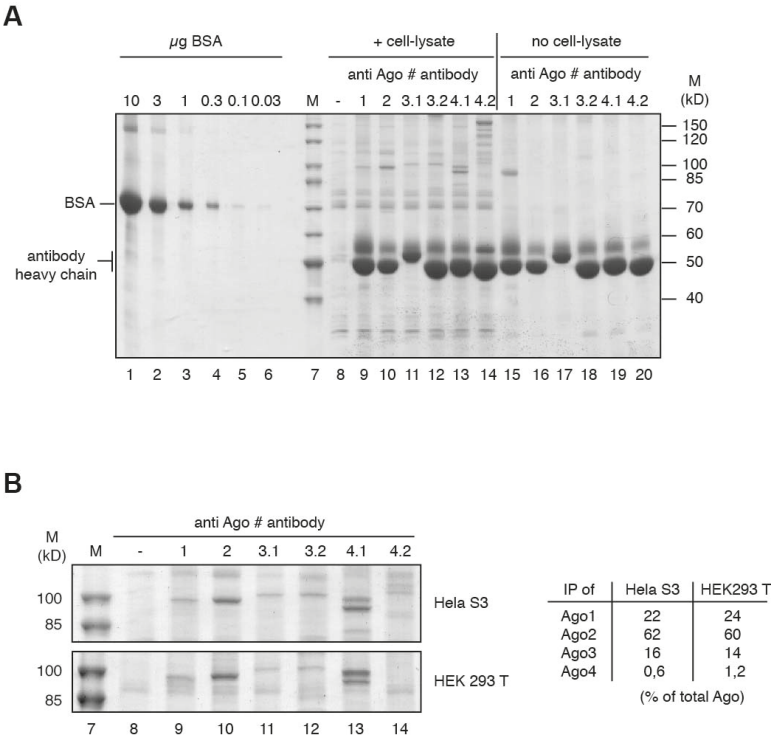
# CHAPTER 2 CHARACTERISTICS OF HUMAN AGO PROTEINS

The human Ago proteins are very similar in sequence and it is a long-standing question in the field whether or not they have specific and/or redundant functions. Besides the knowledge of Ago2 being an endonuclease (Meister et al., 2004), it is not clear what features discriminate the different Ago proteins. In order to characterize human Ago proteins in more detail, we therefore measured their relative cellular abundances, their substrate preference and re-analyzed the cleavage activity of the human Ago proteins. Finally, properties like strand loading and off target activity were examined using siRNAs with modified nucleotides.

The results of the following chapter have been published in the publications (Petri et al., 2011) and (Hauptmann et al., 2013). I contributed the experiments presented in Figures 1, 2E, 5 and 6 (Petri et al.) and Figure 1 (Hauptmann et al.), respectively.

## 2.1 Relative abundance of Ago proteins in human cell lines

In earlier studies, the relative abundances of Ago proteins have only been determined on the mRNA level by quantitative real-time PCR (qRT-PCR) (Gonzalez-Gonzalez et al., 2008; Meister et al., 2004). However, it is crucial for cellular processes as well as for RNAi experiments to know the relative protein levels of non-catalytic Ago1/3/4 and catalytically active Ago2. It has been shown before that the non-catalytic Ago proteins contribute to on- and off-target effects in RNAi experiments (Wu et al., 2008).




| IP of | Hela S3 | HEK293 T |
|---|---|---|
| Ago1 | 22 | 24 |
| Ago2 | 62 | 60 |
| Ago3 | 16 | 14 |
| Ago4 | 0,6 | 1,2 |

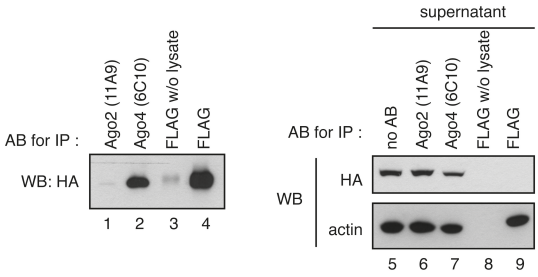
(% of total Ago)

**Figure 2-1 Estimation of Ago1-4 protein amounts in different human cell-lines.** (A) Cell-lysate was prepared from HeLa S3 cells and Ago1-4 proteins were immunoprecipitated using monoclonal antibodies against Ago1 (lane 9), Ago2 (lane 10), Ago3 (two different antibodies 3.1 and 3.2, lanes 11 and 12) and Ago4 (two different antibodies 4.1 and 4.2, lanes 13 and 14). The samples were analyzed on a SDS-polyacrylamide-gel together with a BSA standard (lanes 1-6) and stained with colloidal coomassie. In lanes 15-20, lysate was omitted from the IP. M, molecular weight marker. (B) Cell-lysate was prepared from HEK 293T cells and Ago1-4 proteins were immunoprecipitated as described in (A). The protein bands at a size of about 100 kD were analyzed by mass spectrometry (MS). The signal intensities of all peptides for each Ago protein were used to compare the relative amounts of Ago1-4 with each other. For Ago3, the intensities from both antibody clones 3.1 and 3.2 were averaged. For Ago4 only signals from antibody clone 4.1 were used (table) (Petri et al., 2011).

For measuring cellular Ago protein levels, we immunoprecipitated the endogenous proteins using monoclonal antibodies that discriminate between the human Ago proteins (Figure 2-1) (Beitzinger et al., 2007; Rudel et al., 2008; Weinmann et al., 2009). We isolated Ago1-4 proteins from HeLa S3 and HEK 293T cells and analyzed them by mass spectrometry (Figure 2-1, A and B). In Ago1-3 IPs, the bands with a molecular weight of about 100 kDa were clearly identified as Ago1, Ago2 and Ago3, respectively.

For Ago4, however, antibody 4.1 (clone 6C10) precipitated very little Ago4 and mainly cross reacts with an unrelated protein. Another anti-Ago4 antibody (4.2, clone 3G5) does not show such cross reactivities but endogenous Ago4 was not detectable in the immunoprecipitates suggesting that Ago4 is low abundant in HEK 293T or HeLa S3 cells. To rule out that the low Ago4 levels in IP experiments are due to low antibody affinity, we immunoprecipitated FLAG/HA-tagged (FH) Ago4 from a stably transfected HeLa S3 cell line (Figure 2-2, A). Indeed, the antibody 6C10 immunoprecipitated FH-Ago4 efficiently supporting the idea that the Ago4 protein is rather low abundant in HEK 293T and HeLa S3 cells (Petri et al., 2011).

**A**




**B**

| cell-line | derived from | relative Ago4 MS-signal (compared to HEK 293T) |
|---|---|---|
| HEK 293T | transformed embryonic kidney | 1 |
| Hela CCL2 | cervix carcinoma | 0,03 |
| LNT-229 | glioblastoma cells | 0,04 |
| HCT-116 | colon carcinoma | 0,12 |
| ARPE-19 | retinal pigmented epithelium | 0,61 |
| DG-75 | Burkitt lymphoma | 0,05 |
| HEP-G2 | hepatocellular carcinoma | not identified |
| LNCAP | prostate carcinoma | 1,13 |
| MCF-7 | breast adenocarcinoma | not identified |
| U2OS | bone marrow carcinoma | 0,77 |
| GM5657 | transformed fibroblasts | not identified |
| BJT | transformed fibroblasts | 0,20 |

**Figure 2-2 Analyzing the Ago4 antibody specificity and Ago4 level in different cell lines** (A) Lysate from a cell line stably expressing FLAG-HA-tagged Ago4 was prepared as described in Figure 2-1, A. The IP of Ago4 by the anti-Ago4 antibody (lane 2) was compared to the IP of tagged Ago4 with anti-FLAG antibody (lane 4). Anti-Ago2 (11A9) IP from FH-Ago4-expressing cells served as control. The experiment was analyzed by Western blotting using an antibody against the HA-tag. In addition, the supernatants of the IP samples were analyzed using the anti-HA-antibody and an anti-beta-actin antibody (right panel). Anti-Ago2 (11A9) IP from FH-Ago4 expressing cell lysates served as control (lane 1). (B) Ago4 was immunoprecipitated from lysates of different cell-lines with antibody 6C10 as described in Figure 2-1, A. The IPs were analyzed by MS. Sums of the intensities of all peptides identified for Ago4 were compared among the different cell lines as described in Figure 2-1, B. The peptide intensity sum from HEK 293T was defined as 1 (corresponding to about 1 % of all Ago in HEK 293T).

Using the same approach, we set out to measure relative Ago protein levels in HEK 293T and HeLa S3 cells. Defined amounts of BSA served as standard and we estimated the intracellular concentration of Ago2 as approximately 100 nM (Figure 2-1, A). A comparison of the overall peptide signal intensities of each Ago protein in our mass spectrometry experiments revealed that Ago2 represents about 60%, Ago1 20-25%, Ago3 about 15% and Ago4 about 1% of the total Ago protein pool in HEK 293T or HeLa S3 cells (Figure 2-1, B). In search for Ago4 expressing cells, we analyzed 13 different cell lines and from none of them we could immunoprecipitate higher amounts of Ago4 (Figure 2-2, B), indicating that contribution of Ago4 to RNAi on- and off-target effects in the analyzed cell lines might be negligible.

In summary, we have measured cellular Ago protein levels and found that Ago2 is the most abundant Ago protein followed by Ago1 and Ago3 in the commonly used cell lines HEK 293T and HeLa S3. In many cell lines, only little Ago4 protein is found.

## 2.2 Cleavage activity of human Ago proteins

It has been well established that the human Ago2 is an endonuclease, cleaving not only a perfect complementary target to the guide strand (Meister et al., 2004), but also the passenger strand of an RNA duplex to facilitate RISC maturation (Leuschner et al., 2006; Matranga et al., 2005; Rand et al., 2005). Ago1, Ago3 and Ago4 were shown to be cleavage-incompetent in these studies. In the last few years, however, the cleavage activity of Ago1 has been under debate. Novina and colleagues showed for Ago1 (recombinantly expressed as GST-fusion protein in *E.coli*) passenger strand cleavage (but not mRNA target cleavage) in a minimal *in vitro* system (Wang et al., 2009), although Ago1 does not harbor an active catalytic tetrad (DEDR instead of DEDH).

To reanalyze the cleavage activity and affinity to ss or dsRNA of human Ago1-4, we set up an *in vitro* system using similar buffer conditions to Novina and colleagues (Wang et al., 2009).

Human Ago1-4 were purified from insect cells, the tag was removed and the quality examined by electrophoresis (Figure 2-3, A). Equal amounts of protein were incubated with either a fully complementary siRNA containing a radiolabelled 5' phosphate on the passenger strand (Figure 2-3, B) or first incubated with guide strand followed by an incubation with the radiolabelled passenger strand (Figure 2-3, C). The results clearly show that Ago2 is cleavage active in both conditions (lane 3), whereas Ago1, Ago3 and Ago4 are not. Additionally, the affinity for a single strand RNA is clearly higher than for the pre-annealed double stranded duplex.

Thus, our data clearly shows that Ago1 cleaves neither target RNAs nor passenger strands within an siRNA duplex.

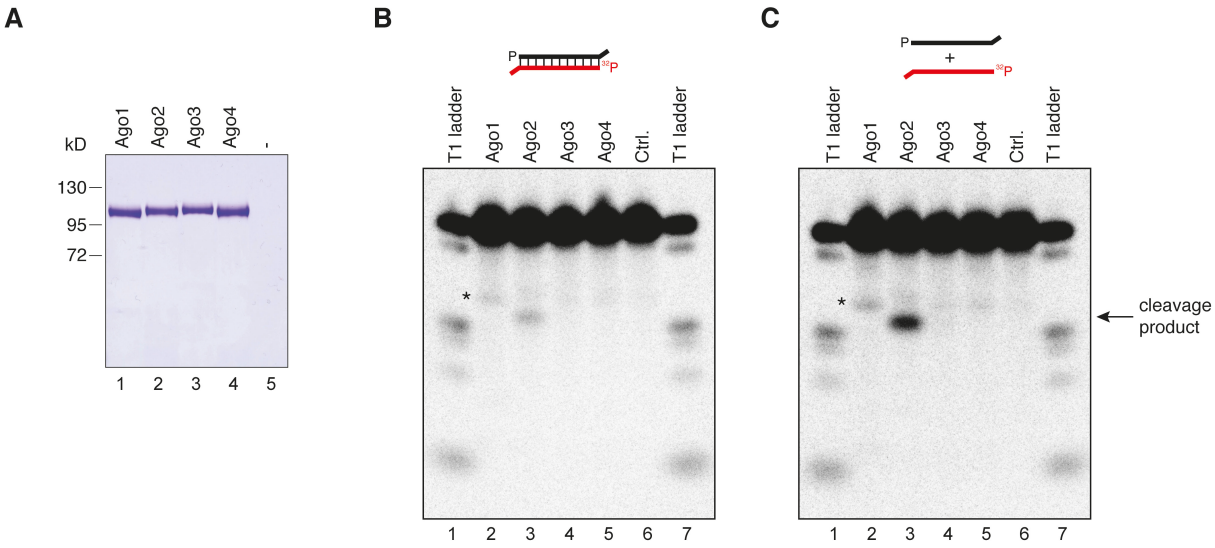



**Figure 2-3 Passenger strand cleavage activity.** (A) SDS-PAGE showing the recombinant human Ago1-4 preparations used for these experiments. (B) Ago proteins were incubated with a pre-annealed siRNA containing a radiolabeled passenger strand (red) or (C) a single stranded guide strand that was supplemented with the passenger strand after Ago-binding. Processing products were analyzed on denaturing RNA gels. The asterisk indicates an unspecific product. -/ Ctrl., no protein was added (Hauptmann et al., 2013).

## 2.3 SiRNA strand incorporation into different Ago proteins

We established above that Ago2 is the most abundant Ago protein in HEK 293T and HeLa cells and also harbors the only catalytic activity in the human Ago clade. Since Ago2 is able to cleave the passenger strand of a miRNA or siRNA duplex, the question arose if there are differences in the loading of siRNA duplexes. We therefore used our immunoprecipitation approach (see 2.1) to analyze siRNA strand loading into the four different human Ago proteins in detail. A thermodynamically asymmetric siRNA (one strands is preferentially incorporated into RISC, siRNA#1, Figure 2-4, A) was transfected into HEK 293T cells and the individual Ago proteins were immunoprecipitated as described above. Small RNAs were extracted and analyzed by Northern blotting using probes against the guide or the passenger strand (Figure 2-4, B). Since different probes have different binding affinities, we loaded the corresponding synthetic siRNA strand onto each Northern blot for signal normalization. As expected for an asymmetric siRNA, the guide strand is incorporated into the different Ago proteins according to their protein expression levels (Figure 2-4, B and C, upper panel). The passenger strand, however, is present in Ago1- and Ago3- but not in Ago2-containing RISC complexes (Figure 2-4, B and C, lower panel). Similar results were obtained for several other siRNA duplexes (data not shown).

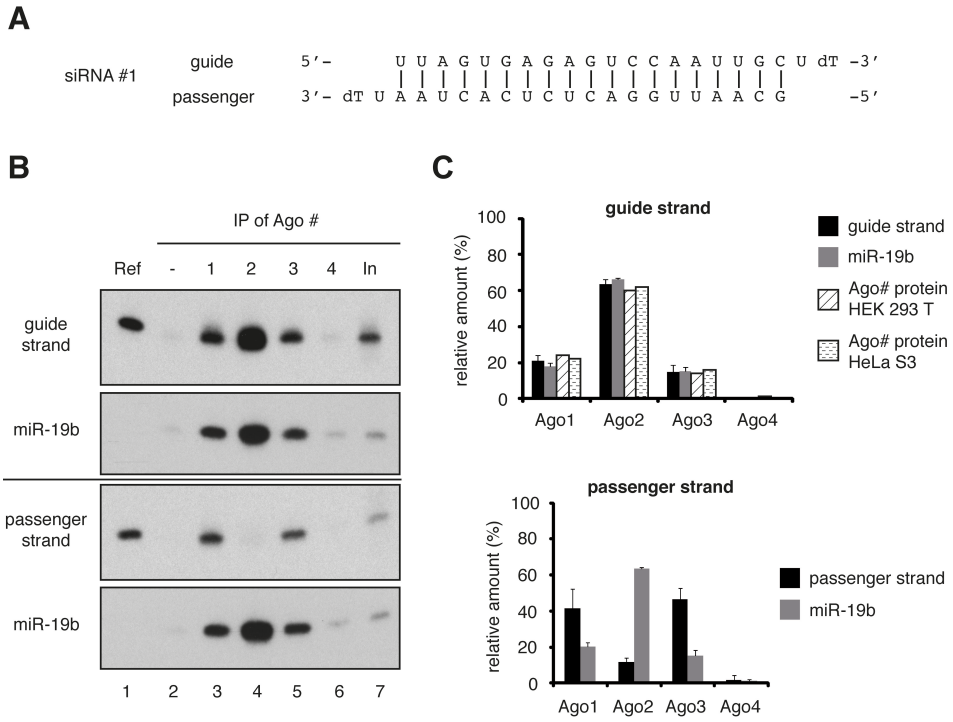



**Figure 2-4 Analysis of siRNA strand loading into different Ago proteins.** (A) The thermodynamically asymmetric siRNA#1 used for the following experiments. (B) Cells were lysed 20h after transfection and Ago1 (lane 3), Ago2 (lane 4), Ago3 (lane 5) and Ago4 (lane 6) were immunoprecipitated from HEK 293T lysates. The co-immunoprecipitated siRNA#1 guide strand (upper panel) as well as the passenger strand (lower panel) was analyzed by Northern blotting. Blots were stripped and re-probed for miR-19b. Ref, 3 pmol of the respective siRNA single strand; In, 2% of the input sample. (C) The radioactive signals from Northern blots of three independent experiments as shown in (B) were quantified using a phosphoimager system and the relative signal intensities of the siRNA#1 guide strand (upper graph, black columns), the passenger strand (lower graph, black columns) and miR-19b (grey columns) were compared between the four Ago proteins. Standard deviations of the signals are indicated. In addition, the relative protein amounts of the individual Ago proteins in HEK 293T and HeLa S3 cells as calculated in Figure 2-1, A, B are included in the upper graph (diagonally and horizontally hatched columns).

The presence of the passenger strand on Ago1 and Ago3 could be due to unwound double stranded siRNA or would also be in agreement with the idea that Ago1 and Ago3 do not follow the asymmetry rule and load both strands equally well. We therefore designed an experiment that differentiates between a bound single strand and a bound duplex. When a single strand is incorporated, it is accessible for target recognition and binding. On the other hand, if a duplex is incorporated, both strands are naturally already binding an RNA molecule (binding to itself) and can thus not interact with a target RNA. We made use of this notion and cloned target sites for the passenger strand and the guide strand into the 3' untranslated region (UTR) of a firefly luciferase reporter. To prevent cleavage by Ago2, the target site is not complementary to the individual siRNA strand at positions 10 and 11 of the siRNA (Figure 2-5, A), since that is Ago2's natural cleavage site. Reporter constructs were co-transfected with siRNA#1 into HEK 293T cells. Ago1 and Ago2 were immunoprecipitated using monoclonal antibodies and co-precipitated luciferase reporter mRNA was analyzed by qRT-PCR (Figure 2-5, B). As expected, the anti-Ago2 antibody predominantly co-immunoprecipitated the mRNA targeted by the siRNA guide strand. Antibodies directed against Ago1, however, did not immunoprecipitate significant amounts of target mRNA suggesting that Ago1 carries mainly siRNA duplexes thereby preventing interactions with the corresponding target RNA (Figure 2-5, B). An

artificial target of the endogenous miR-19b, however, was co-immunoprecipitated by the anti-Ago1 and the anti-Ago2 antibody with similar efficiencies. The weaker signal for anti-Ago1 IPs is due to lower Ago1 protein levels (Figure 2-5, B).

Taken together, our experiment indicates that non-catalytic Ago proteins only inefficiently discard fully complementary passenger strands and cannot associate with target RNA. The overall thermodynamic stability of an siRNA duplex seems therefore to influence passenger strand removal of non-catalytic Ago proteins.

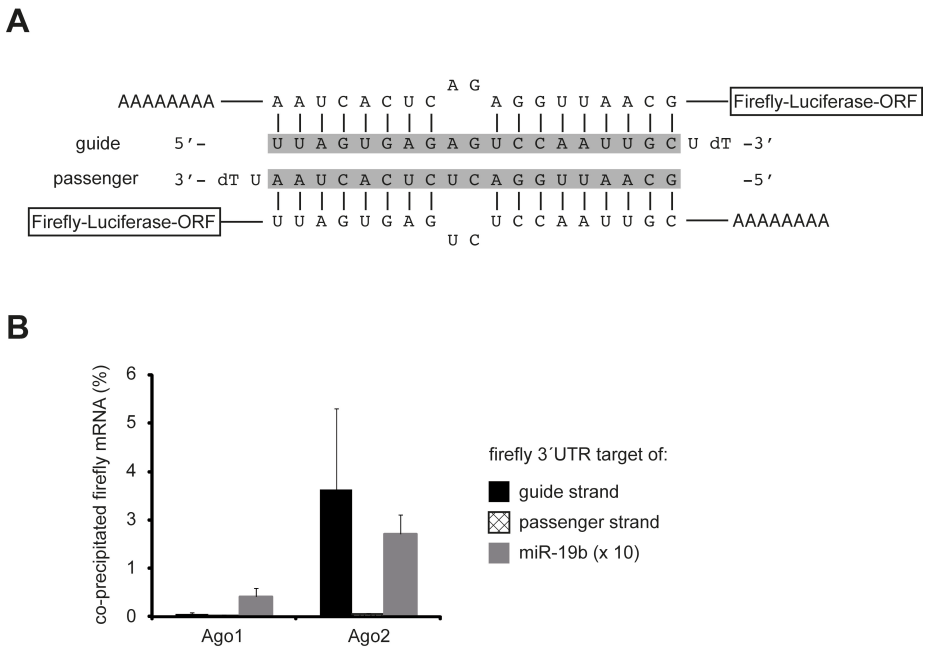



**Figure 2-5 Analysis of the nature of strand binding.** (A) Targeting sequences for the guide (upper part) or the passenger strand (lower part) of siRNA#1 were cloned into the 3´ UTR of the firefly luciferase mRNA of a pMIR-REPORT vector. The siRNA targeting sequences contain mismatches in the annealing-position 10 and 11 of the respective siRNA strands in order to avoid cleavage by Ago2. As reference, a construct containing a targeting sequence for miR-19b also with mismatches in annealing-positions 10 and 11 of the miRNA was used (not shown). (B) HEK 293T cells were transfected with siRNA#1 and either a pMIR-REPORT vector containing a targeting site for the guide strand (black columns), the passenger strand (cross hatched columns) or miR-19b (grey columns). 20h post-transfection, cells were lysed and Ago1 or Ago2 immunoprecipitated from the lysates. The co-immunoprecipitated RNA was isolated, transcribed into cDNA and the relative amount of firefly luciferase mRNA in each sample was analyzed by qRT-PCR. The graph analyzes the relative amounts (%) compared to the input of the respective mRNAs co-immunoprecipitated with the Ago-siRNA or Ago-miRNA complexes.

## 2.4 Analyzing off-target effects using thermodynamically stabilized siRNA duplexes

The experiments above showed that non-catalytic Ago proteins cannot discard the passenger strand by cleavage and indicate that a fully complementary siRNA is bound as a duplex rather than as the guide strand alone. Performing a kinetic of passenger strand displacement over 3 days with Ago1-3, Petri et al. further show that non-catalytic Ago proteins remove passenger strands inefficiently (Petri et al., 2011). Ago2 almost immediately discards the passenger strand, probably due to its cleavage

activity, whereas Ago1 and Ago3 take 2-3 days to fully displace the passenger strand. To test whether Ago2's cleavage activity is responsible for the efficient removal of the passenger strand, the authors used a cleavage-inactive mutant that is still able to bind small RNAs. This mutant protein behaves very similarly to Ago1, namely that it cannot remove the passenger strand efficiently, confirming that the catalytic activity of Ago2 is responsible for passenger strand displacement (Petri et al., 2011). Next, Petri et al. investigated the influence of modified nucleotides on strand incorporation. For the thermal stabilization of siRNA duplexes modifications such as 2' O-methylated nucleotides (Cummins et al., 1995; Majlessi et al., 1998), locked nucleic acids (LNAs) (Kumar et al., 1998; Obika et al., 1997) or unlocked nucleic acids (UNAs) (Fluiter et al., 2009; Laursen et al., 2010) have successfully been used. Petri and colleagues used LNA modified passenger strands to show that this stabilization traps non-catalytic Ago proteins on double-stranded siRNAs. In addition to modified nucleotides, the increase of the GC content of an siRNA duplex led to further selective incorporation of guide strand into Ago2, whereas Ago1 and Ago3 are trapped with duplex (Petri et al., 2011).

In order to analyze whether the reduced strand displacement of Ago1 and Ago3 leads to reduced off-target effects due to Ago1 and Ago3 (but not Ago2), we performed luciferase reporter experiments in different cell lines (Figure 2-6). Three partially complementary binding sites for the guide strand of the siRNA#2 duplex (Figure 2-6, A) were cloned behind a firefly luciferase reporter (Figure 2-6, B). The reporter plasmid was transfected into HEK 293T cells together with two different concentrations (1.6 or 40 nM) of either the unmodified (unstabilized) or the LNA-modified (stabilized) version of siRNA#2 (Figure 2-6, A). At both concentrations, the LNA-modified siRNA#2 repressed the off-target reporter less efficiently. The effects, although rather mild, were nevertheless significant (Figure 2-6, C, left panel).

In order to distinguish between Ago2- and Ago1/3/4-mediated off-target effects, we changed our experimental system and switched to Ago2-deficient mouse embryonic fibroblasts (MEFs). Off-target effects observed in these cells are caused by Ago1, Ago3 and/or Ago4 only. The off-target reporter was transfected together with two concentrations of either the unmodified or the LNA-modified siRNA#2 into wild type (wt) or Ago2-deficient MEFs (Figure 2-6, C, middle and right panel). Indeed, Ago1/3/4-mediated off-target effects were not measureable for the LNA-modified version of siRNA#2 compared to the unmodified siRNA#2 at the higher siRNA concentration (40 nM) in Ago2-deficient MEFs, although there was a pronounced off-target effect of the unmodified siRNA#2 (Figure 2-6, C, right panel). Similar results were obtained at later time points (data not shown). Not much change was observed when low concentrations (1.6 nM) of both variants of siRNA#2 were transfected, most likely due to reduced overall activity. Off-target effects are similar between wt and Ago2-deficient MEFs suggesting that mainly Ago1/3 and 4 might cause the observed off-target effects in the wt MEFs as well as in the Ago2-deficient MEFs, while Ago2 might cause a major portion of the off-target effects observed in HEK 293T cells. In summary, our data suggest that thermodynamically stabilized siRNAs duplexes can almost completely eliminate Ago1/3/4-mediated off-target effects.

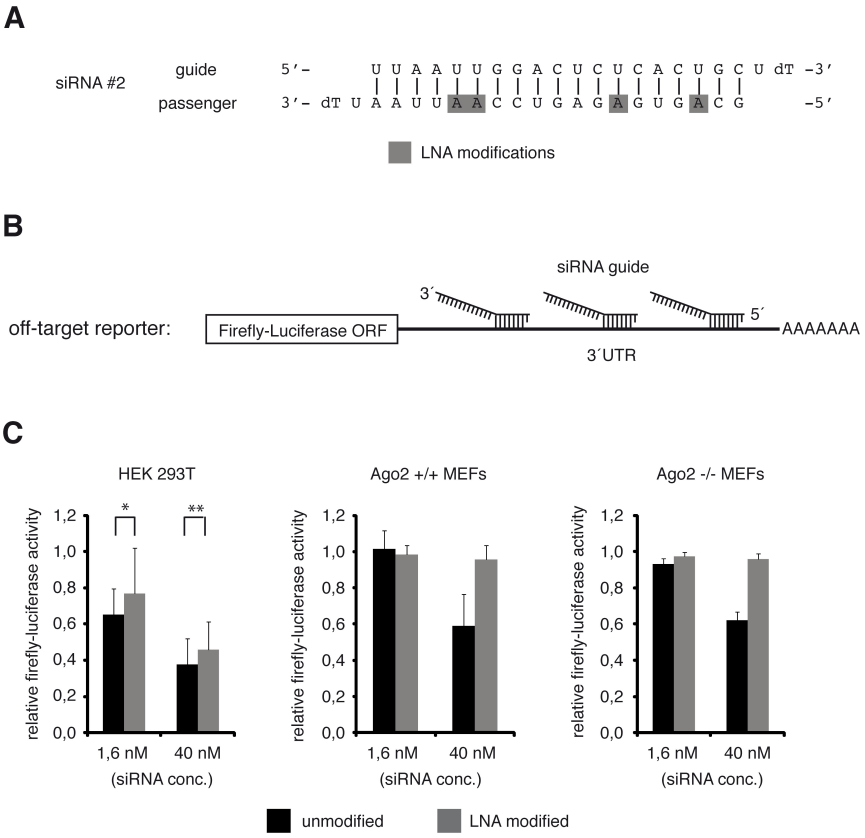



**Figure 2-6 Measurement of off-target activity of an siRNA with modified nucleotides.** (A) The thermodynamically asymmetric siRNA#2 used for the following experiments. LNA modified nucleotides are shaded in grey. (B) Scheme of the luciferase based off-target reporter to measure siRNA#2 guide strand off-target activity. The 3´UTR of the Firefly-luciferase contains three targeting-sites complementary to the seed region of siRNA#2 guide strand. (C) Off-target reporter activity in the presence of indicated concentrations of unmodified siRNA#2 (black bars) and LNA-modified siRNA#2 (grey bars) was measured 48h post transfection in HEK 293T cells, Ago2$^{+/+}$ MEFs and Ago2$^{-/-}$ MEFs. For (C), P-values were calculated (t-test): * - 0,109. ** - 0,007.

## 2.5 Discussion

Although our knowledge on miRNA biogenesis and function has been greatly expanding in the last decade, many aspects of Argonaute proteins are still unknown. The use of siRNAs to deplete cellular mRNAs and evidence of unwanted side-effects have made the analysis of Argonaute properties even more important.

To interpret the intrinsic protein properties, it is important to obtain a reliable assessment of the relative amounts of the human Ago proteins in our cellular systems. The analysis of relative Ago protein amounts yielded the expected result that Ago2 is by far the most abundant Ago protein in HEK 293T and HeLa cells while Ago1 and Ago3 are present in only minor amounts. Although many different cell lines were tested for the expression of Ago4, none could be identified to express any considerable amount. Ago4 mRNA levels have been quantified before and mRNA levels significantly differ from the Ago4 protein levels measured here (Gonzalez-Gonzalez et al., 2008; Meister et al., 2004). Therefore, it is very likely that Ago4 expression is to some extent regulated post-transcriptionally. Our results are supported by the measurements of other groups in different systems (Valdmanis et al., 2012; Wang et al., 2012).

The endonuclease activity of Ago2 has been well established (Meister et al., 2004; Wang et al., 2012). Additionally, we now show that the other Ago proteins definitely do not contain such an activity (Figure 2-3). The reason for this inactivity lies in the amino acid sequence and structure of Ago1 and Ago3. Changing the backbone of the respective protein by insertion of Ago2 pieces, two recent sudies succeeded in activating these otherwise silent Argonautes (Hauptmann et al., 2013; Schurmann et al., 2013).

The community has been aware for quite some time of off-target effects (Svoboda, 2007). Generally, sequence-specific off-target activity is thought to originate from the seed regions (nucleotides 2-8) of the passenger and the guide strand and affects mRNAs containing sequence motifs complementary to passenger and guide strand seed regions (Birmingham et al., 2006; Jackson et al., 2006; Lin et al., 2005). In order to avoid siRNA passenger strand off-target activity and favor guide strand loading, several studies have analyzed siRNA strand selection and established siRNA design concepts that very efficiently help to avoid loading of the passenger strands into functional RISC complexes (Bramsen et al., 2007; Bramsen et al., 2009; Chen et al., 2008; Khvorova et al., 2003; Schwarz et al., 2003). The slicer activity of Ago2 is the desired action for performing on-target siRNA-mediated knock down of mRNAs. Although we have established that Ago2 represents over 60% of the total Ago pool (Figure 2-1), the remaining ~40% of Ago1 and Ago3 cannot contribute to silencing the mRNA target by cleavage and therefore potentially generate off-target activity. However, taking into account the intrinsic protein properties of Ago1, Ago3 and Ago4, we achieved to dramatically lower the off-target activity of siRNAs by basically excluding Ago1/3/4 from participating. An increased stability of the siRNA duplex, i.e. by increasing the GC content (Petri et al., 2011) or by using modified nucleotides (Figure 2-6), leads to an incorporation of the duplex into Ago1/3/4, but they fail to unwind the duplex

(and are unable to cleave the passenger strand) and thus are incapable to exert their off-target activity with this siRNA.

In addition, it could be very interesting to combine our approach of siRNA passenger strand modification to generate thermodynamically stabilized siRNA duplexes with guide strand modifications. Promising combinations could be siRNA passenger strand modifications in the seed region which have been described to cause reduced siRNA off-target activity, for example 2'-O-methyl nucleotides (Jackson et al., 2006) or UNA nucleotides (Bramsen et al., 2010; Laursen et al., 2010; Vaish et al., 2010). Since siRNAs are being used as drugs our siRNA selection parameters may also have impact on siRNA-based therapy (Petri et al., 2011).

Our analyses have shed light on several intrinsic properties of Argonaute proteins in human cells. The results are not only important for understanding cellular processes involving all Ago proteins, but also can help improving technical applications such as siRNA-mediated knock down of endogenous mRNAs.

# CHAPTER 3 ANALYSIS OF MIRNAS ASSOCIATED WITH THE DIFFERENT HUMAN AGO PROTEINS

The specific sequence of miRNAs binding to individual Argonaute proteins naturally has a high impact on the nature of regulation (reviewed in (Czech and Hannon, 2011)). Whereas a sorting of miRNAs in plants and *Drosophila* is evident and some rules for the sorting and loading of miRNAs have been shown (Dueck and Meister, 2014), evidence and reports in humans are sparse. Besides the general identity of the miRNA being bound by an Ago protein, it has become evident that mature miRNAs with slightly differing lengths could be observed for most animal miRNAs. Therefore, it is not only critical to understand the association of miRNAs with endogenous Ago proteins but also to comprehend the extent of different length isoforms.

In this chapter, I analyzed the association of miRNAs with the four different Ago proteins in an endogenous setting. After investigating putative sorting of miRNAs in humans, I focused on miRNA variations and the extent of binding to each Ago protein. MiR-451 is a unique example for a miRNA that is processed Dicer-independently by Ago2 (Figure 1-2, left) (Cheloufi et al., 2010; Cifuentes et al., 2010; Yang et al., 2010a). I analyzed the sorting of this miRNA as well as determined features critical for Ago2-processing of this miRNA.

During the course of our analyses we found that many Ago-associated miRNAs are shortened or extended. This “isomiR” phenomenon has been observed before (Kuchenbauer et al., 2008) and for a better understanding of the following results, the underlying principles are briefly introduced.

This chapter is based on the publication (Dueck et al., 2012). A. Dueck conducted all experiments with help from Ch. Ziegler on figure 3. A. Eichner helped with preparing single-cell suspensions from mouse organs. E. Berezikov performed bioinformatic analyses on deep sequencing data (Figures 4-6). A. Dueck and G. Meister designed the experiments and wrote the manuscript.

## 3.1 Introduction – Mechanisms of isomiR generation

Any given miRNA exists as more than one specific sequence. These sequence and length variations are called isomiRs and arise through several different mechanisms. The observed differences of isomiRs include a different 5’ end, tailing or trimming of the 3’ end and RNA editing. The next paragraphs will summarize known mechanisms generating isomiRs and their impact on target regulation.

### 3.1.1 Influence of RNase III enzymes Drosha and Dicer on miRNA length and identity

During the biogenesis of miRNAs in mammals, two RNase III enzymes, Drosha and Dicer, process first the pri- and then the pre-miRNA to its mature form (see 1.1.1). Drosha catalyzes the first cleavage from the pri-miRNA to the stem-loop pre-miRNA (Figure 1-1), thus generating the 5' end of the 5' arm (5p) and the 3' end of the 3' arm (3p) (the ends at the base of the hairpins). When Dicer processes the pre-miRNA to the mature miRNA duplex, it produces the 3' end of the 5p arm and the 5' end of the 3p arm (the ends close to the loop) (Figure 1-1). In conclusion, Drosha defines the seed of the 5p miRNA and Dicer defines the seed of the 3p arm. One main difference between the enzymes is that Drosha needs an RNA binding protein partner (DGCR8) to recognize and bind pri-miRNAs whereas Dicer does not. Due to a relaxed specificity and the different precursor structures, these RNase III enzymes introduce 5' and 3' end variations.

By examining length patterns of endogenous mature and pre-miRNAs, it has been proposed that Drosha processes miRNAs more accurate than Dicer due to its additional RNA binding partner DGCR8 (Calabrese et al., 2007; Starega-Roslan et al., 2011). The relationship between primary miRNA sequence and processing pattern has not been studied so far, though.

The structure of the precursor stem-loop seems to have the most influence on the stringency of processing by Dicer. The more mismatches (single-base) or bulges (multiple bases that do not base-pair) a precursor stem contains, the more flexible it becomes. In addition, in their study on pri-miRNA and pre-miRNA processing, Krzyzosiak and colleagues propose a model that unmatched nucleotides (bulges) accommodated in the substrate channel of Dicer are often not or not accurately counted by Dicer. The more structural imperfections a precursor accumulates the more it becomes prone to generate heterogeneous miRNA sequences. Since Dicer as a protein is also flexible in itself, these precursors tend to be more inaccurately processed than precursors with more perfectly complementary stems (Starega-Roslan et al., 2011).

### 3.1.2 5' end miRNA isoforms

Many studies note that the 5' end of the mature miRNA is less prone to end heterogeneity than the 3' end, no matter of its location in the precursor (Landgraf et al., 2007; Morin et al., 2008; Ruby et al., 2006; Seitz et al., 2008; Warf et al., 2011). Since 5' end variations would change the seed and thus the target spectrum, selection against deleterious 5' end alterations will occur. Indeed, less than 10% of mature miRNAs in deep sequencing libraries have a shortened 5' end (Chiang et al., 2010; Ruby et al., 2006; Ruby et al., 2007b).

The generation of 5' end isoforms cannot only happen through processing by Drosha or Dicer (see 3.1.1). Several pathways produce miRNAs in a non-canonical manner (see 1.1.1.2). The biogenesis of miRtrons is independent of Drosha but requires exonucleases to produce the substrate for Dicer processing. In mammals, miRtrons are predominantly 5' tailed with few exceptions (Ladewig et al., 2012; Valen et al., 2011). Therefore, the 5' ends of the 5p arm of the hairpin are more prone to have a

heterogeneous end. SnoRNA-derived miRNAs and tRNA-derived miRNAs both rely on different exonucleases that can introduce 5' end heterogeneity (Ender et al., 2008).

### 3.1.3 Isoforms produced by editing

Adenosine deaminases (adenosine deaminase acting on RNA, ADARs) are enzymes converting adenosines to inosine through desamination and act on dsRNA (reviewed in (Nishikura, 2009)). Since miRNAs are also intermediately double-stranded as precursors and miRNA duplexes, they are potential targets of ADARs. Indeed, editing events could be shown for both precursor and mature miRNA (Blow et al., 2006; Kawahara et al., 2007a; Kawahara et al., 2007b; Luciano et al., 2004). The edited inosine pairs with cytosine, so editing events are sequenced as A to G mutations. Although editing occurs more often in brain than any other tissue, which is consistent with the fact that ADARs are almost exclusively expressed in brain, the mutation rate of A to G is rarely or barely elevated above background mutation level (Chiang et al., 2010). In conclusion, RNA editing does not seem to be a general feature of regulation of miRNA function or stability, but could impact specific miRNAs in specific settings.

### 3.1.4 3' end miRNA isoforms

The highest variability of miRNAs is observed at their 3' ends. The two different types of 3' end variation can be explained by a trimming of the mature miRNA sequence or by an addition of non-templated nucleotides to the 3' end.

The observation that miRNAs have variable 3' ends dates some years back (Rajagopalan et al., 2006; Ruby et al., 2006; Ruby et al., 2007b; Seitz et al., 2008; Wu et al., 2007), however, the importance and the mechanisms and regulation of 3' end variants are only incompletely understood. The trimming of mature miRNAs probably occurs after they are incorporated into Ago proteins. There is not much known about the identity of the exoribonucleases involved, but, for example, in *Drosophila*, the 3'-5' exonuclease Nibbler was found to alter the length of a subset of miRNAs while binding to AGO1 (Han et al., 2011; Liu et al., 2011). Depletion of Nibbler causes severe developmental defects such as sterility, which could at least be partially due to these 3' variants. In *Neurospora crassa*, the 3'-5' exonuclease QIP acts together with the exosome to trim miRNA-like small RNAs called milRNAs (Xue et al., 2012). Similar mechanisms likely also exist in vertebrates. Future research will uncover if homologs of these enzymes also participate in the generation of miRNA variability.

The addition of non-templated nucleotides is commonly referred to as tailing. The enzymes responsible for adding nucleotides are terminal nucleotidyl transferases (TNTases) that usually add adenosine or uridine (Rissland et al., 2007). While there have been reports on the function of uridylation of miRNA precursors (Hagan et al., 2009; Heo et al., 2009), a general mechanism

underlying these additions to mature miRNAs is yet lacking. However, there have been studies in flies reporting a massive change of tailing during early development (Fernandez-Valverde et al., 2010).

### 3.1.5 Consequences of miRNA length and sequence heterogeneity

So far, only few studies have shed light on the function of miRNA length and sequence variations. 5' isomiRs seem to have the highest impact, probably because the seed and therefore the target spectrum is shifting. For example, the conserved human miRNAs miR-124 (neuronal miRNA) and miR-133a (expressed in heart and muscle) show each considerable expression of a 5' shifted isoform (~10%) thus influencing their target spectra (Chiang et al., 2010; Lagos-Quintana et al., 2002; Landgraf et al., 2007).
As described above, the trimming of a particular miRNA can influence its function. However, since the 3' part of a miRNA is usually thought to contribute only very little to target recognition and regulation (at least in animals), more studies and effort are needed to explain the few cases of proven effect on developmental processes. Although some features of isomiR generation seem to be random or by chance, there certainly are specific factors and signals underlying the modification of miRNAs that have not been identified before.

## 3.2 Profiling and characterization of miRNAs associated with endogenous Ago proteins

Ago-associated miRNAs have been described before (Chi et al., 2009; Ender et al., 2008). However, these studies have used the overexpression of Ago proteins, thus perturbing intracellular balances. The deleterious effects of alterations of Ago protein levels are underscored in Ago2-deficient cells, where miRNAs are globally down regulated (Diederichs and Haber, 2007). As we could show above (see 2.1), endogenous Ago proteins are expressed not equally well but in a rather stabile ratio to each other. Since our lab had generated and validated monoclonal antibodies directed against each human Ago protein, we made use of these tools to analyze miRNAs associated with the different endogenous Ago proteins.

### 3.2.1 Establishment and generation of small RNA libraries

In order to analyze small RNAs associated with endogenous Ago proteins, I had to setup and establish a protocol for cloning and deep sequencing. As this method is of importance to the results in this section, I will briefly introduce the main steps.
The general protocol was adapted and modified from Pfeffer et al. (Pfeffer et al., 2005). In short, small RNAs bearing a 5' phosphate and a 3' OH were ligated in a first step to a preadenylated adapter (3'

adapter) using a truncated T4 RNA ligase 2 (amino acids 1-249) (Ho and Shuman, 2002; Ho et al., 2004) (see Figure 3-1).

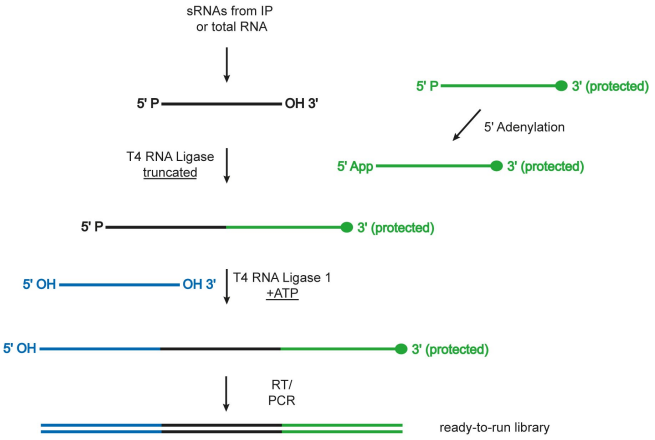



**Figure 3-1** Protocol for the generation of libraries for Illumina Deep Sequencing. The RNA (small RNAs or fractionated RNA) is first selectively ligated to an adenylated adapter with a blocked 3' end using a truncated version of T4 RNA Ligase. In a subsequent step, the 5' adapter is ligated with the conventional T4 RNA Ligase 1. The adapter-RNA complex is revers-transcribed into cDNA and amplified by PCR. After purification by Urea-PAGE the library is ready to load onto an Illumina chip.

This elegant solution minimizes unwanted side products by concatenation of the small RNAs. The second ligation step with the 5' adapter was mediated by T4 RNA Ligase 1 and the addition of ATP. The RNA construct was reverse-transcribed and amplified by PCR. To exclude ligated adapters without any insert, the PCR products were separated on a urea PAGE and only the band having the correct size was isolated. The DNA library was extracted from the gel and was now ready to be analyzed by deep sequencing.

The adapters of the library are compatible with the Illumina platform. In short, the double-stranded library is denatured and the single strands can anneal via the adapters to oligonucleotides bound to a chip. After a round of PCR amplification so-called clusters are formed, each bearing only one unique insert sequence. The insert sequence is now determined by synthesizing the second strand. In each sequencing cycle, one of four fluorescently labeled nucleotides is incorporated into the strand (corresponding to the sequence of the insert) and a laser reads out the fluorescent dye, thus reading out the identity of the base. The fluorescent dye is chemically removed and a new sequencing cycle begins (for further information on sequencing technology and sequencing platform, please refer to www.illumina.com).

### 3.2.2 Analysis of putatively sorted miRNAs associated with the different Ago proteins

In order to analyze small RNAs associated with the different Ago proteins, we performed endogenous Ago immunoprecipitations (IP) from HeLa S3 lysates. The used monoclonal antibodies are highly specific and have been characterized by members of our laboratory (see also 2.1) (Petri et al., 2011; Rudel et al., 2008; Weinmann et al., 2009). After extraction of Ago-bound RNAs and cloning, we

created three technical replicates by PCR and performed deep sequencing. The raw data were processed by a custom analysis pipeline by Eugene Berezikov (University of Groningen). For Ago1, Ago2, Ago3 IP and the total RNA (input) sample, between 1.5 million and 3.5 million reads in total were obtained matching to the human genome. Ago4 and control IP, however, were comparable and only few reads in total were retrievable that matched to the genome (~27000 and ~23000, respectively). Since we already showed that Ago4 is not expressed in HeLa cells (see 2.1) (Petri et al., 2011), the low amount of reads likely reflects background noise, as is certainly the case for the control. Therefore, the Ago4 and control IP libraries were not examined further. The composition of each library is depicted in Figure 3-2, A. As expected, Ago1-Ago3 libraries are highly enriched in miRNAs (around 90% of all reads).

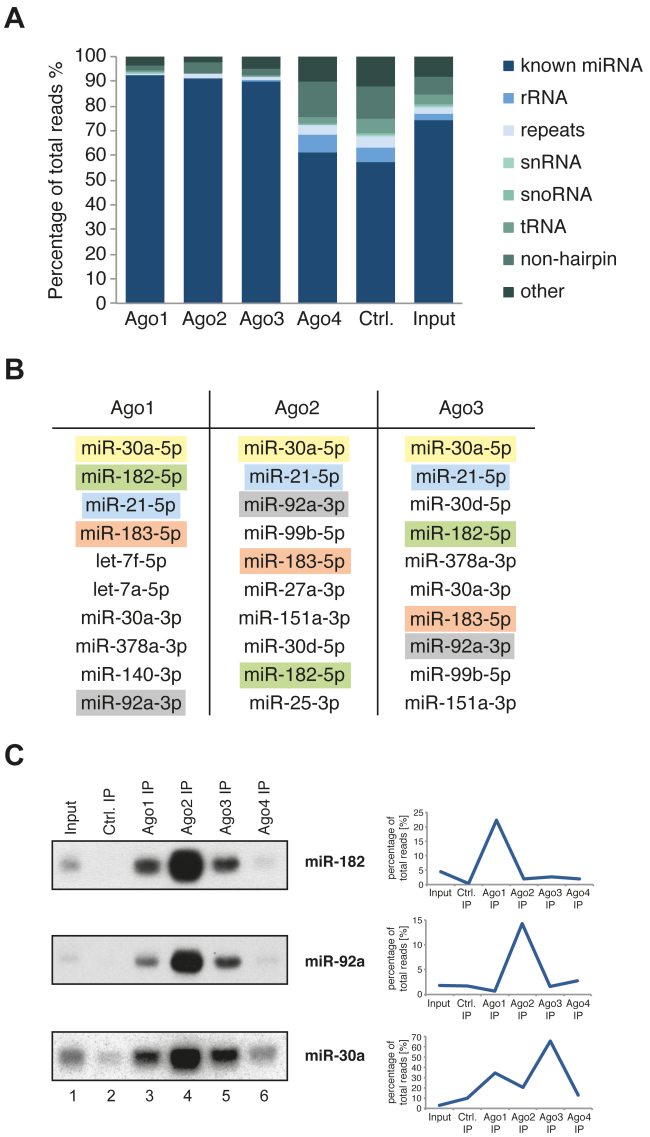



**Figure 3-2 Deep sequencing analysis of small RNAs associated with endogenous Ago proteins.** (A) Composition of the small RNA libraries from the input sample and immunoprecipitated Ago1, Ago2, Ago3, Ago4 and control (Ctrl.). (B) Table of the ten most abundant miRNAs associated with Ago1, Ago2 or Ago3. The colors indicate the same miRNA. (C) Comparison of the sorting profile from the deep sequencing data (right side) for miR-182 (top panel), miR-92a (middle panel, right) and miR-30a (lower panel) with profiles resulting from Northern blotting (left side).

Next, we analyzed the miRNAs that were highly abundant in the Ago1-3 libraries. When comparing the ten most abundant miRNAs, many are shared between the different Agos (Figure 3-2 B, highlighted in color). After examining expression profiles in more detail, however, we found several miRNAs that were associated more with one Ago than the two others. These included miR-182 (associated with Ago1), miR-92a (associated with Ago2) and miR-30a (associated with Ago3) (Figure 3-2 C). To validate these sorting effects, Northern blotting was performed, probing for these three miRNAs. Each time the pattern of miRNA association is the same, with the strongest signal observable in the Ago2 IP. This pattern simply reflects the amount of each Ago protein in the cell (see 2.1 and (Petri et al., 2011)), where Ago2 accounts for approximately 65%, Ago1 for 20% and Ago3 for 15% of the total Ago pool.

The sorting observed in the deep sequencing data cannot be reproduced with the more direct Northern blotting method and therefore we conclude that no clear sorting is evident in human cells.

### 3.2.3 Sequence features of Ago-bound miRNAs

Although a sorting mechanism of individual miRNAs could not be verified by Northern blotting, other features of miRNAs could influence binding to mRNA target sites. MiRNAs have long been shown to exist not as exactly one length, but as isomiRs (see 3.1), varying by 1-3nt (reviewed in (Ameres and Zamore, 2013)). Therefore, the global length distribution in our data sets for each Ago protein was analyzed (Figure 3-3 A). It is striking that the overall profiles of bound miRNAs are quite dissimilar between Ago1, Ago2 and Ago3. The predominant miRNA lengths bound by Ago1 are 22 and 24nt (Figure 3-3, A, left), whereas Ago2 peaks at 22nt (Figure 3-3, A, middle) and Ago3 at 23nt (Figure 3-3, A, right).

Since non-templated nucleotide addition is a major source of miRNA length variation (see 3.1.4), we analyzed this phenomenon in our miRNA libraries from endogenous Ago isolations. First, the total abundance of non-templated nucleotides in the libraries of Ago1, Ago2 and Ago3 was determined (Figure 3-3, B, left panel). While the largest part of miRNAs is not modified (dark grey), a considerable amount was extended at the 3' end. This portion ranged from ~10% (Ago3) up to ~20% (Ago2). Looking closer at the nature of the 3' extensions, the preference for the addition of 1 adenine is readily observable (Figure 3-3, B, right panel). Where Ago2- and Ago3-bound miRNAs seem to also be modified with 1 uridine (red) or more than one nucleotide (other, light grey), Ago1-bound miRNAs are almost exclusively extended by one adenine (blue, >80%). The addition of a G or C seems to be rather rare for all Ago-bound miRNAs.

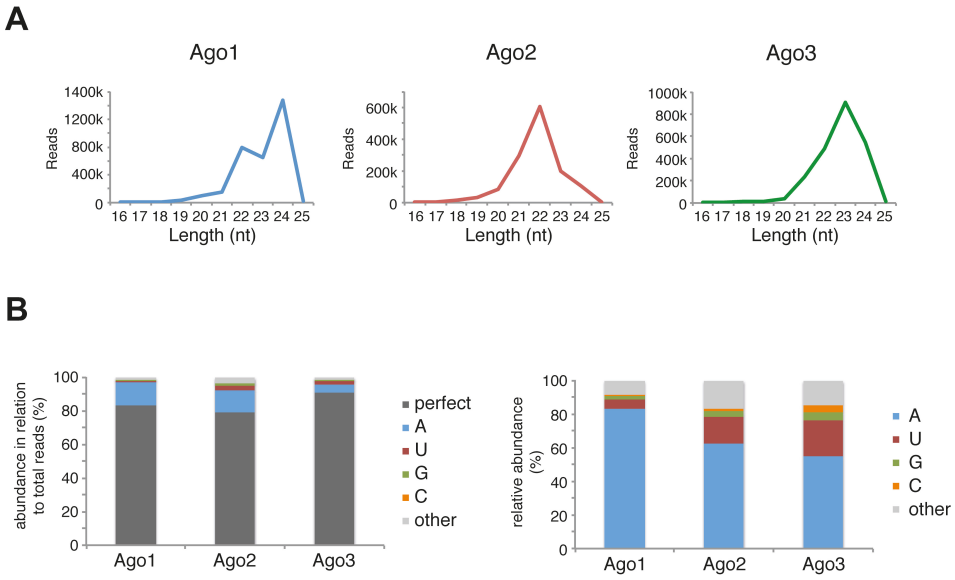



**Figure 3-3 Ago proteins show different global small RNA binding patterns and non-templated 3' extension profiles.** (A) Length distribution patterns of Ago1 (left), Ago2 (middle) and Ago3 (right). (B) Left side: Ago1, Ago2 and Ago3 associated miRNAs were analyzed for non-templated 3' extensions, values were calculated in relation to all miRNA reads of the respective library. Perfect: no non-templated 3' extensions; A, U, G, C: amount of single nucleotide 3' extensions; other: sum of all other 3' extensions (2 nt or longer). Right side: Each single non-templated 3' extension was set in relation to the total number of non-templated 3' extensions in the respective library (Ago1, Ago2 or Ago3).

In conclusion, the characteristics of Ago-bound miRNAs are diverse between Ago1, Ago2 and Ago3. While the length pattern of each Ago shows distinct and different peaks, the tolerance for addition of non-templated nucleotides is more similar, with the preference for 1 A.

After analyzing the global length differences of miRNAs bound to the different Agos, we next focused on individual miRNAs.

For each Ago, the length pattern of individual miRNAs was extracted from the deep sequencing data as well as validated by immunoprecipitation followed by Northern blotting (Figure 3-4). When comparing the two methods, the overlap was very strong. This was even more remarkable since the sequencing results regarding differential Ago-association could not be confirmed by Northern blotting (Figure 3-2). According to the profiles, mainly two groups or types of miRNAs emerged. One group, including miR-23a, miR-30a and miR-21, showed similar profiles for Ago1 and Ago3, while Ago2 seemed to bind also shorter variants of these miRNAs (Figure 3-4, A). The other group, with examples being let-7a and miR-27a, showed a clear, sharp peak for all three Agos (Figure 3-4, B).

In summary I could show that the endogenous four Ago proteins associate to over 90% with known miRNAs. Although there is no clear sorting observable, differences can be seen when analyzing the global length profiles of Ago-bound miRNAs. The investigation of non-templated nucleotide additions to miRNAs showed that the portion of bound and modified miRNAs is very similar between the Ago proteins. Additionally, the Ago proteins display on the level of individual miRNAs a diverse binding preference.

Recent studies have uncovered an unusual processing pathway for the hematopoietic miRNA miR-451. This miRNA is not expressed in either HEK 293T or HeLa cells thus being not present in our deep sequencing libraries. Since the hairpin of miR-451 is processed only by Ago2, the question arose if the mature miRNA would only be bound by Ago2 as well. Also, a Dicer-independent system of generating siRNAs could prove as a powerful tool to knockdown genes with more specificity, since they would only be bound to the endonuclease Ago2. We therefore wanted to analyze miR-451 processing and binding preferences in more detail.

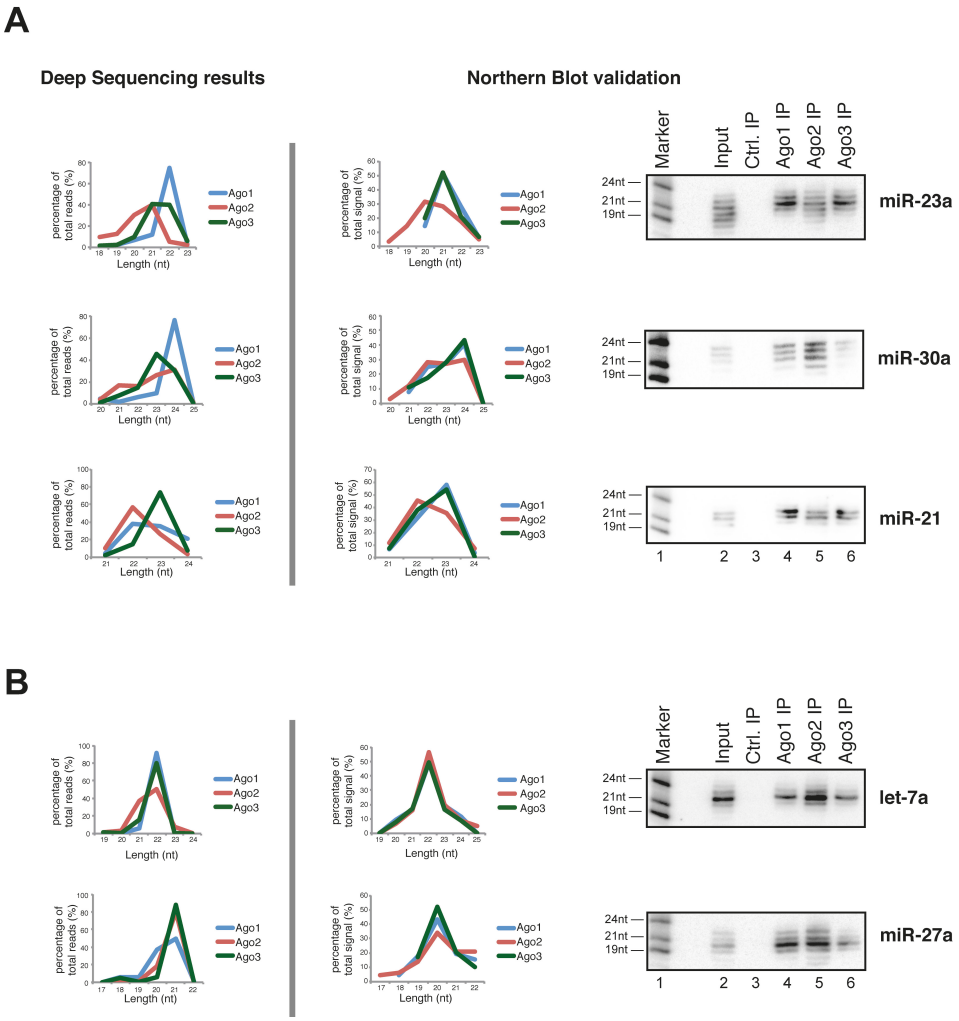



**Figure 3-4 Ago proteins show distinct length profiles for individual miRNAs.** The length pattern of individual miRNAs was extracted from the deep sequencing data (left side) and compared against the length profiles seen by Northern blotting (right side). For this, Northern blotting signals for each miRNA band were quantified; the total signal from each lane/each Ago was set to 100%. The length of a miRNA in Northern blotting was assesses by comparison with the size marker. MiRNA length profiles were separated into: (A) miRNA length profiles showing divergent Ago association and (B) miRNA length profiles showing matching Ago1-3 association.

## 3.3 Characterization of miR-451 processing

miR-451 is a miRNA that is processed in a unique way. Two studies (Cheloufi et al., 2010; Cifuentes et al., 2010) showed that the peculiar precursor of miR-451 is not processed by Dicer, but bound directly by Ago2. The endonucleolytic activity of Ago2 leads to the cleavage of the strand opposite the mature miRNA – similar to passenger strand cleavage – and the resulting 30nt fragment is further matured by Poly(A)-specific ribonuclease (PARN) (Yoda et al., 2013) (see Figure 3-5).

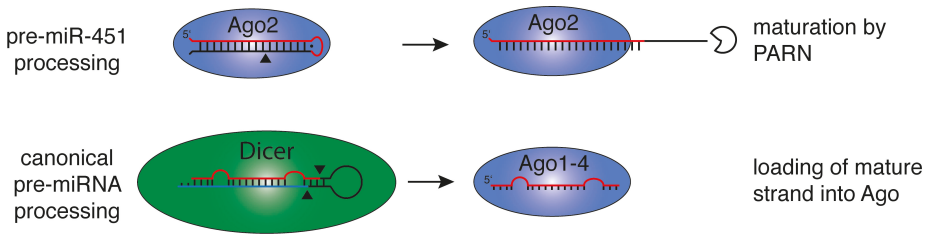



**Figure 3-5 Human miRNA processing pathways.** Schematic representation of Ago2-mediated miR-451 processing (upper part) and canonical pre-miRNA processing by Dicer (lower part).

### 3.3.1 Endogenous miR-451 is associated with Ago2 only

As mentioned above, miR-451 is expressed in hematopoietic cells only. After testing several common human cell lines without success, we studied the expression in various mouse organs. miR-451 was found to be highly expressed in liver and kidney, but was also detectable in spleen, brain and lymph nodes (Figure 3-6, A).

To test the association of miR-451 with the different Ago proteins, endogenous Ago IPs were performed using lysates from mouse liver and spleen. The RNA associated with mouse Ago1 (mAgo1) and mouse Ago2 (mAgo2), respectively, was isolated and analyzed by Northern blotting (Figure 3-6, B). Expectedly, the detection of miR-19b, a miRNA that is canonically processed (Figure 3-5), shows an association with both mAgo1 and mAgo2 (lower panel). miR-451, however, is clearly exclusively bound by mAgo2.

In conclusion, miR-451 is not only processed by Ago2, but is retained in Ago2. This also directly shows that there is no exchange at least of miR-451 between Ago2 and the other Agos.

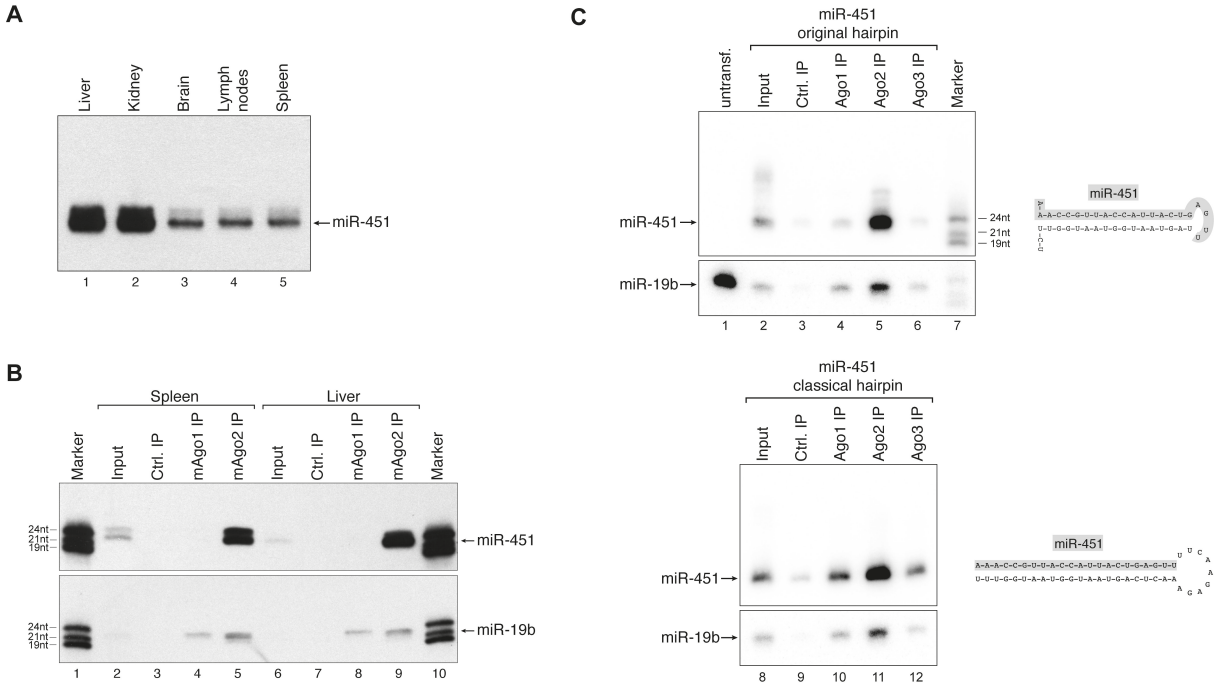



**Figure 3-6 miR-451 associates with Ago2 only.** (A) RNA was extracted from the indicated mouse tissues and miR-451 was detected by Northern blotting. (B) Endogenous mAgo1 (lanes 4 and 8) or mAgo2 (lanes 5 and 9) were immunoprecipitated from spleen or liver and co-immunoprecipitated miR-451 (upper panel) or miR-19b (lower panel) was analyzed by Northern blotting. Lanes 3 and 7 show control immunoprecipitations, lanes 2 and 6 input samples and lanes 1 and 10 size markers. (C) Constructs expressing wt pre-miR-451 (upper panel) or miR-451 embedded into a classical Dicer-dependent hairpin (lower panel) were transfected into HEK 293T cells. Ago1 (lanes 4 and 10), Ago2 (lanes 5 and 11) and Ago3 (lanes 6 and 12) were immunoprecipitated using specific monoclonal antibodies. Co-immunoprecipitated miRNAs were analyzed by Northern blotting against miR-451. Lanes 3 and 9 show control immunoprecipitations, lanes 2 and 8 show input samples and lane 7 a size marker. For controlling IP efficiency, miR-19b was blotted, as well.

Since the unusual structure of the pre-miR-451 hairpin (short stem, short loop) is an obvious explanation for its unique processing, we wanted to explore whether this processing can be recapitulated with an exogenous expression system and which characteristics of miR-451 lead to the sole processing of and association with Ago2.

The original pre-miR-451 hairpin was cloned into a modified pSuper vector, bearing a Polymerase III Promoter (H1). This construct was transfected into HEK 293T, which do not express endogenous miR-451, and Ago1-3 IPs were performed using highly specific monoclonal antibodies against the endogenous proteins followed by detection by Northern blotting. Ago4 was not included in this study as we have already shown that it is expressed in only very low levels in a wide range of human cell lines (see 2.1 and Figure 2-2) (Petri et al., 2011). As shown in Figure 3-6 (C, upper panel), miR-451 is expressed and processed from this construct (lane 2) and again only associated with Ago2 (lane 5). The slight background seen in the Ago1, the Ago3 and the Control Immunoprecipitation is likely due to the high level of overexpression of miR-451. To find out whether the sequence of mature miR-451 alone directs it to be processed and bound by Ago2, we generated an expression construct using the classical structure for siRNA expression. The hairpin was built to harbor the mature sequence of miR-451 on the 5' arm of the hairpin followed by a generally used loop sequence and completed by the perfectly complementary sequence to miR-451. When transfecting this construct into HEK 293T cells

and immunoprecipitating Ago1-3, mature miR-451 can now be found to bind to all three Ago proteins (Figure 3-6, C, lower panel).

Taken together, I could show that endogenous miR-451 is not only processed by Ago2, but also exclusively bound by it. Additionally, these results provide evidence that there is no exchange of mature miRNAs between the different human Ago proteins, a fact that has been controversially discussed in the field until now (Diederichs and Haber, 2007; Grishok et al., 2001; Janas et al., 2012; O'Carroll et al., 2007; Vaucheret et al., 2004).

### 3.3.2 Using the miR-451 hairpin system for expression of siRNAs and miRNAs

Since Ago2 is the catalytic Ago protein and required for RNAi, it is desirable to load siRNAs into Ago2 only to prevent off-target effects caused by non-catalytic Ago proteins (see Chapter 2). A miR-451-based expression system would be the ideal backbone for such siRNAs.

Therefore, we tested if any small RNA (miRNA or siRNA) can be expressed and processed in a miR-451-like manner. An siRNA against GRK4 (G protein-coupled receptor kinase 4, see (Bramsen et al., 2010; Lin et al., 2005)) was used to generated analogous hairpins to the miR-451 experiment: one hairpin mimics the precursor of pre-miR-451, one hairpin is designed in a classical siRNA way (Figure 3-5).

Both constructs were transfected into HEK 293T cells and expression was monitored after Ago1-3 IPs and Northern blotting (Figure 3-7, A). As is the case for pre-miR-451, the siGRK4 mimic hairpin is expressed and processed. The single stranded siRNA can only be seen associated with Ago2 (lane 5), however, the general expression of this hairpin in quite low. On the other hand, the level of expression of the classical GRK4 hairpin is much higher (lane 8). The distribution of the siGRK4 expressed from the classical hairpin shows an equal association of the siRNA with all three Agos.

To compare the knockdown ability of both hairpins, we transfected H1299 cells (lung carcinoma cell line expressing GRK4 mRNA) with both constructs, monitored the expression of the siRNAs by Northern blotting and the mRNA level of GRK4 by qRT-PCR. As was the case in HEK 293T cells, the expression of the mimic hairpin siRNA was considerably lower than that of the classic hairpin siRNA (Figure 3-7, B). The knockdown of GRK4, measured after 42 h, 120 h and 140 h, was comparable between both shRNAs (Figure 3-7, C). Therefore, the miR-451 like hairpin can generally be used for knockdown of target-mRNAs.

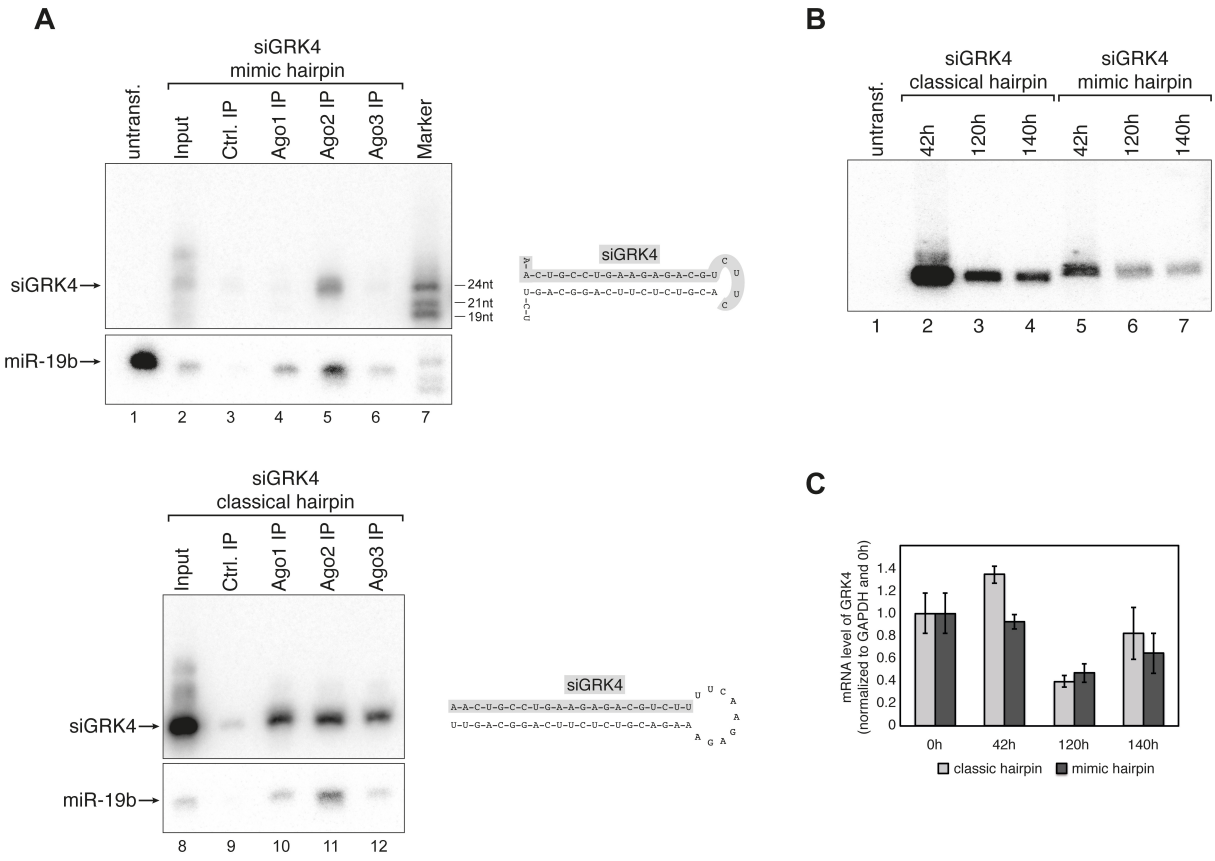



**Figure 3-7 Pre-miR-451-based shRNA expression.** (A) An siRNA against GRK4 was either embedded into a pre-miR-451-like hairpin (upper panel) or a classical Dicer-dependent hairpin (lower panel). Both constructs were transfected into HEK 293T cells and Ago1 (lanes 4 and 10), Ago2 (lanes 5 and 11) or Ago3 (lanes 6 and 12) was immunoprecipitated from the lysates. GRK4 siRNA incorporation was analyzed by Northern blotting. Lanes 3 and 9 show control immunoprecipitations, lanes 2 and 8 input samples and lane 7 a size marker. For controlling IP efficiency, miR-19b was blotted, was well. (B) Both GRK4 shRNAs described in (A) were transfected into HEK 293T cells. Total RNA was extracted at the indicated time points and GRK4 siRNA production was analyzed by Northern blotting using a probe complementary to the guide strand of the GRK4 siRNA. (C) Samples described in (B) were analyzed by qRT-PCR for knock down of GRK4 at the indicated time points.

The low expression of the GRK4 mimic shRNA led us to test if other sequences were also prone to this low level of expression. Two different shRNA mimic clones of an siRNA against Importin 8 (Weinmann et al., 2009) were transfected into HEK 293T cells and their expression compared with the classical shRNA construct. Both mimic clones did not express any siRNA product (Figure 3-8, A). We next compared the expression of let-7c constructs in HEK 293T cells. Let-7c is endogenously expressed only at very low levels, as shown by Northern blotting (Figure 3-8, B, lane 1/5). Three different constructs were compared: the Drosha product of the hairpin encoded in the genome (let-7c genomic), the classical shRNA and the miR-451 mimic shRNA. The expression level of the classical hairpin is considerably higher than that of the other hairpins (Figure 3-8, B, lane 3/7) while the expression of the mimic hairpin is almost at the endogenous level (lanes 4 and 8).

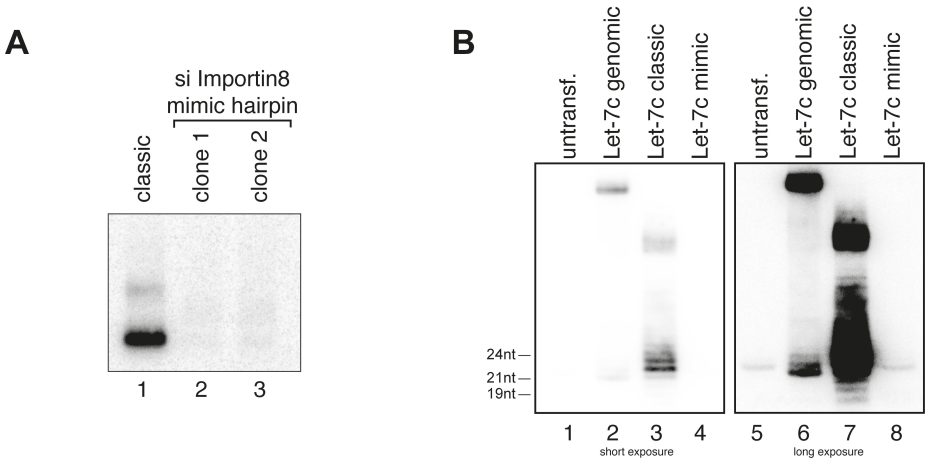



**Figure 3-8 Pre-miR-451-based shRNA expression of other sequences.** (A) SiRNAs against Importin 8 were embedded into a pre-miR-451-like hairpin (two different clones were tested, lanes 2 and 3) or a classical Dicer-like shRNA (lane 1). The constructs were transfected into HEK 293T cells and siRNA production was analyzed by Northern blotting. (B) Relative expression levels of let-7c were tested by transfecting the genomically encoded ("genomic") (lanes 2 and 6), the classical Dicer-like shRNA ("classic") (lanes 3 and 7) or a pre-451-like hairpin ("mimic") (lanes 4 and 8). As a background control, RNA from untransfected HEK 293T was used (lanes 1 and 5).

In conclusion, the structure alone of the hairpin is sufficient to direct the processing by Ago2, however, the low expression of the mimic constructs also points to other factors influencing expression.

### 3.3.3 Structural requirements for Ago2-mediated processing of pre-miR-451

We have shown in the previous paragraphs that pre-miR-451 is efficiently expressed and processed by Ago2, both *in vivo* and *in vitro*. Since using other sequences – either a miRNA or different siRNAs – did not lead to a satisfactory expression, we performed mutational studies on the miR-451 hairpin in order to find the requirements for efficient processing by Ago2.

The sequence of the precursor and therefore of the mature miR-451 bears an adenosine at the 5' end. While an adenosine is not highly unusual, human miRNAs primarily contain a uridine at this position (Frank et al., 2010). When the first nucleotide is mutated from an A to a U, the expression level of the miRNA is maintained, but not elevated (Figure 3-9, A, mutant #1). Notably, the length of the mature form shifts by approximately one nucleotide, which could be explained by a different anchoring to Ago2.

Next, the importance of the loop structure was examined. Gradual extension of the loop sequence showed a drastic reduction of processing when adding 1-3 nt (Figure 3-9, B, lanes 2-4). When adding four or five additional nucleotides, processing was almost completely lost (lanes 5 and 6).

In order to analyze the stem sequence of pre-miR-451, we mutated four different positions, two of them next to the cleavage site of Ago2. The mutations were performed on both strands so as to preserve the double-stranded nature of the stem. All mutations did not have an effect on either expression level or length of the processed RNA as seen by Northern blotting (Figure 3-9, C).

The wt miR-451 precursor contains two G:U wobble base pairs, one counting 6 bases from the start of miR-451 and one on the distal stem, directly in front of the loop. We mutated each of the two G:U wobble base pairs to a Watson-Crick G-C base pair. While mutating the first wobble pair had no effect on miR-451 processing, mutating the wobble in front of the loop diminished the processed product to about the half of wt level (Figure 3-9, D). The presumed requirement for a wobble base pair at this position strongly limits the choice of sequences and therefore the applicability of this system for siRNA-mediated knockdown of mRNAs.

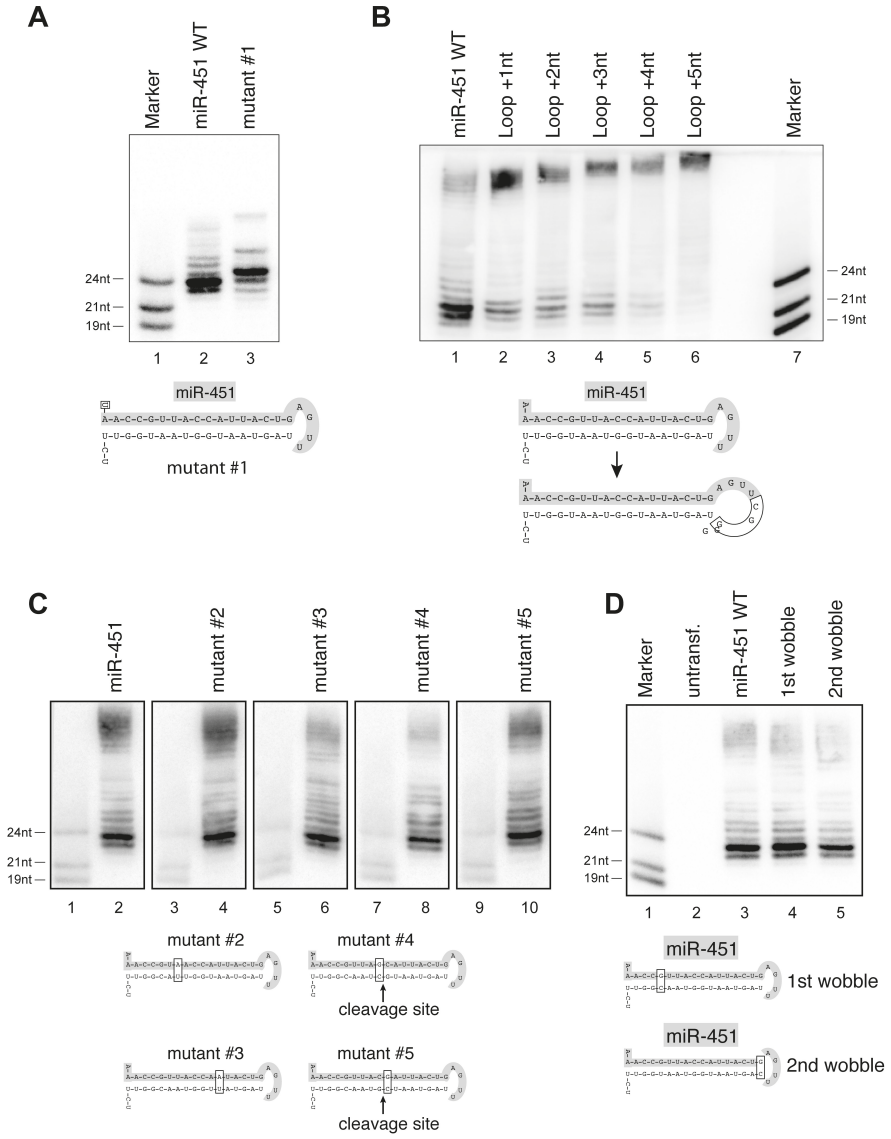



**Figure 3-9 Structure and sequence requirements of miR-451 for Ago2-mediated cleavage.** (A) A wt miR-451 construct (lane 2) or a mutant with a changed 5' end (lane 3, A to U mutation, mutant #1) were transfected into HEK 293T cells and processing was analyzed by Northern blotting. Lane 1 shows a size marker. (B) Nucleotides were inserted into the loop of pre-miR-451 (lanes 2 to 6) and the constructs were analyzed as described in (A). (C) Mutants of the wt pre-miR-451 hairpin were constructed to test the sequence specificity of Ago2-mediated processing. Constructs were transfected into HEK 293T cells and analyzed on individual Northern blots, since each mutation alters the resulting mature sequence. Equal amounts of size marker were loaded to provide comparability. (D) G/U wobble base pairs at the 5' end (lane 4) and near the loop (lane 5) were mutated to G/C Watson-Crick base pairs and the constructs analyzed as in (A).

Taken together, miR-451 is exclusively associated with Ago2 and is not exchanged between the Ago proteins. The unique structure of pre-miR-451 ensures its optimal processing. Substitution of the sequence of miR-451 with other sequences leads to a dramatic reduction in processing, while the exclusive loading into Ago2 remains. Analysis of the pre-miR-451 structure revealed the requirement for a short loop and a wobble base pair next to the terminal loop.

## 3.4 Discussion

Using endogenous Ago protein isolation followed by cloning and deep sequencing, we found that miRNAs are not sorted into different human Ago proteins. It has been reported before that several miRNAs show distinct Ago association patterns (Burroughs et al., 2011). These studies used the over expression of tagged Ago proteins which might change miRNA association and relied on deep sequencing data. Validation using direct Northern blotting had not been performed. It has been demonstrated that small RNA cloning might introduce both ligation and PCR biases into the library composition (Hafner et al., 2011). This might explain our different findings. However, it is also conceivable that miRNAs might associate with Ago protein sequence specifically in other cell types or cell stages.

Our deep sequencing approach of Ago-associated miRNAs revealed that Ago identity impacts the length of the bound miRNA. A very recent study showed that miRNAs are generally shortened during mammalian brain development (Juvvuna et al., 2012). The shortening is due to a change in Ago protein expression suggesting that Ago identity affects the length of miRNAs. These findings are consistent with our results. However, in our study some miRNAs are characterized by a clear and distinct length (e.g. let-7a, Figure 3-4) independently of the Ago protein, which they bind to. These findings indicate that Ago identity is not the sole determinant of the miRNA length. So far unidentified factors might associate with Ago proteins under specific conditions and trim or extend the 3' end of the miRNA (Dueck et al., 2012).

A closer investigation of endogenous Ago-associated miRNAs revealed that a significant portion of miRNAs contain non-templated A additions at the 3' ends. It has been demonstrated that uridylyl transferases such as TUT4 add poly-U stretches to precursors of members of the let-7 family leading to pre-miRNA degradation (Hagan et al., 2009; Heo et al., 2008; Heo et al., 2009). However, this mechanism seems to be let-7-specific and has not been found for other miRNA species. In contrast, mono- or di-nucleotides are often added to the 3' end of other miRNAs (Cloonan et al., 2011; Kuchenbauer et al., 2008; Pantano et al., 2010; Wyman et al., 2011). It has been shown recently, that miRNAs carrying a single A addition are preferentially bound by over expressed myc-tagged Ago1 (Burroughs et al., 2010). Although we find a stronger enrichment of A addition in Ago1-associated miRNAs, extensions of Ago2- and Ago3-associated miRNAs is also dominated by A (Figure 3-3, B). Therefore, we conclude that A and U are the predominant nucleotides that are added to the 3' end of the miRNA. However, Ago protein identity has only a minor effect on nucleotide addition to miRNAs.

An unsolved problem in RNAi experiments is that transfected synthetic or shRNA-derived siRNAs are loaded into all four Ago proteins (see also Chapter 2) (Jackson and Linsley, 2010; Petri et al., 2011; Wu et al., 2008). Only Ago2 is catalytically active and needed for efficient gene knock down. Ago1, Ago3 and Ago4, however, may cause unwanted off-target effects by binding to partially complementary target sites in unrelated 3' UTRs leading to miRNA-like gene silencing effects (Petri et al., 2011). It has been demonstrated that miR-451 processing is Dicer-independent and pre-miR-451 is cleaved by Ago2 (Cheloufi et al., 2010; Cifuentes et al., 2010). Here I show that miR-451 is not only

processed by Ago2 but also loaded exclusively into Ago2-containing RISC complexes. This experimental observation immediately points to an attractive method for specific knock down experiments. ShRNAs mimicking the pre-miR-451 hairpin in size and structure but containing any given siRNA sequence might only interact with Ago2. We have tested this hypothesis and find that indeed siRNAs can be generated from pre-miR-451-like shRNAs. However, processing as well as silencing activity was not as efficient as shRNAs that are processed by Dicer. Although pre-miR-451-like shRNAs are indeed loaded into Ago2-containing RISCs, we conclude from our study that such constructs are less efficient than classical Dicer substrates (Dueck et al., 2012).

In *Drosophila*, small RNAs are sorted into different Ago proteins (Forstemann et al., 2007; Tomari et al., 2007). However, miRNA sorting is less clear in mammals so far. Furthermore, it is not known whether miRNAs are able to dissociate from Ago proteins after processing and RISC loading allowing for an exchange of mature miRNAs between Ago proteins. It has been demonstrated that a protein termed hnRNPE2 can interact with a specific miRNA and sequester it from Ago proteins (Eiring et al., 2010). Also, MacRae and colleagues showed that highly complementary target-mRNAs can promote the release of the Ago-bound miRNA (De et al., 2013). Two studies have been published claiming that miRNAs are up to 13fold more abundant than Ago proteins and can be deposited on target-RNAs without an Ago protein (Janas et al., 2012; Stalder et al., 2013). We find that mature miR-451 associates exclusively with Ago2 indicating that dissociation and re-association with Ago proteins does not occur in case of miR-451. It is tempting to speculate that mature miRNAs are tightly anchored into Ago proteins and miRNA exchange between human Ago proteins is rather unlikely.

# CHAPTER 4 A miR-155-RULED MICRORNA HIERARCHY IN DENDRITIC CELL MATURATION AND MACROPHAGE ACTIVATION

The following chapter is based on the publication (Dueck et al., 2014). All experiments were conducted by A. Dueck, the cells used in this chapter were isolated by A. Eichner. The manuscript was written by A. Dueck and G. Meister.

## 4.1 Introduction

MiRNA-guided gene regulation has been implicated in literally all cellular pathways (for a description of miRNA biogenesis and function see 1.1). In addition, most cell types including immune cells such as macrophages or dendritic cells (DCs) contain specific miRNA signatures that help to establish and maintain gene expression programs. Consequently, these miRNA signatures change during cell stimulation or differentiation and miRNAs in conjunction with transcriptional regulation lead to alterations in global gene expression. Macrophages and DCs are not only essential for the immune system but have also important functions during embryonic development and are involved in numerous diseases including metabolic disease, atherosclerosis and cancer (Busch and Zernecke, 2012; Squadrito et al., 2013; Urbich et al., 2008). Both cell types originate from myeloid progenitor cells in the bone marrow (Shi and Pamer, 2011). A common progenitor that can give rise to DCs and macrophages is the monocyte. Monocytes are circulating white blood cells that can differentiate into tissue-resident macrophages or DCs in steady state and also in response to inflammation (Murray and Wynn, 2011). Among other white blood cells, macrophages and DCs perform phagocytosis and remove necrotic or apoptotic cells, pathogens or cell debris. In addition, macrophages contain specific receptors allowing for the uptake and clearance of different forms of low-density lipoproteins (LDL) such as oxidized LDL (oxLDL) or enzymatically hydrolyzed LDL (eLDL), which is frequently produced in plaques found in atherosclerosis (Figure 4-1) (Moore et al., 2013; Moore and Tabas, 2011).
Opposed to macrophages, DCs have an essential function for the initiation of adaptive immunity and upon specific stimuli, e.g. contact to a pathogen, immature DCs mature and migrate to the lymph nodes where they activate a T cell response against the pathogen. In *in vitro* systems, DC maturation is often achieved by lipopolysaccharide (LPS), a cell wall component of gram-negative bacteria. DCs can also be matured by LDL. However, the physiological function of this effect remains elusive.

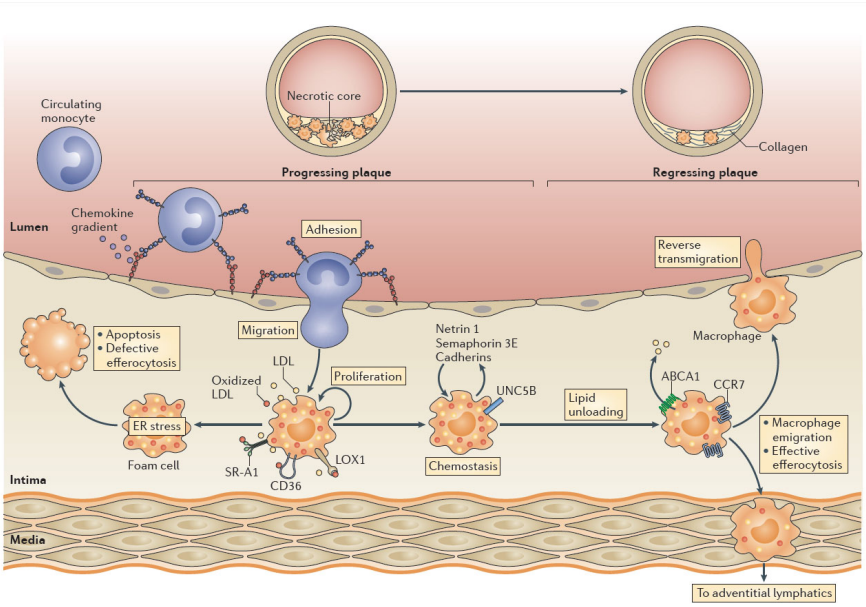



**Figure 4-1 Pathways involved in arteriosclerotic plaque formation.** Circulating monocytes are attracted by chemokines to sites of inflammation. They adhere to the endothelial cell layer and migrate into the intima, where they differentiate into macrophages. The macrophages take up low-density lipoproteins (LDL) and modified LDL, like oxidized and enzymatically hydrolyzed LDL. These macrophages transform into so-called foam cells and express retention molecules (such as netrin 1 and its receptor UNC5B, semaphorin 3E and cadherins). A foam cell can react to this state in two ways: In a condition of constant endoplasmic reticulum (ER) stress caused by uptake of LDL and inflammatory signaling, the cell undergoes apoptosis which leads to the formation of the necrotic core when coupled with effective efferocytosis typical for advanced arteriosclerosis. On the other hand, the cell can up-regulate migration factors and down-regulate retention factors resulting in an unloading of lipids and migration into the lumen or the lymph. CCR7, CC-chemokine receptor 7; ABCA1, ATP-binding cassette subfamily A member 1; LDL, low-density lipoprotein; LOX1, lectin-like oxidized LDL receptor 1; SR-A1, scavenger receptor A1. Figure was taken from Moore et al., 2013.

MiRNAs play important roles in the hematopoietic system. For example, many roles for miR-155 in the immune system have been reported. MiR-155 knock out studies have shown that miR-155 is important for T helper cell differentiation and thus an optimal T cell-dependent antibody response (Thai et al., 2007). In addition, it has been demonstrated that miR-155 is also important for the function of B lymphocytes and DCs (Rodriguez et al., 2007). Genome-wide miR-155 target identification approaches identified regulated mRNAs important for B cell regulation, lymphocyte homeostasis as well as cytokines, chemokines and transcription factors (Loeb et al., 2012; Rodriguez et al., 2007). MiR-155 function has also been implicated in disease such as autoimmune inflammation (O'Connell et al., 2010) or atherosclerosis (Nazari-Jahantigh et al., 2012a). In macrophages, miR-155 is up regulated during inflammatory response (O'Connell et al., 2007) and in DCs, miR-155 together with miR-221 regulate cell development, apoptosis and interleukin-12 (IL-12) production. In addition to miR-155, many other miRNAs play important roles in macrophage and DC biology (Busch and Zernecke, 2012; Nazari-Jahantigh et al., 2012b; Zernecke, 2012). For example, miR-147 regulates inflammatory responses (Liu et al., 2009) and miR-210 regulates the production of proinflammatory cytokines in murine macrophages (Qi et al., 2012).

Although several profiling studies aiming at the identification of miRNAs that are involved in macrophage or DC functions have been performed, the interplay between individual miRNAs in such miRNA regulatory networks has not been investigated. We therefore wanted to analyze miRNA networks and hierarchies in these immune cells in more detail.

## 4.2 MiRNA expression patterns in differentially matured DCs

Immature DCs mature upon contact with various stimuli. To analyze whether different stimuli cause different miRNA expression responses in monocyte-derived murine DCs, we matured DCs with LPS, oxLDL, eLDL and LDL and small RNAs were extracted, cloned and sequenced (Figure 4-2).

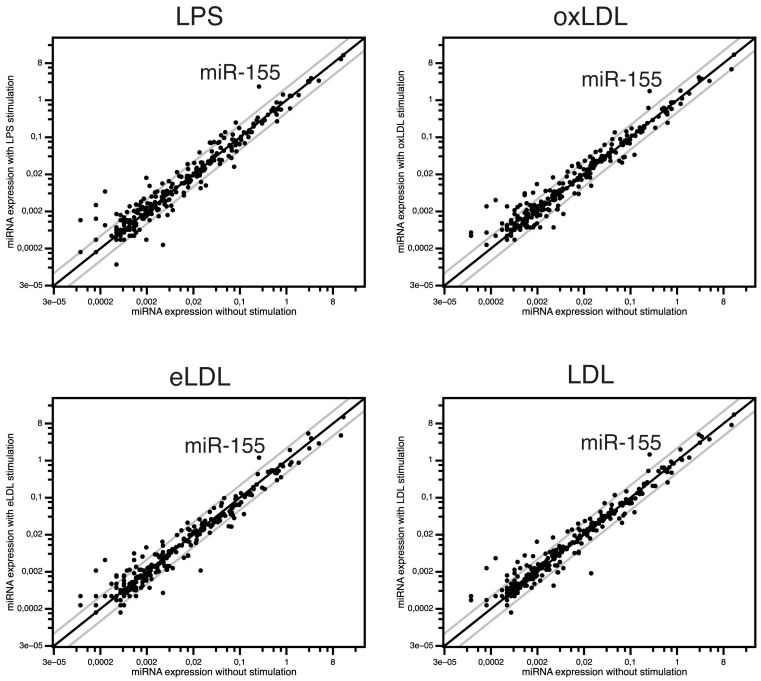



**Figure 4-2** MiRNA expression levels of wild type (wt) immature dendritic cells (DCs) are plotted against the miRNA expression levels of matured DCs (mDCs). Stimuli used were LPS (top left), oxLDL (top right), eLDL (bottom left) and LDL (bottom right).

While many miRNAs are expressed upon DC maturation independently of the specific maturation agent, several miRNAs appeared to be stimuli-specific (Figure 4-3, Table 4-1).

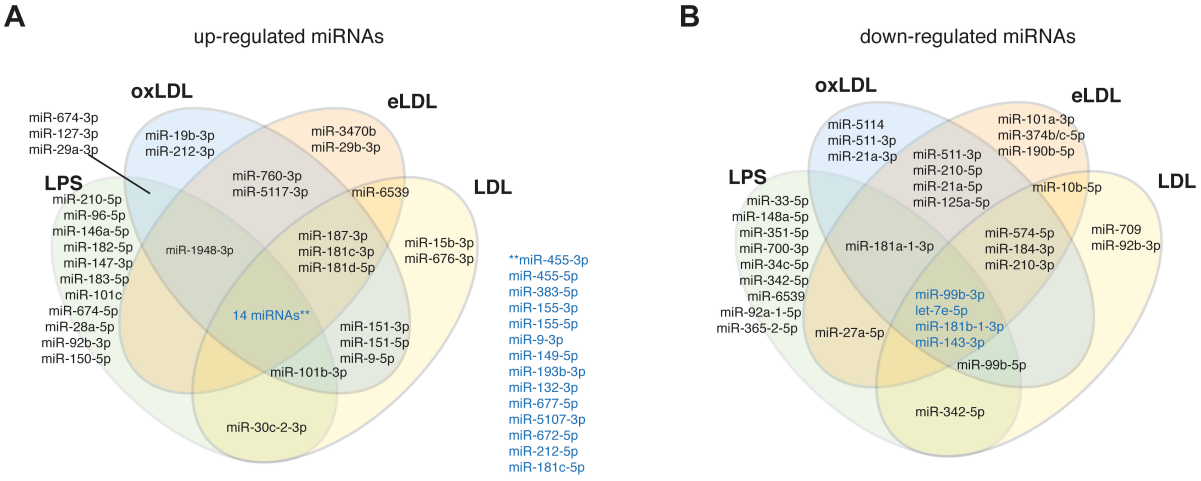



**Figure 4-3** (A) Venn diagram showing up regulated (> 2-fold up) miRNAs in DC maturation. (B) Venn diagram showing down regulated (> 2-fold down) miRNAs in DC maturation (Dueck et al., 2014).

For example, 14 miRNAs including miR-155 (both strands), miR-383-5p, miR-455 (both strands), miR-9-3p or miR-149-5p were up-regulated under all four conditions while miR-99b-3p, miR-143-3p, miR-181b-1-3p and let-7e-5p were down-regulated (Figure 4-3, A and B, shown in blue). Generally, the miRNA profiles obtained upon maturation with LDL, oxLDL or eLDL overlap stronger compared to maturation with LPS. Nevertheless, we identified also stimuli-specific miRNAs (e.g. miR-19b-3p or miR-21a-3p for oxLDL, miR-29b-3p or miR-101a-3p for eLDL and miR-15b-3p or miR-92b-3p for LDL maturation (Figure 4-3, A and B)). LPS-maturation revealed the largest set of specific miRNAs including miR-210-5p, miR-96-5p, which are up regulated (Figure 4-3, A) and miR-33-5p, miR-148a-5p or miR-351-5p that are down-regulated (Figure 4-3, B). Taken together, our comprehensive miRNA sequencing approach identifies miRNA signatures that are specific to different stimuli leading to DC maturation (Dueck et al., 2014).

**Table 4-1 Overview of libraries sequenced from primary murine dendritic cells.** DC, dendritic cells; Δ/Δ miR155, miR-155-deficient cells

| | **WT DC immature** | **WT DC +LPS** | **WT DC +oxLDL** | **WT DC +eLDL** | **WT DC +LDL** | **Δ/Δ miR155 DC immature** | **Δ/Δ miR155 DC +LPS** | **Δ/Δ miR155 DC +oxLDL** | **Δ/Δ miR155 DC +eLDL** | **Δ/Δ miR155 DC +LDL** |
|---|---|---|---|---|---|---|---|---|---|---|
| Valid reads (after trimming and filtering) | 13489698 | 14995685 | 17559327 | 15196800 | 22796027 | 19498613 | 16195524 | 15538770 | 15874994 | 15327433 |
| Sum of annotated miRNA reads | 7649900 | 8441379 | 9212941 | 7044314 | 12670159 | 10238279 | 8437759 | 8276415 | 7843454 | 7577882 |
| % of miRNAs in total library | 56.7 | 56.3 | 52.5 | 46.4 | 55.6 | 52.5 | 52.1 | 53.3 | 49.4 | 49.4 |

## 4.3 MiRNA expression patterns in differentially stimulated macrophages

Not only can DCs be matured by LPS, LDL, oxLDL or eLDL but macrophages can be stimulated with these agents, too. To identify miRNA regulatory networks in macrophages, we analyzed miRNA expression in differentially stimulated murine bone marrow-derived macrophages (Figure 4-4).

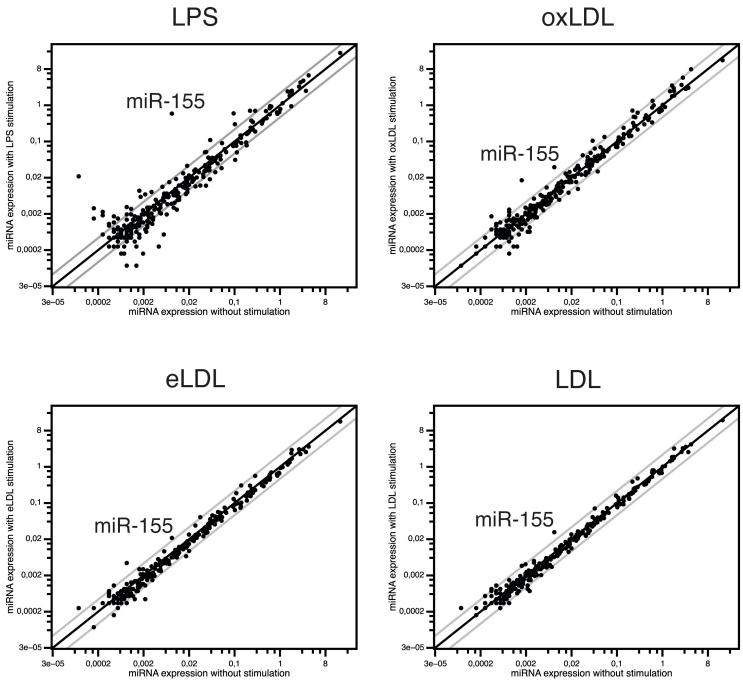



**Figure 4-4** MiRNA expression levels of wild type (wt) unstimulated macrophages are plotted against the miRNA expression levels of stimulated macrophages. Stimuli used were LPS (top left), oxLDL (top right), eLDL (bottom left) and LDL (bottom right).

Macrophages were treated with LPS, oxLDL, eLDL and LDL, small RNAs were isolated, cloned and analyzed by deep sequencing (Table 4-2). Similar to DCs, several miRNAs were differentially expressed (Figure 4-5). Interestingly, only miR-155 was up regulated and miR-3473b/e was down regulated under all four conditions (Figure 4-5, A and B, shown in blue).

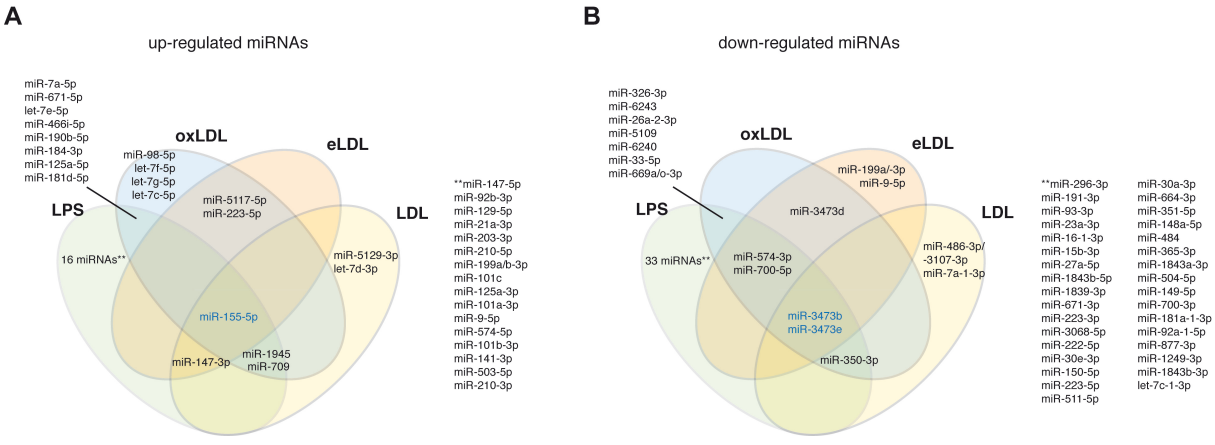



**Figure 4-5** (A) Venn diagram showing up regulated (> 2-fold up) miRNAs in stimulated macrophages. (B) Venn diagram showing down regulated (> 2-fold down) miRNAs in stimulated macrophages.

We also identified miRNAs that are specifically regulated by the individual stimuli in macrophages. For example, let-7f, g and c (all 5p strand) as well as miR-98-5p are up-regulated upon oxLDL stimulation, let-7a and e (both -5p) are also up-regulated as observed for LPS stimulation. Again, LPS stimulation

had the strongest effect on miRNA expression: miR-147-5p, miR-92b-3p, miR-129-5p or miR-21a-3p were up- und miR-93-3p, miR-191-3p or miR-27a-5p were down-regulated. Similar to DCs, our comprehensive sequencing analysis uncovered a number of stimuli-specific miRNAs upon macrophage stimulation (Dueck et al., 2014).

**Table 4-2 Overview of libraries sequenced from primary murine macrophages.** MΦ, macrophages; unstim., unstimulated; Δ/Δ miR155, miR-155-deficient cells

| | WT MΦ unstim. | WT MΦ +LPS | WT MΦ +oxLDL | WT MΦ +eLDL | WT MΦ +LDL | Δ/Δ miR155 MΦ unstim. | Δ/Δ miR155 MΦ +LPS | Δ/Δ miR155 MΦ +oxLDL | Δ/Δ miR155 MΦ +eLDL | Δ/Δ miR155 MΦ +LDL |
|---|---|---|---|---|---|---|---|---|---|---|
| Valid reads (after trimming and filtering) | 12804807 | 13888791 | 13793274 | 10574992 | 13188765 | 14653213 | 17146942 | 15190123 | 11980695 | 10254997 |
| Sum of annotated miRNA reads | 7471355 | 9687261 | 9389018 | 5632347 | 7570936 | 9048738 | 11572739 | 9678657 | 6990817 | 6490581 |
| % of miRNAs in total library | 58.3 | 69.7 | 68.1 | 53.3 | 57.4 | 61.8 | 67.5 | 63.7 | 58.4 | 63.3 |

Finally, we compared stimuli-specific miRNA regulation between DCs and macrophages to identify miRNAs that are generally affected by the used stimuli. Indeed, several miRNAs overlapped between DCs and macrophages upon LPS, oxLDL and eLDL treatment (e.g. miR-155-5p, miR-210-5p, miR-147-3p, miR-33-5p or miR-27a-5p; Table 4-3).

**Table 4-3 miRNAs commonly regulated in macrophages and DCs (> 2 fold)**

| LPS (up) | LPS (down) | oxLDL (up) | oxLDL (down) | eLDL (up) | eLDL (down) | LDL (up) | LDL (down) |
|---|---|---|---|---|---|---|---|
| miR-155-5p | miR-33-5p | miR-155-5p | no overlap | miR-155-5p | no overlap | miR-155-5p | no overlap |
| miR-210-5p | miR-27a-5p | miR-5117-5p | | miR-5117-5p | | | |
| miR-147-3p | miR-148a-5p | miR-181d-5p | | | | | |
| miR-101c | miR-351-5p | | | | | | |
| miR-101b-3p | miR-700-3p | | | | | | |
| miR-92b-3p | miR-223-5p | | | | | | |
| | miR-181a-1-3p | | | | | | |
| | miR-92a-1-5p | | | | | | |

## 4.4 Identification of miR-155-dependent miRNAs in DCs and macrophages by comparison of fold changes

MiR-155 is one of the most important miRNAs in the hematopoietic system. We hypothesized that miR-155 not only affects the expression of mRNAs but maybe also of other miRNAs. In order to identify DC miRNAs that are dependent on miR-155 and are regulated upon maturation, small RNAs were cloned and sequenced from bone marrow-derived DCs originating from wt or miR-155-deficient mice (Figure 4-6). Wt cells and miR-155 deficient cells were analyzed separately and their fold changes between immature cells and cells treated with the different stimuli were plotted against each other to find miR-155-dependent miRNAs (Figure 4-6). A threshold of 50 reads per million of total reads was chosen to reduce low abundant and probably biologically irrelevant miRNAs. In such plots, miR-155-independent miRNAs are found between the two grey lines (indicating a 2-fold change), while

miR-155-dependent miRNAs are found as outliers (colored dots). Several miRNAs were up or down regulated in a miR-155-dependent manner in LPS (Figure 4-6, upper left), oxLDL (upper right), eLDL (lower left) or LDL (lower right) matured DCs. For example, miR-10b-5p was mis-regulated in miR-155 deficient cells in all four different stimuli. MiRNAs miR-184-3p and miR-187-3p were affected in three out of four conditions (Dueck et al., 2014).

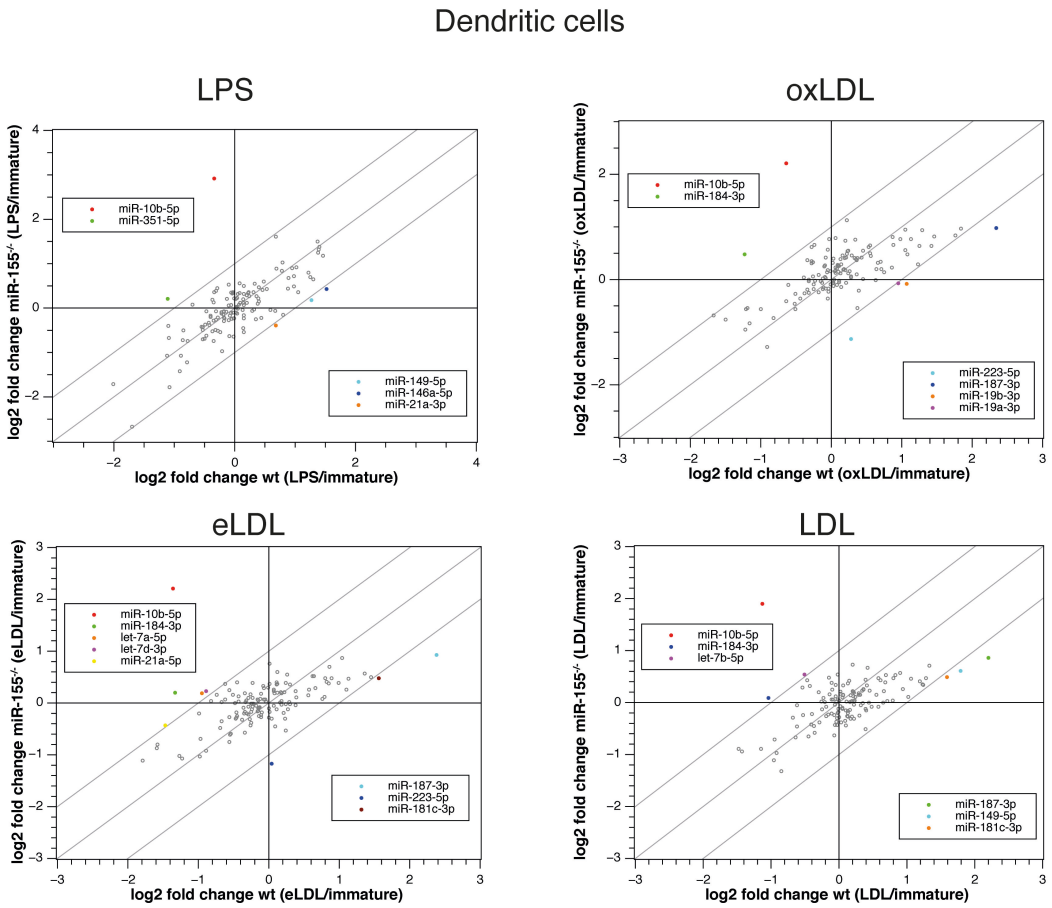



**Figure 4-6** Scatter plots comparing fold changes in the wild type and the miR-155 deficient condition in DCs. Outer lines mark a difference of 2-fold mis-regulation.

We next analyzed miR-155-dependent miRNAs during macrophage stimulation (Figure 4-7). To identify miR-155-dependent miRNAs, we stimulated wt and miR-155 deficient macrophages with LPS, LDL, oxLDL and eLDL, isolated small RNAs and analyzed them by deep sequencing. Similar to DCs, we plotted the fold changes (stimulated vs. unstimulated) of all miRNAs found in miR-155 deficient cells against the fold changes of miRNAs found in wt macrophages (Figure 4-7). Again, we found several miRNAs whose expression depends on miR-155. MiR-709 was affected under all four conditions and miR-223-5p was mis-regulated in two out of four conditions. Again our data indicate that a miR-155-dependent miRNA network exists in macrophages as well. Our sequencing data suggests that several miRNAs might be affected by miR-155 during the maturation of bone marrow-derived murine DCs and macrophages suggesting a hierarchical system directed by miR-155.

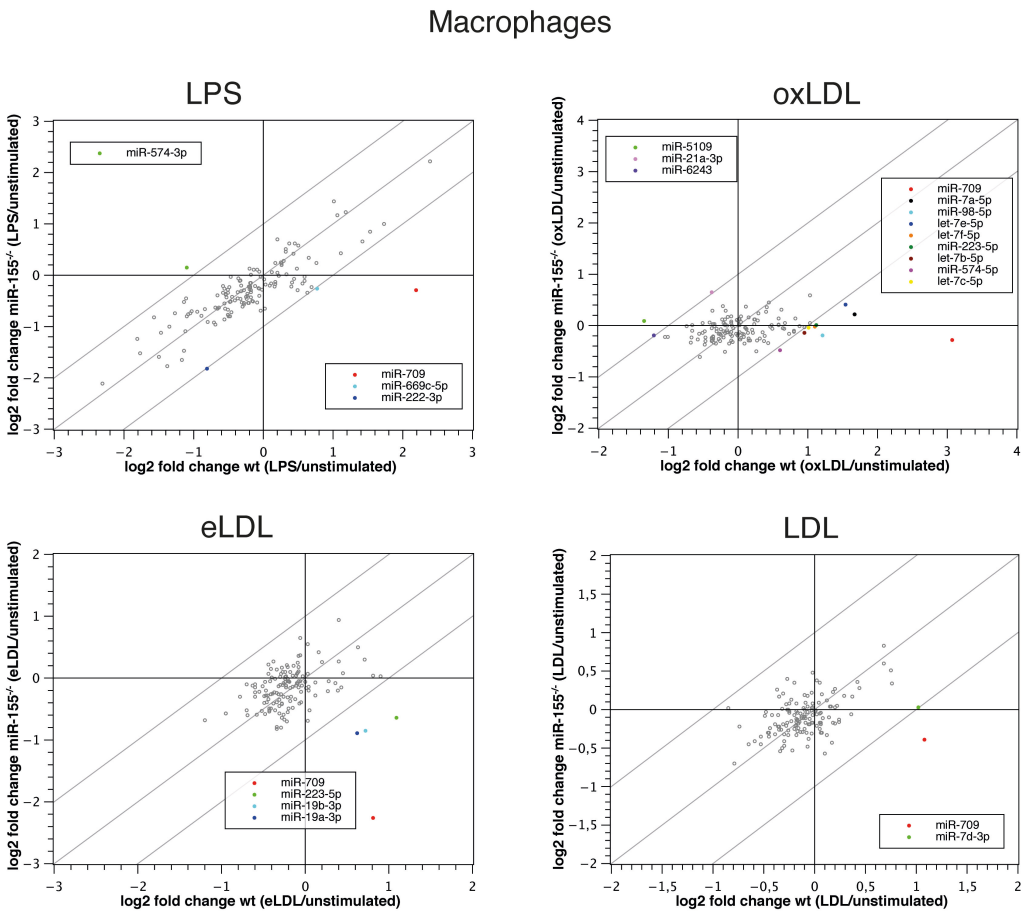



**Figure 4-7** Scatter plots comparing fold changes in the wild type and the miR-155 deficient condition in macrophages. Outer lines mark a difference of 2-fold mis-regulation.

## 4.5 Identification of miR-155-dependent miRNAs in DCs and macrophages by absolute expression levels

Comparing fold changes with each other might lead to a loss of miRNAs that do not pass our filters but could nevertheless be significantly up or down regulated based on absolute expression levels. Therefore, we compared expression levels between wt and miR-155-deficient DCs (Figure 4-8, A, left) and macrophages (Figure 4-8, A, right) with each other. Indeed, we identified several miRNAs that are differentially expressed, among them are miR-455-3p, miR-143-3p and miR-210-3p.

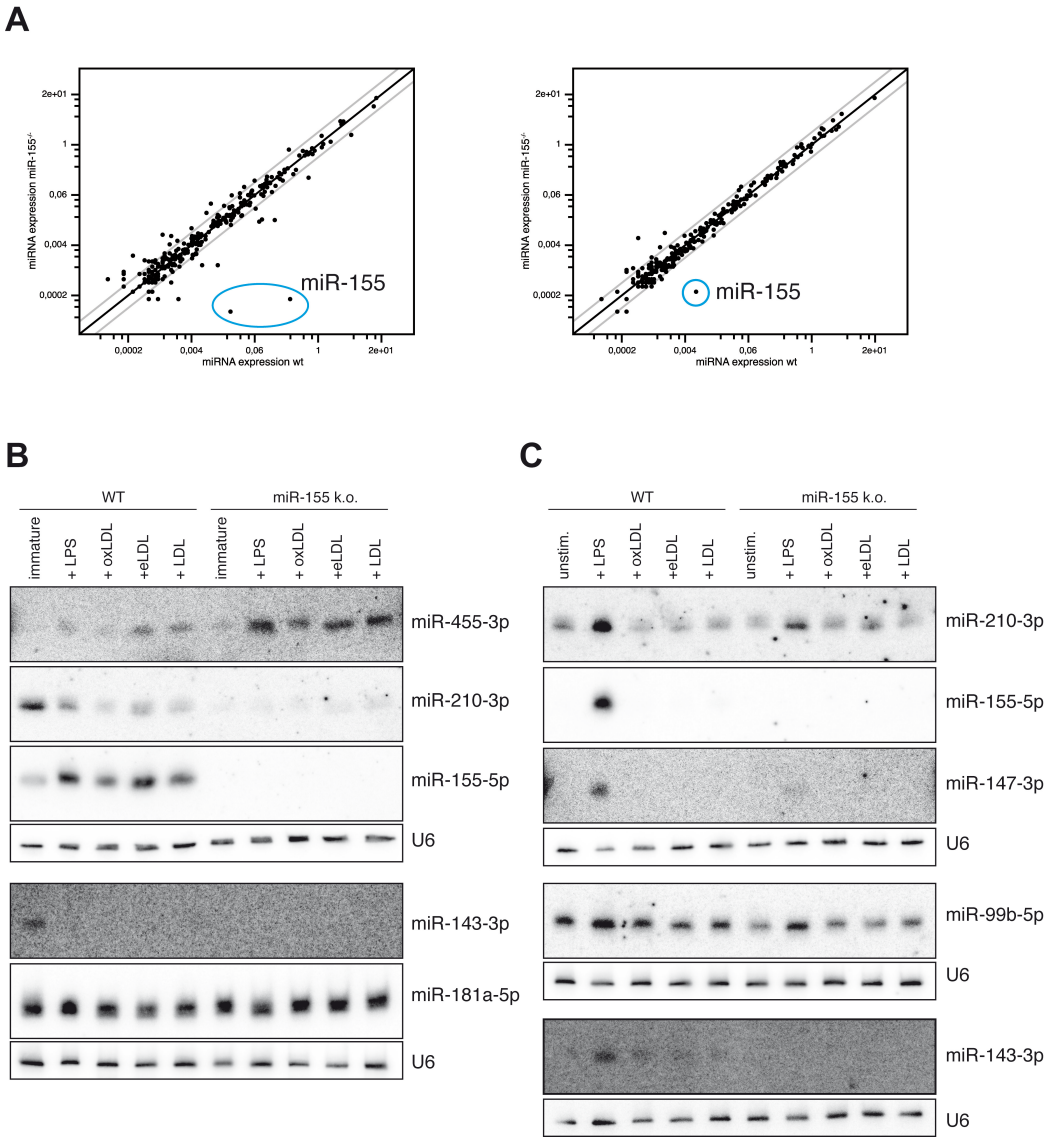



**Figure 4-8** (A) Scatter plots comparing miRNA expression in wt versus miR-155 deficient immature DCs (left panel) and unstimulated macrophages (right panel). Outer lines mark a 2-fold up- or down-regulation. The blue circles highlight miR-155 expression (both 5p and 3p for dendritic cells, 5p for the macrophages). (B) and (C) Expression profiles of individual miRNAs in DCs and macrophages, respectively, validated by Northern blotting. Wt and miR-155 deficient DCs (B) or wt or miR-155 deficient macrophages (C) were treated with LPS, oxLDL, eLDL and LDL and individual miRNAs were analyzed by Northern blotting. Northern blots using porbes specific to U6 were used as loading control for each specific membrane. The Northern blots were performed using RNA from one experiment (n=1).

To consolidate our model of miR-155-dependent miRNA expression, we validated the expression levels of several miRNAs in DCs by Northern blotting (Figure 4-8, B). Analysis of miR-155 confirmed that the miR-155 knock out DCs are indeed miR-155 deficient. MiR-210-3p is down regulated upon all four stimuli in wt cells. However, in miR-155 deficient cells, miR-210-3p expression is significantly reduced indicating that miR-155 influences miR-210-3p expression in DCs. Similar results were obtained for miR-143-3p. Furthermore, miR-455-3p is expressed at rather low levels in immature wt and miR-155 deficient DCs. Upon stimulation, miR-455-3p is much stronger expressed when miR-155 is not present indicating that miR-155 inhibits the expression of miR-455-3p under these conditions. In contrast, miR-181a-5p is rather independent of miR-155 expression.

We next validated our miRNA sequencing results by Northern blotting (Figure 4-8, C). While miR-155, miR-147-3p and miR-143-3p were not found in miR-155-deficient macrophages, miR-99b-5p was slightly down regulated and miR-210-3p rather unaffected. Therefore, we clearly show that several miRNAs are affected by the expression of miR-155 in macrophages as well. These results confirm a miR-155-dependent miRNA hierarchy during DC maturation and macrophage stimulation.

## 4.6 MiR-155 targets C/EBPβ leading to miR-455 regulation

In order to identify the molecular mechanisms underlying miR-155-regulated miRNAs, we investigated miR-455-3p expression, which is up regulated in miR-155 deficient cells (Figure 4-8, B). It has been demonstrated before that miR-155 regulates the transcription factor C/EBPβ and affects adipocyte differentiation (Chen et al., 2013). Therefore, we searched for putative C/EBPβ binding sites close to potential *MiR-455* promoters (Figure 4-9, A). Indeed, three potential sites are found in the *MiR-455* containing intron of the *Col27a1* gene. In addition, another four binding sites are present at the *Col27a1* promoter. Since miR-155 is induced the strongest in cells stimulated with LPS, we next analyzed C/EBPβ levels after LPS maturation in wt and miR-155 knock out DCs (Figure 4-9, B). Upon LPS treatment, C/EBPβ is down regulated, which is consistent with the observed miR-155 increase. Strikingly, C/EBPβ is strongly elevated in miR-155 knock out cells matured with LPS, which in turn might lead to up regulation of miR-455-3p (Figure 4-9, B and C) (Dueck et al., 2014).

## 4.7 Discussion

MiRNAs are regulatory molecules that help to establish gene expression profiles by regulating protein-coding genes. Such regulatory networks lead to the manifestation of cell identity. Thus, miRNA expression profiles are specific to cell types and also the target repertoire is often cell type-specific. Although very often not well understood in molecular detail, the general effects of miRNAs on mRNA target expression are rather clear. In contrast, consequences of miRNAs on the expression of other miRNAs have not been analyzed. Here, we have characterized miRNA expression profiles in wt DCs or macrophages and compared them with profiles from miR-155$^{-/-}$ DCs or macrophages. This approaches allows for the establishment of a miRNA hierarchy directed by miR-155.

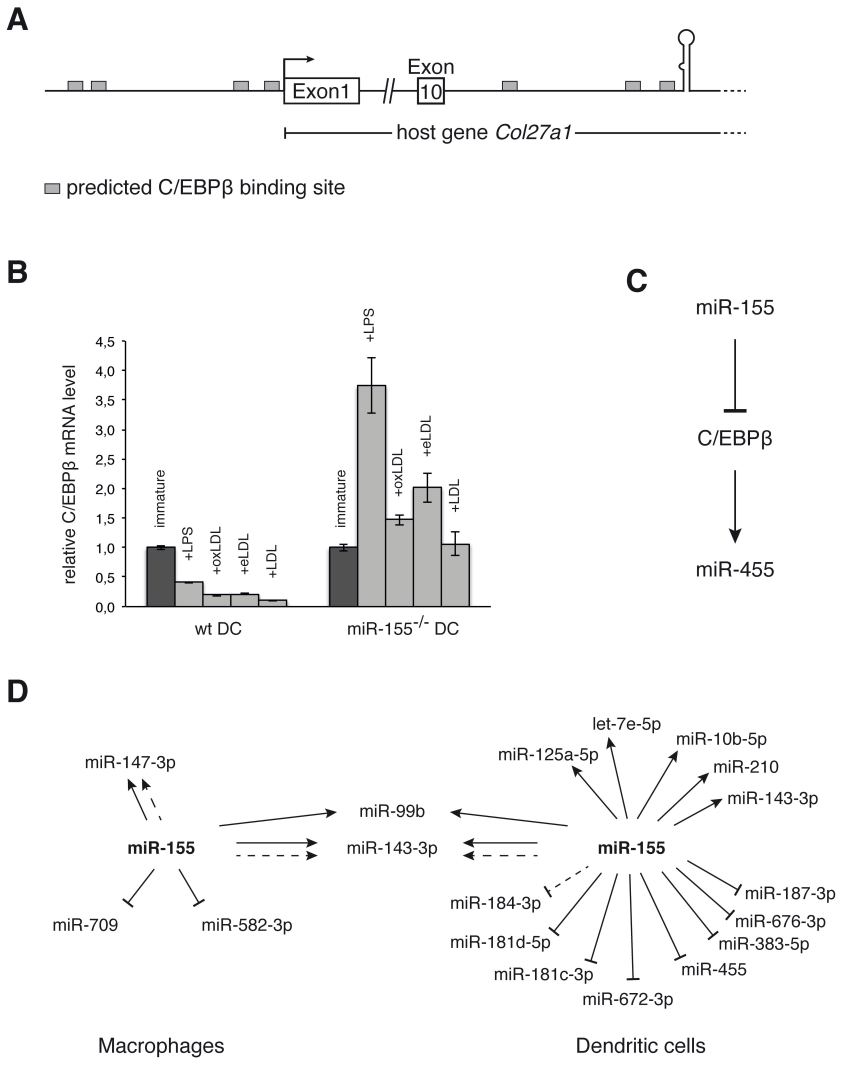



**Figure 4-9** (A) Schematic representation of the promoter region of the putative host gene of miR-455, *Col27a1*, and the upstream intronic sequence of *miR-455*. Predicted binding sites for the transcription factor C/EBPβ are shown as grey boxes. (B) qRT-PCR measurement of *C/EBPβ* mRNA in immature and matured (LPS, oxLDL, eLDL, LDL) dendritic cells, both in wild type and in miR-155 deficient cells. Levels were normalized to the *36B4* housekeeping gene and the respective level in immature cells. Dark grey, immature condition; light grey, matured cells. (C) Proposed regulatory model of miR-155 influenced miR-455 expression. (D) Summary of miRNAs regulated by miR-155. Solid arrows represent supportive roles of miR-155 in expression levels, while dotted arrows represent supportive roles in expression changes (data shown in Figure 4-6). Lines with a bar represent the inhibition by miR-155 in either expression level (solid line) or expression change (dotted line). Left, regulated miRNAs in macrophages, right, regulated miRNAs in DCs.

In macrophages, the expression of miR-99b-5p, miR-147-3p and miR-143-3p is stimulated by miR-155, while miR-709 and miR-582-3p are repressed (Figure 4-9, D). In DCs, miR-99b (both strands), miR143-3p, miR-125a-5p, let-7e-5p, miR-10b-5p, miR-210 (both strands) and miR143-3p are stimulated and miR-187-3p, miR-676-3p, miR-383-5p, miR-455 (both strands), miR-672-3p, miR-181c-3p, miR-181d-5p or miR-184-3p are repressed (Figure 4-9, D) (Dueck et al., 2014).

How is such a regulation achieved? Most likely, either transcriptional activators or repressors that bind to miRNA promoters and regulate miRNA expression are directly targeted by miR-155 leading to stimulation or inhibition of expression. Supporting this model, we find that *MiR-455* contains several C/EBPβ binding sites at its putative promoter regions. Interestingly, C/EBPβ is a direct miR-155 target. However, miR-455 is mildly up regulated upon maturation of wt DCs as well suggesting that additional pathways might contribute to overall miR-455 expression levels. Other transcription factors that are miR-155 targets and could affect miRNA expression are Ship1 (O'Connell et al., 2009), Pu.1 (Vigorito et al., 2007) or Socs1 (Androulidaki et al., 2009), for example. In addition, it is also conceivable that regulatory proteins that directly interact with miRNA processing intermediates could be targeted by miR-155. Such regulatory factors are often cell-type-specific. This might explain why we observe considerable differences between miR-155-dependent miRNAs in DCs and macrophages. A better understanding of the miR-155 target spectrum will help elucidating how such a miRNA hierarchy is established and maintained.

Recently, it has been shown that miR-709 regulates miR-15a and miR-16 (Tang et al., 2012). In this study, murine miR-709 seems to localize to the nucleus where it directly binds to the pri-miR-15a/16 transcript and inhibits its expression. Although rather low abundant, miR-709 is also among the miRNAs that are stimulated by miR-155 in our study. However, in our system, we do not observe effects on miR-15a/16 suggesting again cell-type-specific effects. Furthermore, miR-155 is not only important for macrophage or DC function but also in many other cells of the hematopoietic system such as CD8+ T cells (Dudda et al., 2013) or T helper cells (Thai et al., 2007). In addition, miR-155 is important for brown fat tissue differentiation (Chen et al., 2013) and has also been implicated in a number of different types of cancer (Garzon et al., 2009). It will be interesting to see whether miR-155 ruled miRNAs also exist in other tissues and what the consequences are for cell function and disease.

# CHAPTER 5 CHARACTERIZATION OF SMALL RNAS IN THE GREEN ALGA *VOLVOX CARTERI*

The results of this section are part of two publications (Dueck et al., manuscript in preparation and Evers et al., manuscript in preparation). A. Dueck performed most experiments and analyzed the data. Maurits Evers (Department of Functional Genetics, University of Regensburg) designed the miRNA identification pipeline and performed most of the bioinformatics analyses. Eugene Berezikov (University of Groningen) assembled RNA Seq data to a preliminary transcriptome. Katharina Unger and Norbert Eichner helped with sample preparations, setup and cloning of expression constructs and qRT-PCR measurements. Rainer Merkl (Institute for Biophysics and Physical Biochemistry, University of Regensburg) performed transcriptomic search for putative RNAi-related proteins. A. Dueck and Gunter Meister designed the experiments and wrote the manuscript.

## 5.1 Introduction

Many members of the green algae have been used to study a variety of different processes. Two of its members, unicellular *Chlamydomonas reinhardtii* and multicellular *Volvox carteri* serve as model organisms for studying evolutionary aspects such as the development of multicellularity or division of labor between soma and germ line. While there have been studies on small RNAs acting in *Chlamydomonas reinhardtii*, the existence and nature of small RNAs in *Volvox carteri* and the extent of the mediated regulation has not been analyzed so far. As the current knowledge on the biogenesis and action of small RNAs in plants has been summarized in the general introduction (see 1.1.2), the following paragraphs will therefore focus on an introduction to *Volvox carteri* as a model organism.

### 5.1.1 The green alga *Volvox carteri*

*Volvox carteri* (referred to as *Volvox* from hereon) is a green alga belonging to the group of volvocine algae. These, in turn, belong to the group of Chlorophyta, which are only distantly related to the Streptophyta (including land plants) and the Rhodophyta (red algae).

*Volvox* is a freshwater alga living in puddles, ponds or deep lakes. It lives as an obligate photoautotrophic organism, its growth is therefore photosynthesis limited (Kirk, 1997). *Volvox* has two genetically distinct genders, male and female, which both can reproduce asexually or sexually. During the asexual reproduction cycles, the genders are morphologically very similar (Kirk, 1997) (Figure 5-1).

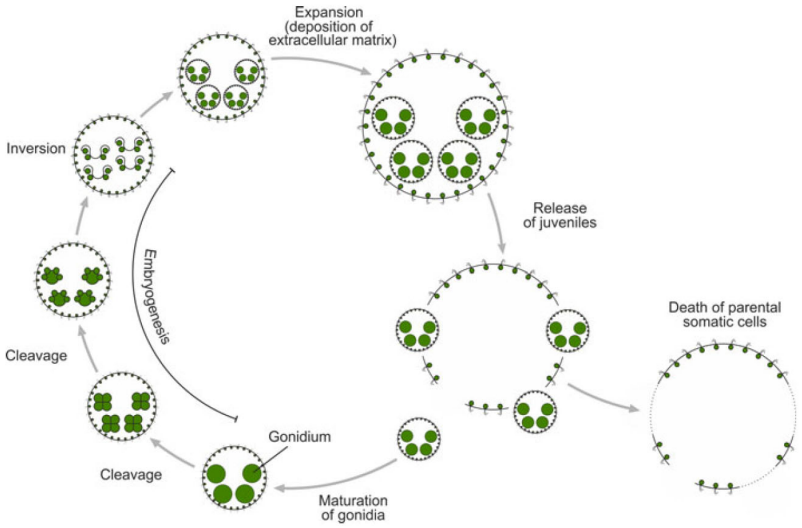



**Figure 5-1 Asexual life cycle of *Volvox carteri*.** A mature spheroid contains ~16 gonidia (reproductive cells) and ~2000 somatic cells. During embryogenesis, each gonidium undergoes cleavage with 11-12 divisions (some of them asymmetric) until all cells that will be present in the adult are produced. The layout of the cells is inside-out with respect to the adult organism and therefore the embryos need to undergo inversion. After inversion, the spheroids begin to produce extracellular matrix, leading to a size expansion. When the juvenile daughter colonies have reached a certain size, they are released from the mother colonies, the parental somatic cell hull dies after a few days. This life cycle usually takes ~48h. Figure taken from Hallmann, 2011 (Hallmann, 2011).

In its asexual form, *Volvox* consists of ~2000 somatic cells and ~16 gonidia (reproductive cells). The somatic cells form a sphere that is filled with extracellular matrix. The gonidia are anchored in the somatic cell hull, but are facing the inside of the sphere (Figure 5-2, A and B).

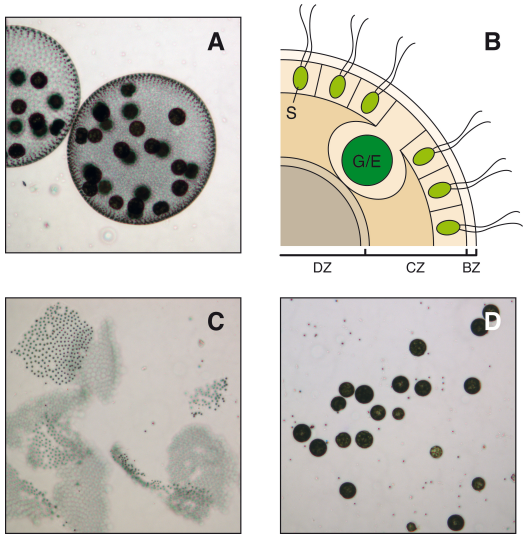



**Figure 5-2 Morphology of *Volvox carteri*.** (A) An adult *Volvox* spheroid consists of ~2000 somatic cells and ~16 gonidia. (B) Structure and layers of the extracellular matrix (ECM) (Kirk et al., 1986). The outermost boundary zone (BZ) is followed by the central zone (CZ). Here, somatic cell and gonidia are embedded in distinct compartments. The inner core is called deep zone and constitutes more than half of the volume of a spheroid. (C) and (D) By shearing *Volvox* spheroids, somatic cell sheets (C) can efficiently be separated from gonidia (D).

To reproduce asexually, each gonidium undergoes a series of rapid cleavages (Figure 5-1). At the stage of 32 cells, the cells divide asymmetrically to form the daughter gonidia. The cell fate during asymmetric cell division is determined by the amount of cytosol (quantity) rather than the content of cytoplasm (quality). If a cell is > 8$\mu$m at the end of the cleavage event, it will always become a reproductive cell (gonidia) whereas smaller cells develop into somatic cells (Kirk et al., 1993). After all cells of the daughter spheroid have been formed, it enlarges by producing extracellular matrix and undergoes inversion in order to position the gonidia to the inner side of the spheroid. The process of inversion is necessary because the fully-cleaved *Volvox* embryo is a hollow sphere with all potential flagellar ends of the somatic cells pointing inwards. First, a "dent" appears causing the sphere to open up. Through a wave of motions, the embryo everts itself thus turning its inside out (Figure 5-3; for a detailed description see (Kirk, 1997)). By further production of extracellular matrix, the spheroid reaches its adult size.

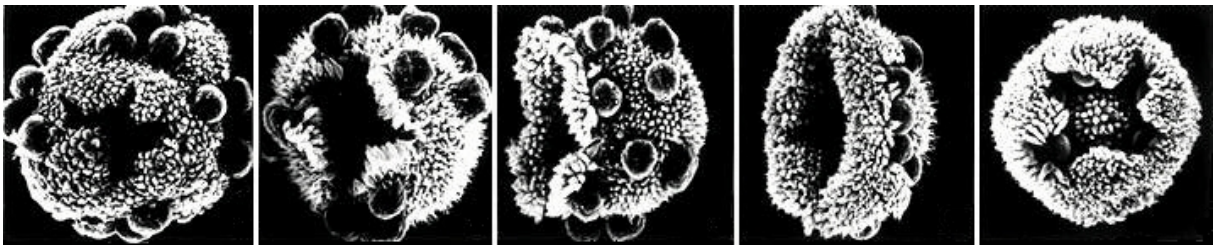


**Figure 5-3 Inversion of a *Volvox* embryo.** The fully-cleaved *Volvox* embryo is a hollow ball with gonidia pointing outwards (big cells) and the flagellar side of the small somatic cells pointing inward. After a rift appears at one side of the embryo, a wave of motion flows through the embryo. The movement of the somatic cells causes a rolling motion on itself thus turning its inside out. Picture was taken from (Kirk et al., 1982).

The sexual life cycle is induced by the sex-inducing pheromone. It is an approximately 30 kD glycoprotein and is highly effective, that means able to induce sexual development, at concentrations of $10^{-16}$ M (100 attomolar) or below (Starr and Jaenicke, 1974). Upon induction, female and male strains undergo a modified life cycle. In the female line, the modified gonidia give rise to embryos that later produce 32 egg cells (one more round of asymmetrical cleavage). In the male line, the modified gonidia produced spheroids containing so-called androgonidia. These androgonidia initiate another round of cleavage events and sperm packets are formed and released into the water.

#### *5.1.1.1 Small RNAs in the green alga Chlamydomonas reinhardtii*

The closest unicellular relative of *Volvox* is *Chlamydomonas reinhardtii* (referred to as *Chlamydomonas* from hereon). While the group of Chlorophyta (including Volvocales) and Streptophyta (including land plants) diverged over 1.2 billions of years ago, *Volvox* diverged from *Chlamydomonas* only ~236 million years ago (Herron et al., 2009). Two laboratories, Qi and colleagues and Baulcombe and colleagues, have analyzed the small RNA repertoire of *Chlamydomonas* in detail (Molnar et al., 2007; Zhao et al., 2007).

*Chlamydomonas* predominantly contains small RNAs of a size of 20-22nt, bearing mostly U (and to a lesser amount A) at their 5' end. 4-11 loci were found that produced phased siRNAs that presumably

could be similar to *trans*-acting siRNA in *Arabidopsis*. Both groups found a number of potential miRNA sequences that were shown to contain a 5' phosphate group (Zhao et al., 2007) and to be modified at the 3' end with a 2'O-methyl group (Molnar et al., 2007). These features are shared with miRNAs in land plants. The precursor structure differs, however. While a general rule for processing of miRNA precursors in higher plants is the recognition of a basal structure and the processing occurring mostly from bottom to top, many *Chlamydomonas* miRNAs are located 18-24 nt away from the tip of the precursor (Molnar et al., 2007). This implies an importance of structural elements for the processing of the precursor. Another part of precursors is characterized by very long, double-stranded hairpins that hybridize almost perfectly complementary, yielding putatively more than one mature miRNA (Molnar et al., 2007; Zhao et al., 2007). This suggests that these miRNA precursors might be considered as young miRNA in an evolutionary context. Up to now, no conservation of *Chlamydomonas* miRNAs and miRNAs in land plants or green algae has been detected (see also this work).

Both groups could show that miRNAs can induce cleavage of (some of) their predicted targets and that the cleavage activity corresponded to a miRNP particle size of ~200kDa. Until now, three Dicer-like proteins and at least two complete Argonaute proteins (with PAZ and Piwi domain) have been annotated in *Chlamydomonas* (Schroda, 2006), but a detailed analysis of proteins of the RNAi machinery is still lacking.

So far, 86 miRNAs of *Chlamydomonas* contained in 50 hairpins are deposited in miRBase (release 20, June 2013) and this number is likely to increase with the examination of miRNA profiles under diverse conditions and life stages using an ever growing sequencing depth (Shu and Hu, 2012).

### 5.1.2 Model organism for the evolution of multicellularity

The incident of the evolution of multicellularity can be found a number of times during the history of life on earth (Baldauf et al., 2000; Bonner, 2001). Much is known about the molecular mechanisms underlying the division of labor and cellular differentiation, but since most cellular systems are far too complex to analyze, little knowledge was obtained so far on the mechanism of *de novo* evolution of multicellularity. The group of the volvocine algae, including *Chlamydomonas* and *Volvox*, offers a simple system, which is highly useful in more than one way. First, both organisms are widely used as experimental models. Second, in the family of volvocine algae multicellularity has evolved several times. *Chlamydomonas* with its unicellular nature is one extreme while the other is *Volvox* with full division of labor between soma and germ line and third, *Volvox* and *Chlamydomonas* have diverged from their common ancestor only about 200 million years ago (Herron et al., 2009). By comparison, animals and plants have evolved from their respective unicellular ancestors >600 and ~450 million years ago (Herron et al., 2009; Kirk, 1997).

*Volvox* does not only exhibit complete division of labor, it also evolved asymmetric cell division with small somatic cells and large gonidial precursors. In addition, it acquired a complex extracellular matrix, a sexual program with egg- and sperm cells (oogamy) that is very different from the isogametic

program (gametes of similar morphology) of *Chlamydomonas* and embryo inversion (see also Figure 5-1) (Nishii and Miller, 2010).

The timing of the different steps leading to the evolution of multicellularity is not yet clearly defined. David Kirk postulated a twelve-step program for the evolution of multicellularity and division of labor with reference to the volvocine algae (Kirk, 2005). This model was further refined (order of events/gains and losses of traits in some lineages) by Michod and colleagues (Herron et al., 2009) using a multiple gene data set with multiple fossil calibrations. In short, there are twelve major steps needed to recapitulate the evolution from *Chlamydomonas* to *Volvox*. These acquired characteristics include, for example, establishment of organ polarity, partial or complete inversion of the embryo, germ/soma division of labor and asymmetric cell division (Figure 5-4).

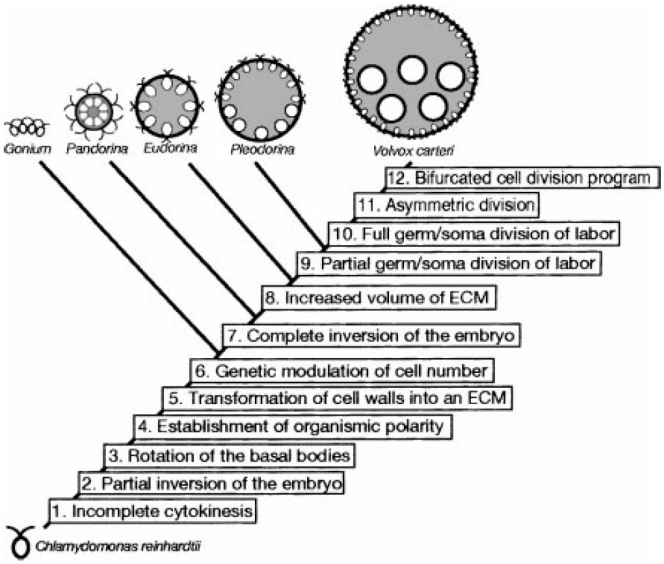



**Figure 5-4 The twelve-step program postulated by Kirk (Kirk, 2005).** Each step displays a requirement for the acquisition of complexity from a *Chlamydomonas*-like ancestor to modern *Volvox carteri*. Steps 1-6 are all represented in *Gonium pectorale*, the basal member of the genus *Gonium*. For each further step, one exemplary species can be found. Note that this scheme depicts a simplified volvocine cladogram. For a detailed description of each step and classification, please refer to Kirk, 2005 and Herron et al., 2009 (Herron et al., 2009; Kirk, 2005).

In the last decade, many publications have increased and deepened the knowledge of the links and connections leading from a unicellular to a multicellular organism. The exact mechanisms and involved proteins remain to be analyzed in depth. The technical progress in sequencing methods and therefore the availability of genome and transcriptome data will help to find the genetic differences responsible for evolutional steps in the volvocine algae.

### 5.1.3 Relationship between *Volvox* and *Chlamydomonas*

As mentioned earlier, *Volvox* and *Chlamydomonas* are evolutionarily only about 200 millions of years apart. They are closely related in terms of genomic coding potential, but *Volvox* is visibly capable of

skills that *Chlamydomonas* is not. The following paragraph will summarize the current knowledge on genes of *Volvox*, which are essential for the accomplishment of certain steps, e.g. germ-soma division of labor, asymmetric division and inversion, and compare their conservation and function with *Chlamydomonas*.

#### *5.1.3.1 Genes essential for inversion – the inversion genes*

Three important genes for inversion have been characterized so far, the inversion genes *invA*, *invB* and *invC*. InvA is a kinesin that localizes to the cytoplasmic bridges connecting neighboring cells and provides the force to complete this inside-out motion (Nishii et al., 2003). InvB and InvC are important proteins for the enlargement of the embryo. Without this enlargement, the embryo would be too small for inversion to take place (Ueki and Nishii, 2008, 2009). All three genes have highly similar orthologs in *Chlamydomonas*.

#### *5.1.3.2 Asymmetric cell division – the gonidialess genes*

One major class of genes in *Volvox* that is essential for asymmetric cell division is the group of *gonidialess* genes (Kirk et al., 1991). Interestingly, these genes are not required for symmetric cell divisions. The first gene cloned and characterized is *glsA* (*gonidialessA*), which encodes a protein that contains multiple potential protein binding sites including a chaperone-like domain, the J domain (Miller and Kirk, 1999). This domain can also be found in Hsp40-class chaperones (Caplan et al., 1993) and functions there as an activator domain for Hsp70 chaperones. Mutations in this domain lead to a loss of function of GlsA. The molecular mechanism of GlsA function has not been elucidated so far. It is known that GlsA localizes to the region of the mitotic spindle in early embryo blastomeres (Miller and Kirk, 1999). The *Chlamydomonas* genome also encodes a GlsA-like protein that was named Gar1 (GlsA related 1). GlsA and Gar1 are highly similar, sharing over 80% of amino acid sequence. Indeed, it could be shown that *gar1* is able to rescue *glsA* mutants (Cheng et al., 2003). This finding indicates that the ability to perform asymmetric cell division is not due to encoding a GlsA like protein, rather, the shifting of the GlsA protein interaction network seems to responsible.

#### *5.1.3.3 Germ-soma division of labor – RegA*

Where GlsA is important for the production of gonidia, the protein RegA (somatic regenerator A) is a master regulator for the maintenance of somatic cells. It is a putative DNA-binding transcription factor that possibly functions through the repression of (nuclear) genes that are required for chloroplast biogenesis and function (Duncan et al., 2006; Kirk et al., 1999; Meissner et al., 1999). The chloroplast of one gonidium is distributed to all cells during cleavage. In the end, only a small portion (<0.05%) of the original chloroplast is contained in one somatic cell. Many adult somatic cells do not contain any chloroplast DNA and are therefore clearly incapable of chloroplast biogenesis (Coleman and Maguire, 1982). Therefore, RegA is at least one key player in limiting the growth of somatic cells by the

inhibition of chloroplast synthesis. *regA*$^-$ mutants appear to differentiate normally into somatic cells, but later on, these cells dedifferentiate to display reproductive behavior. The cells become fully functional gonidia and produce offspring with a similar phenotype (Huskey and Griffin, 1979; Kirk et al., 1987; Starr, 1970). The *regA* gene belongs to the family of VARL (volvocine algal *regA* like) domain genes and is part of a tandem array together with 4 other VARL genes (Duncan et al., 2007). While *Volvox* possesses at least 14 members of these genes, *Chlamydomonas* has 12 members. An examination of the evolutionary origins of the *regA* gene revealed that it probably stems from a common unicellular ancestor. However, this proto-gene was retained in *Volvox*, but lost in *Chlamydomonas* (Duncan et al., 2007). Although there exist orthologs of *regA* in *Chlamydomonas* (*RLS1*, *rlsD*), none of these genes are conserved in this tandem array structure. Therefore, regA seems to be unique to *Volvox*.

Taken together, there are already several genes known that are crucial for the life cycle of *Volvox*, regulating important processes such as asymmetric cell division, inversion and maintenance of somatic cell fate. The majority of these genes can also be found in *Chlamydomonas*, with at least one notable exception: the regulator of somatic cell fate, RegA. Future studies will certainly uncover more master regulators that will give further hints on the global networks regulating multicellularity and division of labor.

## 5.2 Transcriptome Analysis

The genome of *Volvox carteri* has not long been published (Prochnik et al., 2010). Therefore, it is not surprising that its transcriptome is only poorly characterized. Not only are the transcripts themselves incompletely annotated, there is also very little information available concerning the expression levels of each transcript at a given point of the life cycle.

Since it is our aim to characterize small RNA expression in *Volvox carteri*, it is important to have expression data to match our small RNA samples in order to verify transcripts that encode for protein essential for small RNA biogenesis pathways as well as correlating mRNA expression with small RNA expression.

We sequenced 6 different samples of *Volvox carteri* cells in two biological replicates: *Volvox* cultures were either grown in the vegetative phase, induced for 16 h with the sex inductor or fully differentiated to the female organism. Of each state, the two different cell types (somatic cells and reproductive, i.e. gonidia or egg cells, respectively) were divided by shearing and filtering and measured independently. After cloning, all twelve libraries were sequenced in single-end runs as well as in paired-end runs on a HiSeq2000 sequencer from Illumina. Paired-end runs contain more information concerning longer stretches (both ends of a long insert are sequenced), while single-end runs can generate more reads for lower cost.

The resulting reads were mapped against the *Volvox* genome. For single-end runs, the mapping efficiency was ~60-85% of all reads, whereas for paired-end runs it was 54-84% of all reads (Table

5-1). Since for the paired-end reads both of the reads need to map to the genome in order to gather information, the mapping efficiency drops around 5-10%. In total, we sequenced 696,367,042 (~700 million) reads that could be mapped against the genome (Table 5-1).

**Table 5-1 RNA sequencing of *Volvox carteri* transcripts.** Three states of *Volvox carteri* life cycle were sequenced in two biological replicates. For each state (vegetative grown, induced for 16 h with the sex inductor, fully differentiated female organism) the two cell types (somatic cells and reproductive cells) were separated. For every library, the mappable reads and the overall alignment rate are shown.

**single-end runs (SE)**

| | Replicate I | | | | | |
|---|---|---|---|---|---|---|
| Sample | vegetative (asexual) | | induced | | female | |
| | somatic | reproductive | somatic | reproductive | somatic | reproductive |
| internal alias | V1 | V2 | V3 | V4 | V5 | V6 |
| reads | 42310577 | 35102808 | 35406055 | 492286 | 31825797 | 10784620 |
| overall read alignment rate | 85.5% | 84.4% | 76.4% | 67.5% | 83.4% | 74.8% |

| | Replicate II | | | | | |
|---|---|---|---|---|---|---|
| Sample | vegetative (asexual) | | induced | | female | |
| | somatic | reproductive | somatic | reproductive | somatic | reproductive |
| internal alias | V7 | V8 | V9 | V10 | V11 | V12 |
| reads | 23192676 | 16170619 | 16357814 | 1381019 | 19521246 | 25897216 |
| overall read alignment rate | 70.8% | 81.7% | 57.3% | 64.5% | 58.8% | 85.7% |

| | |
|---|---|
| sum SE reads all libs | 258442733 |

**paired-end runs (PE)**

| | Replicate I | | | | | |
|---|---|---|---|---|---|---|
| Sample | vegetative (asexual) | | induced | | female | |
| | somatic | reproductive | somatic | reproductive | somatic | reproductive |
| internal alias | V1 | V2 | V3 | V4 | V5 | V6 |
| left read | 25867065 | 25564529 | 17852607 | 4551549 | 16237650 | 18483124 |
| right read | 26095153 | 25811073 | 17976999 | 4579082 | 16353446 | 18606384 |
| overall read alignment rate | 84.0% | 82.7% | 75.1% | 63.8% | 81.2% | 72.3% |
| concordant pair alignment rate | 79.0% | 77.9% | 69.1% | 59.3% | 75.7% | 67.2% |
| sum reads per lib | 51962218 | 51375602 | 35829606 | 9130631 | 32591096 | 37089508 |

| | Replicate II | | | | | |
|---|---|---|---|---|---|---|
| Sample | vegetative (asexual) | | induced | | female | |
| | somatic | reproductive | somatic | reproductive | somatic | reproductive |
| internal alias | V7 | V8 | V9 | V10 | V11 | V12 |
| left read | 22964413 | 18886474 | 15026595 | 2913777 | 16448096 | 33281127 |
| right read | 23171148 | 19057621 | 15150829 | 2929293 | 16545092 | 33571183 |
| overall read alignment rate | 67.0% | 76.2% | 54.4% | 59.2% | 55.6% | 80.3% |
| concordant pair alignment rate | 60.6% | 70.4% | 46.5% | 54.8% | 48.1% | 74.9% |
| sum reads per lib | 46135561 | 37944095 | 30177424 | 5843070 | 32993188 | 66852310 |

| | |
|---|---|
| sum PE reads all libs | 437924309 |

| | |
|---|---|
| sum ALL reads | 696367042 |

To generate a transcriptome, we used the well-established program Cufflinks (Trapnell et al., 2012). This tool first maps all reads to the genome and then assembles the transcripts. Also, differential expression between two samples can be calculated.

To ensure that the preparations of the different cell types were clean and to verify the expression data we obtained, the transcript level of different mRNAs was measured. The genes *ssgA* and *regA* are only expressed in somatic cells (Ertl et al., 1989; Kirk et al., 1999; Tam and Kirk, 1991). In agreement

with that, we could measure their transcripts in all three somatic cell samples, but not in reproductive cells (Figure 5-5). Our gene expression data shows expression of *regA* mRNA in somatic cells, but not in reproductive cells (transcript 8391.5). Therefore, no fold change can be specified. For *ssgA*, two transcripts cover the annotated mRNA (transcripts 7237 and 7238). Both are significantly stronger expressed in each somatic cell sample compared to the reproductive cells with fold changes lying between 70 and 770.

A marker gene for reproductive cells, however, is not well established. We therefore searched for highly up regulated transcripts in our transcriptome data. Transcript 22037 is strongly up regulated in vegetative gonidia and induced gonidia (48 and 128 fold, respectively), when compared to their corresponding somatic cells. Although the difference in transcript level measured by qRT-PCR is not as high as measured by RNA-Seq, a strong increase is nevertheless detectable (Figure 5-5).

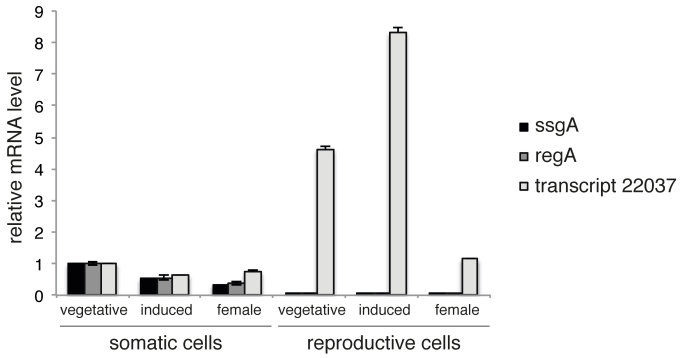



**Figure 5-5 Measurement of mRNA levels of *Volvox carteri* marker genes.** The transcript levels of marker genes for somatic cells (*ssgA* and *regA*) and for reproductive cells (transcript 22037) were measured by qRT-PCR. Values were normalized to ActinA and vegetative somatic cells. Error bars represent the standard deviation for three technical replicates.

Taken together, the RNA-Seq data is in very good agreement with published gene expression results. Additionally, a transcript that was chosen because of its high expression in gonidia in our RNA-Seq data was verified by qRT-PCR measurements.

## 5.3 Analysis of proteins involved in small RNA biogenesis and function

To assess small RNA origins and functions, we set out to identify proteins involved in small RNA biogenesis or small RNA function. As earlier introduced, plants require a Dicer-like protein to excise both the precursor and subsequently the mature miRNA duplex. Plant miRNAs are methylated at their 3' end to protect them from degradation. This task is carried out by a methyltransferase, HEN1 (Park et al., 2002). For executing the miRNA-guided function, at least one Argonaute protein is necessary (Meister, 2013). For the amplification of siRNAs, which is often observed in plants, the additional

activity of an RNA-dependent RNA polymerase (RdRP) is necessary in order to synthesize the second strand (Cogoni and Macino, 1999; Dalmay et al., 2000; Smardon et al., 2000).

All of these proteins are characterized by at least one conserved domain and thus can be searched using specific features. For all proteins, we employed several approaches. First, the full-length protein sequences from *Arabidopsis thaliana* were used as template input for an algorithm that can predict protein domains (HHblits (Remmert et al., 2012), see also methods section 8.4.3.3). Next, we searched with protein sequences of conserved domains. Thirdly, the genomic area around potential hits was analyzed for the occurrence of additional domains of the respective protein. We used two datasets for searching RNAi proteins: translated transcripts annotated by the JGI (Joint Genome Institute, U.S. Department of Energy, University of California; (Prochnik et al., 2010)) and translated transcripts annotated using RNA Seq data that we generated in our lab (see also 5.2).

The small RNA factors RdRP, Dicer, HEN1 and Argonaute have all been analyzed and are presented in the following paragraphs.

*RNA-dependent RNA polymerase.* We could identify one transcript (Vocar20000861m, (Prochnik et al., 2010)) that harbors an RdRP domain (Pfam: PF05183). The transcript encodes for a protein with 1679 amino acid residues and a molecular weight of 174.5 kDa. The RdRP domain spans from residue 431 to residue 1055. The genomic locus and the domain structure are shown in Figure 5-6.

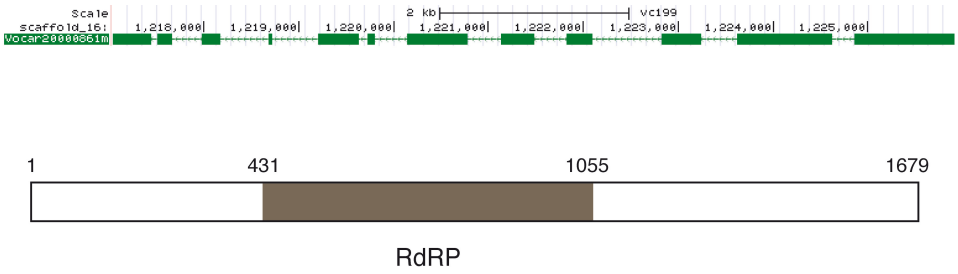



**Figure 5-6 Identification and domain structure of an RdRP from *Volvox carteri*.** The upper panel shows the CDS structure. The lower panel depicts the protein in a domain view. The transcript was identified by HHBlits using RDR6 (*Arabidospis thaliana*) as template input (E-value 1.9E-97).

*HEN1 homologue.* Proteins of the HEN1 family are characterized by their similarity to the *Arabidopsis thaliana* protein HEN1, i.e. containing one or more double-stranded RNA binding motifs (dsrm, Pfam: PF00035) and a methyltransferase domain. The HEN1 homologue of *Volvox carteri* is encoded in the transcript Vocar20003300m (HHBlits with *A.thaliana*, E-value 7.8E-73). It is comprised of 1430 residues and a molecular weight of 143.1 kDa. It contains a dsrm and a region highly similar to the methyltransferase domain of HEN1 (amino acids 955 to 1129), which was identified by Panther (Mi et al., 2013) (E-value 7.9E-54) (Figure 5-7).

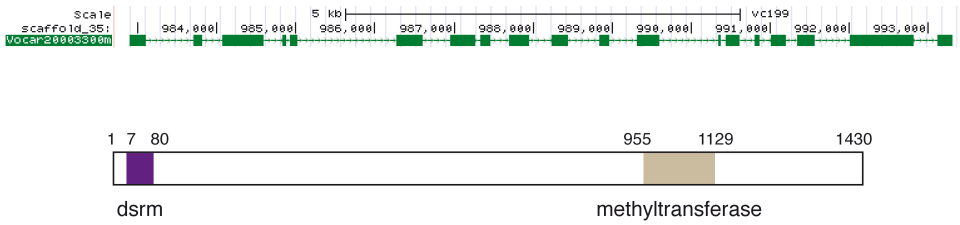



**Figure 5-7 Identification and domain structure of a HEN1 homologue from *Volvox carteri*.** The upper panel shows the exon structure of the coding sequence (CDS) of the HEN1 homologue Vocar20003300m. The lower panel depicts the domain structure of Vocar20003300m, showing a double-stranded RNA binding motif (amino acids 7-80) and a methyltransferase domain (amino acids 955-1129) which could be annotated by HHBlits.

*Dicer-like protein.* Dicer and Dicer-like proteins in higher plants and mammals are large proteins with several domains. We found transcripts encoding all necessary domains in very close proximity to each other, namely a DEAD/DEAH box helicase (Pfam: PF00270), a Helicase C (Pfam: PF00271), a Dicer dimerization domain (Pfam: PF03368) and two RNase III domains (Pfam: PF00636). We could not find a transcript encoding for double-stranded RNA binding activity in close proximity. Although there is not one fully annotated transcript containing the Dicer-like protein but several smaller transcripts, this kind of domain organization is clearly indicative of a Dicer-like gene (Figure 5-8). A rough overlay of the transcripts would yield a coding sequence (CDS) that encodes for a protein with about 2300 amino acids resembling a molecular weight of ~250 kDa (Figure 5-8, lower panel).

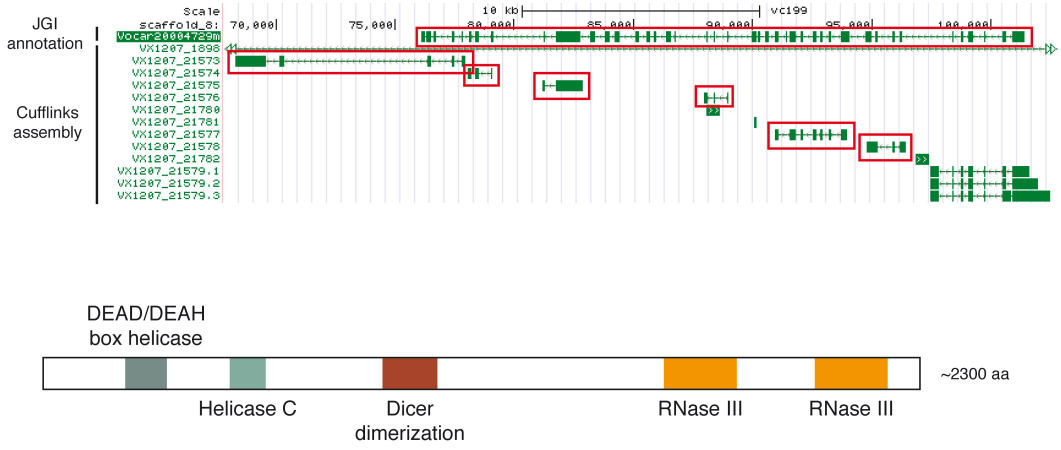



**Figure 5-8 Identification and domain structure of a putative Dicer-like protein from *Volvox carteri*.** Transcripts that are probably contained in the Dicer-like protein are highlight with red boxes (upper panel). JGI annotation, transcripts annotated by the JGI. Cufflinks assembly, transcripts from the RNA Seq data generated by our group, assembled by Cufflinks (Trapnell et al., 2012). Overlay of the transcripts would yield a CDS that encodes for a protein with about 2300 amino acid residues (lower panel).

*Argonaute proteins.* Argonaute proteins are a conserved family of proteins and are characterized by a PAZ and a Piwi domain. We found two loci in the *Volvox* genome matching these criteria. The first Argonaute had been annotated before as an Argonaute-like protein (XM_002952848, “AGO3”) and was verified by us through manual curation. For this, we compared the annotated transcript with the transcript annotated by the JGI, Vocar20011215m (Prochnik et al., 2010), and the RNA Seq data generated by ourselves. The resulting CDS is located on scaffold 52 on the plus strand, spanning

nucleotides 424991 to 435955. The protein contains a DUF1785, a PAZ and a Piwi domain (Figure 5-9, A), is comprised of 1056 amino acids and has a molecular weight of 111 kDa (for the protein sequence see Supplement 1, Appendix).

A second Argonaute-like protein fragment had also been annotated before (XM_002952044, "AGO2"; Vocar20001549m). These annotated sequences, however, are fragmented, containing only a small piece of the PAZ domain and the Piwi domain. Our transcriptome assembly clearly shows that an overlapping transcript exists that encodes for the N-terminal part of the protein (Figure 5-9, B). Consequently, this Argonaute protein has a very similar domain structure compared to the first Argonaute described above, containing a DUF1785, a PAZ and a Piwi domain. It contains ~1000 amino acids and has a molecular weight of 110 kDa.

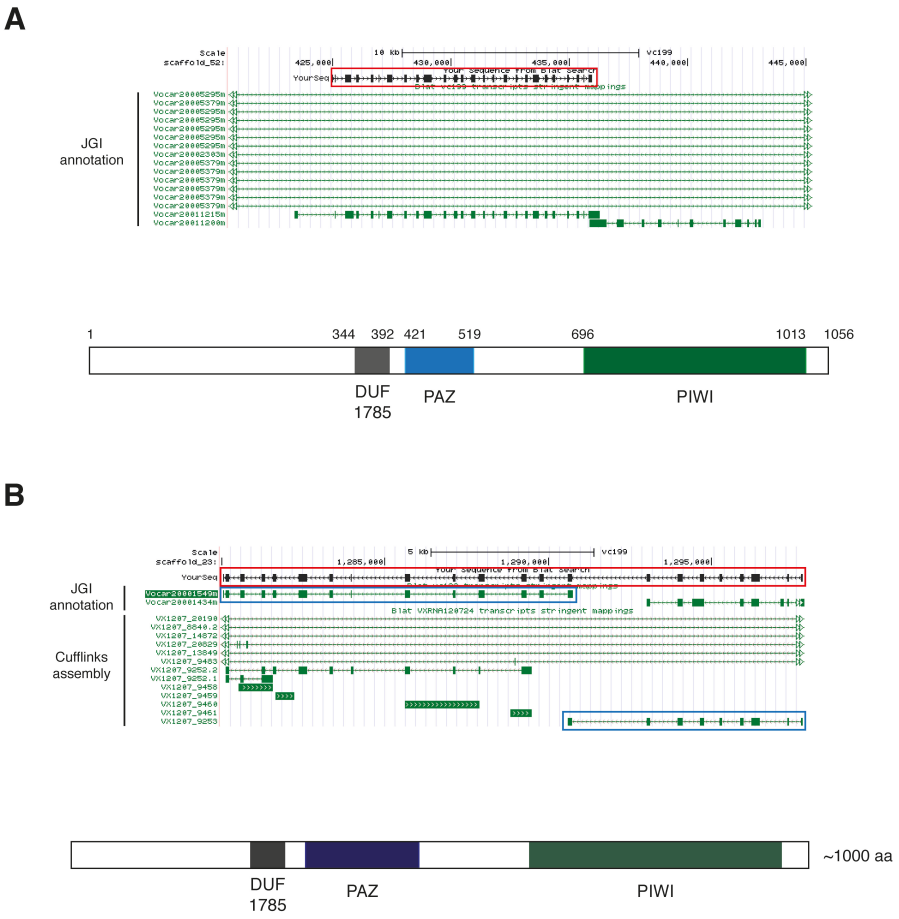



**Figure 5-9 Transcript identification and domain organization of *Volvox carteri* Argonautes.** (A) Curated sequence of an Argonaute protein (AGO3) of *Volvox carteri*. Transcript sequence had been annotated before and was adjusted using the RNA Seq data generated by our group. The final CDS is shown in the Genome Browser and highlighted with a red box (upper panel). Domain structure of the Argonaute protein (lower panel): Pfam domain search resulted in three annotated domains, the DUF1785 (E-value 1.7E-11), the PAZ (E-value 1.1E-15) and the Piwi domain (E-value 2.5E-89). (B) The second Argonaute (AGO2) that was identified spreads currently over two transcripts (see blue boxes, upper panel). The overlapped sequence is shown in a red box. The domain structure is similar to the Argonaute protein described in (A), namely a DUF1785 (E-value 5.7E-07), the PAZ (E-value 3.7E-13) and the Piwi domain (E-value 1.2E-70).

For convenience, the initial nomenclature of the Argonaute proteins from *Volvox* in the NCBI database will be retained, i.e. the first Argonaute protein described here will be termed AGO3 and the second one AGO2 from hereon.

Some but not all Argonaute proteins contain intrinsic nuclease activity. This catalytic activity is comprised in the Piwi domain and requires the presence of four catalytic residues, which canonically are DEDH. For some Argonautes, e.g. *Arabidopsis thaliana* AGO2, endonucleolytic activity could also be assigned to a DEDD motif (Carbonell et al., 2012). Knowledge about the catalytic residues of an Argonaute protein, therefore, can already point to a possible regulatory mechanism. To this extent, an alignment of the *Volvox*, *Chlamydomonas* and *Arabidopsis* Argonaute Piwi domains was constructed (Figure 5-10). The AGO3 contains a DEDD motif, while the AGO2 lacks the last residue, containing only DED. Of the *Chlamydomonas* Argonautes, only AGO3 contains a full motif, DEDD (Figure 5-10). With this finding, the prerequisite of endonucleolytic activity has been met. Whether or not AGO3 is indeed capable of cleaving its targets has to be validated experimentally.

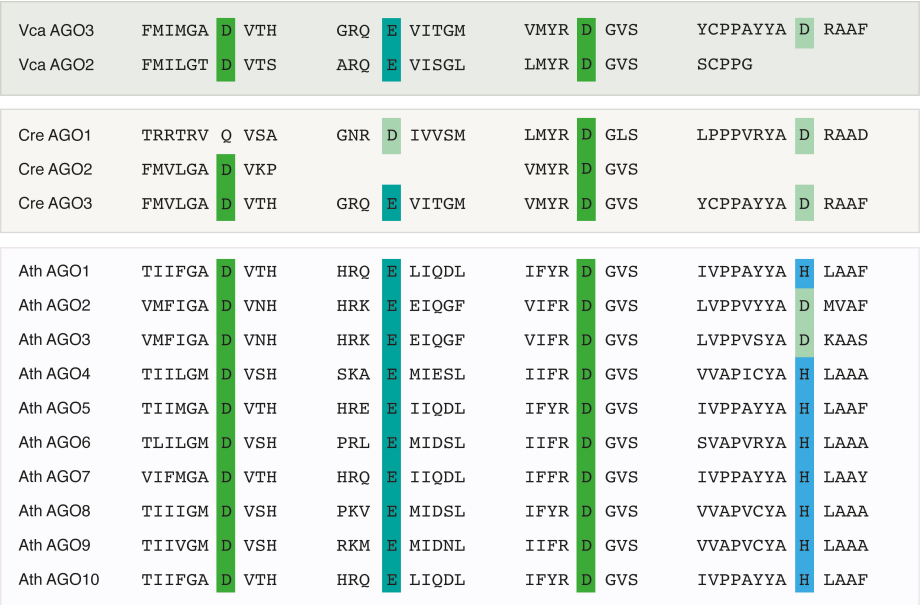



**Figure 5-10 Alignment of the catalytic amino acid residues of the Piwi domains of different organisms.** The Argonaute proteins of *Volvox carteri*, *Chlamydomonas reinhardtii* and *Arabidopsis thaliana* were aligned using Clustal Omega (Sievers and Higgins, 2014; Sievers et al., 2011). The regions around each catalytic residue are shown, the catalytic residues themselves are highlighted in color. The canonical motif for active endonuclease activity is DEDH, other motifs such as DEDD can still sustain cleavage activity. Abbreviations: Vca, *Volvox carteri*; Cre, *Chlamydomonas reinhardtii*; Ath, *Arabidopsis thaliana*. Following sequences were used for the alignment: AGO3 (see Supplement 1, Appendix); Vocar20001549m (containing the Piwi domain of AGO2); XP_001694840.1 (Cre AGO1); XP_001698670.1 (Cre AGO2); XP_001698906.1 (Cre AGO3); NP_175274.1 (Ath AGO1); NP_174413.2 (Ath AGO2); NP_174414.1 (Ath AGO3); NP_565633.1 (Ath AGO4); NP_850110.1 (Ath AGO5); NP_180853.2 (Ath AGO6); NP_177103.1 (Ath AGO7); NP_197602.2 (Ath AGO8); NP_197613.2 (Ath AGO9); NP_001190464.1 (Ath AGO10).

In conclusion, we show that critical proteins for the biogenesis of miRNAs as well as secondary siRNAs are present in the genome of *Volvox carteri*. Furthermore, AGO3 contains the catalytic tetrad DEDD that is a prerequisite for endonucleolytic activity. Future studies will reveal further characteristics of these proteins.

## 5.4 Cloning and expression of AGO3

The characterization of small RNAs in an organism with little information on annotation of RNAs as well as biogenesis pathways needs to be undertaken with great care. Analyzing the total pool of small RNAs can be helpful, but also makes it difficult to differentiate between functional small RNAs and degradation products specific for the analyzed sample. Additionally, some plant Argonaute proteins are highly specialized and bind to only one or a subset of small RNA types, e.g. *Arabidopsis* AGO1, which binds almost exclusively to miRNAs. The characterization of subsets of small RNAs might therefore also lead to a more detailed characterization of the protein binding partner. Thus, the identification of at least one complete Argonaute protein raised interesting possibilities. The isolation of AGO3 cDNA and a characterization of the encoded protein with its associated small RNAs would not only indicate the existence of a functional RNAi machinery but would also provide insight into the function of the AGO3-associated small RNAs themselves.

First, we measured the endogenous mRNA level of AGO3 in different stages of the *Volvox* life cycle (Figure 5-11, A). While the level in somatic cells in all three stages is nearly identical, in reproductive cells, however, there seems to be a higher level in “induced gonidia”, which are gonidia after 16 h of induction with the sex pheromone. Since the overall expression doesn’t change more than one order of magnitude, further experiments are needed to analyzed the impact of this potential AGO3 regulation.

After establishing that this Argonaute is indeed expressed in *Volvox*, we set out to isolate its cDNA. The cDNA was amplified in pieces and then assembled to yield the full sequence. Only the very N-terminal part could not be isolated by PCR and was added by using synthesized oligos.

Since no antibody was available for performing cell biology experiments, we made use of an established vector system for the transformation of *Volvox* with foreign DNA (Jakobiak et al., 2004), containing an integration cassette with promoter elements and a polyadenylation signal from *Volvox* genes. For our purposes, we slightly changed the vector for expressing the transgene to now contain a 6xmyc-tag positioned N-terminal of a new multiple cloning site (MCS).

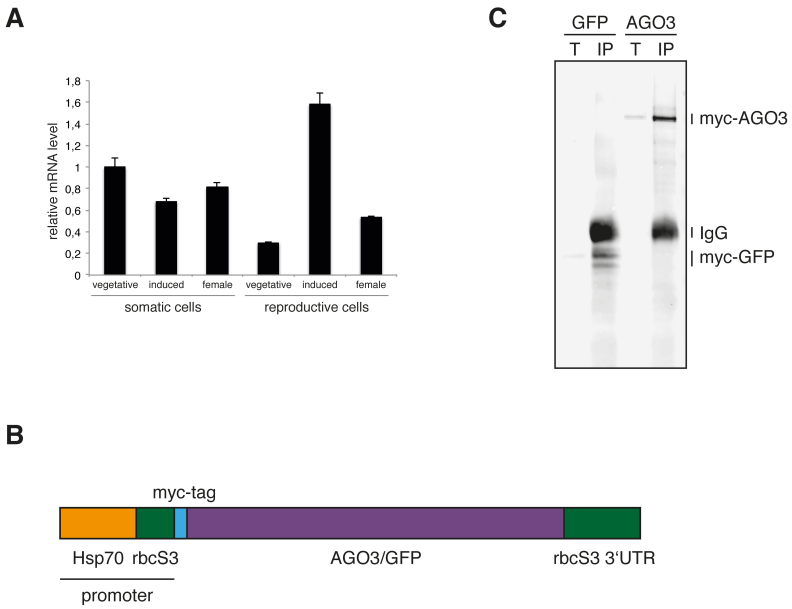



**Figure 5-11 Endogenous and exogenous expression of AGO3.** (A) Endogenous levels of AGO3 mRNA were measured in three different stages of *Volvox* life cycle using qRT-PCR. Error bars represent the standard deviation of three technical replicates. (B) Expression cassette used for the transformation of *Volvox*. As a promoter, the strong Hsp70 promoter together with promoter region of the small subunit of Ribulose-1,5-bisphosphate-carboxylase, RuBisCo (*rbcS3*) was used. The 3'UTR of *rbcS3* functions as polyadenylation signal. (C) Western blot of exogenously expressed GFP and AGO3. *Volvox* grown in vegetative phase were transformed, cultured with selection marker and finally lysed to perform immunoprecipitations. For detection, anti-myc antibody was used.

The vector with this expression cassette was now applied to generate *Volvox* clones expressing either AGO3 or a codon-optimized GFP (Fuhrmann et al., 1999) (Figure 5-11, B and C). After transformation and several days of selection, asexual grown *Volvox* cultures were harvested and tested for the genomic insertion of the respective genes by PCR (data not shown). Since both cultures were positive, the samples were used for the isolation of expressed proteins. Visualization by Western blotting (Figure 5-11, C) shows for both proteins expression of the respective construct and bands of the expected size. Using this system we can therefore successfully over express AGO3 in *Volvox* cultures. However, after several generations the protein expression was turned off. This phenomenon of transgene silencing is well known, although the gene is inserted into the genomic DNA (Fuhrmann et al., 1999; Neupert et al., 2009). It will therefore be difficult to stably express AGO3 and later test developmental stages of *Volvox*. Nevertheless, it was possible to transiently express AGO3 and isolate enough material for subsequent cloning and sequencing of AGO3-associated small RNAs.

## 5.5 Characterization of small RNAs associated with AGO3

The non-stable culture of AGO3-expressing *Volvox* was used to perform an anti-AGO3 IP and isolate associated small RNAs (see Figure 5-11, C). The RNA was cloned and sequenced using Illumina sequencing technology. The following paragraphs will first describe the general content of the library and subsequently specify the results and analysis for the relevant small RNA categories.

### 5.5.1 General features of small RNAs associated with AGO3

The AGO3 library contains 2,682,842 reads in total. Of these, 2,392,956 reads (89.2%) could be mapped to the *Volvox* genome. First, the global length distribution and nucleotide preference of AGO3 was analyzed. Most AGO3-associated small RNAs are 21 or 22 nt in length (Figure 5-12, A). Although there are small RNAs that are 24 nt long, they do not represent a major part of the total small RNA content. The nucleotide preference of AGO3 shows a clear bias towards a 5' U (T) (Figure 5-12, B). These findings are remarkable since it shows that we have very likely isolated a fully functional Argonaute with clear binding properties and bound small RNA subset.

Subsequently, all reads were examined according to their genomic origin. The following categories were investigated: miRNAs, tRNAs, rRNAs, repeat-associated, mRNAs and phased RNA. Some categories were analyzed using (incompletely) annotated sequences (mRNA, tRNA, rRNA, repeats), miRNAs and phased RNAs, however, were categorized by applying various bioinformatics tools (see 5.5.2 and 5.5.3).

The majority of reads cannot be assigned to any category so far. As RNA transcripts in general are only poorly annotated and characterized in *Volvox*, this might not be surprising. 38% of the reads could be annotated as mRNAs, miRNAs, rRNAs, phased small RNAs or repeat-associated small RNAs (Figure 5-12, C).

Taken together, the very strong preference for U at the 5' end bound by AGO3 implies that functional RNP complexes were isolated. Although we could not fully annotate the small RNA content yet, this probably is due to lacking annotation paired with little knowledge about small RNA pathways.

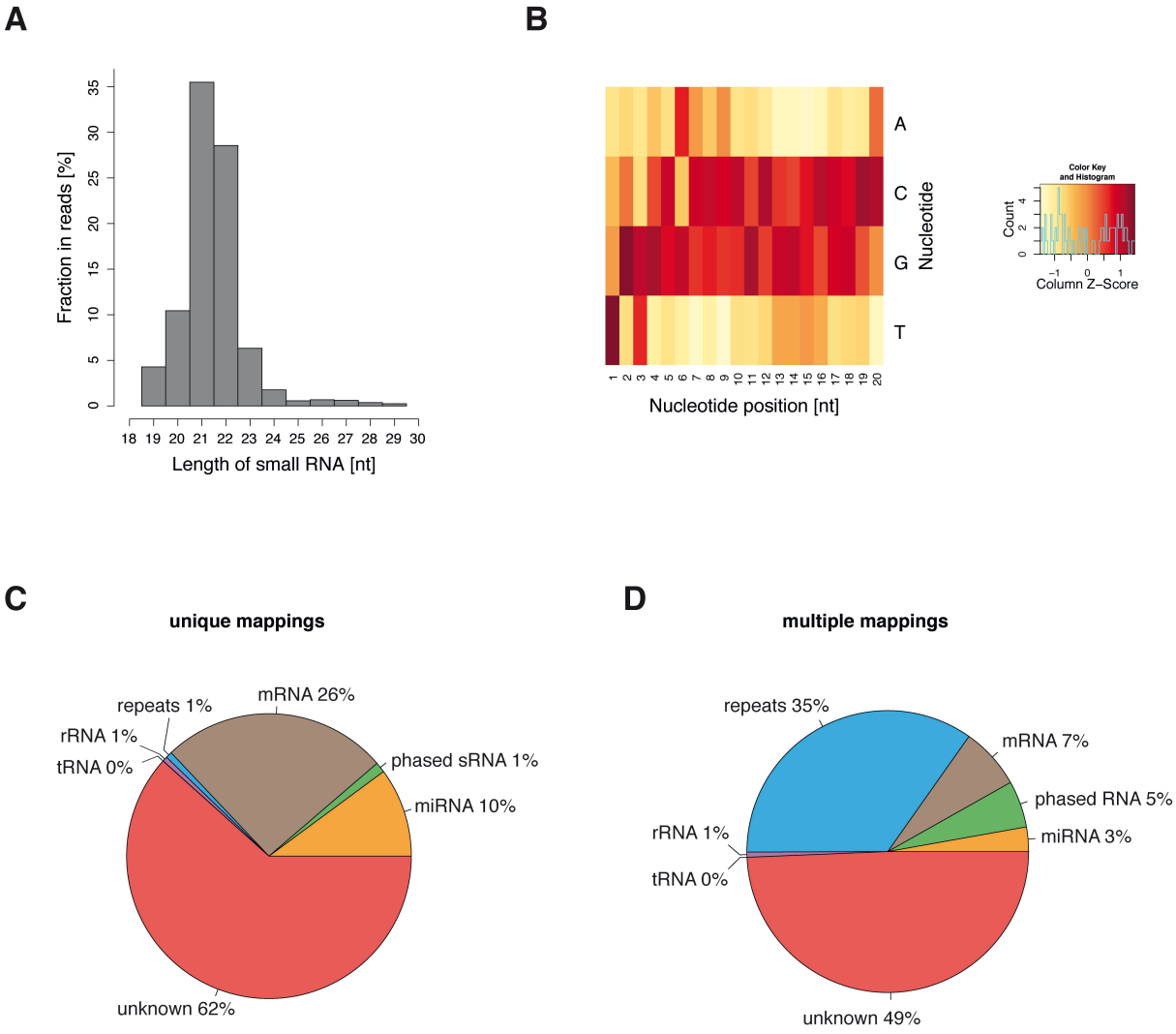



**Figure 5-12 Library characteristics and composition of AGO3-associated reads.** (A) Global length profile of AGO3-associated small RNAs. (B) Nucleotide preference of AGO3-associated small RNAs. All reads were used to generate this plot of the first 20 bases. A key for the colors is provided to the right of the heatmap (higher occurrence = darker color). (C) Respective fraction of reads mapping uniquely to mRNAs, miRNAs, phased RNAs, rRNAs, tRNAs or repeat regions. Reads that did not fall into either category were assigned as unknown. (D) Same categories as in (C) but allowing for multiple mapping of reads, i.e. each mapping per read is counted as one read.

### 5.5.2 Analysis and expression levels of miRNAs

#### *5.5.2.1 Assembly of a de novo miRNA identification pipeline*

MiRNA precursors are a heterogeneous population in plants in terms of structure and length (see 1.1.2). In contrast to animals, miRNA prediction is rather difficult and evolutionary conservation is an important determinant for miRNA identification. This approach, however, will be unsuccessful in species that do not contain conserved or homologous miRNAs at all. Two papers have already shown that *Chlamydomonas reinhardtii* does not contain conserved miRNAs with land plants (Molnar et al.,

2007; Zhao et al., 2007). Since *Chlamydomonas* and *Volvox* share a common ancestor, such observations are probably true for *Volvox* as well.

In order to characterize miRNAs independent of conservation, we set up a miRNA identification pipeline in collaboration with Dr. Maurits Evers (Evers et al., manuscript in preparation). The next paragraphs will shortly summarize the different steps and criteria of our pipeline. Since Argonaute proteins associate with only distinct small RNA populations, e.g. miRNAs, and do not incorporate cellular degradation products, we chose the small RNA sequencing data set of the AGO3 immunoprecipitation as input data to identify miRNAs.

First, all reads (after quality filtering, adapter removal etc.) are mapped to the *Volvox* genome. Next, regions of interest (contigs) are selected on clustering of reads. If these contigs contain at least 100 reads, they are extended by 200 nt on each side (flanks) and the full sequence (in total <600 nt) is extracted to assess structural parameters.

In order to identify a tentative miRNA precursor, all potential sequences within a contig are tested for folding, i.e. the sequence is varied in the extended flanks (sliding window) and every time, a structure is generated using the minimum free energy algorithm of Zuker et al. (Zuker and Stiegler, 1981) as implemented in RNAfold (Schuster et al., 1994). Since there is a huge amount of structures generated, an algorithm is applied to evaluate structures for having the overall lowest free energy after folding. For this, we used simulated annealing as optimization method as it can be applied when searching for a good approximation to the global optimum in a reasonable time. The following criteria need to be fulfilled for the putative precursor to pass the structure and folding assessment: The structure of the precursor is not allowed to contain more than 3 hairpins, one segment of double-stranded stretch (allowing for up to two mismatches within) should be at least 20 nt long (otherwise it cannot harbor a miRNA) and finally the minimum free energy per nucleotide should be lower than -0.4 kcal/mol/nt. If these criteria are met, a statistical significance test is performed to determine the probability for the structure of the putative precursor being the result of folding a random sequence (null hypothesis). If the test results in a p-value below 0.01, the structure is considered to be a potential miRNA precursor.

In the last step, we investigate whether expression from potential precursors is consistent with precursor processing into at least one mature miRNA. To this extent small RNA sequencing data are mapped to the potential precursors in order to identify the location and optimal sequence of the small RNA on the precursor. The coverage of reads per nucleotide is analyzed, and the sequence lying within the main expression block is identified as a valid small RNA sequence. The putative mature miRNA sequence needs to fulfill the following criteria: a) a precise processing pattern should be observed, meaning a sharp edge of expression at the 5' end of the small RNA, b) the sequence should be equal to or more than 18 nt and no longer than 30 nt in length, c) the fraction of paired nucleotides of the small RNA should exceed 67%, meaning e.g. more than 14 base pairs in a 21 nt long sequence need to be paired, d) the small RNA sequence must not fold back on itself, e) the number of adjacent mismatches is lower than four mismatches and f) the number of adjacent mismatches at the 5' and 3' ends of the small RNA should not exceed two nucleotides.

For all passed small RNA sequences, an output file is generated that contains the most important characteristics, a coverage plot for the putative miRNA precursor and a structure plot with highlighted small RNA sequences.

With this method, we have generated a small RNA read-based miRNA "detection" tool optimized for plant miRNAs.

### *5.5.2.2 Features of Volvox miRNAs*

Applying the pipeline described above (see 5.5.2.1), we identified 288 distinct miRNA sequences (including expressed star sequences), a full list can be found in the appendix (see Supplement 3, Table, Appendix). For 103 mature sequences we could assign a corresponding star strand. Since it is known that miRNAs can show a certain bias for their 5' nucleotide due to a nucleotide-specific binding pocket of the associated Argonaute protein (see 1.3), the nucleotide composition of all AGO3-bound miRNAs was analyzed (Figure 5-13). A U at the 5' end is clearly preferred, in fact, all predicted miRNA sequences start with a U. The remainder of the sequence seems to represent the genome composition rather than any other preference or motif.

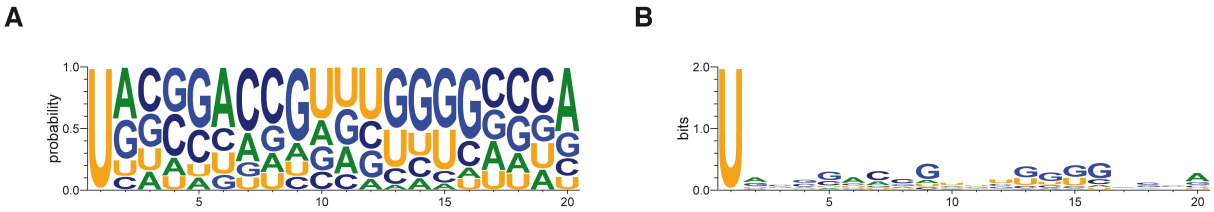



**Figure 5-13 Nucleotide preference and nucleotide distribution of novel miRNAs.** (A) The first 20 bases of all newly annotated miRNAs are depicted by their probability. The height of each nucleotide represents the fraction of that nucleotide on the total. (B) As in (A), but here, the conservation at each position is presented. Logos were created using Weblogo 3.3 (Crooks et al., 2004).

The general features of miRNAs have been extensively studied in many mammalian species and also in some land plants. There is only little knowledge on miRNAs in the simple organisms of green algae. Therefore, the length of the mature miRNAs and their precursors were compared to important organisms or groups. Additionally, the amount of pairing of the miRNA precursors was investigated.

The profile of *Volvox* mature miRNAs shows a clear peak at a length of 21 nt and 22 nt (Figure 5-14, top left). A more pronounced length of 21 nt is seen for *Chlamydomonas* and even more so, *Arabidopsis* (precursor and mature miRNA sequences for *Chlamydomonas* and *Arabidopsis* were obtained from miRBase, see also 8.4.3.3). A distinct class of miRNAs with the length of 24 nt ("long" miRNAs) has been described in *Arabidopsis* and rice (Axtell, 2013), but these long miRNAs do not seem to be a recognizable fraction in *Volvox* in this particular sample as well as in *Chlamydomonas*. Overall, the length profile of mature miRNAs of *Volvox* is more similar to plants than that of mammals, which allow for more length variation (Figure 5-14).

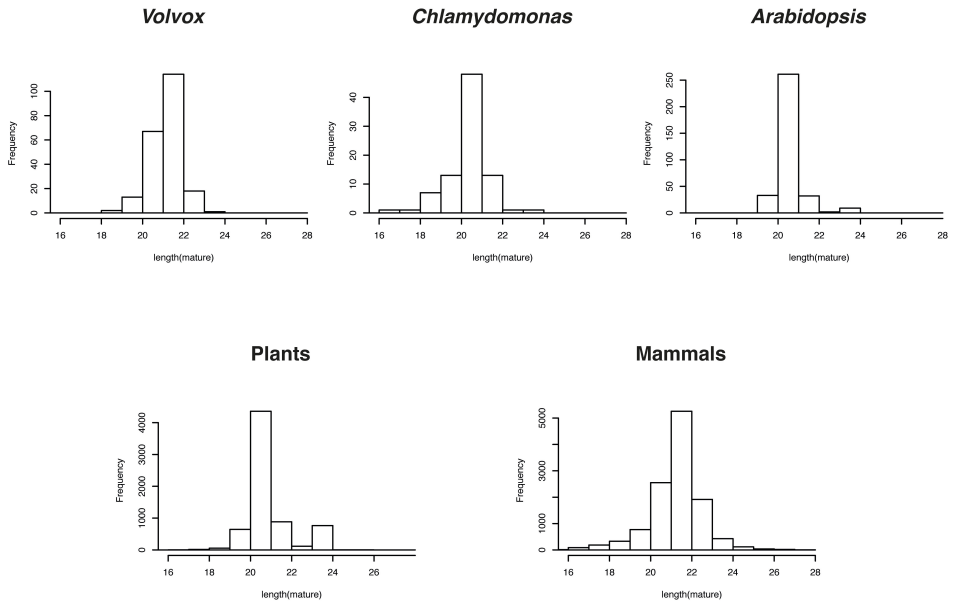



**Figure 5-14 Length of mature miRNAs.** The length of mature miRNAs was plotted against their frequency for *Volvox carteri* (top left), *Chlamydomonas reinhardtii* (top middle), *Arabidopsis thaliana* (top right), all plants (bottom left) and all mammals (bottom right). *Volvox* sequences resulted from our identification pipeline, all other sequences were downloaded from miRBase (release 20, June 2013).

Plants in general tend to have longer and more variable miRNA precursors than mammals (Figure 5-15). For both supergroups, most of the precursors are about 100 nt long. However, while mammalian hairpins are not longer than 150 nt, plant precursors are often around 200 nt or longer (Figure 5-15, bottom panels). The same holds true for both *Volvox* and *Chlamydomonas*, which have precursors from about 100 nt to over 400 nt. *Chlamydomonas* and *Arabidopsis* have several very long precursors (> 800 nt). Since our pipeline restricts precursors to maximum 600 nt length (see 5.5.2.1), potential longer precursors will not be detected and will have to be characterized by other means in the future.

One interesting aspect of miRNA precursor features is the amount of base-pairing within a precursor (Figure 5-16). Evolutionary models of miRNA gene origin are gene duplications or random hairpins present in the genome, for example (see 1.2). Therefore, if miRNA precursors show a generally higher degree of base-pairing, this might be an indicator for the existence of relatively young miRNA genes. Mammals not only tend to have shorter precursors, but also seem to contain more bulges compared to plants (Figure 5-16, bottom panels). This effect is even more pronounced in *Volvox* and *Chlamydomonas*, with *Volvox* precursors being over 80% base-paired on average and *Chlamydomonas* containing many miRNAs with around or over 90% of base-pairing. *Arabidopsis* is a higher land plant with quite complex small RNA pathways. It also shares some ancient miRNA genes with mosses, for example. Most miRNA precursors of *Arabidopsis* display base-pairing values between 65-80%.

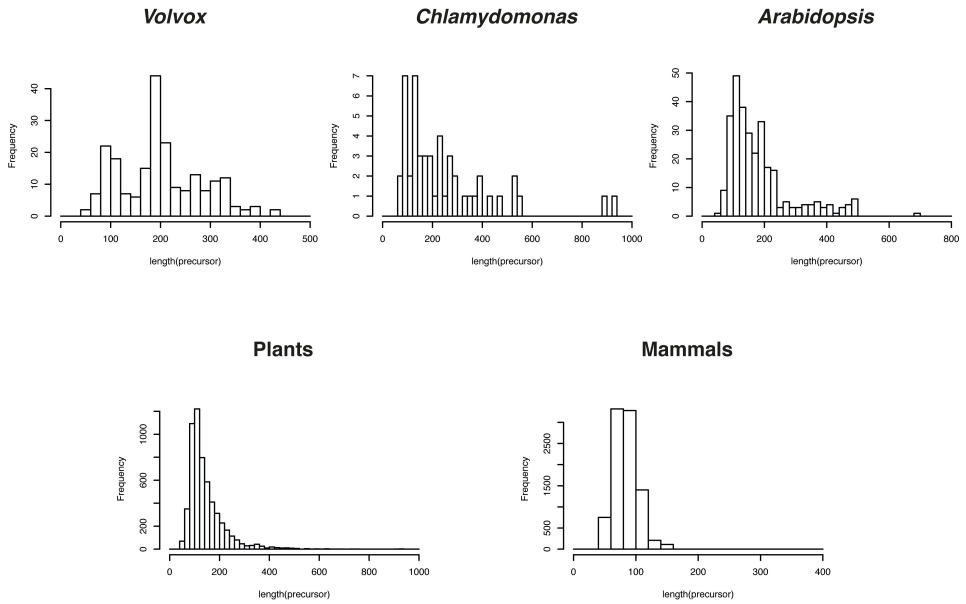



**Figure 5-15 Length of miRNA precursors.** The length of miRNA precursors was plotted against their frequency for *Volvox carteri* (top left), *Chlamydomonas reinhardtii* (top middle), *Arabidopsis thaliana* (top right), all plants (bottom left) and all mammals (bottom right). *Volvox* sequences resulted from our identification pipeline, all other sequences were downloaded from miRBase (release 20, June 2013).

In conclusion, mature miRNAs of *Volvox* and their precursors are more alike to annotated miRNAs of plants than to those of mammals. However, green algae seem to represent a distinct subset of plants, since their precursor properties are different to *Arabidopsis thaliana* and plants in general.

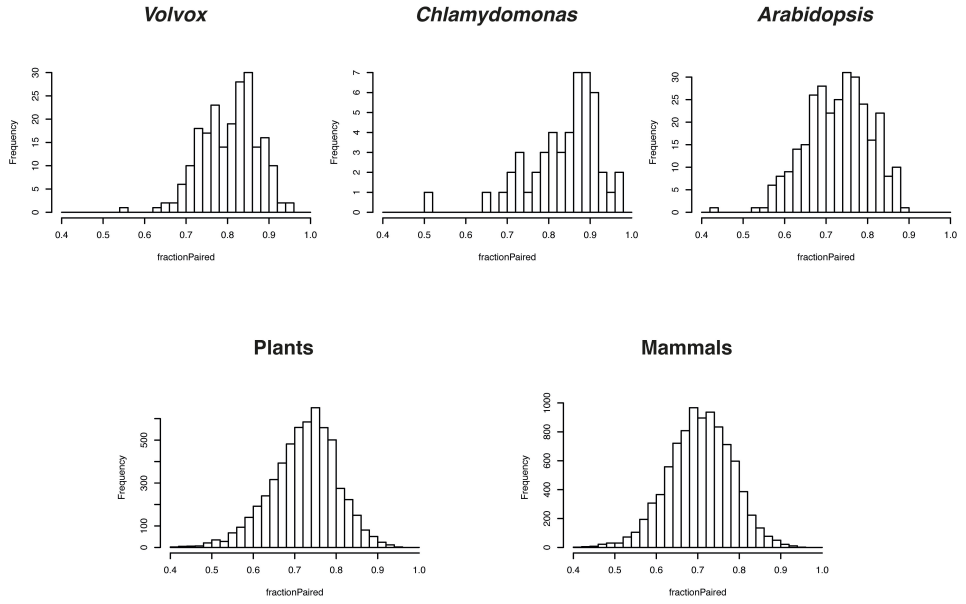



**Figure 5-16 Base-pairing of miRNA precursors.** The fraction of base-pairing of miRNA precursors (1.0 being 100%) was plotted against their frequency for *Volvox carteri* (top left), *Chlamydomonas reinhardtii* (top middle), *Arabidopsis thaliana* (top right), all plants (bottom left) and all mammals (bottom right). *Volvox* sequences resulted from our identification pipeline, all other sequences were downloaded from miRBase (release 20, June 2013).

### *5.5.2.3 miRNAs are 2'O-methylated*

MiRNAs in plants are usually methylated to protect them from degradation (see 1.1.2.1). As shown above, the *Volvox* genome encodes for a putative enzyme that catalyzes this modification, i.e. a HEN1 homologue. To test whether the miRNAs of *Volvox* carry such a modification, a β-elimination reaction was carried out. In this reaction, the RNA is first subjected to a treatment with an oxidizing agent (periodate) and subsequently investigated on a Northern blot. RNA with an unprotected 3' end, namely carrying a 3' OH, will migrate faster in the gel, because of the elimination and subsequent loss of the 3' terminal nucleotide. This shift can be visualized on a Northern blot as seen for the control RNA oligonucleotide that was spiked into the reaction (Figure 5-17, top panel). Two different novel miRNAs of *Volvox* were tested (Cluster_430 and Cluster_23) and none of them reveals a shift, meaning all of them bear protected and therefore probably methylated 3' ends (Figure 5-17).

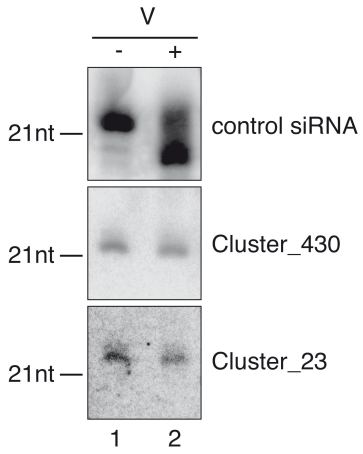



**Figure 5-17 miRNAs of *Volvox carteri* are modified at their 3' end.** Total RNA of *Volvox carteri* in vegetative growth was used to probe the 3' end of different miRNAs for the presence of a modification. After the chemical reaction (β-elimination), the presence of a free hydroxyl group shows in a faster mobility on a Northern blot as seen for the control, a spiked-in synthetic siRNA oligonucleotide (lane 2). The two tested *Volvox* miRNAs do not shift to a smaller size. V, total RNA of vegetative *Volvox samples*.

### *5.5.2.4 miRNAs expression and origin*

The miRNAs that were identified by cloning and sequencing were associated to an overexpressed AGO3 in vegetative colonies. Since miRNAs in plants play important roles particularly during developmental processes, our hypothesis was that at least some miRNAs might be differentially expressed. Therefore, the expression of several endogenous miRNAs was analyzed by Northern blotting (Figure 5-18). Indeed, some of the miRNAs show a differential expression between cell types (e.g. Cluster_5285 and Cluster_2909) or developmental states (Cluster_884) (Figure 5-18).

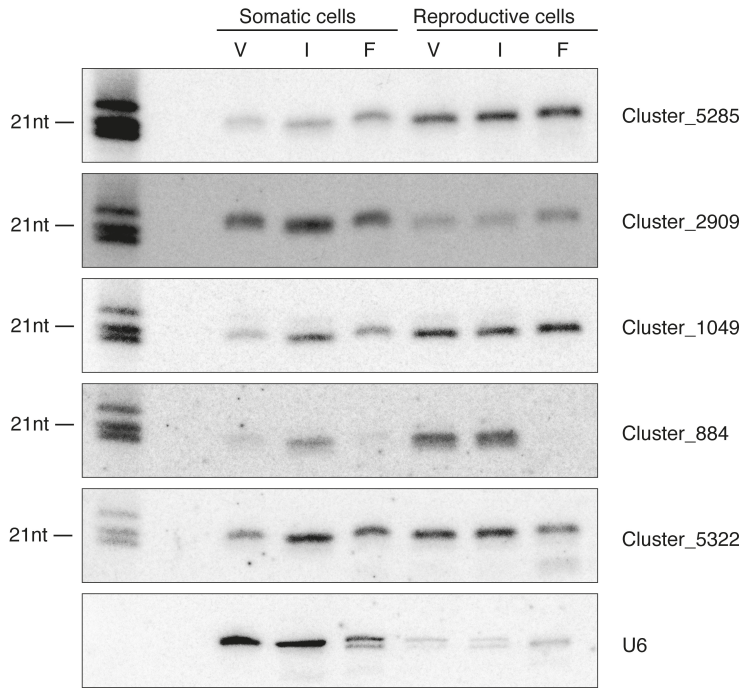



**Figure 5-18 MiRNA expression in *Volvox carteri*.** Several different miRNAs were analyzed for their expression by Northern blotting. The first and last lane shows marker RNA oligonucleotides of the length 19 nt, 21 nt and 24 nt. V, vegetative; I, induced; F, female. U6 was blotted as a loading control (bottom panel).

Next, we investigated the relationship of all identified miRNAs with each other. The nomenclature for plant miRNA families requires that only identical or nearly identical sequences are combined to form a miRNA family (Meyers et al., 2008). To assess this, we calculated a similarity score for each miRNA pair (Levenshtein distance). The resulting scores were clustered hierarchically and plotted as a heatmap, highlighting related miRNA on the diagonal (Figure 5-19, A). The clearly most dominant group of miRNAs contains 51 identical sequences, the second biggest cluster are the matching strands on the other side of the arm. Besides this prominent group, a few other clusters of miRNAs are visible. Based on a similarity of sequences of having 3 or less mismatches, 140 families can be defined (Figure 5-19, B).

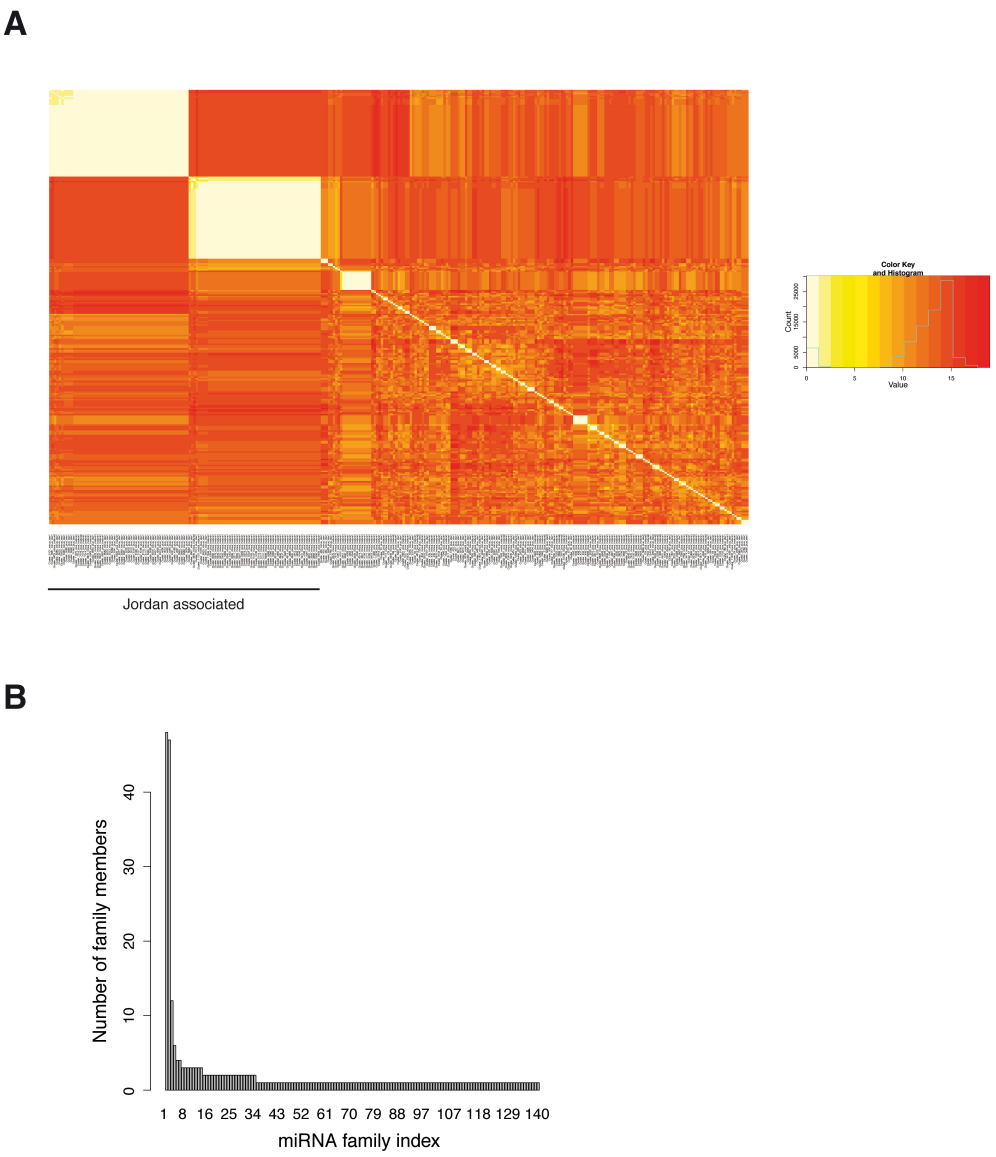



**Figure 5-19 Relationship of *Volvox* miRNAs.** (A) All identified *Volvox* miRNA sequences were compared by calculating a similarity score (Levenshtein distance (Levenshtein, 1965)). Identical sequences obtain the lowest score. The resulting score was plotted to create a heatmap with light yellow = identical sequences and dark red = highly dissimilar. (B) MiRNA families were defined as including sequences with 3 or less mismatches. Diagram shows the distribution of members across the determined 140 miRNA families.

The largest family of miRNAs in *Volvox* originates from the transposable element *Jordan* (Miller et al., 1993). Plant transposable elements can be classified into two classes and three main groups, each having autonomous (independent) and non-autonomous (dependent) members (Figure 5-20, A). Autonomous elements encode for proteins that are required and sufficient for transposition, whereas non-autonomous elements do not encode such capacity, but retain certain *cis*-elements needed for transposition (Feschotte et al., 2002). The difference between the two classes is whether their transposition intermediate is DNA (class I) or RNA (class II). For detailed description on plant transposable elements, their classification and their impact on plant genomes see (Feschotte et al., 2002; Huang et al., 2012; Lisch, 2013).

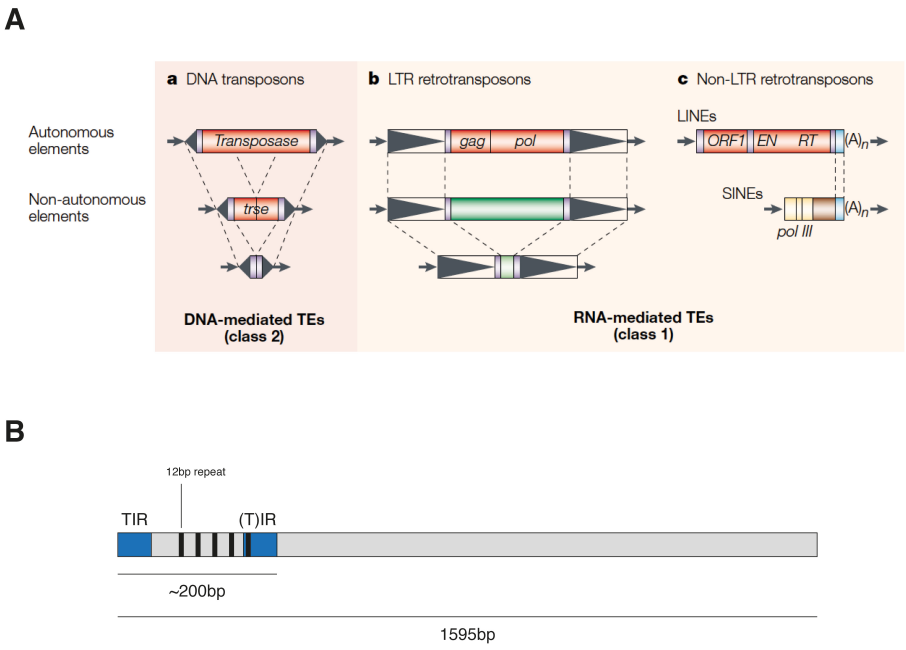



**Figure 5-20 Classification of plant transposable elements.** (A) Plant transposable elements can be roughly divided into DNA transposons (a), LTR retrotransposons (b) and Non-LTR-retrotransposons (c). All groups contain autonomous (independent of other genes/proteins) and non-autonomous, truncated (dependent on other genes/proteins) members. Figure was taken from (Feschotte et al., 2002). (B) The largest group of related miRNAs from Figure 5-19 originates from the *Jordan* transposable element. TIR, terminal inverted repeat (blue); 12 bp repeats are depicted as black boxes (Miller et al., 1993).

*Jordan* is a class II non-autonomous DNA transposon and resembles the *En/Spm*-related transposons from higher plants as well as members of the „*CACTA*“ family. It contains no open reading frame (ORF) longer than 300 bp and does not seem to encode for a transposase (Miller et al., 1993). It can be found with over 40 full-length copies in the genome. *Jordan* is 1596 bp in length and contains 55 bp inverted repeats as well as five 12 bp direct repeats in its first 200 bp (Figure 5-20, B). Due to the inverted repeats, *Jordan* elements form a structured region that obviously can give rise to miRNAs. MiRNAs originating from *Jordan* can be detected very strongly by Northern blotting (Cluster_5285 and Cluster_5322, Figure 5-18).

#### *5.5.2.5 miRNA targets*

MiRNAs in plants target mRNAs differently than in animals. Although there seems to be some contribution to translational inhibition (Brodersen et al., 2008; Iwakawa and Tomari, 2013; Lanet et al., 2009), the main regulatory mechanism works through almost perfectly complementary binding of miRNA and target followed by cleavage and degradation of the target mRNA.

To investigate potential miRNA targets in *Volvox*, we therefore searched for target mRNAs following these rules (adapted from (Molnar et al., 2007)): (a) the binding site of the miRNA should have no more than four mismatches, (b) there should not be a mismatch at positions 10 an 11 of the miRNA, because complementarity here is required for cleavage, (c) there should not be adjacent mismatches at positions 2-12, (d) no more than two adjacent mismatches for positions > 12 are allowed and (e) there should be no bulge in the miRNA.

After applying these criteria, 40 targets for the predicted miRNAs from the AGO3 immunoprecipitation could be identified (Table 5-2). The identity of these mRNA targets is of obvious importance, however, the assembly and annotation of mRNA transcripts in *Volvox* is somewhat lacking. Only part of the transcripts can therefore be judged according to their putative function.

**Table 5-2 Putative targets of *Volvox* miRNAs**

| RNA/protein | Conserved domain(s) | Putative function | Length transcript (nt) | Number of miRNA target sites |
|---|---|---|---|---|
| Gypsy4-I_VC | | Transposable element | 6039 | 1 |
| Gypsy5-I_VC | | Transposable element | 10514 | 2 |
| *JORDAN* | | Transposable element | 1595 | 113 |
| MTF0324 | 2 Vps53-like, N-terminal; DUF2884 | VPS53C_f; Similar to VPS53, a component of the GARP (Golgi-associated retrograde protein) complex. | 7194 | 1 |
| Vocar20000474m | - | - | 1545 | 1 |
| Vocar20001254m | 7 TM | Flagellar associated protein? Based upon similarity to *Chlamydomonas* | 5925 | 57 |
| Vocar20001603m | Transketolase, pyrimidine binding domain; Transketolase, C-terminal domain | Pyruvate dehydrogenase | 2662 | 1 |
| Vocar20001916m | - | - | 656 | 2 |
| Vocar20002009m | - | - | 549 | 2 |
| Vocar20002148m | - | Similar to *Jordan* transposition protein | 858 | 2 |
| Vocar20002561m | - | - | 204 | 2 |
| Vocar20003525m | - | - | 1566 | 1 |
| Vocar20003912m | - | - | 1752 | 1 |
| Vocar20004073m | - | - | 1541 | 1 |
| Vocar20004273m | - | - | 672 | 2 |
| Vocar20004886m | Catalytic domain of the Protein Serine/Threonine Kinase, Phosphoinositide-dependent kinase 1 | 3-phosphoinoside dependent protein kinase [*Chlamydomonas reinhardtii*] | 2263 | 2 |
| Vocar20005024m | - | - | 216 | 2 |
| Vocar20005142m | - | - | 234 | 2 |
| Vocar20005460m | - | - | 399 | 4 |
| Vocar20005563m | - | - | 822 | 2 |
| Vocar20005717m | HEAT-like repeat | - | 1281 | 2 |
| Vocar20006223m | - | - | 363 | 2 |
| Vocar20006432m | - | - | 2631 | 2 |
| Vocar20006713m | - | - | 2733 | 2 |
| Vocar20007112m | - | - | 912 | 58 |
| Vocar20007374m | - | - | 540 | 1 |
| Vocar20008356m | Glycosyl hydrolases family 38 N-terminal domain; Plant protein of unknown function (DUF946); Glycosyl hydrolases family 38 C-terminal domain; Alpha mannosidase, middle domain | Alpha-mannosidase | 5733 | 3 |
| Vocar20009578m | F-box/leucine rich repeat protein | - | 2076 | 1 |
| Vocar20010062m | - | - | 477 | 5 |
| Vocar20011053m | - | - | 1938 | 1 |
| Vocar20011540m | 4 WD domains | WD and tetratricopeptide repeat protein 1 | 2649 | 1 |
| Vocar20011794m | - | - | 996 | 1 |
| Vocar20012325m | 3 Lipoxygenase | Lipoxygenase | 3015 | 1 |
| Vocar20014100m | SOR/SNZ family | Tryptophane biosynthesis protein | 1465 | 1 |
| Vocar20014355m | Adenylate and Guanylate cyclase catalytic domain | | 7713 | 1 |
| Vocar20014616m | Malate synthase (insignificant) | Malate synthase? | 510 | 1 |
| Vocar20014661m | JmjC domain, hydroxylase | - | 1107 | 3 |
| Vocar20015099m | - | - | 8187 | 1 |
| Vocar20015144m | - | - | 237 | 1 |
| Vocar20015311m | - | - | 1483 | 1 |

Most of the mRNA targets, where a tentative functional annotation could be described, encode metabolic or signaling pathway enzymes. Although these enzymes might be interesting in terms of regulation or developmental processes, we do not know enough of the *Volvox* metabolism in order to evaluate that. Two mRNA targets stick out of this list because the high amount of miRNAs targeting them, Vocar20001254m and Vocar20007112m. This phenomenon is caused by the existence of an exon that is highly similar to the 5' end of the *Jordan* transposon. Thus, almost all miRNAs spawned by this region of *Jordan* target these two transcripts. The mRNA Vocar20002148m encodes a protein that is similar to the *Jordan* transposition protein (identified by Blast search). This mRNA might be an attractive target for validation, since *Jordan* needs a helper protein in order to translocate. Last but not least, three transposable elements are predicted to be targeted by *Volvox* miRNAs. *Jordan* itself is showing in this list, but whether it is "only" acting as a miRNA origin or is also targeted by these miRNAs, remains to be shown. However, the two other transposons, Gypsy4-I_VC and Gypsy5-I_VC, are not generating miRNAs from our dataset and are predicted to be targeted by "independent" miRNAs.

In conclusion, we show that our predicted miRNAs from the AGO3 IP have putative targets with potentially interesting functions. Further experiments need to show whether their mRNA levels are regulated by miRNAs *in vivo*.

### 5.5.3 Phased small RNAs and small RNA expression from repetitive elements

When Dicer proteins successively process a long ds RNA molecule, so-called phased siRNAs are generated. This phased expression can be picked up and extracted by bioinformatics tools such as the *ta-si prediction* tool from the UEA small RNA Workbench (Stocks et al., 2012), which is based on the algorithm by Chen et al. (Chen et al., 2007). The tool was created to predict plant trans-acting siRNAs (ta-siRNAs) since they are generated in a phased manner (see 1.4). For *Volvox*, we do not know if such a pathway exists and thus we analyzed our small RNA sequencing data for such RNA species.

After applying this tool, 220 loci with a phased small RNA expression pattern were identified. Many phased loci are intergenic (see example in Figure 5-21, A), however many loci also coincide with described transposable elements. As seen in Figure 5-21, B, an example phased sRNA locus is antisense to the transposable element called *Kangaroo*. This element is ~8.8kb long and belongs to the *DIRS1* group of transposons, a class I non-conventional non-LTR transposon (see Figure 5-20) with split direct repeats (Duncan et al., 2002; Poulter and Goodwin, 2005). Notably, *Kangaroo* can be found 34 times in the genome, 18 of these copies overlap with 31 phased RNA loci antisense to *Kangaroo*.

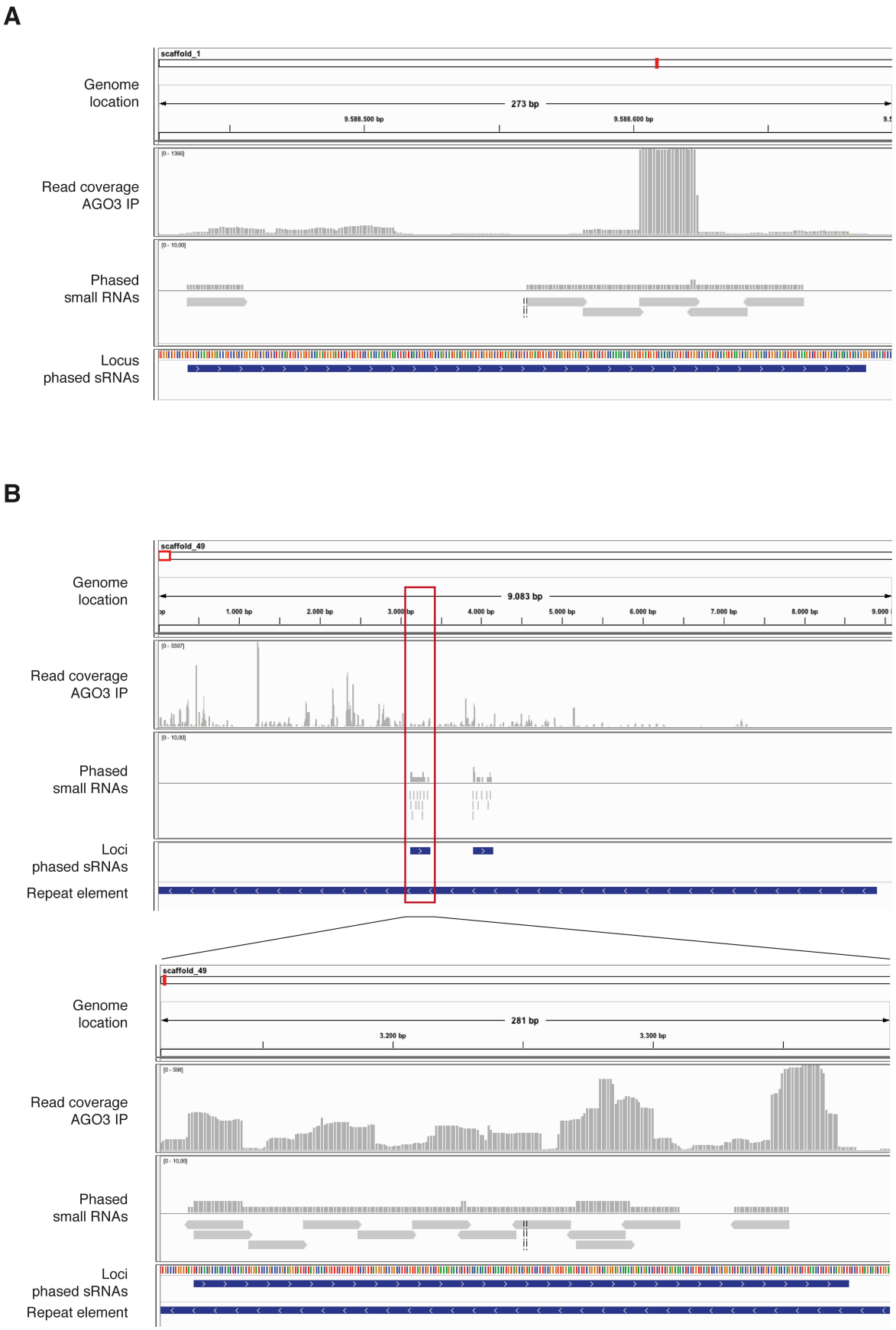



**Figure 5-21 Phased small RNA loci in *Volvox*.** (A) Example of an intergenic phased small RNA locus. Shown are the tracks with reads from the AGO3 IP and the subset of reads that are phased. (B) Example of a phased locus lying antisense to an annotated transposable element, *Kangaroo*. The upper panel shows the full length of *Kangaroo* at this site. The lower panel zooms into one of the two phased siRNA loci (indicated by black lines).

The finding that phased small RNAs can be found in antisense to transposons prompted us to analyze the extent of small RNA expression either sense or antisense to all annotated transposable elements. We therefore retrieved the consensus sequences for *Volvox* transposons from Repbase Update release 19.02 (Jurka et al., 2005) and mapped all reads against these sequences. While not every element shows a strong coverage of reads, quite a few transposons exhibit expression of small RNAs at their 5' end or of basically the whole element (Figure 5-22).

The Ty3-*gypsy* superfamily of transposons can be found throughout the plant kingdom. Ty3-*gypsy* transposons are LTR retrotransposons (see Figure 5-20) and occur usually in a high copy number, sometimes up to a few million in plants with large genomes (Kumar and Bennetzen, 1999). Three members of this family in *Volvox* show visible expression of small RNAs (Figure 5-22, A).

When looking at the expression of small RNAs from the consensus sequence of *Kangaroo*, it can be appreciated that the whole element is covered with small RNAs both in sense and in antisense direction (Figure 5-22, B, left panel).

*LINE1* (*L1*) as well as *RTEX-3* are non-LTR transposons (see Figure 5-20). Whereas in animals L1 can contribute to up to 20% of the whole genome, its abundance in plants is low (Huang et al., 2012). Both elements show expression of sense and antisense small RNAs from their 5' terminus (Figure 5-22, B, middle and right panel).

When analyzing the miRNA content of AGO3 associated small RNAs (see 5.5.2.4), one origin could be identified as being the transposable element *Jordan*. The family of miRNAs originates from the first 200 bp of the element, which forms a hairpin structure due to inverted repeats. Mapping all small RNA reads against *Jordan* reveals not only the expression from the 200 bp region at the 5' end, but also expression from the region between ~200 - ~800 bp (Figure 5-22, C). These two regions have therefore tentatively been named "miRNA region" and "siRNA region" (Figure 5-22, C right panel).

Taken together, repetitive elements show expression of sense and antisense small RNAs. Sometimes, this coincides with a phased expression, e.g. in the case of *Kangaroo*, and might hint at a pathway similar to the ta-siRNA pathway in *Arabidopsis* (see 1.4). More detailed analyses and experimental validation will be needed to elucidate this phenomenon further.

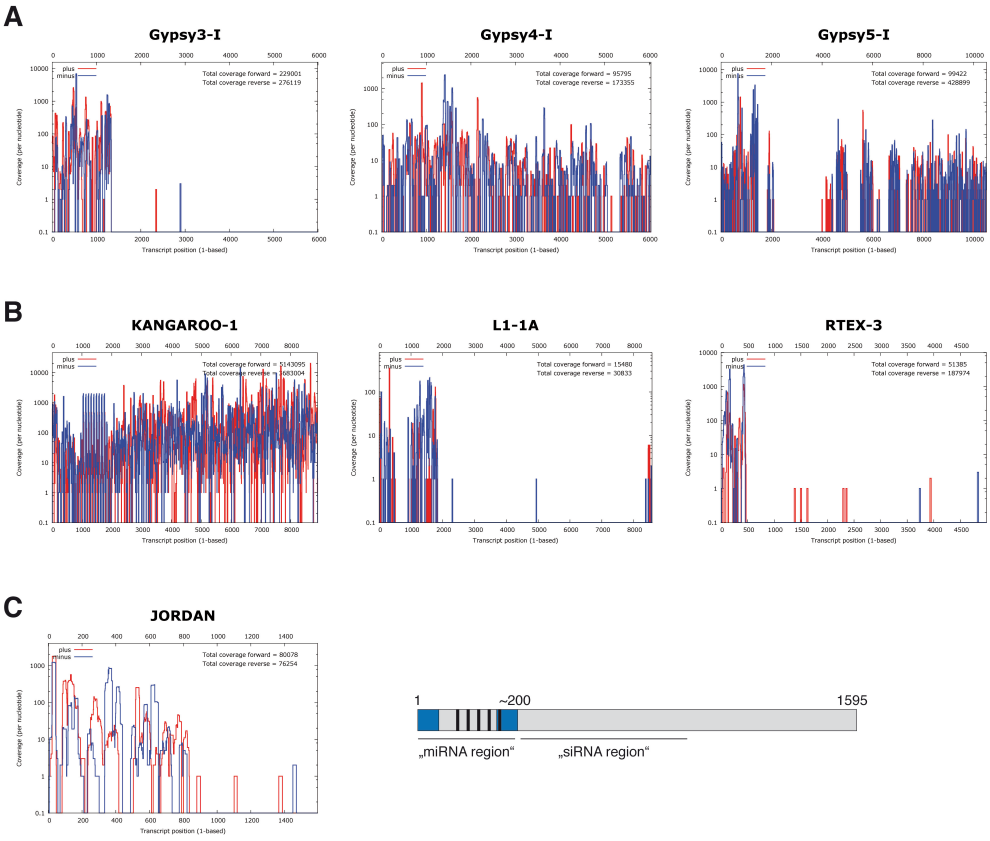
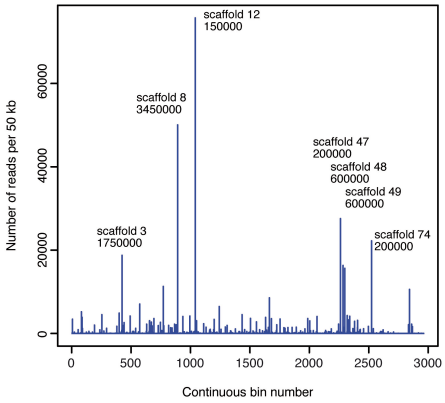



**Figure 5-22 Read coverage for several repetitive elements.** For each coverage plot, small RNAs were plotted against the full length of the respective element. Red, reads matching to the plus strand (sense); blue, reads matching to the minus strand (antisense). (A) Elements of the *Gypsy* family. (B) Three different elements, *Kangaroo*, *LINE1* and *RTEX-3*. (C) The coverage for *Jordan* is shown in a coverage plot (left) and in a schematic view according to hypothesized function (right). Blue, terminal inverted repeat; black bars, 12 bp repeats.

### 5.5.4 Analyzing the “unknown” fraction of small RNAs

The fraction of reads associated with AGO3 that could not be assigned to any functional group is quite large (~60%) (Figure 5-12, C). Possible reasons to explain this are the low degree of annotation of the *Volvox* genome or unassembled sequences, for example. In order to get an idea of the identity of these reads, we investigated all “unknown” reads over the length of the whole genome. The *Volvox* genome is not assembled in chromosomes yet, but rather fragmented in so-called scaffolds. In the current assembly, the 131.2 Mb of the *Volvox* genome are arranged in 434 scaffolds. For simplicity’s sake, all scaffolds were aligned in a row (representing the whole genome, albeit not in its correct order) and divided into partitions (so-called bins) of 50 kb. All reads within each bin were summed up and plotted in order to find potential hotspots of small RNA expression (Figure 5-23, A). There are several peaks that stand out in comparison to the others; the bin location of the most prominent peaks was labeled next to the peak (Figure 5-23, A). These seven peaks were selected for further analysis. It was investigated how many small RNAs within the respective bin contribute to the total amount of reads in that bin and what other features, like the structure of the small RNA locus, could be derived.

**A**




**B**

| sRNA locus | strand | peak of expression | sequence main peak | location | read count | structure |
|---|---|---|---|---|---|---|
| Scaffold 12 | + | 4 sRNAs; 1 main peak | TTGGATAGCAGCCTTGAGCACC | 193609-193630 | 51000 | very long ds |
| Scaffold 8 | - | 1 sRNA | TCGAGGTCGTCTTGGCAGCCG | 3469620-3469640 | 50000 | many hairpins |
| Scaffold 74 | + | 1 sRNA | TTGGCTTCTGAACTGGTTTACC | 223183-223204 | 21000 | many hairpins |
| Scaffold 47 | + | 1 sRNA | TGGGACGAGCATAAGAAGGACT | 239357-239378 | 21000 | very long ds |
| Scaffold 3 | - | 1 sRNA | TGGAAGAGTAACGCGCGCGCACCC | 1767682-1767705 | 13000 | very long ds |
| Scaffold 48 | + | more than 1 sRNA | TGGGAATCGGACATGGTAGCA | 635126-635146 | 8000 | very long ds |
| Scaffold 49 | + | 2 equally expressed main sRNAs + many more sRNAs | TAAGAGGACGTGACAAAAATTC | 617519-617540 | 2500 | many hairpins |
| | + | | TGGCGAACTTGGAAAGAACT | 614384-614403 | 2400 | |
| Σ | | 8 sRNAs | | | 168900 | |

**C**

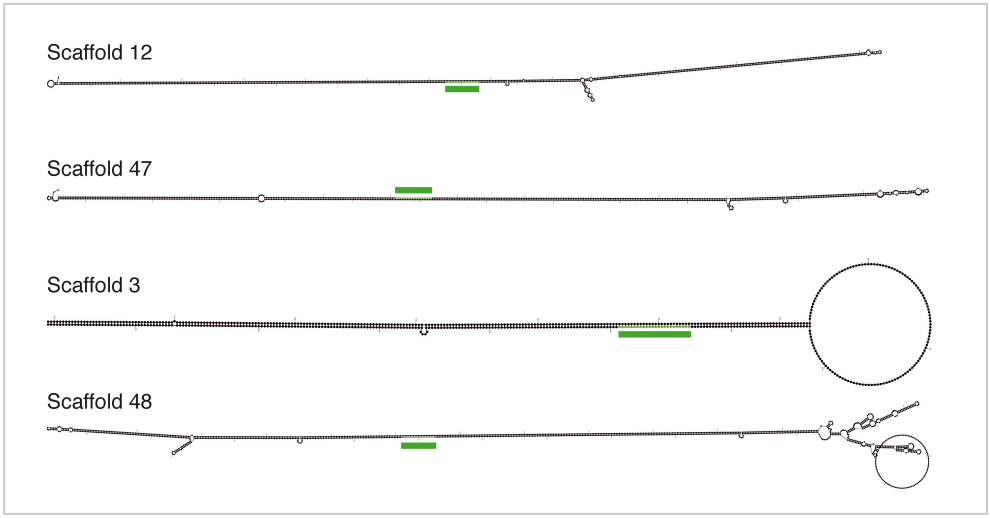



**Figure 5-23 Analysis of unassigned reads associated with AGO3.** (A) All reads that could not be assigned to an RNA category were displayed. For this, all genomic scaffolds were lined up and divided into 50 kb bins. All reads falling into one bin were summed up. Total read numbers for each continuous bin were plotted. (B) Overview over the most prominent peaks from (A). (C) Structures of the four small RNA loci that form very long, almost perfectly complementary foldback structures. The small RNA with the most reads (see also B) is highlighted with a green bar.

The results of this analysis are summarized in Figure 5-23, B. It is apparent that for the majority of peaks, only one small RNA is responsible for the high expression of the 50 kb region. For analyzing the structure of potential precursor molecules, the sequence of the vicinity of the respective small RNA was extracted and folded using mfold version 2.3 at 30°C (Walter et al., 1994; Zuker, 2003).

Surprisingly, four out of these seven loci display very long ds foldback structures (Figure 5-23, C). The pairing of these structures is unusually perfectly complementary reminiscent of inverted duplicated sequences. The expression of the main eight small RNAs alone from the loci mentioned above accounts for 7% of all AGO3 associated reads. Indeed, when analyzing their expression in different stages of the *Volvox* life cycle by Northern blotting, three out of three tested small RNAs were expressed rather abundantly (Figure 5-24). This simple investigation of small RNA expression hotspots does not yield further insight into possible functions of the small RNAs, but their high abundance indicates biogenesis and effector pathways that are unknown yet. However, it was not possible to characterize these small RNAs in more detail in the time frame of this thesis. It will be interesting to see whether these different small RNAs are *Volvox*-specific and have common or distinct functions.

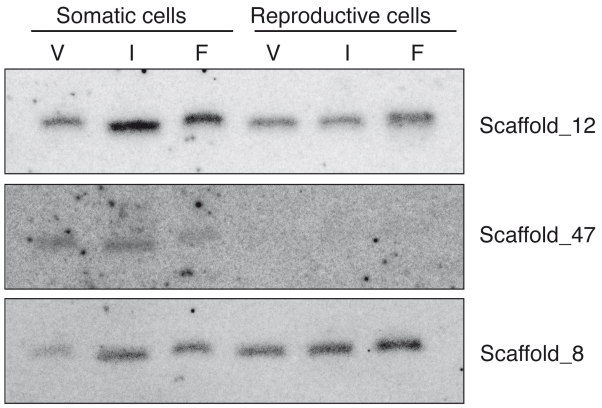



**Figure 5-24 Validation of small RNA expression from functionally unknown loci.** Three of the small RNAs identified by analyzing genome-wide small RNA expression hotspots and shown in Figure 5-23 were investigated by Northern blotting. They were named according to the description in Figure 5-23.

### 5.5.5 Conservation of small RNAs between *Chlamydomonas* and *Volvox*

The unicellular alga *C. reinhardtii* is the closest unicellular relative to multicellular *V. carteri*. Since they are only 200 million of years apart, it would be expected to find a strong overlap and conservation of small RNA expression. We therefore analyzed our deep sequencing data in order to find conserved small RNAs. Surprisingly, we could not detect any small RNAs (miRNAs or other) of *Volvox* to be similar to *Chlamydomonas*' small RNAs. To test this finding, we probed several exemplary small RNAs by Northern blotting of *Chlamydomonas* and *Volvox* total RNA. One small RNA from *Volvox* from our analysis and two small RNAs from *Chlamydomonas* identified by Baulcombe and colleagues (Molnar et al., 2007) were probed (Figure 5-25). This second method confirms that a signal is only visible in the organism where the small RNA was originally identified.

This result was predicted by two groups by genomic analyses (Prochnik et al., 2010; Zhao et al., 2007), and also observed by Qi and colleagues (Li et al., 2014).

Although *Volvox* and *Chlamydomonas* are closely related organisms, their small RNA repertoire seems to have dramatically diverged so that we cannot find any overlap in the data generated so far.

This result is astonishing since their proteomes are predicted to be quite similar (Prochnik et al., 2010) and as mentioned above, proteins of one organism can functionally replace proteins of the other organism (see 5.1.3).

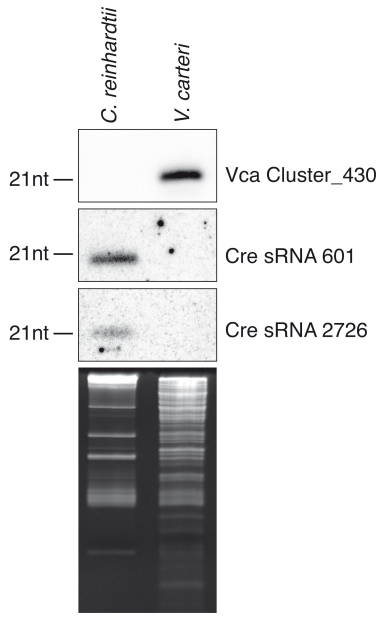



**Figure 5-25 Measure of conservation between *Chlamydomonas* and *Volvox* small RNAs.** Total RNA from *Chlamydomonas reinhardtii* (Cre) and *Volvox carteri* (Vca) was analyzed by Northern blotting. One small RNA from *Volvox* identified by our pipeline and two small RNAs identified by (Molnar et al., 2007) were probed to investigate cross-species conservation of small RNAs (upper panels). An ethidium bromide stain shows total RNA levels (lower panel).

## 5.6 Discussion

The green alga *V.carteri* is a model organism for studying the evolution of multicellularity and division of labor between germ line and soma. Since its most closely related unicellular relative *C.reinhardtii* is also a model organism, both species are rather well described in comparison to other green algae. In 2007, the small RNA and miRNA repertoire of *Chlamydomonas* has been published (Molnar et al., 2007; Zhao et al., 2007), followed by the publication of the first draft genome of *Volvox carteri* in 2010 (Prochnik et al., 2010). This foundation of data prompted us to analyze small RNAs of *V.carteri* and investigate functional and evolutionary questions.

Our search in the *Volvox* genome revealed that *Volvox* possesses all necessary proteins to generate and functionally use small RNAs. These include a Dicer-like protein, two Argonaute proteins, a HEN1 homologue and an RNA-dependent RNA polymerase. We optimized the already annotated sequence of one Argonaute, AGO3, isolated its cDNA and generated a *Volvox* strain that constitutively expressed a tagged version of AGO3. Although the strain only expressed AGO3 for a limited amount of time (several weeks) because of epigenetic silencing of the transgene, we were able to immunoprecipitate AGO3 and isolate its associated small RNAs. These were subsequently cloned and sequenced.

Plant Argonaute proteins have more defined functions than human Argonautes. For example, AGO1 in *Arabidopsis thaliana* exclusively binds miRNAs, whereas AGO4 binds to siRNAs that are 24 nt in length (Dueck and Meister, 2014). In contrast, all four human Ago protein family members bind a very similar set of miRNAs (this work, (Dueck et al., 2012)). Since AGO3 is one of two Argonaute genes in *Volvox*, we anticipated that by identifying the associated small RNAs, we could deduce a role for this Argonaute gene in *Volvox*. The length profile of all bound small RNAs is predominantly 21 and 22 nt, with no striking peak at 24 nt (Figure 5-12, A). The biogenesis of 24 nt long hc-RNAs is conserved across flowering plants and some mosses (Axtell, 2013). Hc-RNAs are necessary for suppression of transposable elements by *de novo* induction of repressive chromatin modifications (see also 1.4). Although it cannot be ruled out that under different conditions like stress or starvation this kind of small RNAs would appear, under normal growth conditions during the asexual, vegetative phase there is no evidence for this class of small RNAs. A similar profile has been seen by Li et al. (Li et al., 2014) and by Montes et al. (Montes et al., 2014) in total RNA from *Volvox*, confirming that there is not one single length for small RNAs but a double peak.

The composition of AGO3 associated reads is somehow surprising at first glance. Most Argonaute genes bind one predominant small RNA species, e.g. human Agos miRNAs, *Arabidopsis* AGO4 hc-RNAs, *Drosophila* Ago2 siRNAs etc. In the case of *Volvox* AGO3, however, we see a more heterogeneous picture. We do find putative miRNAs bound to AGO3 and can validate them as being expressed and modified (Figure 5-17 and Figure 5-18), but only 10% of all bound small RNAs are predicted miRNAs. Additionally, small RNAs stemming from mRNAs (in sense) that are incorporated into an Argonaute protein are rather rare (so far) and then, a high amount of small RNA reads cannot

be categorized at all. The answers to these questions are probably due to one of three circumstances (or parts of all):

First, the *Volvox* genome is still of minor quality (assembly status, gaps, sequencing errors) and consequently the annotation status of mRNA and other transcripts. The main genome assembly is about 131.2 Mb big and arranged in 434 scaffolds. 14,971 total loci were mapped that contain protein-coding transcripts. 4.4% of the genome is currently unknown and labeled as gaps (source: JGI (Prochnik et al., 2010)). The fraction of small RNAs falling into our categories "mRNA" or "unknown" is quite large. There might be a lot of mRNAs misannotated as protein-coding that could in truth be non-coding RNA.

A second reason might be the sensitivity of our miRNA identification pipeline. While we carefully evaluated all parameters to identify *Volvox* miRNAs, miRNA genes might be missed due to constraints. For example, we set a limit at 600 bp for a potential miRNA precursor and a constraint of a minimum free energy value per nucleotide. With more information about the nature of *Volvox* miRNAs, parameters can be adjusted and additional miRNA genes might be found. An already interesting example that doesn't conform to our criteria are the small RNA hotspots that we identified in the "unknown" fraction of reads (Figure 5-23). These small RNAs originating from probably very long precursors might actually be miRNAs with an unusually processing and expression pattern. Future experiments and studies will help to uncover their specific functions and gene structures.

Third, the evolutional stage of *Volvox* and its Argonaute genes has to be taken into account. *Arabidopsis thaliana* is the most well studied plant organism in terms of small RNAs and Argonaute genes and functions. This plant, also known as mouse-ear cress, is a small flowering plant native to Europe and Asia and a model organism for many biological aspects. Due to its very small genome size of approximately 135 Mb (for a flowering plant), it was the first plant genome that was sequenced (Initiative, 2000). Although its 10 Argonaute proteins have not been analyzed in full yet, most of our knowledge of small RNA functions and mechanisms in plants originates from research in *Arabidopsis*. The common ancestor of *Arabidopsis* and the green algae dates back 1200 MYA (million of years ago) (Herron et al., 2009), implying that these two groups are quite far away in evolutionary terms. Additionally, while *Arabidopsis* is a higher plant, *Volvox carteri* is one of the simplest multicellular organisms with only two cell types. In *Arabidopsis*, many developmental stages like growth of roots or flowering exist, which are not present in *Volvox* at all. Therefore, it is possible that the function of Argonaute proteins diverged and knowledge from *Arabidopsis* cannot be transferred to *Volvox*.

When looking at the structure of the identified miRNAs in *Volvox*, it becomes apparent that these structures do not represent end points of miRNA evolution. Most *Volvox* precursors are around 200 nt long, many even longer (Figure 5-15). The fraction of base-pairing is also quite high (Figure 5-16), highlighting the relatively high complementarity between the two arms of the precursor. According to the general idea that miRNAs can evolve from inverted gene duplication (Voinnet, 2009), *Volvox* likely contains quite young miRNA genes. A relatively similar scenario can be observed for *Chlamydomonas* (Molnar et al., 2007; Zhao et al., 2007), however, it seems that *Chlamydomonas* miRNAs are relatively unique (86 miRNAs, with maximum two members per family) while *Volvox* currently possesses 288

miRNAs in 140 families with many families containing more than three members. *Volvox* and *Chlamydomonas* do not share any small RNAs. This surprising circumstance might be caused by the driving force of evolution. For the establishment of multicellularity and the acquirement of additional pathways, *Volvox* might have needed to evolve new regulatory small RNAs.

Small RNAs and transposable elements in *V.carteri* seem to be somehow connected. We have discovered a large miRNA family originating from the transposon named *Jordan*. Inverted repeats form a foldback structure that can obviously be processed by the Dicer-like enzyme of *Volvox*. This miRNA family is predicted to target mRNAs as well as the *Jordan* transcript itself (Table 5-2). In case of the mRNAs, the very last exon contains a fragment of *Jordan* and thus is targeted. It is conceivable that a regulatory mechanism has arisen that utilizes miRNAs originating from *Jordan* to regulate transposon-independent mRNAs. Further analyses of the RNA Seq data and experimental validation need to show whether these miRNAs target *Jordan* itself or whether *Jordan* is only their source. A proposed evolutionary origin of miRNAs has been degenerate transposon sequences called miniature inverted-repeat transposable elements (MITEs) (Felippes et al., 2008; Piriyapongsa and Jordan, 2008) (see 1.2). However, this is the first time miRNAs were shown to be likely generated from a full length transposable element.

No other transposable element could be detected to spawn miRNAs like *Jordan*. However, when mapping all small RNA reads onto consensus sequences of all *Volvox* transposons, many transposable elements show small RNA coverage in sense as well as in antisense (Figure 5-22). In a recent publication, Martienssen and colleagues have shown in *Arabidopsis* that miRNA-mediated cleavage of transposon transcripts can initiate siRNA production. These siRNAs were called epigenetically activated siRNAs (easiRNA) as they are only activated during reprogramming of the germ line (Creasey et al., 2014). A similar mechanism might be observable in *Volvox*. We could already identify three transposons that are targeted by miRNAs, namely *Jordan*, *Gypsy4-I* and *Gypsy5-I* (Table 5-2). All three show small RNA expression in sense and antisense. These small RNAs are incorporated into AGO3, therefore implying to be functional downstream. The exact function and consequences, however, have to be determined by further analyses.

In conclusion, we have identified a functional small RNA biogenesis and effector machinery in *Volvox carteri*. The Argonaute protein AGO3 binds small RNAs with the length of 21-22 nt and seems to have more than one function. We could identify putative miRNAs bound to AGO3 as well as small RNAs that are phased, mapping to mRNAs or repetitive elements. Although many aspects could already be analyzed, a wealth of information is still hidden in our data sets.

Additionally, we have identified a second Argonaute gene (*AGO2*) in *Volvox*. The encoded protein might complement AGO3 function by binding to different subsets of small RNAs. The analysis of AGO2-bound small RNAs as well as total small RNA profiles of different stages of the life cycle of *Volvox* will elucidate small RNA pathways in *Volvox* further. It will be highly interesting, what new aspects and functions of small RNAs in *Volvox carteri* can be uncovered.

# CHAPTER 6 COMPREHENSIVE SUMMARY AND OUTLOOK

This thesis describes Argonaute protein properties as well as miRNA expression and characteristics in several different organisms and cellular systems.

The binding partners and effector proteins of miRNAs are Argonaute proteins (reviewed in (Ameres and Zamore, 2013; Hutvagner and Simard, 2008; Meister, 2013)). The identity, expression pattern and cellular abundance are of obvious importance for miRNA function. This work investigated Argonaute mRNA and protein characteristics such as abundance, cleavage activity or incorporation of siRNA duplexes in human HeLa and HEK 293T cells (Chapter 2). The human Ago proteins are expressed at very different levels relative to each other with Ago2 being the most abundant Ago protein in the investigated cell lines (Figure 2-1). Further, I could show that Ago2 is indeed the only human Ago protein capable of cleavage by testing passenger strand as well as target cleavage (Figure 2-3) (Hauptmann et al., 2013). Because of this cleavage activity, Ago2 can more easily displace the passenger strand of perfectly complementary siRNAs than Ago1 or Ago3, even when the passenger strand is stabilized by incorporated modified nucleotides (such as LNAs) (Figure 2-5). In fact, we showed that by stabilizing an RNA duplex, the on-target activity of Ago2 could be increased, whereas off-target effects caused by Ago1 and Ago3 could be diminished (Petri et al., 2011). This result provides the basis for improving RNAi effectiveness and specificity by exploiting differential Ago protein characteristics.

Besides the Ago proteins present in the cell, the relative and absolute levels of miRNAs are important factors, since the more molecules are present, the higher is the impact on mRNA regulation. The total miRNA profiles of murine macrophages as well as of murine dendritic cells isolated from bone marrow were analyzed under different culture conditions (Chapter 4) (Dueck et al., 2014). Not only were the miRNA profiles characterized when these cells are induced or stimulated, but also a miRNA hierarchy was described using the corresponding miR-155-deficient cells (untreated or stimulated/matured) (Figure 4-9, C). The effect of one miRNA or one small RNA is only minor, rather, miRNAs and small RNAs in general act in a more cumulative or concerted manner. Therefore, it is crucial to understand small RNA networks and their impact on a cellular level. Using a mouse model deficient in miR-155, I investigated the consequences of miRNA loss on the remaining miRNA pool. The miRNA profiles of macrophages as well as of DCs change substantially upon loss of miR-155. We additionally identified miRNAs, that are regulated in the same manner in miR-155-deficient cells (Dueck et al., 2014).

The deep sequencing data of macrophages and dendritic cells was validated by Northern blotting and both methods were in quite good agreement (Figure 4-8). In this case, the libraries were constructed out of the total pool of small RNAs. For the analysis of miRNAs associated with the different Ago proteins, in contrast, the libraries were constructed from IPs. The miRNA amounts incorporated into human Ago proteins showed differences between the distinct Agos in our deep sequencing libraries, however, these differences could not be verified by Northern blotting. It seems that the method of

enrichment by IP coupled with cloning and deep sequencing introduces biases that render these libraries incomparable to each other in relation to the quantitative amounts of bound miRNAs.

Several specific miRNA-Ago interactions were analyzed. The investigation of human Ago-miRNA complexes showed that Ago1-3 bind different lengths of miRNAs (Figure 3-3, A). Ago1 and Ago3 preferentially bind longer miRNA isoforms than Ago2, which also binds to relatively short sequences (down to ~18 nt). Although some slight differences could be seen in relation to non-templated nucleotide addition as well, the addition of one adenosine to Ago-bound miRNAs seems to be the most prevalent (Figure 3-3, B). Focusing on individual miRNAs, two different types of miRNA-association were observable for individual Ago proteins. One type of miRNA-association showed highly similar length patterns, i.e. Ago1-3 bind a miRNA in the same way (of the same length). The other type shows differences mainly between Ago2 and Ago1/3 (Figure 3-4). The difference in the bound individual length profiles is consistent with the observed global length profiles, namely that Ago1 and Ago3 bind to longer isomiRs, whereas Ago2 also binds to shorter isomiRs. Not only is Ago2 an exception in this case, Ago2 was also shown to be the only Ago to bind miR-451 (Figure 3-6). Mutational studies confirmed that the unique structure and also some sequence features determine correct miR-451 processing by Ago2. In contrast, foreign sequences built into the miR-451 backbone expressed only weakly (Figure 3-7 and Figure 3-8). Therefore, the processing of and the interaction with miR-451 seems to be optimized for Ago2 (Dueck et al., 2012).

The current knowledge and the above mentioned results underline the importance of understanding small RNA features as well as their protein binding partner. Therefore, when seeking to analyze small RNA expression and function in a so far uncharacterized organism, the analysis of involved proteins can yield critical insight. In the multicellular alga *V.carteri*, I investigated the Argonaute protein AGO3 in detail. Its mRNA seems to be distributed in almost equal levels between somatic cell and reproductive cells (Figure 5-11). The AGO3 protein has a molecular mass of 111 kDa and harbors a PAZ and a Piwi domain, which are key domains for all Argonaute proteins characterized so far (Figure 5-9). The N-terminal domain is glycine-rich, which is also observed in some AGO proteins of *Chlamydomonas* and *Arabidopsis*. The amino acid residues that are crucial for target cleavage are DEDD in AGO3, thus rendering AGO3 potentially cleavage active (as, for example, AGO2 of *Arabidopsis* (Carbonell et al., 2012)). Additionally, we have described a full-length second Argonaute protein (AGO2) that will be interesting to study further. In our current annotation, AGO2 bears the catalytic motif DED, with the last amino acid entirely missing (Figure 5-10). Future experiments will show, if AGO3 is indeed cleavage active and if the annotated AGO2 mRNA and protein sequence can be validated *in vivo*.

Like human Ago and many other Argonaute proteins (see Supplement 2, Table, Appendix) (Dueck and Meister, 2014; Frank et al., 2010), *Volvox* AGO3 predominantly binds to small RNAs with a 5' U and these are, in contrast to mammals, of a specific length of 21 and 22 nt. We analyzed the small RNA content bound to AGO3 and could assign different small RNA categories, e.g. miRNAs, repeat-associated small RNAs or phased small RNAs. We generated our own bioinformatics identification pipeline for the prediction of miRNAs that made it possible to define parameters fitting to the

characteristics of the group of green algae (includes *Chlamydomonas reinhardtii*, see also (Molnar et al., 2007; Zhao et al., 2007)). Several of the predicted miRNAs were validated by Northern blotting and shown to be modified at their 3' end. However, the biggest part of AGO3-associated small RNAs could not be identified functionally. The *Volvox* genome is only poorly annotated and future re-analyses will shed more light onto AGO3-bound small RNAs as well as their potential targets.

Small RNA networks and profiles of certain cells or cellular conditions are not static, unchanging structures. Rather, they also represent snapshots of evolutionary development of the investigated organism. Our analysis of *Volvox* small RNAs uncovered a very low degree of conservation to *Volvox'* closest unicellular relative, *Chlamydomonas*. From an evolutionary point of view, this finding is highly interesting, since both organisms have separated only ~200 million years ago (Herron et al., 2009). The miRNA genes, that we can detect and predict, show features of being quite young, e.g. having highly complementary hairpins (Voinnet, 2009). It would be an attractive model to hypothesize that small RNAs are a major evolutionary driving force for the development of multicellularity and division of labor between somatic and reproductive cells, since it was already shown that the transcriptomes and proteomes of the two organisms are highly similar (Ferris et al., 2010; Prochnik et al., 2010). Functional studies on small RNAs and small RNA pathways in *Volvox* will certainly uncover if small RNAs regulate key aspects of *Volvox* development.

# CHAPTER 7 MATERIALS

## 7.1 Chemicals and enzymes

Unless stated otherwise, chemicals were purchased from Amersham Biosciences (Buckinghamshire, UK), Applichem (Darmstadt, Germany), Bio-Rad (Hercules, USA), Merck (Darmstadt, Germany), Qiagen (Hilden, Germany), Roth (Karlsruhe, Germany) and Sigma-Aldrich (Munich, Germany).
Radiochemicals were purchased from Perkin Elmer (Waltham, USA) or Hartmann Analytics (Braunschweig, Germany), restriction enzymes and enzymes for RNA and DNA modifications (T4 PNK, ligases etc.) were purchased from Fermentas (Thermo Fisher Scientific, Waltham, USA) or New England Biolabs (Ipswich, USA).
DNA oligonucleotides were ordered from Metabion (Martinsried, Germany), RNA oligonucleotides were produced by in-house service facilities or ordered from Biomers (Ulm, Germany).

## 7.2 Plasmids

| | |
|---|---|
| pIRES-VP5 | Encodes for an N-terminal Flag/HA tag (Meister et al., 2004) |
| pMIR-REPORT | Encodes a firefly luciferase (*Phototinus pylaris*) under the control of a mammalian promoter/terminator system, with an miRNA target cloning region downstream of the luciferase translation sequence (Thermo Fisher Scientific, Waltham, MA, USA) |
| pMIR-RL | Is modified from the commercially available pMIR-REPORT vector (Thermo Fisher Scientific, Waltham, MA, USA; (Beitzinger et al., 2007)). Besides the firefly luciferase, a *Renilla reniformis* luciferase (termed renilla) under the control of a SV40 promotor was was inserted into the *SspI* site of pMIR-REPORT. The additional luciferase can be used for transfection normalization. |
| pSuper | Allows for the expression of shRNAs under the control of an RNA polymerase III promoter (oligoengine, Seattle, USA). The backbone is modified as used in (Zhu et al., 2009). |
| pCS2-myc6 | Encodes for an N-terminal myc6-tag |
| pPmr3 | This plasmid encodes for the bacterial Paromomycin resistance gene, *aphH,* which can be used as a dominant selectable marker in *Volvox carteri* (Jakobiak et al., 2004). Its GenBank accession number is AY429514. |

## 7.3 Antibodies

| Antigen | Source | Dilution | Application | Reference / Manufacturer |
|---|---|---|---|---|
| α-humanAgo1 [1C9] | rat hybridoma supernatant, monoclonal | 1:10 | WB | (Beitzinger et al., 2007) |
| α-humanAgo1 [4B8] | rat hybridoma supernatant, monoclonal | | IP | (Beitzinger et al., 2007) |
| α-humanAgo2 [11A9] | rat hybridoma supernatant, monoclonal | 1:50 | WB/ IP | (Rudel et al., 2008) |
| α-humanAgo3 [5A3] | rat hybridoma supernatant, monoclonal | | IP | (Weinmann et al., 2009) |
| α-humanAgo3 [4A11] | rat hybridoma supernatant, monoclonal | | IP | (Petri et al., 2011) |
| α-humanAgo4 [6C10] | rat hybridoma supernatant, monoclonal | | IP | (Weinmann et al., 2009) |
| α-humanAgo4 [3G5] | rat hybridoma supernatant, monoclonal | | IP | (Petri et al., 2011) |
| α-mouseAgo2 [6F9] | rat hybridoma supernatant, monoclonal | | IP | (Zhu et al., 2010) |
| α-ratRmC | rat hybridoma supernatant, monoclonal | | IP | Elisabeth Kremmer |
| α-beta-actin [ab6276] | mouse, monoclonal | 1:5000 | WB | Biozol, Abcam |
| α-myc (C3956) | rabbit, monoclonal | 1:1000 | WB | Sigma-Aldrich |
| α-alpha-tubulin | mouse, monoclonal | 1:5000 | WB | Sigma-Aldrich |
| α-HA [16B12] | mouse, monoclonal HA.11 | 1:1000 | WB | Covance |
| goat α-rabbit | IR Dye, 800CW | 1:10000 | WB | LI-COR |

## 7.4 Kits, membranes and reagents

| Name | Supplier |
|---|---|
| 1.0µm (Ø) Gold (Microcarrier) | Bio-Rad (Hercules, CA, USA) |
| 2x MESA Green qPCR MasterMix Plus | Eurogentec (Belgium) |
| 900psi Rupture Disks | Bio-Rad (Hercules, CA, USA) |
| Amersham Hybond-ECL | GE Healthcare (Freiburg, Germany) |
| Amersham Hybond-N | GE Healthcare (Freiburg, Germany) |
| First Strand cDNA Synthesis Kit | Thermo (Rockford, IL, USA) |
| HE Transcreen screen | Kodak (Rochester, USA) |
| High Sensitivity DNA Assay | Agilent Technologies (Böblingen, Germany) |
| illustra MicroSpin G-25 Columns | GE Healthcare (Freiburg, Germany) |
| Imaging Screen K (35x43cm) | Bio-Rad (Hercules, CA, USA) |
| Kodak BioMax MR Film | Sigma-Aldrich (Steinheim, Germany) |
| Kodak BioMax MS Film | Sigma-Aldrich (Steinheim, Germany) |

| | |
|---|---|
| Lipofectamine 2000 | Invitrogen (Karlsruhe, Germany) |
| Macrocarrier | Bio-Rad (Hercules, CA, USA) |
| Passive lysis buffer | Promega (Madison, USA) |
| Phusion polymerase | New England Biolabs (Ipswich, USA) |
| Protein G Sepharose Beads | GE Healthcare (Freiburg, Germany) |
| SsoFast™ EvaGreen® Supermix | Bio-Rad (Hercules, CA, USA) |
| Stopping Screens | Bio-Rad (Hercules, CA, USA) |
| SuperScriptIII First Strand cDNA SuperMix | Invitrogen (Karlsruhe, Germany) |

## 7.5 Instruments

| **Name** | **Supplier** |
|---|---|
| 2720 Thermal Cycler | Applied Biosystems (Foster City, CA, USA) |
| Agilent 2100 Bioanalyzer | Agilent Technologies (Böblingen, Germany) |
| Avanti J-20 XP Centrifuge | Beckman Coulter (Krefeld, Germany) |
| Biolistic PDS-1000/He Particle Delivery System | Bio-Rad (Hercules, CA, USA) |
| Branson Sonifier 450 | Heinemann (Schwäbisch Gmünd, Germany) |
| Centrifuge 5415D | Eppendorf (Hamburg, Germany) |
| CP 1000 | AGFA (Mortsel, Belgium) |
| Electrophoresis Power Supply EV233 | Consort (Turnhout , Belgium) |
| Geiger Counter Model LB123 | EG&G Berthold (Bad Wildbad, Germany) |
| GeneAmp PCR System 9700 | Applied Biosystems (Foster City, CA, USA) |
| Heraeus Biofuge pico | Thermo Scientific (Rockford, IL, USA) |
| Heraeus Cytoperm | Thermo Scientific (Rockford, IL, USA) |
| Heraeus Fresco 17 | Thermo Scientific (Rockford, IL, USA) |
| Heraeus HeraCell 240i $CO_2$ Incubator | Thermo Scientific (Rockford, IL, USA) |
| Heraeus Megafuge 40R | Thermo Scientific (Rockford, IL, USA) |
| HeraSafe KS | Thermo (Rockford, IL, USA) |
| Hybond Shake'n'Stack | Thermo (Rockford, IL, USA) |
| Hybridization oven Type T 5042 | Heraeus (Hanau, Germany) |
| IKA MS3 | Agilent Technologies (Böblingen, Germany) |
| Incubator Model B6200 | Heraeus (Hanau, Germany) |
| Innova 44 Incubator Shaker Series | New Brunswick Scientific (Edison, NJ, USA) |
| Mastercycler gradient | Eppendorf (Hamburg, Germany) |
| Microscope Diavert | Leitz (Wetzlar, Germany) |
| Milli-Q PLUS | Millipore (Billerica, MA, USA) |
| MiSeq | Illumina (San Diego, USA) |
| Mithras LB 940 luminometer | *Berthold Technologies* (Bad Wildbad, Germany) |
| Mixers Thermomixer comfort | Eppendorf (Hamburg, Germany) |

| | |
|---|---|
| MyiQ SingleColor Real-Time PCR Detection System | Bio-Rad (Hercules, CA, USA) |
| NanoDrop Spectrophotometer ND-1000 | NanoDrop Technologies (Wilmington, DE, USA) |
| Odyssey Infrared Imager Model 9120 | LI-COR (Bad Homburg, Germany) |
| PMI Personal Molecular Imager FX | Bio-Rad (Hercules, CA, USA) |
| Polymax 2040 | Heidolph (Schwabach, Germany) |
| PowerPac HC Power Supply | Bio-Rad (Hercules, CA, USA) |
| Quantum ST4 | PeqLab (Erlangen, Germany) |
| Screen Eraser-K | Bio-Rad (Hercules, CA, USA) |
| Thermomixer compact | Eppendorf (Hamburg, Germany) |
| Thermostat 5320 | Eppendorf (Hamburg, Germany) |
| Trans-Blot SD Semi-dry transfer cell | Bio-Rad (Hercules, CA, USA) |
| Ultraspec 3300 pro UV/Visible Spectrophotometer | Amersham Biosciences (Little Chalfont, UK) |
| UV Stratalinker 2400 | Stratagene (La Jolla, CA, USA) |
| Vortexer REAX top | Heidolph (Schwabach, Germany) |

## 7.6 Cell lines and bacterial strains

### 7.6.1 Mammalian cell lines

| Cell line | Description |
|---|---|
| HEK 293T | transformed embryonic kidney |
| Hela CCL2 | cervix carcinoma |
| Hela S3 | cervix carcinoma, adapted to suspension culture |
| H1299 | human non-small cell lung carcinoma cell line |
| LNT-229 | glioblastoma cells |
| HCT-116 | colon carcinoma |
| ARPE-19 | retinal pigmented epithelium |
| DG-75 | Burkitt lymphoma |
| HEP-G2 | hepatocellular carcinoma |
| LNCAP | prostate carcinoma |
| MCF-7 | breast adenocarcinoma |
| U2OS | bone marrow carcinoma |
| GM5657 | transformed fibroblasts |
| BJT | transformed fibroblasts |
| MEF wt | mouse embryonic fibroblasts, wild type |
| MEF Ago2 $^{-/-}$ | mouse embryonic fibroblasts with Ago2 deletion (Liu et al., 2004) |

### 7.6.2 Bacteria

| Strain | Genotype |
|---|---|
| XL1-Blue | endA1 gyrA96(nal$^R$) thi-1 recA1 relA1 lac glnV44 F'[ ::Tn10 proAB$^+$ lacI$^q$ Δ(lacZ)M15] hsdR17($r_K^-$ $m_K^+$) |

## 7.7 Cell culture media

For cultivation of cell lines, the following medium was used:

*DMEM complete*

| | |
|---|---|
| 500 ml | DMEM (PAA, Pasching, Austria) |
| 10% | Fetal bovine serum (Sigma-Aldrich) |
| 1% | Penicillin/Streptomycin (PAA, Pasching, Austria) |

*Joklik's medium complete*

| | |
|---|---|
| 11.02 g/L | Minimum essential medium Joklik's modified (Sigma-Aldrich, Steinheim) |
| 2 g/L | NaHCO3 |
| 0.29 g/L | L-Glutamine (Sigma-Aldrich, Steinheim) |
| 1 % (v/v) | MEM non-essential amino acids (Gibco, Langley, OK, USA) |
| 1 % (v/v) | Penicillin/Streptomycin (PAA, Pasching, Austria) |
| 5 % (v/v) | Fetal bovine serum (Sigma-Aldrich) |

adjust pH to 7.0 with NaOH or HCl

OptiMEM (*Invitrogen*)

*R10 culture medium*

| | |
|---|---|
| RPMI 1640 | (PAA, Pasching, Austria) |
| 10 % (v/v) | Fetal bovine serum (Sigma-Aldrich) |
| 1 % (v/v) | Penicillin/Streptomycin (PAA, Pasching, Austria) |

## 7.8 Buffers and solutions

*ACK lysis buffer*

| | |
|---|---|
| 150 mM | $NH_4Cl$ |
| 1 mM | $KHCO_3$ |
| 100 mM | $Na_2EDTA$ |

*5x borate buffer (pH 8.6)*

| | |
|---|---|
| 148 mM | borax |
| 148 mM | boric acid |

*1x borax buffer (pH 9.5)*

| | |
|---|---|
| 30 mM | borax |
| 30 mM | boric acid |
| 50 mM | NaOH |

*Cell lysis buffer 1*

| | |
|---|---|
| 150 mM | KCl |
| 25 mM | Tris pH 7.5 |
| 2 mM | EDTA |
| 1 mM | NaF |
| 0.5 % | NP-40 |
| freshly add (if necessary): | |
| 0.5 mM | DTT |
| 0.5 mM | AEBSF |

*Cell lysis buffer 2 (= IP wash buffer 2)*

| | |
|---|---|
| 600 mM | KCl |
| 25 mM | Tris pH 7.5 |
| 2 mM | EDTA |
| 1 mM | NaF |
| 0.6 % | NP-40 |

*Cell lysis buffer 3 (= IP wash buffer 3)*

| | |
|---|---|
| 300 mM | KCl |
| 25 mM | Tris pH 7.5 |
| 2 mM | EDTA |
| 1 mM | NaF |
| 0.3 % | NP-40 |

*Colloidal Coomassie solution*

| | |
|---|---|
| 8.3% | phosphoric acid |
| 8.3% | $(NH_4)_2SO_4$ |
| 20% | Methanol |
| 1 g/L | Coomassie Brilliant Blue G250 (Sigma-Aldrich) |

*50X Denhardt's Solution*

| | |
|---|---|
| 1% | Albumin fraction V |
| 1% | Polyvinylpyrrolidon K30 |
| 1% | Ficoll 400 |

First add water and then add the ingredients.

*DNA loading dye (5x)*

| | |
|---|---|
| 15 g | Saccharose |
| 50 ml | $H_2O$ |
| 0.025% | Xylene cyanol |

*EDC –Crosslinking (1 membrane)*

| | |
|---|---|
| 2 ml | $H_2O$ |
| 61.25 µl | Methylimidazol |
| 75 µl | HCl (1 M) |
| 188.25 mg | EDC (N-(3-Dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride) |
| up to 6 ml | $H_2O$ |

*Elution buffer*

| | |
|---|---|
| 300 mM | NaCl |
| 2 mM | EDTA |

*Firefly buffer*

| | |
|---|---|
| 470 mM | D-Luciferine (P.J.K. GmbH, Kleinblitterdorf, Germany) |
| 530 mM | ATP (P.J.K. GmbH) |
| 270 mM | Coenzyme A (P.J.K. GmbH) |
| 20 mM | Tricine |
| 5.34 mM | Magnesiumsulfate heptahydrate |
| 0.1 mM | EDTA |
| 33.3 mM | DTT |

*2x HEPES (for calcium phosphate transfection)*

| | |
|---|---|
| 274 mM | NaCl |
| 54.6 mM | Hepes |
| 1.5 mM | $Na_2HPO_4$ |

*Hutner's trace elements*

| | | |
|---|---|---|
| 50 g | EDTA disodium salt | 250 ml |
| 22 g | $ZnSO_4$ x 7 $H_2O$ | 100 ml |
| 11.4 g | $H_3BO_3$ | 200 ml |
| 5.06 g | $MnCl_2$ x 4 $H_2O$ | 50 ml |
| 1.61 g | $CoCl_2$ x 6 $H_2O$ | 50 ml |
| 1.57 g | $CuSO_4$ x 5 $H_2O$ | 50 ml |
| 1.10 g | $(NH_4)_6Mo_7O_{24}$ x 4 $H_2O$ | 50 ml |
| 4.99 g | $FeSO_4$ x 7 $H_2O$ | 50 ml |
| Water to 1 l | | |

This solution was prepared according to a complex protocol (Hutner, 1950). In short, every salt is first solved in water. After mixing all salts, the solution is boiled, cooled down to 70°C and the pH is adjusted to 7.0 with KOH. The solution is incubated another 1-2 weeks and the forming precipitate is removed by filtering.

*Hybridization Solution*

| | |
|---|---|
| 50 ml | 20x SSC |
| 4 ml | 1M $Na_2HPO_4$ pH 7.2 |
| 140 ml | 10% SDS |
| 4 ml | 50X Denhardt's Solution |

*IP wash buffer 1*

| | |
|---|---|
| 300 mM | KCl |
| 50 mM | Tris pH 7.4 |
| 1 mM | $MgCl_2$ |
| 0.1 % | NP-40 |

*LB agar plates*

1.5% agar in LB medium

*LB (lysogenic broth) medium*

| | |
|---|---|
| 1 % (w/v) | Trpyton |
| 1 % (w/v) | NaCl |
| 0.5 % (w/v) | Yeast extract |

*Lysis buffer for genomic DNA preparation from Volvox carteri*

| | |
|---|---|
| 100 mM | NaOH |
| 2 M | NaCl |
| 0.5% | SDS |

*Northern Blot Wash Solution I*

| | |
|---|---|
| 5X | SSC |
| 1% (v/v) | SDS |

*Northern Blot Wash Solution II*

| | |
|---|---|
| 1X | SSC |
| 1% (v/v) | SDS |

*P-IV metal solution*

| | |
|---|---|
| 5.1 mM | EDTA |
| 0.72 mM | $FeCl_3$ x 6 $H_2O$ |
| 0.41 mM | $MnCl_2$ x 4 $H_2O$ |
| 0.073 mM | $ZnCl_2$ |
| 0.017 mM | $CoCl_2$ x 6 $H_2O$ |
| 0.039 mM | $Na_2MoO_4$ |

*Phosphate buffered saline (PBS)*

| | |
|---|---|
| 130 mM | NaCl |
| 774 mM | $Na_2HPO_4$ |
| 226 mM | $NaH_2PO_4$ |

*Phosphate solution for Chlamydomonas medium*

| | |
|---|---|
| 28.8 g | $K_2HPO_4$ |
| 14.4 g | $KH_2PO_4$ |
| Water to 100 ml | |

*Protein sample buffer (5x)*

| | |
|---|---|
| 300 mM | Tris pH 6.8 |
| 50 % | Glycerol |
| 10 % | SDS |
| 0.01 % | Bromophenol blue |

*Proteinase K buffer*

| | |
|---|---|
| 300 mM | NaCl |
| 200 mM | Tris pH 7.5 |
| 25 mM | EDTA |
| 2 % | SDS |

*Proteinase K storage buffer*

| | |
|---|---|
| 20 mg/ml | Proteinase K |
| 50 mM | Tris pH 8.0 |
| 1 mM | $CaCl_2$ |
| 50 % (v/v) | Glycerol |

*Recombinant Ago buffer*

| | |
|---|---|
| 20 mM | Hepes/NaOH pH 7.5 |
| 300 mM | NaCl |
| 1 mM | DTT |
| 2 mM | $MgCl_2$ |

*Renilla buffer*

| | |
|---|---|
| 2.2 mM | EDTA |
| 220 mM | $K_2PO_4$ pH 5.1 |
| 0.44 mg/ml | BSA |
| 1.1 M | NaCl |
| 1.3 mM | $NaN_3$ |
| 1.43 mM | Coelenterazine (P.J.K. GmbH, Kleinblitterdorf, Germany) |

*RNA loading dye (1x)*

| | |
|---|---|
| 99.95 % | Formamide, deionized |
| 0.025 % | Xylene cyanol |
| 0.025 % | Bromophenol blue |

*SDS running buffer (1x)*

| | |
|---|---|
| 200 mM | Glycine |
| 25 mM | Tris pH 7.5 |
| 25 mM | SDS |

*10 % separation gel (SDS-PAGE)*

| | |
|---|---|
| 10 % | Acrylamide-Bis solution (37.5:1, 30 % w/v) (Serva) |
| 400 mM | Tris-HCl pH 8.8 |
| 0.1 % | SDS |
| 0.1 % | APS |
| 0.05 % | TEMED |

*2x siRNA annealing buffer*

| | |
|---|---|
| 60 mM | Hepes pH 7.4 |
| 4 mM | Magnesium acetate |
| 200 mM | Potassium acetate |
| filter sterile | |

*20X SSC (pH7.0)*

| | |
|---|---|
| 3 M | NaCl |
| 0.3 M | Sodium citrate |

*5 % stacking gel (SDS-PAGE)*

| | |
|---|---|
| 5 % | Acrylamide-Bis solution (37.5:1, 30 % w/v) (Serva) |
| 75 mM | Tris-HCl pH 6.8 |
| 0.1 % | SDS |
| 0.1 % | APS |
| 0.05 % | TEMED |

*TAP medium*

| | |
|---|---|
| 2.42 g | Tris |
| 25 ml | TAP salts |
| 0.375 ml | phosphate solution |
| 1 ml | Hutner's trace elements |
| 1 ml | glacial acetic acid |
| Water to 1 l | |

*TAP salts*

| | |
|---|---|
| 15.0 g | $NH_4Cl$ |
| 4.0 g | $MgSO_4$ x 7 $H_2O$ |
| 2.0 g | $CaCl_2$ x 2 $H_2O$ |
| Water to 1 l | |

*TBE buffer (1x)*

| | |
|---|---|
| 89 mM | Tris pH 8.3 |
| 89 mM | Boric acid |
| 2.5 mM | EDTA |

*Towbin buffer (1x, for semi-dry western blotting)*

| | |
|---|---|
| 38,6 mM | Glycine |
| 48 mM | Tris |
| 0.0037 % (w/v) | SDS |
| 20 % | Methanol |

*Translation Mix*

| | |
|---|---|
| 1 mM | ATP |
| 0.2 mM | GTP |
| 10 U/ml | RiboLock (Fermentas) |
| 100 mM | KCl |
| 1.5 mM | $MgCl_2$ |
| 0.5 mM | DTT |

*Tris-buffered saline (1x TBS)*

| | |
|---|---|
| 10 mM | Tris pH 7.5 |
| 150 mM | NaCl |

*Urea buffer for T1 digestion*

| | |
|---|---|
| 10 M | Urea |
| 1.5 mM | EDTA |
| 0.05 % | Bromophenol blue |
| 0.05 % | Xylene cyanol |

*Volvox medium*

| | |
|---|---|
| 0.51 mM (0.12 g/l) | $Ca(NO_3)_2$ x 4 $H_2O$ |
| 0.16 mM (0.04 g/l) | $MgSO_4$ x 7 $H_2O$ |
| 0.67 mM (0.05 g/l) | KCl |
| 0.30 mM (0.064 g/l) | Disodiumglycerophosphate x 6 $H_2O$ |
| 3.78 mM (0.5 g/l) | Glycylglycine |
| 0.15 mM (10 $\mu$l) | Biotin (1.5 mg/100 ml) |
| 0.25 mM (10 $\mu$l) | Vitamin B12 (2.5 mg/100 ml) |
| 3 ml | P-IV metal solution |

Water to 1 l, use 1 N NaOH to adjust to pH 8.0

*Wash buffer (for western blotting)*

| | |
|---|---|
| 10 mM | Tris pH 7.5 |
| 150 mM | NaCl |
| 0.1 % | Tween-20 |

# CHAPTER 8 METHODS

## 8.1 Molecular biological methods

### 8.1.1 General methods

If not otherwise described below, general methods like DNA and RNA gel electrophoresis, PCR, purification and extraction of DNA and RNA etc. were performed according to (Sambrook et al., 1989) or as described in the according manufacturer's manuals.
For DNA isolation from *E.coli*, the NucleoSpin® Plasmid (*Macherey-Nagel*, Düren, Germany) or NucleoBond® XtraMidi-Kit (*Macherey-Nagel*, Düren, Germany) was used. For the elution of DNA fragments from agarose gels, the NucleoSpin® Gel and PCR-cleanup (*Macherey-Nagel*, Düren, Germany) was used.

### 8.1.2 Generation of DNA plasmids

#### *8.1.2.1 pMIR-REPORT and pMIR-RL*

For pMIR-REPORT constructs, sequences complementary to siRNA#1 guide strand, passenger strand and human miR19b were cloned into the 3' UTR of the firefly luciferase of a pMIR-REPORT vector (Ambion). The vector was cut with the restriction enzymes *SacI* and *NaeI*. DNA-oligonucleotides encoding the siRNA/miRNA targeting sites were pre-annealed and 5'-phosphorylated with PNK:

| | |
|---|---|
| pMir-siRNA#1 guide target | 5'-CATCGCCACCTTGTTTAAGCCGCAATTGGAGACTCACTAAATTAGACCTACGCACTCCAGGCC-3' |
| | 5'-GGCCTGGAGTGCGTAGGTCTAATTTAGTGAGTCTCCAATTGCGGCTTAAACAAGGTGGCGATGAGCT-3' |
| pMir-siRNA#1 passenger target | 5'-CATCGCCACCTTGTTTAAGCCTTAGTGAGACACCAATTGCATTAGACCTACGCACTCCAGGCC-3' |
| | 5'-GGCCTGGAGTGCGTAGGTCTAATGCAATTGGTGTCTCACTAAGGCTTAAACAAGGTGGCGATGAGCT-3' |
| pMir-miR-19b target | 5'-CATCGCCACCTTGTTTAAGCCTCAGTTTTGCATCCATTTGCACAATTAGACCTACGCACTCCAGGCC-3' |
| | 5'-GGCCTGGAGTGCGTAGGTCTAATTGTGCAAATGGATGCAAAACTGAGGCTTAAACAAGGTGGCGATGAGCT-3' |

For luciferase-based reporters, the following pre-annealed DNA-oligonucleotides were cloned into pMIR-RL plasmids in a similar way as described for pMIR-REPORT cloning above:

| | |
|---|---|
| siRNA#2 guide strand (gs) on-target reporter | 5'-CATCGCCACCTTGTTTAAGCCGCAGTGAGAGTCCAATTAAATTAGACCTACGCACTCCAGGCC-3' |
| | 5'-GGCCTGGAGTGCGTAGGTCTAATTTAATTGGACTCTCACTGCGGCTTAAACAAGGTGGCGATGAGCT-3' |
| siRNA#2 gs off-target reporter | 5'-CATCCAATTACGCCACCTACTGTTTAAGCCCCAATTAATTAGACCTACTAGCACTCCCCAATTATGGCC-3' |
| | 5'-GGCCATAATTGGGGAGTGCTAGTAGGTCTAATTAATTGGGGCTTAAACAGTAGGTGGCGTAATTGGATGAGCT-3' |

### *8.1.2.2 pSuper*

Expression of shRNAs was achieved by using a modified pSuper plasmid (oligoengine, Seattle, USA). All sequences were cloned as oligonucleotides according to the manufacturer's protocol using the restriction enzymes *BglII* and *HindIII*. All clones were verified by sequencing. The Imp8 shRNA has been reported before (Weinmann et al., 2009).

| **Primer** | **Sequence (5' – 3')** |
|---|---|
| miR-451 original | GATCTCCAAACCGTTACCATTACTGAGTTTAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTAAACTCAGTAATGGTAACGGTTTGGA |
| miR-451 classical | GATCTCCAAACCGTTACCATTACTGAGTTTTCAAGAGAAACTCAGTAATGGTAACGGTTTTTTTTA |
| | AGCTTAAAAAAAACCGTTACCATTACTGAGTTTCTCTTGAAAACTCAGTAATGGTAACGGTTTGGA |
| siGRK4 classical | GATCTCCAACTGCCTGAAGAGACGTCTTTTCAAGAGAAAGACGTCTCTTCAGGCAGTTTTTTTA |
| | AGCTTAAAAAAACTGCCTGAAGAGACGTCTTTCTCTTGAAAAGACGTCTCTTCAGGCAGTTGGA |
| siGRK4 mimic | GATCTCCAACTGCCTGAAGAGACGTCTTCACGTCTCTTCAGGCAGTCTTTTTTA |
| | AGCTTAAAAAAGACTGCCTGAAGAGACGTGAAGACGTCTCTTCAGGCAGTTGGA |
| siImp8 classical | ACAGAGATCTCCTTAGTGAGAGTCCAATTAATTCAAGAGATTAATTGGACTCTCACTAATTTTTAAGCTTACAG |
| | CTGTAAGCTTAAAAATTAGTGAGAGTCCAATTAATCTCTTGAATTAATTGGACTCTCACTAAGGAGATCTCTGT |
| siImp8 mimic | GATCTCCTTAGTGAGAGTCCAATTAACCATAATTGGACTCTCACTACTTTTTTA |
| | AGCTTAAAAAAGTAGTGAGAGTCCAATTATGGTTAATTGGACTCTCACTAAGGA |
| mutant #1 | GATCTCCTAACCGTTACCATTACTGAGTTTAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTAAACTCAGTAATGGTAACGGTTAGGA |
| mutant #2 | GATCTCCAAACCGTAACCATTACTGAGTTTAGTAATGGTTATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATAACCATTACTAAACTCAGTAATGGTTACGGTTTGGA |
| mutant #3 | GATCTCCAAACCGTTACCAATACTGAGTTTAGTATTGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCAATACTAAACTCAGTATTGGTAACGGTTTGGA |
| mutant #4 | GATCTCCAAACCGTTAGCATTACTGAGTTTAGTAATGCTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTAGCATTACTAAACTCAGTAATGCTAACGGTTTGGA |
| mutant #5 | GATCTCCAAACCGTTACGATTACTGAGTTTAGTAATCGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACGATTACTAAACTCAGTAATCGTAACGGTTTGGA |
| miR-451 loop +1nt | GATCTCCAAACCGTTACCATTACTGAGTTCTAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTAGAACTCAGTAATGGTAACGGTTTGGA |
| miR-451 loop +2nt | GATCTCCAAACCGTTACCATTACTGAGTTCGTAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTACGAACTCAGTAATGGTAACGGTTTGGA |
| miR-451 loop +3nt | GATCTCCAAACCGTTACCATTACTGAGTTCGGTAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTACCGAACTCAGTAATGGTAACGGTTTGGA |
| miR-451 loop +4nt | GATCTCCAAACCGTTACCATTACTGAGTTCGGGTAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTACCCGAACTCAGTAATGGTAACGGTTTGGA |
| miR-451 loop +5nt | GATCTCCAAACCGTTACCATTACTGAGTTCGGGGTAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTACCCCGAACTCAGTAATGGTAACGGTTTGGA |
| 1st wobble | GATCTCCAAACCGTTACCATTACTGAGTTTAGTAATGGTAACGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCGTTACCATTACTAAACTCAGTAATGGTAACGGTTTGGA |
| 2nd wobble | GATCTCCAAACCGTTACCATTACTGAGTTCAGTAATGGTAATGGTTCTTTTTTA |
| | AGCTTAAAAAAGAACCATTACCATTACTGAACTCAGTAATGGTAACGGTTTGGA |

| Primer | Sequence (5‘ – 3‘) |
|---|---|
| let-7c genomic | GATCTCCTGAGGTAGTAGGTTGTATGGTTTAGAGTTACACCCTGGGAGTTAACTGTACAACCTTCTAGCTTTTTTTTA |
| | AGCTTAAAAAAAAGCTAGAAGGTTGTACAGTTAACTCCCAGGGTGTAACTCTAAACCATACAACCTACTACCTCAGGA |
| let-7c classical | GATCTCCTGAGGTAGTAGGTTGTATGGTTTTCAAGAGAAACCATACAACCTACTACCTCATTTTTA |
| | AGCTTAAAAATGAGGTAGTAGGTTGTATGGTTTCTCTTGAAAACCATACAACCTACTACCTCAGGA |
| let-7c mimic | GATCTCCTGAGGTAGTAGGTTGTATGGTTGTACAACCTACTACCTCATTTTTA |
| | AGCTTAAAAATGAGGTAGTAGGTTGTACAACCATACAACCTACTACCTCAGGA |

### *8.1.2.3 Cloning of an expression plasmid for use in Volvox carteri*

For the generation of an expression construct for *Volvox carteri* proteins, the backbone of the pPmr3 vector (Jakobiak et al., 2004) was used. The Flag-tagged *aphH* gene was excised by *BamHI* and *PstI* and an alternative multiple cloning site (MCS) with an N-terminal 6x myc-tag was inserted. The myc-tag was amplified from the pCS2 plasmid and the MCS was added using several DNA oligonucleotides:

| Primer | Sequence (5‘ – 3‘) |
|---|---|
| MCS primer 1 (*forward*) | CGCTGGATCCATTTAAAGCTATGGAGCAAAAGC |
| MCS primer 2 (*reverse*) | CCGGCCGGCGCGCCTCCGGAGACGTCCGGCCTTGAATTCAAGTCCTCTTCAG |
| MCS primer 3 (*reverse*) | CGCACTGCAGGCATGCCCCGGGCCATGGGGCCGGCCGGCGCGCCTCCGGAGACGTCCGGCCTTGAATT |

Primers 1 and 2 were used for the amplification of the 6x myc tag of pCS2 by PCR. 50 ng of the resulting product were subjected to a second PCR using primer 1 as forward and primer 3 as reverse primer. The ends of this product carry restriction sites for *BamHI* (5’ end) and *PstI* (3’ end) and could therefore be used to clone this fragment into pPmr3 backbone (Figure 8-1).

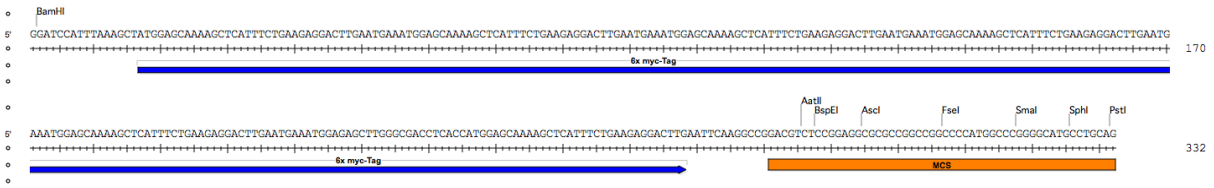



**Figure 8-1 Fragment containing 6x myc-tag and MCS.** The 6x myc-tag was amplified from the pCS2 plasmid. Further sequences were added by PCR using long DNA oligonucleotides. Unique restriction enzymes are labeled above the sequence.

### *8.1.2.4 Isolation of the cDNA of AGO3*

The amplification of the cDNA of AGO3 could not be accomplished in a single PCR run, probably due to the length of the gene and the general high GC content of the *Volvox carteri* genome. The CDS was therefore split in seven pieces. All fragments but the 5’ end could be amplified successfully. The 5’ part was constructed with the use of synthesized DNA oligonucleotides. The different fragments were designed in a way to contain overlaps that could be used to join the fragments into the full CDS. A final

amplification step by PCR added the restriction sites necessary to transfer the AGO3 CDS into the modified expression vector (see 8.1.2.3).

### 8.1.3 RNA Extraction from tissue samples or cultured cells

#### *8.1.3.1 Extraction of RNA from mouse organs*

The mouse strain C57BL/6J was used for experiments with mouse tissues. For isolation of RNA, liver, brain, kidneys and spleen were shock frozen in liquid nitrogen and pulverized in a mortar with a pestle. The crushed tissue was lysed immediately in peqGOLD TriFast (peqlab), rigorously mixed and incubated 5-10 min at room temperature. The RNA was isolated according to the manufacturer's manual. After precipitation over night at -20°C, the RNA was pelleted for 30 minutes, 17,000*g* and 4°C. After washing with 1 ml ice-cold ethanol (80%), the RNA was centrifuged again for 30 min, 17,000*g* and 4°C. Pellets were dried a short time at room temperature, suspended in $ddH_2O$ and solved at 70°C for 5 minutes while shaking (1000 rpm). RNA was stored at -80°C.

#### *8.1.3.2 Extraction of RNA from cultured cells*

Cells were harvested by scraping off, trypsinization or centrifugation. They were subsequently washed with cold PBS and immediately lysed in peqGOLD TriFast (peqlab) or Trizol (Life Technologies) reagent following the manufacturer's instruction. After incubation at room temperature for 5 to 10 min, 200 µl of chloroform per milliliter of reagent were added, the components were mixed and phases were separated by centrifugation (10-15 min, 12,000*g*, 4°C). The aqueous phase containing the RNA was precipitated by the addition of isopropanol to a final concentration of 45% (v/v) over night at -20°C. When expecting a low amount of RNA, 20 µg of glycogen (RNA grade, Fermentas) per milliliter of initial reagent were added. The RNA was recovered by centrifugation (17,000*g*, 4°C, 30 min), followed by a washing step (1 ml ice-cold 80% EtOH). After another centrifugation step (17,000*g*, 4°C, 15 min), the RNA pellet was air-dried for 2-3 min and solved in ultra pure water (5 min, 70°C, shaking at 1000 rpm). RNA was stored at -80°C.

### 8.1.4 Mass spectrometry

Samples were separated by 10% denaturing SDS-PAGE. For mass spectrometry analysis, the gels were stained afterward with colloidal Coomassie solution. Bands of interest were cut out from the gel and analyzed by the mass spectrometry group of the Microchemistry Core Facility, Max-Planck-Institute of Biochemistry (Martinsried, Germany). All intensities of peptides belonging to an Ago in its respective IP sample were added, and the sums of the peptide intensities were compared among the

different Ago-IPs in order to roughly calculate the relative amounts of the individual Ago proteins (Petri et al., 2011).

### 8.1.5 Preparation of single-cell suspensions from mouse organs

For preparation of single cell suspensions, livers and spleens of two mice were isolated. The livers were pressed through a cell strainer into a petri dish containing PBS. Remaining tissue was discarded. The spleens were crushed between the roughened surfaces of two microscope slides. The cells were washed off the slides with PBS and subsequently pipetted through a cell strainer (BD Falcon, REF 352350, 70 $\mu$m nylon). Cells in both suspensions were collected by centrifugation at 200*g*, 4°C for 5 min. To remove contaminating erythrocytes, cells were resuspended in 10 to 20 ml of ACK lysis buffer (150 mM $NH_4Cl$, 1 mM $KHCO_3$, 100 mM $Na_2EDTA$) and incubated for 5 minutes at room temperature. Remaining liver and spleen cells, respectively, were collected again by centrifugation (200*g*, 5 min, 4°C). Lysed (red) erythrocytes remained in the supernatant and were discarded. White cell pellets were washed once with cold PBS, before cells were lysed in cell lysis buffer 1 (25 mM Tris/HCl pH 7.4, 150 mM KCl, 0.5% (v/v) NP-40, 2 mM EDTA, 1 mM NaF) for 20 min on ice. Cell debris was removed by centrifugation at 17,000*g*, 4°C for 45 min. The supernatant was shock frozen with liquid nitrogen and stored at -80°C.

### 8.1.6 Ago complex purification

Monoclonal antibodies were coupled to Protein G Sepharose (GE Healthcare) beads over night at 4°C. Antibody-decorated beads were washed twice with lysis buffer.
Different lysis buffers were used depending on the experiment:
For IPs from mouse tissue, lysates from liver and spleen of two mice were applied (Cell lysis buffer 1, Chapter 3). For IPs from transfected HEK 293T, cells were grown on 15 cm plates to full density, washed with cold PBS and lysed in 1 ml of Cell lysis buffer 1 per plate (Chapter 3).
For analyzing the cellular amounts of Ago proteins in HEK 293T and HeLa S3, cells were lysed on ice in 5 volumes / weight (ml/g) Cell lysis buffer 2 (Chapter 2). For the immune-precipitation of Ago proteins in order to co-precipitate and analyze siRNA and miRNA, cells were lysed in 8 volumes / weight (ml/g) Cell lysis buffer 3 (Chapter 2).
The lysates were cleared by centrifugation at 17,000 *g* for 30 minutes, before addition to the beads. As a control for the IPs, beads without coupled antibody or coupled with an unspecific antibody (anti-RmC) were used. After 4 h of rotation at 4°C, the beads were washed four times with IP wash buffer 1 (Chapter 3), IP wash buffer 2 (Chapter 2, determination of cellular amounts) or IP wash buffer 3 (Chapter 2, analysis of siRNAs and miRNAs) and once with cold PBS.
For RNA extraction, the immunoprecipitates were treated with Proteinase K (40 $\mu$g per sample, in 200 $\mu$l 300 mM NaCl, 25 mM EDTA, 2% SDS, 200 mM Tris pH 7.5) at 65°C for 15 minutes. The RNA

was extracted with 200 µl of acidic phenol and precipitated with 2.5 volumes pure ethanol and 20 µg glycogen (Fermentas) at -20°C overnight. For input samples, 100 µl of the respective lysate was denatured by addition of 1 ml of TriFast (Peqlab) or Trizol (Life Technologies) reagent. The RNA was subsequently extracted according the manufacturer's instructions. After pelleting the RNA, it was washed once with cold 80% ethanol, dried and solved in $H_2O$ (see 8.1.3.2).

### 8.1.7 Ago cleavage reaction

The non-5'-phosphorylated RNA strands (guide 5'-UCGAAGUAUUCCGCGUACGUdT-3', passenger 5'-CGUACGCGGAAUACUUCGAUdT-3') were phosphorylated in a T4 polynucleotide kinase (T4 PNK; Fermentas) reaction. 20 pmol each were labeled either with 20 µCi of $\gamma$-$^{32}$P-ATP (passenger strand) or 1 mM of cold ATP (guide strand) in a 20 µl reaction. The RNAs were gel-purified and mixed in equimolar amounts. For annealing, the mixture was heated at 95°C for 3 min, cooled first at 37°C and then on ice.

For passenger strand cleavage reaction, 5 pmol of recombinant Argonaute protein was incubated with 2 Bq/cm$^2$ of annealed siRNA in a 20 µl reaction (20 mM Hepes/KOH pH 7.5, 3 µg yeast RNA, 3 mM $MgCl_2$, 10% (v/v) glycerol, 1 mM DTT, 0.1 mg/ml BSA). For target cleavage, the guide strand was preincubated for 30 min at 30°C before the passenger strand was added. In both cases, cleavage was performed for 1 h at 30°C and stopped by Proteinase K digestion (0.4 µg/µl Proteinase K, 300 mM NaCl, 25 mM EDTA, 2% SDS, 200 mM Tris pH 7.5). The RNA was extracted using phenol/chloroform/isoamylalcohol (25:24:1, Roth) and precipitated over night with 20 µg of glycogen RNA grade (Fermentas). The resulting pellet was washed once with ice-cold 80% ethanol, solved in RNAse-free water and loaded onto an 18% urea gel (UreaGel System, National diagnostics). For visualization, the gels were wrapped in saran, exposed to a screen and scanned with the PMI (Bio-Rad).

To assess the purity of the recombinant proteins, 2 µg of each protein was loaded onto a NuPAGE® 4-12% Bis-Tris Gel (Invitrogen, Life technologies) and visualized with Coomassie.

### 8.1.8 RNA polyacrylamide gel electrophoresis

RNA polyacrylamide gel electrophoresis was performed using the UreaGel System (National diagnostics). Components were mixed according to the chosen percentage and the manufacturer's protocol and gels were polymerized for at least 1 h. Before loading the samples, the gels were pre-run at 250-350 V for about 15-30 min using 1x TBE buffer, until gels were hand-warm. The wells were rinsed several times to get rid of diffusing urea. Samples were heated at 95°C for 2-3 min and loaded into the freshly rinsed wells. Depending on the samples, a voltage between 250-500 V was chosen in a way that the gel would not get hot. Disassembled gels were subjected to an ethidium bromide stain (5 µl EtBr in 100 ml 1x TBE) and/or to Northern blotting (see 8.1.9).

### 8.1.9 Northern blotting

Northern blotting was performed as described before (Dueck et al., 2012; Pall and Hamilton, 2008). In short, either 10-20 $\mu$g of total RNA or all of immunoprecipitated RNA were mixed with equal amounts of RNA loading dye, heated at 95°C for 2-3 min and separated on 12% or 18% urea gels (UreaGel System, National diagnostics). 18% gels were used for single nucleotide resolution of small RNAs the size 15-30 nt (siRNAs and miRNAs). As a size marker, ribooligonucleotides with a length of 19, 21 and 24 nucleotides (nt) were labeled with $^{32}$P prior to loading. If a reference oligo was loaded, 3 pmol of the respective siRNA strand was used. After running the gel for 1 h (12%) or 3-4 h (18%) at 350-450 V, the RNA was semi-dry blotted onto an Amersham Hybond-N membrane (GE Healthcare) at 20 V for 30 minutes. The membrane was subsequently cross-linked using EDC-solution and incubated for 1 h at 50°C. The membrane was rinsed in water and subjected to pre-hybridization with hybridization solution. For labeling of a probe antisense to the respective miRNA or siRNA, 20 pmol of a DNA oligonucleotide were incubated with 20 $\mu$Ci of $\gamma^{32}$P-ATP in a T4 PNK reaction (Fermentas) for 30-60 min at 37°C. The reaction was stopped with 30$\mu$l of 30 mM EDTA. The labeled oligonucleotide was purified with an illustra MicroSpin G-25 Column (GE Healthcare), added to the membrane and together were incubated over night at 50°C. The membrane was washed twice with 5xSSC, 1% SDS, once with 1xSSC, 1% SDS and wrapped in saran. Signals were detected by either exposure to a screen and scanning with the PMI (Bio-Rad) or exposure to BioMax MS films (Kodak) using an intensifying screen (GE Healthcare). Signals were quantified using the Quantity One software (Bio-Rad).

| Detection of | Sequence probe (5' -> 3') |
|---|---|
| let-7a (human) | AACTATACAACCTACTACCTCA |
| let-7c (human) | AACCATACAACCTACTACCTCA |
| miR-19b (human and mouse) | TCAGTTTTGCATCCATTTGCACA |
| miR-21 (human) | TCAACATCAGTCTGATAAGCTA |
| miR-23a (human) | GGAAATCCCTGGCAATGTGAT |
| miR-27a (human) | GCGGAACTTAGCCACTGTGAA |
| miR-30a (human) | CTTCCAGTCGAGGATGTTTACA |
| miR-92a (human) | ACAGGCCGGGACAAGTGCAATA |
| miR-99b-5p (mouse) | CGCAAGGTCGGTTCTACGGGTG |
| miR-143-3p (mouse) | GAGCTACAGTGCTTCATCTCA |
| miR-147-3p (mouse) | TAGCAGAAGCATTTCCGCACAC |
| miR-155-5p (mouse) | ACCCCTATCACAATTAGCATTAA |
| miR-181a-5p (mouse) | ACTCACCGACAGCGTTGAATGTT |
| miR-182 (human) | TGTGAGTTCTACCATTGCCAAA |
| miR-210-3p (mouse) | TCAGCCGCTGTCACACGCACAG |
| miR-451 (human and mouse) | AACTCAGTAATGGTAACGGTTT |
| miR-455-3p (mouse) | GTGTATATGCCCGTGGACTGC |
| miR-455-5p (mouse) | CGATGTAGTCCAAAGGCACATA |

| Detection of | Sequence probe (5' -> 3') |
|---|---|
| siGRK4 guide | AAGACGTCTCTTCAGGCAGTT |
| siImportin8 | TTAATTGGACTCTCACTAA |
| siRNA#1 guide | TTAGTGAGAGTCCAATTGCTT |
| siRNA#1 passenger | GCAATTGGACTCTCACTAATT |
| U6 (human and mouse) | GAATTTGCGTGTCATCCTTGCGCAGGGGCCATGCTAA |
| U6 (*Volvox*) | TCCAATTTTAACAGTTGTGCCGAAGCAC |
| Cluster_884 (*Volvox*) | GGTGCAGGGCATTCTCGCCGAA |
| Cluster_5285 (*Volvox*) | AATGGGCCCCAAACGGTCGGTA |
| Cluster_2909 (*Volvox*) | CCCCAGTGGATTAAGCCCTTCA |
| Cluster_1049 (*Volvox*) | AGTTTTCCAAAGCTCACAGCCCA |
| Cluster_5322 (*Volvox*) | GTTACCGACCGTTTGGGGCCCA |
| sRNA Scaffold 8 (*Volvox*) | CGGCTGCCAAGACGACCTCGA |
| sRNA Scaffold 12 (*Volvox*) | GGTGCTCAAGGCTGCTATCCAA |
| sRNA Scaffold 47 (*Volvox*) | AGTCCTTCTTATGCTCGTCCCA |
| Cre sRNA 601 (*Chlamydomonas*) | TGACAATAGTAGTACGGCCA |
| Cre sRNA 2726 (*Chlamydomonas*) | CACCTGGTCCCGCTACCTGAA |

#### *8.1.9.1 Stripping of Northern blot membranes*

In order to be able to detect more than one small RNA on a Northern blot, the membranes were stripped of the current probe. Water was boiled (typically 500 ml for one membrane), then an SDS solution was carefully added to reach a final concentration of 0.1% SDS (5 ml of 10% SDS). The membrane was rolled up with the RNA facing the inside and inserted into the bottle with hot water. After an incubation of at least 30 min on a rocker, the membrane was wrapped in saran and exposed to a screen for at least over night to control the removal of the probe.

### 8.1.10 Reverse transcription and mRNA detection by quantitative real-time PCR (qRT-PCR)

For measurement of siRNA knockdown efficiencies in murine cells by quantitative real-time PCR (qRT-PCR), cDNA synthesis was carried out using 1 $\mu$g of total RNA with the First Strand cDNA Synthesis Kit (Fermentas), according to the manufacturer's instructions. The transcribed cDNA was diluted 1:10 and 5$\mu$l of this dilution was mixed with 2 pmol of forward and reverse primer and 10 $\mu$l of 2x MESA Green qPCR MasterMix Plus (Eurogentec) in a total volume of 20 $\mu$l. Samples were analyzed on a Bio-Rad MyiQ analyzer.

Primers used:

GRK4 forward 5' AGGAGGAGAACCCTTCCAAA,
reverse 5' TTTCCAGCCATTTCCATTGT;

| | |
|---|---|
| GAPDH | forward 5' AATGGAAATCCCATCACCATCT, |
| | reverse 5' CGCCCCACTTGATTTTGG. |
| C/EBPbeta | forward 5' ATCGACTTCAGCCCCTACCT, |
| | reverse 5' GGCTCACGTAACCGTAGTCG; |
| 36B4 | forward 5' ATGTGCAGCTGATAAAGACTGG, |
| | reverse 5' AGGCCTTGACCTTTTCAGTAAG. |

For measurement of reporter mRNA by qRT-PCR, HEK 293T were transfected using calcium phosphate, as mentioned above. The cells were harvested 22–24 h after transfection and were lysed as described under 8.1.6. After IP of Ago proteins and RNA precipitation, the RNA was washed once with ice-cold 80% ethanol and solved in 25 µl water for 5 min at 70°C. Seven microliters of RNA was subjected to *DNase I* (Fermentas) digestion for 30 min at 37°C. 2 µl of the RNA was transcribed into cDNA with a First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer's instructions using Random Hexamer Primer. The resulting cDNA was analyzed by qRT-PCR. It was diluted 1:10; 5 µl per sample of this dilution was mixed with 2 pmol each of forward and reverse primer and 10 µl of 2x MESA Green qPCR MasterMix Plus (Eurogentec) in a total volume of 20 µl. We used the following primers directed against firefly-luciferase mRNA: forward, 5'-GTGTTCGTCTTCGTCCCAGT-3'; reverse, 5'-GCTGGGCGTTAATCAGAGAG-3'. Samples were analyzed on a Bio-Rad MyiQ analyzer.
For measuring gene expression in *Volvox carteri*, cDNA was produced in slightly different way. 1 µg of total RNA was reverse transcribed using 4 µM of each gene specific primer and 1.5 µl of Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. To control for DNA contamination, a reaction without the reverse transcriptase was performed in parallel. The cDNA was diluted 1:20 and measured using the SsoFast™ EvaGreen® Supermix (Bio-Rad). For this, 8 µl diluted cDNA was mixed with 4 pmol of forward and reverse primer and 10 µl of 2x mix in a 20 µl reaction. The measurement was performed in triplicate on a C1000 Thermal Cycler with a CFX96 Real-Time System (Bio-Rad). Primers for ActA, ssgA and regA have been described before (Nematollahi et al., 2006).

| | |
|---|---|
| ActA | forward 5' TGAGAAGACGTACGAGCTGC, |
| | reverse 5' CCTCCATGCCGATTAGGCTA; |
| ssgA | forward 5' TTCGCATCGTGAAGGACCTT, |
| | reverse 5' CCGTTAACGTCCATGAACAG; |
| regA | forward 5' CAATGGCAGCAAATGGATGTC, |
| | reverse 5' GTTCCAAATCAGGCAACACG; |
| transcript22037 | forward 5' GCTTTCCTCTGTCGGCTGAT, |
| | reverse 5' ATCCTTCACATCACGGCTCC; |
| AGO3 | forward 5' CTAAGCCCCAGCGTCTAGTG, |
| | reverse 5' CACAATGAACGTGATGGCGG; |

### 8.1.11 SDS-PAGE and Western blotting

To assess the purity of the recombinant proteins, 2 µg of each protein was loaded onto a NuPAGE® 4-12% Bis-Tris Gel (Invitrogen, Life technologies) and visualized with Coomassie.
For performing a Western blot, samples were shortly heated at 95°C and then loaded onto a 10% SDS-PAGE gel. After separation, the proteins were blotted onto an Amersham Hybond-ECL membrane (GE Healthcare) using Towbin buffer for 2 h at 1.5 mA/cm$^2$. The membrane was blocked in TBS containing 0.1% Tween20 and 5% milk powder and subsequently incubated with the first antibody over night at 4°C. After three washing steps with TBS containing 0.1% Tween 20, the secondary antibody was added for 1 h at RT. After a second washing step (as before), the membrane was scanned on a LI-COR reader.

## 8.2 Cellular biological methods

### 8.2.1 Culture and transformation of bacterial cells

*E.coli* strain XL1-Blue (Agilent Technologies, Böblingen) was cultured under standard conditions (37°C in LB medium). The genotype of XL1-Blue is endA1 gyrA96(nal$^R$) thi-1 recA1 relA1 lac glnV44 F'[ ::Tn10 proAB$^+$ lacI$^q$ Δ(lacZ)M15] hsdR17($r_K^-$ $m_K^+$) which renders it a strain to use for DNA amplification. Transformations were performed with chemically competent XL1-Blue cells. For this, one aliquot of cells was mixed with 100 ng of DNA or a ligation reaction and incubated on ice for 20 min. Heat shock was performed for 1 min at 42°C followed by a cool down on ice for 1 min. One milliliter of LB medium without antibiotics was added and cells were incubated at 37°C for 15-30 min while shaking. Depending on the type of experiment, 10 µl (re-transformation of pure DNA) or 100-900µl of the culture (ligation reaction) were plated onto LB agar plates supplemented with the suitable antibiotic (100 µg/mL ampicillin or 30 µg/mL kanamycin).

### 8.2.2 Culture of mammalian cells

Most cell lines mentioned in this thesis were cultivated under standard conditions: Dulbecco's modified Eagle's medium (DMEM; PAA) supplemented with 10% fetal bovine serum (GIBCO), 100 U/mL penicillin, and 100 mg/mL streptomycin (GIBCO) at 37°C in 5% $CO_2$ atmosphere. The Ago2$^{-/-}$ MEFs (Liu et al., 2004) as well as the Hela S3 cell line stably expressing FH-tagged Ago4 have been described before (Meister et al., 2004).
For passaging the cells, the medium was removed and cells washed once with PBS. Trypsin was added and incubated varying on the particular cell line. Part of the suspension of seeded onto a new dish. Cells were passaged usually every 2-3 days.

Primary murine dendritic cells and macrophages were kept in R10 at 37°C in 5% $CO_2$ atmosphere.
For growing HeLaS3 cells in suspension culture, 10x15 cm plates of adherent cells were detached by adding trypsin, pelleted by centrifugation (200 *g*, 4°C, 5 min) and used to inoculate 150 ml Joklik's medium contained in a spinner flask (Corning). Cells were cultivated while rotating at 37°C with closed lids. After cells had adapted to suspension culture, they were passaged every 2-3 days by discarding part of the culture and adding fresh medium.

### 8.2.3 Transfection of HEK 293T cells with siRNA and plasmids

Cells were transfected using calcium phosphate mediated transfection.
For transfection with plasmids for shRNA production, cells were plated and transfected 5 to 6 h later with 5 $\mu$g or 15 $\mu$g DNA for 10 cm and 15 cm plates, respectively. Cells were incubated for 40 h and subsequently harvested.
For transfection of plasmids and siRNAs, cells were plated about 15 h prior to transfection at 20% confluency. For all transfections of either pure siRNA or plasmid to a cell culture dish with a diameter of 15 cm the following mix was added: 37,5 µl of a 20 µM siRNA stock solution or 10 µg of plasmid DNA was diluted in 645 µl $H_2O$ and 91,5 µl 2M $CaCl_2$. To this mixture was added 750 µl of 2× HEPES-buffered saline (274 mM NaCl, 1.5 mM $Na_2HPO_4$, 54.6 mM HEPES-KOH pH 7.1).
For transfections of siRNA and pMIR-REPORT plasmid at the same time, the following mix was used: 7.5 $\mu$g plasmid and 25 $\mu$l of 20 $\mu$M of siRNA #1 were diluted in 860 µl $H_2O$ and 122 µl 2 M $CaCl_2$. To this mixture was added 1000 µl of 2× HEPES-buffered saline.
The transfection mix was incubated for 10 min and was then added drop-wise onto the cells.
siRNA#1, guide strand 5'-UUAGUGAGAGUCCAAUUGCUT-3' and passenger strand 5'-GCAAUUGGACUCUCACUAAUT-3',
siunbias (control) guide strand 5'-UUGUCUUGCAUUCGACUAAUT-3' and passenger strand 5'-UUAGUCGAAUGCAAGACAAUT-3'.

### 8.2.4 Transfection of H1299 cells

H1299 cells were transfected using 2 $\mu$g DNA and 5 $\mu$l Lipofectamine 2000 (Invitrogen) per 6well according to the manufacturer's manual. For longer incubation periods, cells were expanded when necessary.

### 8.2.5 Mouse strains/Preparation of primary cells

Wildtype (C57BL/6) and mir-155$^{-/-}$ (Thai et al., 2007), B6.Cg-Mir155tm1.1Rsky/J, JAX) mice used in this study were bred and maintained in a conventional animal facility according to local regulations, and sacrificed at 8 to 12 weeks of age for use in experiments.

DCs were generated as described previously (Lutz et al., 1999). In brief, flushed bone marrow from femur and tibia was cultured in petri dishes (∅ = 10 cm) and R10 culture medium (RPMI 1640 supplemented with 10 % fetal calf serum, L-Glutamine and Penicillin/Streptomycin, all from Life Sciences) supplemented with GM-CSF hybridoma supernatant. At day 8-9 of culture, 5x10$^5$ DCs were transferred to a cell culture dish (∅ = 6 cm) and matured by overnight stimulation with either 200 ng/ml LPS (Sigma-Aldrich), 4 $\mu$g/ml LDL, oxLDL or eLDL. LDL and derivatives were a kind gift from Prof. Gerd Schmitz (University Clinic Regensburg, Germany). Macrophages were differentiated according to the DC culture but using M-CSF hybridoma supernatant instead of GM-CSF (Rutherford and Schook, 1992). At day 8-9 of macrophage culture, confluent petri dishes were stimulated such as the DCs. For each condition, one sample was prepared for both deep sequencing and validation by Northern blotting.

### 8.2.6 Luciferase reporter assays

On- or off-target dual luciferase pMIR-RL reporter plasmids were co-transfected with siRNA reversely into HEK 293T cells, wt MEFs, or Ago2$^{-/-}$ MEFs (Liu et al., 2004) by the use of Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. SiRNA was co-transfected at indicated concentrations together with 10 ng of reporter plasmid into HEK 293T cells or together with 50 ng of reporter plasmid into wt MEFs or Ago2$^{-/-}$ MEFs per well of a 96-well plate. Each sample was transfected as a technical triplicate per experiment. Cells were lysed in Passive Lysis Buffer (Promega) at indicated time points after transfection. Firefly and Renilla luciferase activities were measured essentially as described before (Hock et al., 2007), but as separate samples. Renilla luciferase signals were used to normalize firefly luciferase signals; siRNA effects were compared to the effects of an unrelated control siRNA.

## 8.3 Culture and treatment of *Chlamydomonas reinhardtii* and *Volvox carteri*

If not specified in this section, methods and protocols above were used for the preparation of samples from *Volvox carteri*.

### 8.3.1 Culture of *Chlamydomonas reinhardtii*

The cell-wall deficient strain CW15 mt- was obtained from Jörg Nickelsen (University of Munich) and cultivated on agar plates from TAP medium at room temperature on a shelf exposed to natural sunlight. TAP medium and its ingredients were made as described at chlamy.org, originally described in Gorman et al. (Gorman and Levine, 1965). Liquid cultures were grown in TAP medium in Erlmeyer flasks at room temperature and exposed to natural sunlight.

### 8.3.2 Strains and culture conditions of *Volvox carteri*

The female *Volvox carteri f. nagariensis* wild-type strain HK10 was originally obtained from R.C. Starr (Culture Collection of Algae, University of Texas, Austin, Texas, USA). The strain Vol6♀ was obtained from A. Hallmann (University of Bielefeld, Germany). Synchronous cultures were grown in *Volvox* medium (Provasoli and Pintner, 1959) at 28°C under an 8 h dark/16 h light (10,000 lux) cycle (Starr and Jaenicke, 1974). The sex-inducing pheromone was used as described by Haas and Sumper (Haas and Sumper, 1991).

### 8.3.3 Separation of cell types

About 5000 asexual growing *Volvox* spheroids at the stage shortly before the onset of embryogenesis were collected by filtration on a 100 $\mu$m mesh nylon screen and broken by passing them through a 0.6 mm hypodermic needle. The suspension was filtered on a 100 $\mu$m mesh nylon screen, which allows free reproductive cells (gonidia), free somatic cells and small fragments of the somatic cell layer to pass but retaining larger fragments of somatic cell layers nearly free from gonidia. The filtrate was passed in a second step through a 40 $\mu$m mesh nylon screen which retained only gonidia and small fragments of the somatic cell layer. The resuspended residue was allowed to settle down repeatedly in a small volume separating the reproductive cells from somatic cell layers.

The separation of gonidia and somatic cell layers of sexual induced *Volvox* spheroids at the stage shortly before the onset of embryogenesis and 16 h after application of the sex-inducing pheromone was done in the same manner as described above.

The separation of egg cells and somatic cells of sexual growing *Volvox* spheroids 64 h after application of the sex-inducing pheromone was done in a modified procedure: Egg cells were set free by passing the spheroids twice through a 0.5 mm hypodermic needle. A first filtration step on a 40 $\mu$m mesh nylon screen retained large fragments of the somatic cell layer. Egg cells which pass through were collected on a 10 $\mu$m mesh nylon screen and purified further by letting them settle down repeatedly in a small volume. The residue of the first filtration step was dissociated once more by passing through a 0.4 mm hypodermic needle. Fragments of the somatic cell layer were collected on a 40 $\mu$m mesh nylon screen.

### 8.3.4 Production of stable *Volvox* strains using the gold particle gun

#### *8.3.4.1 Precipitation of DNA (plasmids) onto gold-microcarriers*

For each transformation, 10 µg of plasmid DNA encoding the selection marker (pPmr3, see (Jakobiak et al., 2004)) and 10 µg plasmid encoding the target gene were used. The DNA solution was added to 0.5 µmol gold particles (Bio-Rad) while mixing vigorously. Continuing the shaking, 125 µmol $CaCl_2$ (50 µl of 2.5 M) and 2 µmol spermidin (20 µl of 100 mM) were added. The mixture was further incubated at 4°C for 30 min under continuous shaking. To precipitate the DNA onto the gold, 200 µl of 100% ethanol were added, followed by a short centrifugation step (3-4 s at 8000 rpm). The gold particles were washed three times with ice cold 100% ethanol and taken up in a total volume of 40 µl. The particles were stored on ice until transformation.

#### *8.3.4.2 Nuclear transformation*

Nuclear transformation was carried out according to Jakobiak et al., 2004 (Jakobiak et al., 2004). One aliquot of gold particles (see 8.3.4.1) were used to transform *Volvox* spheroids from one Fernbach flask. For this, the gold particle mixture was spread over the center of 6 macrocarriers (fixed in their metal carriers). The macrocarriers were warmed on a heating block at 37°C to evaporate residual ethanol. One macrocarrier, one stopping screen and a rupture disk (900 psi) were placed into the apparatus. The *Volvox* spheroids were harvested using a sieve and spread in its center. The sieve was placed directly below the macrocarriers. Transformation was carried out under vacuum using the Biolistic PDS 1000/He particle gun (Bio-Rad Laboratories, Hercules, USA). For one transformation, the *Volvox* spheroids were six times spread on the sieve, bombarded with gold particles and submerged in *Volvox* medium during change of the carriers and disks.
Following bombardment, the spheroids were split into ten petri dishes (∅ = 10 cm) containing 30 ml medium each. Two days after transformation, 30 µg/ml paromomycin sulfate (selection marker, Sigma-Aldrich) was added to each plate. After selection of paromomycin positive clones, the concentration was decreased to 10 µg/ml.

### 8.3.5 Preparation of genomic DNA

*Volvox* spheroids were either harvested by sieving or spinning down. In a volume of approximately 1 ml *Volvox* medium, spheroids were dissociated with a 0.4 mm needle, pulling up and down 3 times. After a centrifugation of 2 min at 17000 *g*, the supernatant was taken off. Centrifugation was repeated and after taking off the supernatant completely, pre-warmed lysis buffer (100 mM NaOH, 2 M NaCl, 0.5% SDS) was added in a 1:1 ratio. The cell pellet was thoroughly resuspended and incubated at 95°C for 5 min, immediately followed by adding 20 volumes of 50 mM Tris/HCl pH 7.5. 2 µl of the diluted genomic DNA sample was used for detection of gene insertion by PCR.

Primers used for the amplification by PCR:

| | | |
|---|---|---|
| myc-GFP (*Chlamydomonas*) | forward | 5'-TACGGTGTGCAGTGCTTCTC-3' |
| | reverse | 5'-GTTGTGGCGGATCTTGAAGT-3' |
| myc-AGO3 | forward | 5'-CTTGAATGAAATGGAGAGCTTGG-3' |
| | reverse | 5'-CTCCGTCAAATCCACCACCT-3' |

To differentiate between endogenous and exogenous *AGO3*, the forward primer anneals in the tag.

### 8.3.6 β-elimination of small RNAs

Total RNA from vegetative somatic cells and vegetative gonidia was mixed with 20pmol of a random oligo RNA (5'-UUAGUGAGAGUCCAAUUAAUU-3', Biomers). Beta-elimination was performed as described previously (Vagin et al., 2006).

13.5 µl of total RNA (10-20 µg, vegetative somatic cells or vegetative gonidia) were mixed with 4.5 µl of 5x borate buffer (148 mM borax, 148 mM boric acid, pH 8.6) and 2.5 µl of freshly dissolved 200 mM $NaIO_4$. After incubation for 10 min at room temperature, 2 µl of glycerol were added to quench unreacted $NaIO_4$. Samples were incubated for another 10 min at room temperature and then dried by vacuum centrifugation for 1 h at room temperature.

Samples were dissolved in 50 µl 1x borax buffer (30 mM borax, 30 mM boric acid, 50 mM NaOH, pH 9.5) and incubated at 45°C for 90 min. 20 µg glycogen was added to each sample and the RNA was precipitated with 2.5 volumes of ethanol at -20°C over night. RNA was collected by centrifugation at 17000 *g*, 4°C, 30 min. Pellets were directly dissolved in RNA loading dye.

## 8.4 Next generation sequencing (Deep sequencing)

### 8.4.1 Generation of small RNA libraries

#### *8.4.1.1 Small RNA libraries for Chapter 3*

Isolated RNA was ligated to a barcoded, adenylated 3' adapter by a truncated T4 RNA Ligase 2 (expressed and purified in our laboratory) (Ho and Shuman, 2002), the 5' RNA adapter was added in a second ligation step by T4 RNA Ligase 1 (NEB). The product was reverse-transcribed using the SuperScriptIII First Strand Synthesis Super Mix (Invitrogen) using a specific primer, followed by a PCR amplification. The samples were run on a 6% Urea-PAGE, the bands corresponding to small RNA containing ligation products were cut out and eluted over night in elution buffer (300 mM NaCl, 2 mM EDTA). The libraries were precipitated with ethanol over night at -20°C, then collected by centrifugation and solved in water.

The libraries were measured on a Genome Analyzer GAIIx (Illumina) by Fasteris SA (Geneva, Switzerland) in 1x38bp single end runs. The raw data can be accessed under the GEO accession GSE45506.

Primers/adapters used:

Adenylated 3' adapter 5'-Phospho-T-4nt-Barcode-CGTATGCCGTCTTCTGCTTG-(C7amino)-3'

5' RNA adapter 5'-GUUCAGAGUUCUACAGUCCGACGAUC-3'

Specific primer for reverse transcription 5' CAAGCAGAAGACGGCATACGA-3'

5' PCR primer 5'-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGACGATC-3'

3' PCR primer 5'-CAAGCAGAAGACGGCATACGA-3'

#### *8.4.1.2 Small RNA libraries for Chapter 4*

2µg of isolated, total RNA was ligated to an adenylated 3' adapter by a truncated T4 RNA Ligase 2 (Ho and Shuman, 2002), the 5' RNA adapter was added in a second ligation step by T4 RNA Ligase 1. The product was reverse-transcribed using the SuperScriptIII First Strand Synthesis Super Mix (Invitrogen) and a specific primer, followed by a PCR amplification. The 3' PCR primer contains a sample-specific 6 nt long index and can be sequenced separately in a second sequencing run directly following the actual insert sequencing (TruSeq technology, Illumina). The samples were run on a 6% Urea-PAGE (National Diagnostics), the bands corresponding to small RNA containing ligation products were cut out and eluted over night in 300 mM NaCl, 2 mM EDTA. The libraries were precipitated with ethanol over night at -20°C, then collected by centrifugation and solved in water. Libraries were sequenced on a HiScan (Illumina) by the KFB (Kompetenzzentrum für fluoreszente Bioanalytik, Regensburg, Germany) in 1x50 bp single end runs. The raw data can be accessed under the GEO accession GSE48404.

Primers/adapters used:

Adenylated 3' adapter 5'-Phospho-TGGAATTCTCGGGTGCCAAGG-(C7amino)-3'

5' RNA adapter 5'-GUUCAGAGUUCUACAGUCCGACGAUC-3'

Specific primer for reverse transcription 5'-GCCTTGGCACCCGAGAATTCCA

5' PCR primer 5'-AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA-3'

3' PCR primer 5'-CAAGCAGAAGACGGCATACGAGAT-6nt Barcode-GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA-3'

#### *8.4.1.3 Small RNA library for Chapter 5*

The RNA from the immunoprecipitation of AGO3 was used for the generation of a small RNA library. Cloning of the library was performed as described in 8.4.1.2. The library was sequenced on a MiSeq (Illumina) in a 1x66 bp run. The raw data can be accessed under the GEO accession GSE59469.

### 8.4.2 RNA Seq for transcriptome analysis

The RNA Seq libraries were generated from 4 µg total RNA per sample. The TruSeq Sample Preparation Kit (Illumina) was used according to the manufacturer's instructions. The only derivation of the protocol was the substitution of the Superscript II reverse transcriptase with the Superscript III reverse transcriptase (both Life Technologies). Accordingly, the cDNA synthesis temperature was raised from 42°C to 50°C.

Libraries of two biological replicates were generated on different days and sequenced on a HiSeq 2000 at the facilities of Illumina in a 100 bp paired end run. Additionally, the same libraries were sequenced in single runs (100 bp) with GATC (Konstanz, Germany).

For the generation of strand-specific libraries, samples of all cells in all life stages were pooled. The library was constructed at the lab of Eugene Berezikov (Utrecht, Netherlands) and sequenced on an ABI SOLiD.

### 8.4.3 Data analysis

#### *8.4.3.1 Data analysis for Chapter 3*

The initial analysis of the deep sequencing libraries was done by Eugene Berezikov. Barcode splitting and adapter trimming were performed using custom scripts. Reads were mapped to *H.sapiens* genome assembly (build GRCh37) using megablast software.

Analysis of RNA categories and miRNA analysis were performed using the miR-Intess™ pipeline (InteRNA Genomics B.V., Netherlands). Analysis of 3' extensions of miRNAs was performed using custom scripts.

#### *8.4.3.2 Data analysis for Chapter 4*

Obtained deep sequencing data was analyzed using an in house written script. The data was mapped against miRBase 19 (August 2012) of mouse miRNAs (*mus musculus*). The minimal length of a read was set to 18 nucleotides, no mismatch was allowed. The reads for each miRNA were normalized against the total read number of the respective library.

For analysis of the libraries, a threshold was set to eliminate very low abundant miRNAs. For all macrophage and all dendritic cell libraries (10 libraries each), an average read number per million of total reads was calculated for all miRNAs. The cutoff was set to 5 reads per million of total reads for expression level analyses and to 50 for fold change analyses. For scatterplots with expression data, all miRNAs above threshold were plotted. For scatterplots on fold changes, again, all miRNAs above the cutoff were plotted. Venn diagrams were created using a threshold of two fold up and two fold down regulation, respectively.

Analysis of transcription factor binding sites was performed using AliBaba2.1. 5kb upstream of the transcription start site of *Col27a1* (NM_025685.3) and the intron sequence upstream of *miR-455* (NR_030477.1) were tested for binding sites of C/EBPβ.

#### *8.4.3.3 Data analysis for Chapter 5*

*RNA Seq for transcriptome analysis*

The transcriptome assembly was built by Eugene Berezikov using Cufflinks. Differential expression was calculated also by employing Cufflinks.

*Search for and analysis of proteins implicated in small RNA pathways*

HHblits (Remmert et al., 2012), which is part of the HHsuite, was used to search for fragments of RNAi processing proteins in the transcriptome of *Volvox carteri*. We used the current transcript assembly v.2 of the JGI (Joint Genome Institute), which is freely available at the Phytozome 10 database (Prochnik et al., 2010). In addition, the transcriptome data generated by ourselves was used for this analysis (see paragraph above). HHblits requires for each protein to be searched for a hidden-Markov-modell (HHM) to be generated by the user. These HHMs constitute a custom database, which was compiled according to the protocol detailed in chapter 3.4 of the HHSuite User Guide (downloaded from ftp://toolkit.genzentrum.lmu.de). To initiate the compilation of the respective HHMs, the following proteins from *Arabidopsis thaliana* were used as a seed: DCL4 (UniProtKB ID P84634), DCL1 (Q9SP32), HEN1 (Q9C5Q8) and RdRP6 (Q9SG02). For the search of additional Argonaute genes, the modified *Volvox carteri* AGO3 was used.

For each sequence, a multiple sequence alignment was created by using HHblits applied to the database uniprot20_2013_03 from the EBI (http://www.uniprot.org/). As required, secondary structure was predicted by means of psipred_3.5 (McGuffin et al., 2000) and added to the HHblits alignments. The compilation of the database was finalized according to the above mentioned protocol.

Since the open reading frames of annotated *Volvox carteri* proteins were not always supporting full length proteins with start and stop codons, different protein sequences were created. DNA was translated in the six putative reading frames to protein sequences by means of methods from the Biophython package (http://biophython.org). HHblits hits with an E-value ≤ $10E^{-5}$ were considered important and further processed. Next, the genomic locus of each transcript was extracted and the list of transcripts was sorted accordingly. Transcripts were grouped according to coverage of the query protein, e.g. for Argonaute, two transcripts in close proximity with one encoding the N-terminal and the other encoding the C-terminal part were considered to be an important hit. To further validate the candidate genes, the protein search algorithm of Pfam version 27.0 (Finn et al., 2014) and Panther HMM Sequence Scoring (Mi et al., 2013) were employed to identify single important domains in the respective transcripts.

*Categorization of small RNAs associated with AGO3*

Reads overlapping exons of mRNA transcripts were assigned to the mRNA fraction. Since there are no databases listing *Volvox carteri* rRNA and tRNA genes in full, reads were mapped against the database entries for *Chlamydomonas reinhardtii* tRNAs (PlantRNA database, (Cognat et al., 2013)) and rRNAs from *Arabidopsis thaliana* (SILVA database, (Quast et al., 2013)) and the order *Volvocales* (exported from GenBank, NCBI). Repeats were assigned using Repbase Update 19.02 (Jurka et al., 2005), phased RNAs were predicted using the *ta-si prediction* tool from the UEA small RNA Workbench (Stocks et al., 2012), which is based on the algorithm by Chen et al. (Chen et al., 2007). MiRNAs were predicted utilizing our own developed identification pipeline, which is described in detail in the results section (see 5.5.2.1). For the analysis of miRNA parameters from other species than *Volvox carteri,* the entries of miRBase version 20 were used.

*Folding of RNA structures*

For the folding of RNA sequences, mfold 2.3 was employed (Walter et al., 1994; Zuker, 2003). *Volvox carteri* RNAs were folded at 30°C, *Arabidopsis thaliana* RNAs at 23°C and *Homo sapiens* RNAs at 37°C. All other parameters were left at the default setting.

# PUBLICATIONS

Dueck, A. and Meister, G. (2009) TRIMming microRNA function in mouse stem cells. *Nat Cell Biol*, **11**, 1392-1393.

Dueck, A. and Meister, G. (2010) MicroRNA processing without Dicer. *Genome Biol*, **11**, 123.

Petri, S., Dueck, A., Lehmann, G., Putz, N., Rudel, S., Kremmer, E. and Meister, G. (2011) Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects. *Rna*, **17**, 737-749.

Valen, E., Preker, P., Andersen, P.R., Zhao, X., Chen, Y., Ender, C., Dueck, A., Meister, G., Sandelin, A. and Jensen, T.H. (2011) Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat Struct Mol Biol*, **18**, 1075-1082.

Motsch, N., Alles, J., Imig, J., Zhu, J., Barth, S., Reineke, T., Tinguely, M., Cogliatti, S., Dueck, A., Meister, G. *et al.* (2012) MicroRNA profiling of Epstein-Barr virus-associated NK/T-cell lymphomas by deep sequencing. *PLoS ONE*, **7**, e42193.

Dueck, A., Ziegler, C., Eichner, A., Berezikov, E. and Meister, G. (2012) microRNAs associated with the different human Argonaute proteins. *Nucleic Acids Res*, **40**, 9850-9862.

Hauptmann, J., Dueck, A., Harlander, S., Pfaff, J., Merkl, R. and Meister, G. (2013) Turning catalytically inactive human Argonaute proteins into active slicer enzymes. *Nat Struct Mol Biol*, **20**, 814-817.

Dueck, A., Eichner, A., Sixt, M. and Meister, G. (2014) A miR-155-dependent microRNA hierarchy in dendritic cell maturation and macrophage activation. *FEBS Lett*, **588**, 632-640.

Dueck, A. and Meister, G. (2014) Assembly and function of small RNA - argonaute protein complexes. *Biol Chem*, **395**, 611-629.

Dueck, A., Evers, M., Henz, S.R., Wenzl, S., Unger, K., Berezikov, E., Weigel, D., Merkl, R. Engelmann, J. and Meister, G. (2014) The small RNAs of the model organism *Volvox carteri*. Manuscript in preparation.

Evers, M., Dueck, A., Meister, G. and Engelmann, J. (2014) Conservation-independent identification of novel miRNAs. Manuscript in preparation.

# APPENDIX

### Supplement 1: Amino acid sequence of AGO3

>AGO3
MSGRGRGASGGGYGGGGGGRGGGGGYRGEGGGGSRPSSAYGGGGGGGYGGGGGGGGGGYGGGRGGGGGGYEGGRG
GGGGGRGGGFDGGRGGGGGGRGGGGGRGGGGGGGGGVLNPQEGAAMLDAVLRTARNMRANVNVDINSEGRPMSIT
RRPNEGTVGRAVNLFANYFRLQTAPGFPRAAYHYDVTIKSVEEARMGGGGRGGGGRGGGRGGRGVAPEPAGPEAA
AEGEDLPPRLAHRVLKAAATQYKWPDGAWRFDGRKNLYLPGQGIPPEVREWKVTLPPREGDRGDKTKSFVVTTKH
VNVVDLSSLQAYLAQQQQQAPRDAMQVLDVVIRHAFAVDPLCTVLGRGYYYPGDGVEPLTGGAEVWKGFQQSFKL
VESGLMLNLDSSFAAFMSERSLPELLAEMCNTRDLSRVDPSRLRSAARNLSGFKVTFPMKGGHLRKKPMIGLSEQ
GAANTMFHNEAEGRSMSVAEYFKSTGRPLRYPNLPCANVGNRMKPTYIPVELCTVVAGQRRMKLDAKQSAGMISA
AKQDPRTKGDAVVVQARRVQSTLQGSGTEAKWGLKLNTDLMRLPGRLLPTPVLQYGSPVCFDVGPNGSWNLRDVK
FHEARALDSWAVVCCIPKEEVDFDGEYSLWDFLIDMCDNMGKCGMAVVDPVRRGSDAAPPVVFQMGREIPNRGIE
NAMRSAAEAAAKRYKKPAKLLLVILPESLTDEYREIKRVSDIELGIPSQVVAGSKAKVGPKAGPRGGGPQYCANV
AMKINNKLGGVNVTLSGGLRYLPVLGGQGALPFMIMGADVTHPTGAAARADVRDPSVAAVVASLDQSMGRWGSRV
LLQTGRQEVITGMATATKELLLEFYRANRNTKPQRLVMYRDGVSEGQFDQVLAEEYMAIRKACRELEESYRPAIT
FIVVQKRHNTRLLPADGAASDQKGNVLPGTVVDKGIVAPDGFDFYLNSHAGLQGTNKPAHYHVLIDEIGFGADGV
ELLTYWLCYLYQRTTKSVSYCPPAYYADRAAFRGRTLLAATSSASDTASEAGSMRAGQGGASAPATFAGIHRDLS
NVLYFM

**Supplement 2, Table: Argonaute protein and small RNA diversity.**

| Argonaute gene | Molecular function | Bound small RNA/DNA, length, features | References |
|---|---|---|---|
| ***Neurospora crassa*** | | | |
| QDE2 | DsRNA-induced/transgene-induced gene silencing (Quelling*)<br>Induction by DNA damage<br>Gene regulation<br>n.d. | siRNA, 25nt<br>qiRNA, 21nt, 5' U<br>milRNAs, 19nt/25nt, 5' U<br>disiRNAs, 22nt, 5' U | (Catalanotto et al., 2000; Cogoni and Macino, 1997; Lee et al., 2009; Lee et al., 2010) |
| SMS-2 | Silencing of unpaired DNA in meiosis | | (Lee et al., 2003a) |
| ***Schizosaccharomyces pombe*** | | | |
| Ago1 | Heterochromatin silencing, TGS, PTGS, H3K9 methylation | siRNAs, priRNA, 21-24nt, 5' U | (Halic and Moazed, 2010; Noma et al., 2004; Sigova et al., 2004; Verdel et al., 2004; Volpe et al., 2002) |
| ***Thermus thermophilus*** | | | |
| TtAgo | Identifying and targeting foreign DNA | Small DNA molecules, 21-24 nt | (Sheng et al., 2014; Swarts et al., 2014) |
| ***Rhodobacter sphaeroides*** | | | |
| RsAgo | Identifying and targeting foreign DNA | diRNA, 15-19nt, 5' UU/C<br>riDNA, 22-24nt, 3' NNNAA/G | (Olovnikov et al., 2013) |
| ***Tetrahymena thermophila*** | | | |
| Twi1 (Piwi) | DNA elimination | scnRNAs, 27-30nt, 3' end 2'O-methylated | (Kurth and Mochizuki, 2009; Mochizuki et al., 2002; Mochizuki and Gorovsky, 2004) |
| Twi2 (Piwi) | Pseudo-gene derived, phased, high-copy repeat-derived | 23-24nt, 5' U | (Couvillion et al., 2009) |
| Twi3 (Piwi) | Uncharacterized | | |
| Twi4 (Piwi) | | | |
| Twi5 (Piwi) | | | |
| Twi6 (Piwi) | | | |
| Twi7 (Piwi) | Low-copy repeat-derived | 23-24nt, 5' U<br>≈32-34nt | (Couvillion et al., 2009) |

| Argonaute gene | Molecular function | Bound small RNA/DNA, length, features | References |
|---|---|---|---|
| Twi8 (Piwi) | Protein-coding gene-derived | 23-24nt, 5' U, modified 3' end | (Couvillion et al., 2009) |
| Twi9 (Piwi) | Uncharacterized | Heterogeneous size | (Couvillion et al., 2009) |
| Twi10 (Piwi) | Telomere-derived | 23-24nt, 5' U<br>≈33-36nt | (Couvillion et al., 2009) |
| Twi11 (Piwi) | Uncharacterized | 27-30nt | (Couvillion et al., 2009) |
| Twi12 (Piwi) | RNA processing in the nucleus | Heterogeneous size/≈18-22 nt, tRNA fragments | (Couvillion et al., 2012; Couvillion et al., 2009; Couvillion et al., 2010) |
| ***Paramecium tetraurelia*** | | | |
| Ptiwi01 (Piwi) | DNA elimination, developmental genome rearrangement | scnRNA, ≈25nt, 5'UNG | (Bouhouche et al., 2011) |
| Ptiwi02 (Piwi) | Uncharacterized | | |
| Ptiwi03 (Piwi) | | | |
| Ptiwi04 (Piwi) | | | |
| Ptiwi05 (Piwi) | | | |
| Ptiwi06 (Piwi) | | | |
| Ptiwi07 (Piwi) | | | |
| Ptiwi08 (Piwi) | | | |
| Ptiwi09 (Piwi) | DNA elimination, developmental genome rearrangement | scnRNA, ≈25nt, 5'UNG | (Bouhouche et al., 2011) |
| Ptiwi10 (Piwi) | Uncharacterized | | |
| Ptiwi11 (Piwi) | | | |
| Ptiwi12 (Piwi) | dsRNA-induced silencing | siRNA, ≈23nt | (Bouhouche et al., 2011) |
| Ptiwi13 (Piwi) | Transgene- and dsRNA-induced silencing | siRNA, ≈23nt | (Bouhouche et al., 2011) |
| Ptiwi14 (Piwi) | Transgene-induced silencing | siRNA, ≈23nt | (Bouhouche et al., 2011) |
| Ptiwi15 (Piwi) | dsRNA-induced silencing | siRNA, ≈23nt | (Bouhouche et al., 2011) |

| Argonaute gene | Molecular function | Bound small RNA/DNA, length, features | References |
|---|---|---|---|
| ***Oxytricha trifallax*** | | | |
| Otiwi1 (Piwi) | DNA retention | piRNA, ≈27nt, 5'U | (Fang et al., 2012) |
| Otiwi2 | Uncharacterized | | |
| Otiwi3 | | | |
| Otiwi4 | | | |
| Otiwi5 | | | |
| Otiwi6 | | | |
| Otiwi7 | | | |
| Otiwi8 | | | |
| Otiwi9 | | | |
| Otiwi10 | | | |
| Otiwi11 | | | |
| Otiwi12 | | | |
| Otiwi13 | | | |
| ***Arabidopsis thaliana*** | | | |
| AGO1 | miRNA-mediated gene silencing, ta-siRNA production, translational repression, RNA-directed DNA methylation (RdDM) | miRNA, 21nt, 5' U, 22nt, miR173 | (Baumberger and Baulcombe, 2005; Brodersen et al., 2008; Chen et al., 2010; Cuperus et al., 2010; Garcia, 2008; Mi et al., 2008; Montgomery et al., 2008b; Vaucheret et al., 2004; Vazquez et al., 2004b) |
| AGO2 | ta-siRNA, rasiRNA, DNA methylation | siRNA, 21nt, virus-derived 5' A | (Garcia et al., 2012; Mi et al., 2008; Takeda et al., 2008) |
| AGO3 | Uncharacterized | | |
| AGO4 | rasiRNA, RdDM (cytosine methylation), heterochromatin silencing | hc-RNA, 24nt, 5' A | (Havecker et al., 2010; Mi et al., 2008; Zilberman et al., 2003) |
| AGO5 | Intergenic siRNAs, miRNA | siRNA, virus-derived, 21/22/24nt, 5' C | (Mi et al., 2008; Takeda et al., 2008) |
| AGO6 | rasiRNA, RdDM (cytosine methylation), heterochromatin silencing | hc-RNA, 24nt, 5' A | (Eun et al., 2011; Havecker et al., 2010; Zheng et al., 2007) |
| AGO7 | ta-siRNA production, heteroblasty, leaf development | miR390 | (Hunter et al., 2003; Montgomery et al., 2008a) |
| AGO8 | Uncharacterized | | |
| AGO9 | Heterochromatin silencing, female gamete formation | 24nt, 5' A | (Duran-Figueroa and Vielle-Calzada, 2010; Havecker et al., 2010; Olmedo-Monfil et al., 2010) |

| Argonaute gene | Molecular function | Bound small RNA/DNA, length, features | References |
|---|---|---|---|
| AGO10 | Translational repression of a subset of miRNA targets | miRNA | (Brodersen et al., 2008; Ji et al., 2011; Mallory et al., 2009; Varallyay et al., 2010) |
| ***Caenorhabditis elegans*** | | | |
| RDE-1 | Exogenous RNAi | Primary siRNA, 5'-P | (Steiner et al., 2007; Tabara et al., 1999; Yigit et al., 2006) |
| ALG-1 | miRNA-mediated gene silencing, TGS | miRNA | (Grishok et al., 2001; Grishok et al., 2005; Steiner et al., 2007) |
| ALG-2 | miRNA-mediated gene silencing | miRNA | (Grishok et al., 2001) |
| ALG-3 (T22B3.2) | Promotion of thermotolerant male fertility | 26G, 3' unmodified | (Conine et al., 2010; Kamminga et al., 2012) |
| ALG-4 (ZK757.3) | Promotion of thermotolerant male fertility | 26G, 3' unmodified | (Conine et al., 2010; Kamminga et al., 2012) |
| CSR-1 | Chromosome segregation, RNAi, promotion of expression of many male-specific germline genes, RNA-induced epigenetic silencing (RNAa) | 22G, 5' tri-P | (Claycomb et al., 2009; Conine et al., 2013; Gu et al., 2009; Seth et al., 2013; Yigit et al., 2006) |
| WAGO-1 (R06C7.1) | Transcriptional silencing, cytoplasmic | 22G, 5' tri-P | (Claycomb et al., 2009; Gu et al., 2009) |
| WAGO-2 (F55A12.1) | Transcriptional silencing | 22G, 5' tri-P | (Gu et al., 2009) |
| WAGO-3 (PPW-2) | Endogenous and exogenous RNAi; transposon silencing | 22G, 5' tri-P | (Gu et al., 2009; Vastenhouw et al., 2003; Yigit et al., 2006) |
| WAGO-4 (F58G1.1) | Endogenous and exogenous RNAi, transcriptional silencing | 22G, 5' tri-P | (Gu et al., 2009; Yigit et al., 2006) |
| WAGO-5 (ZK1248.7) | Transcriptional silencing | 22G, 5' tri-P | (Gu et al., 2009) |
| WAGO-6 (SAGO-2) | Endogenous and exogenous RNAi, transcriptional silencing | 22G, 5' tri-P | (Gu et al., 2009; Yigit et al., 2006) |
| WAGO-7 (PPW-1) | Endogenous and exogenous RNAi, transcriptional silencing | 22G, 5' tri-P | (Gu et al., 2009; Tijsterman et al., 2002; Yigit et al., 2006) |
| WAGO-8 (SAGO-1) | Endogenous and exogenous RNAi, transcriptional silencing | 22G, 5' tri-P | (Gu et al., 2009; Yigit et al., 2006) |
| WAGO-9 (HRDE1, C16C10.3) | Endogenous and exogenous RNAi, transcriptional silencing, maintenance of multigenerational epigenetic memory, nuclear | 22G, 5' tri-P | (Ashe et al., 2012; Gu et al., 2009; Shirayama et al., 2012; Yigit et al., 2006) |
| WAGO-10 (T22H9.3) | Transcriptional silencing, nuclear | 22G, 5' tri-P | (Gu et al., 2009) |
| WAGO-11 (Y49F6A.1) | Transcriptional silencing, nuclear | 22G, 5' tri-P | (Gu et al., 2009) |
| WAGO-12 (NRDE-3) | Transcriptional silencing, nuclear RNAi | 22G, 5' tri-P | (Burton et al., 2011; Guang et al., 2008; Juang et al., 2013; Shirayama et al., 2012) |
| ERGO-1 (Piwi) | Endogenous RNAi | 26G, 5'-P, 3' modified | (Vasale et al., 2010; Yigit et al., 2006) |

| Argonaute gene | Molecular function | Bound small RNA/DNA, length, features | References |
|---|---|---|---|
| PRG-1 (Piwi) | Germline maintenance, initiation of WAGO-maintained gene silencing, RNAe | piRNA, 21U | (Batista et al., 2008; Cox et al., 1998; Das et al., 2008; Shirayama et al., 2012; Wang and Reinke, 2008) |
| PRG-2 (Piwi) | Little or no function, germline development | piRNA, 21U | (Batista et al., 2008; Cox et al., 1998; Das et al., 2008) |
| ***Drosophila melanogaster*** | | | |
| AGO1 | miRNA-mediated gene silencing | miRNA, 5' U | (Czech et al., 2009; Ghildiyal et al., 2010; Okamura et al., 2004) |
| AGO2 | RNAi | siRNA, 22nt, 5' C; 3' end 2' O-methylated | (Chung et al., 2008; Czech et al., 2009; Ghildiyal et al., 2010; Hammond et al., 2001; Horwich et al., 2007) |
| AGO3 (Piwi) | Transposon silencing | piRNA, ~23-26nt, A at pos.10, 3' end 2' O-methylated | (Brennecke et al., 2007; Gunawardane et al., 2007; Horwich et al., 2007; Saito et al., 2007) |
| PIWI (Piwi) | Transposon silencing, germline stem-cell maintenance, RNAi | piRNA, 23-29nt, 5' U, 3' end 2' O-methylated | (Brennecke et al., 2007; Gunawardane et al., 2007; Horwich et al., 2007; Saito et al., 2006; Saito et al., 2007) |
| Aubergine (Piwi) | Transposon silencing, stellate silencing, RNAi | piRNA, 23-29nt, 5' U, 3' end 2' O-methylated | (Brennecke et al., 2007; Gunawardane et al., 2007; Horwich et al., 2007; Saito et al., 2007; Vagin et al., 2006) |
| **Zebrafish (*Danio rerio*)** | | | |
| Ziwi (Piwi) | Germ-cell maintenance, transposon silencing | piRNA, 5' U | (Houwing et al., 2008; Houwing et al., 2007) |
| Zili (Piwi) | Germ-cell maintenance and differentiation, meiosis | piRNA, A at pos.10 | (Houwing et al., 2008) |
| Ago1 | Not characterized | | |
| Ago2 | | | |
| Ago3a | | | |
| Ago3b | | | |
| Ago4 | | | |
| **Murine/Human** | | | |
| AGO1 | miRNA-mediated gene silencing, heterochromatin silencing | miRNA, 21-24nt, 5'U/A | (Janowski et al., 2006; Kim et al., 2006) (Janowski et al., 2006; Kim et al., 2006; Pillai et al., 2004) |
| AGO2 | RNAi, miRNA-mediated gene silencing, heterochromatin silencing | miRNA, 21-24nt, 5'U/A | (Janowski et al., 2006; Liu et al., 2004; Meister et al., 2004; Pillai et al., 2004) |
| AGO3 | miRNA-mediated gene silencing | miRNA, 21-24nt, 5'U/A | (Pillai et al., 2004; Wu et al., 2008) |
| AGO4 | miRNA-mediated gene silencing | miRNA, 21-24nt, 5'U/A | (Pillai et al., 2004; Wu et al., 2008) |

| Argonaute gene | Molecular function | Bound small RNA/DNA, length, features | References |
|---|---|---|---|
| MIWI (mouse) | Spermatogenesis | piRNA | (Deng and Lin, 2002; Girard et al., 2006; Grivna et al., 2006) |
| MIWI2 (mouse) | Spermatogenesis | piRNA | (Aravin et al., 2006; Carmell et al., 2007) |
| MILI (mouse) | Spermatogenesis | piRNA | (Aravin et al., 2006; Kuramochi-Miyagawa et al., 2004) |
| RIWI (rat) | piRNA | piRNA | (Lau et al., 2006) |

Abbreviations:
diRNA, DNA-interacting RNA; disiRNA, Dicer-independent small RNA; milRNA, miRNA-like RNA; n.d., not determined; priRNA, primal small RNA; qiRNA, QDE-2 interacting RNA; riDNA, RNA-interacting DNA; RNAa, RNA-induced epigenetic gene activation; RNAe, RNA-induced epigenetic gene silencing; RNAi, RNA interference; scnRNA, scan RNA.
* Quelling is a mechanism describing posttranscriptional silencing in *Neurospora crassa*.

## Supplement 3, Table: MiRNAs of *Volvox* identified in this work

| ID | pair assignment | Chromosome | Strand | Start (precursor) | Stop (precursor) | MFE/nt | p-value | Sequence of small RNA | Arm |
|---|---|---|---|---|---|---|---|---|---|
| Cluster_3 | mature 1 | scaffold_13 | plus | 238352 | 238683 | -0,8349 | 0,00E+00 | TTTAAATGCTGTTGCCTTGGCC | 5p |
| | mature 2 | | | | | | | TAAACTTGCAGGACGCTAACCCT | 5p |
| Cluster_23 | mature 1 | scaffold_56 | plus | 431937 | 432245 | -0,9557 | 0,00E+00 | TTGAACAAGTCGAGCGGCACTG | 3p |
| | star 1 | | | | | | | TATTTCACAGTGCCGCTCGACT | |
| | mature 2 | | | | | | | TGTCCAGGGACCTGAAGGGGCT | 5p |
| Cluster_24 | mature 1 | scaffold_12 | plus | 1969908 | 1970111 | -0,8397 | 0,00E+00 | TAGTCTCCGCCTGGCTGTTCACT | 5p |
| | star 1 | | | | | | | TGAACAGCCAGGCGGAGACT | |
| Cluster_50 | mature 1 | scaffold_91 | plus | 102638 | 102990 | -0,7742 | 0,00E+00 | TGAACTCATCCTTGTCGCTGCC | 5p |
| | star 1 | | | | | | | TTGAAGGGCAGCAACAAGGATG | |
| Cluster_57 | mature 1 | scaffold_41 | plus | 653085 | 653474 | -0,8515 | 0,00E+00 | TCGGGATTAAGGTGGCGCTGC | 3p |
| Cluster_129 | mature 1 | scaffold_27 | plus | 1226679 | 1227017 | -0,8469 | 0,00E+00 | TTGATGTAGAGCTTGCGACCGGG | 5p |
| | mature 2 | | | | | | | TATGGTGCCTTCCTGTGAACTCC | 3p |
| Cluster_189 | mature 1 | scaffold_17 | plus | 173107 | 173372 | -0,7113 | 0,00E+00 | TGCGTATGCACAGAGAATCGGAT | 3p |
| | star 1 | | | | | | | TGCGATTCTCTGTGCATACG | |
| | mature 2 | | | | | | | TTTCAATCTCAACTTAGACTT | 3p |
| | star 2 | | | | | | | TATGTCTAAGTTGAGATCGAAAT | |
| Cluster_430 | mature 1 | scaffold_7 | minus | 12352 | 12518 | -0,7575 | 0,00E+00 | TGGAAAACTTAGGCTGTGAAAC | 3p |
| | star 1 | | | | | | | TTCCGAGCTTAAGTTTTCCCAAA | |
| Cluster_432 | mature 1 | scaffold_47 | minus | 10531 | 10697 | -0,7575 | 0,00E+00 | TGGAAAACTTAGGCTGTGAAAC | 3p |
| | star 1 | | | | | | | TTCCGAGCTTAAGTTTTCCCAAA | |
| Cluster_433 | mature 1 | scaffold_172 | plus | 10239 | 10405 | -0,7575 | 0,00E+00 | TGGGCTGTGAGCTTTGGAAAACT | 3p |
| | star 1 | | | | | | | TTTTCCCAAAGCCCTCAGCCCAGT | |
| Cluster_435 | mature 1 | scaffold_192 | minus | 13522 | 13688 | -0,7862 | 0,00E+00 | TGGAAAACTTAGGCTGTGAAAC | 3p |
| | star 1 | | | | | | | TTCCGAGCTTAAGTTTTCCCAAA | |
| Cluster_436 | mature 1 | scaffold_118 | plus | 13226 | 13392 | -0,7862 | 0,00E+00 | TGGGCTGTGAGCTTTGGAAAACT | 3p |
| Cluster_437 | mature 1 | scaffold_115 | plus | 19829 | 19995 | -0,7575 | 0,00E+00 | TGGGCTGTGAGCTTTGGAAAACT | 3p |
| | star 1 | | | | | | | TTTTCCCAAAGCCCTCAGCCCAGT | |
| Cluster_438 | mature 1 | scaffold_936 | plus | 1740 | 1906 | -0,7862 | 0,00E+00 | TGGGCTGTGAGCTTTGGAAAACT | 3p |
| Cluster_617 | mature 1 | scaffold_32 | minus | 1482190 | 1482273 | -0,4833 | 3,11E-04 | TGTGGGACGAGGCAGCAGACC | 3p |
| Cluster_628 | mature 1 | scaffold_41 | plus | 458530 | 458781 | -0,7111 | 0,00E+00 | TGTGGGACGCGACTTATAAGC | 5p |
| | star 1 | | | | | | | TTATAAGTCGCGTCCCATAGC | |
| Cluster_808 | mature 1 | scaffold_1 | plus | 3114929 | 3115354 | -0,8099 | 0,00E+00 | TGATGAACTTGCTTGCTTTACT | 5p |
| | star 1 | | | | | | | TAAAGCAAGCAAGTTCATCAGG | |
| Cluster_884 | mature 1 | scaffold_1 | minus | 12674068 | 12674407 | -0,7809 | 0,00E+00 | TTCGGCGAGAATGCCCTGCACC | 3p |
| | star 1 | | | | | | | TGCAGGGCATTCTCGCCGAATA | |
| Cluster_1049 | mature 1 | scaffold_1 | plus | 14147875 | 14148043 | -0,7509 | 0,00E+00 | TGGGCTGTGAGCTTTGGAAAACT | 3p |
| Cluster_1058 | mature 1 | scaffold_93 | plus | 137719 | 137930 | -0,6821 | 0,00E+00 | TGGGCTGTGAGCTTTGGAAAACT | 3p |
| | star 1 | | | | | | | TTTTCCCAAAGCCCTCAGCCCAGT | |
| Cluster_1117 | mature 1 | scaffold_1 | minus | 3114930 | 3115354 | -0,8028 | 0,00E+00 | TGATGAACTTGCTTGCTTTACT | 5p |
| | star 1 | | | | | | | TAAAGCAAGCAAGTTCATCAGG | |
| Cluster_1581 | mature 1 | scaffold_13 | plus | 682490 | 682588 | -0,4091 | 1,92E-04 | TGAGGTATAGCTGTGGCTGAC | 5p |
| Cluster_2003 | mature 1 | scaffold_18 | minus | 1356052 | 1356368 | -1,006 | 0,00E+00 | TAGGACTCGGGACTGCATGAAC | 3p |
| | star 1 | | | | | | | TCATGCAGTCCCGAGTCCTAGC | |
| | star 2 | | | | | | | ACGGGTGTACGGATGGAGGAGCT | 3p |
| | mature 2 | | | | | | | TACGGGTGTACGGATGGAGGAGC | |
| Cluster_2028 | mature 1 | scaffold_18 | plus | 1356052 | 1356368 | -0,9909 | 0,00E+00 | TAGGACTCGGGACTGCATGAAC | 3p |
| | star 1 | | | | | | | TCATGCAGTCCCGAGTCCTAGC | |
| | mature 2 | | | | | | | TACGGGTGTACGGATGGAGGAGC | 3p |
| Cluster_2122 | mature 1 | scaffold_49 | plus | 114893 | 115038 | -0,4753 | 2,53E-04 | TTTGGCCTGTCTGCGGAGATG | 3p |
| Cluster_2125 | mature 1 | scaffold_5 | plus | 1333482 | 1333549 | -0,5441 | 5,71E-06 | TTTGGCCTGTCTGCGGAGATG | 3p |
| Cluster_2208 | mature 1 | scaffold_4 | plus | 4114604 | 4114936 | -1,0078 | 0,00E+00 | TGGGATGAGGGCTGGGACTGG | 5p |
| | star 1 | | | | | | | TCCAGTCCCAGCCCTCATCCC | |
| Cluster_2209 | mature 1 | scaffold_4 | minus | 4114609 | 4114931 | -1,0022 | 0,00E+00 | TGGGATGAGGGCTGGGACTGG | 5p |
| | star 1 | | | | | | | TCCAGTCCCAGCCCTCATCCC | |
| Cluster_2294 | mature 1 | scaffold_27 | minus | 1728260 | 1728568 | -0,9094 | 0,00E+00 | TTGTCTGCGGACGAGGTCGAGA | 5p |
| | star 1 | | | | | | | TGCTCTTCTCGACCTCGTCCGC | |
| Cluster_2335 | mature 1 | scaffold_8 | plus | 1385272 | 1385546 | -0,5454 | 0,00E+00 | TTAAGAACAATGACCGGCATA | 3p |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | star 1 | | | | | | | TATGCCGGTCATTGTTCTTAAG | |
| Cluster_2380 | mature 1<br>star 1 | scaffold_8 | minus | 1385263 | 1385527 | -0,5219 | 0,00E+00 | TTAAGAACAATGACCGGCATAT<br>TATGCCGGTCATTGTTCTTAAG | 3p |
| Cluster_2458 | mature 1 | scaffold_38 | minus | 594473 | 594770 | -0,7024 | 0,00E+00 | TTTGGTAGCTGTGTTGAGACA | 3p |
| Cluster_2588 | mature 1<br>star 1 | scaffold_60 | minus | 218198 | 218396 | -0,7789 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_2634 | mature 1<br>star 1 | scaffold_31 | plus | 411353 | 411535 | -0,7994 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_2658 | mature 1<br>star 1 | scaffold_27 | minus | 568379 | 568580 | -0,7752 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_2668 | mature 1<br>star 1 | scaffold_5 | minus | 86659 | 86860 | -0,7752 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_2905 | mature 1 | scaffold_21 | minus | 2139251 | 2139577 | -0,8599 | 0,00E+00 | TGCCGCCCTCGTCCTTCTGG | 3p |
| Cluster_2909 | mature 1 | scaffold_24 | minus | 874558 | 874756 | -0,7528 | 0,00E+00 | TTGAAGGGCTTAATCCACTGG | 3p |
| Cluster_2962 | mature 1 | scaffold_40 | plus | 145536 | 145839 | -0,6211 | 0,00E+00 | TTGACTTGTCGCAAATGGACC | 5p |
| Cluster_3232 | mature 1 | scaffold_38 | plus | 594449 | 594794 | -0,8032 | 0,00E+00 | TTTGGTAGCTGTGTTGAGACA | 3p |
| Cluster_3403 | mature 1<br>star 1 | scaffold_16 | plus | 2354022 | 2354116 | -0,4305 | 1,35E-06 | TGCATCAGGAAGTGGGACTTC<br>TTGTACAAGTTCTTCGATCCAG | 5p |
| Cluster_3404 | mature 1<br>star 1 | scaffold_17 | plus | 2072273 | 2072362 | -0,4422 | 1,70E-06 | TGCATCAGGAAGTGGGACTTC<br>TTGTACAAGTTCTTCGATCCAG | 5p |
| Cluster_3406 | mature 1<br>star 1 | scaffold_3 | plus | 8581 | 8670 | -0,4422 | 8,86E-06 | TGCATCAGGAAGTGGGACTTC<br>TTGTACAAGTTCTTCGATCCAG | 5p |
| Cluster_3474 | mature 1<br>star 1<br>star2<br>mature 2 | scaffold_25 | minus | 132327 | 132630 | -0,8447 | 0,00E+00 | TGGGACCGCTGATACGAAGCT<br>TTCGTATCAGCGGTCCCAGG<br>TCGGAAATCCTCGATCCAGATG<br>TCTGGATCGAGGATTTCCGAGC | 3p |
| Cluster_3681 | mature 1<br>star 1 | scaffold_11 | minus | 1036057 | 1036255 | -0,6965 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACAGTCGGTAAC | 5p |
| Cluster_3788 | mature 1<br>star 1 | scaffold_58 | plus | 120407 | 120575 | -0,4272 | 3,51E-05 | TATGGATGTGTGTGAGATGCC<br>TATGTCGCCACGTTTCTGGATG | 5p |
| Cluster_3866 | mature 1<br>star 1<br>mature 2<br>star | scaffold_78 | plus | 52109 | 52277 | -0,4272 | 1,31E-04 | TATGGATGTGTGTGAGATGCC<br>TATGTCGCCACGTTTCTGGATG<br>TGTGTAGGGACGAAGGGAGAAG<br>CTTTTCCTCCGTCTGGGAATGC | 5p |
| Cluster_3970 | mature 1<br>star 1<br>mature 2 | scaffold_15 | plus | 1067293 | 1067461 | -0,4266 | 3,42E-05 | TATGGATGTGTGTGAGATGCC<br>TATGTCGCCACGTTTCTGGATG<br>TGTGTAGGGACGAAGGGAGAAGTT | 5p<br><br>3p |
| Cluster_4061 | mature 1 | scaffold_2 | plus | 4788034 | 4788298 | -0,4596 | 2,76E-03 | TGGAGTTGGATTGGCAGGGG | 3p |
| Cluster_4091 | mature 1<br>star 1 | scaffold_31 | plus | 691973 | 692175 | -0,7621 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4171 | mature 1 | scaffold_64 | minus | 27609 | 27707 | -0,7414 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4181 | mature 1 | scaffold_51 | plus | 577972 | 578070 | -0,7414 | 0,00E+00 | TCTCTGCACTGGGCTGTGAGCT | 3p |
| Cluster_4185 | mature 1 | scaffold_32 | minus | 39278 | 39376 | -0,7414 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4188 | mature 1<br>star 1 | scaffold_100 | plus | 67970 | 68168 | -0,7327 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4199 | mature 1 | scaffold_53 | minus | 3603 | 3701 | -0,7414 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4201 | mature 1 | scaffold_15 | minus | 123468 | 123554 | -0,8218 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4214 | mature 1 | scaffold_8 | minus | 29409 | 29510 | -0,6892 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4240 | mature 1 | scaffold_97 | plus | 123143 | 123241 | -0,7859 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4241 | mature 1<br>star 1 | scaffold_31 | minus | 411353 | 411535 | -0,7087 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4243 | mature 1 | scaffold_82 | minus | 984 | 1082 | -0,7414 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4249 | mature 1 | scaffold_74 | plus | 244785 | 244871 | -0,8218 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4253 | mature 1 | scaffold_31 | plus | 1552231 | 1552329 | -0,7414 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4268 | mature 1 | scaffold_93 | plus | 140793 | 140894 | -0,6892 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4290 | mature 1 | scaffold_71 | minus | 22645 | 22743 | -0,7859 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4349 | mature 1<br>star 1 | scaffold_5 | plus | 86659 | 86860 | -0,6728 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4351 | mature 1 | scaffold_50 | minus | 83332 | 83430 | -0,7414 | 0,00E+00 | TCTGCACTGGGCTGTGAGCTT | 3p |
| Cluster_4452 | mature 1<br>star 1 | scaffold_27 | plus | 568398 | 568580 | -0,7066 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4512 | mature 1 | scaffold_60 | plus | 218198 | 218396 | -0,6794 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT | 5p |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | star 1<br>mature 2 | | | | | | | TGGGCCCCAAACGGTCGGTAAC<br>TTAACCCAGCTGATCGTTAACCC | 5p |
| Cluster_4561 | mature1<br>star1 | scaffold_15 | minus | 1962905 | 1963143 | -0,6607 | 0,00E+00 | TTCTGGATATCAGAGGACACACA<br>TGTGTCCTCTGATATGCAGAAT | 5p |
| Cluster_4698 | mature 1 | scaffold_49 | plus | 122432 | 122585 | -0,5318 | 6,77E-03 | TGCACGACGATGCCCAGGACG | 5p |
| Cluster_4880 | mature 1 | scaffold_47 | minus | 84781 | 84979 | -0,6327 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT | 5p |
| Cluster_4896 | mature 1<br>star 1 | scaffold_95 | plus | 75912 | 76110 | -0,6472 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4913 | mature 1<br>star 1 | scaffold_11 | minus | 2924105 | 2924325 | -0,7109 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4920 | mature 1<br>star 1 | scaffold_20 | plus | 2215648 | 2215847 | -0,68 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4937 | mature 1<br>star 1 | scaffold_42 | minus | 98365 | 98533 | -0,5775 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_4940 | mature 1<br>mature 2 | scaffold_53 | minus | 485815 | 486046 | -0,5746 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>ACCGACCGTTTGGGGCCCATTA | 5p<br>3p |
| Cluster_4990 | mature 1<br>star 1 | scaffold_68 | plus | 255497 | 255695 | -0,7226 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGAGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5019 | mature 1<br>star 1 | scaffold_31 | minus | 452107 | 452305 | -0,6588 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5065 | mature 1<br>star 1 | scaffold_1 | plus | 3120005 | 3120203 | -0,7226 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGAGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5086 | mature 1<br>star 1 | scaffold_95 | plus | 28956 | 29154 | -0,7226 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGAGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5088 | mature 1<br>star 1 | scaffold_14 | plus | 2823446 | 2823644 | -0,6663 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5096 | mature 1<br>star 1 | scaffold_47 | minus | 256992 | 257190 | -0,7482 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5101 | mature 1<br>star 1 | scaffold_23 | minus | 117104 | 117303 | -0,61 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGTCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5117 | mature 1<br>star 1 | scaffold_115 | minus | 11621 | 11803 | -0,7667 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5119 | mature 1<br>star 1 | scaffold_15 | minus | 124943 | 125208 | -0,7583 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5135 | mature 1<br>star 1 | scaffold_47 | plus | 256973 | 257190 | -0,6454 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5137 | mature 1<br>star 1 | scaffold_95 | minus | 75912 | 76110 | -0,7437 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5168 | mature 1<br>star 1 | scaffold_14 | minus | 1996255 | 1996453 | -0,6663 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5197 | mature 1<br>star 1 | scaffold_141 | plus | 8454 | 8640 | -0,7203 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5210 | mature 1<br>star 1 | scaffold_115 | plus | 11621 | 11803 | -0,7246 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5219 | mature 1<br>star 1 | scaffold_31 | plus | 452107 | 452305 | -0,7337 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5222 | mature 1<br>star 1 | scaffold_51 | plus | 123877 | 124081 | -0,8712 | 0,00E+00 | TGTAGCGATGAGAGGAAAGAGGC<br>TCTTTCCTCTCATCGCTACAGC | 3p |
| Cluster_5230 | mature 1<br>star 1 | scaffold_51 | minus | 123877 | 124081 | -0,858 | 0,00E+00 | TAGCGATGAGAGGAAAGAGGC<br>TCTTTCCTCTCATCGCTACAGC | 3p |
| Cluster_5232 | mature 1<br>star 1 | scaffold_14 | plus | 1996255 | 1996453 | -0,7276 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5261 | mature 1<br>star 1 | scaffold_31 | minus | 691973 | 692175 | -0,6921 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5269 | mature 1<br>star 1 | scaffold_14 | minus | 2823446 | 2823644 | -0,7176 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5271 | mature 1<br>star 1 | scaffold_20 | minus | 2215648 | 2215846 | -0,597 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5285 | mature 1<br>star 1 | scaffold_62 | minus | 340214 | 340412 | -0,6492 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAATGGTCGGTAAC | 5p |
| Cluster_5322 | mature 1 | scaffold_62 | plus | 340214 | 340412 | -0,698 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC | 3p |
| Cluster_5341 | mature 1 | scaffold_42 | plus | 98367 | 98531 | -0,6685 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT | 5p |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | star 1 | | | | | | | TGGGCCCCAAACGGTCGGTAAC | |
| Cluster_5345 | mature 1<br>star 1 | scaffold_84 | minus | 117121 | 117319 | -0,7226 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGAGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5362 | mature 1<br>star 1 | scaffold_11 | plus | 2924122 | 2924325 | -0,7745 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5492 | mature 1 | scaffold_95 | minus | 21818 | 22000 | -0,7743 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT | 5p |
| Cluster_5512 | mature 1<br>star 1 | scaffold_11 | minus | 2513058 | 2513256 | -0,697 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGCGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5533 | mature 1<br>star 1 | scaffold_100 | minus | 67970 | 68168 | -0,6307 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5578 | mature 1<br>star 1 | scaffold_141 | minus | 8437 | 8640 | -0,773 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_5803 | star 1<br>mature 1 | scaffold_8 | minus | 1277838 | 1278126 | -0,8678 | 0,00E+00 | TCAGTCTGGGCCTTGAAGCGG<br>TCAAGGCCCAGACTGAAGACC | 3p |
| Cluster_5993 | mature 1<br>star 1<br>mature 2<br>star 2 | scaffold_51 | plus | 270573 | 270864 | -0,7075 | 0,00E+00 | TGGACACGATATCGTCTGGCT<br>TAGCCAGACGATATCGTGTCCATG<br>TGAGATTCTCGGCTGTAATTGA<br>TACAGCCAAGAATTTCATGGAC | 5p |
| Cluster_6100 | star 1<br>mature 1 | scaffold_12 | minus | 115981 | 116328 | -0,4606 | 0,00E+00 | TGCACTAATCCCCTTTCTACCC<br>TAGAAAGGGGATTAGTGCACC | 3p |
| Cluster_6143 | mature 1<br>star 1 | scaffold_8 | plus | 1277854 | 1278157 | -1,0148 | 0,00E+00 | TCAAGGCCCAGACTGAAGACC<br>TTCAGTCTGGGCCTTGAAGCGG | 5p |
| Cluster_6503 | mature 1<br>star 1 | scaffold_15 | minus | 530684 | 530950 | -0,8116 | 0,00E+00 | TGTTCAGATGAATCGTCCGTC<br>TGGACGGACGATTCATCTGACC | 3p |
| Cluster_6582 | mature 1<br>star 1 | scaffold_14 | minus | 103105 | 103404 | -0,8017 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_6626 | mature 1 | scaffold_20 | minus | 541204 | 541297 | -0,4394 | 7,53E-03 | TTTGCAAGACGGTTGGAGTGGG | 3p |
| Cluster_6643 | mature 1 | scaffold_31 | plus | 265211 | 265304 | -0,4394 | 7,31E-03 | TTTGCAAGACGGTTGGAGTGGG | 3p |
| Cluster_6715 | mature 1<br>star 1 | scaffold_27 | plus | 1290035 | 1290357 | -0,7557 | 0,00E+00 | TCCTTGTCATCTTTCTCGTCGCCC<br>TGACGAGAAAGATGACAAGGAAG | 3p |
| Cluster_6763 | mature 1 | scaffold_24 | minus | 1864000 | 1864265 | -0,6763 | 0,00E+00 | TCTTGTCCCGAACCTTGTCCT | 3p |
| Cluster_6862 | mature 1 | scaffold_24 | plus | 1864016 | 1864252 | -0,6481 | 0,00E+00 | TCTTGTCCCGAACCTTGTCCT | 3p |
| Cluster_6891 | mature 1 | scaffold_11 | plus | 2513058 | 2513256 | -0,6472 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC | 3p |
| Cluster_6907 | mature 1<br>star 1 | scaffold_95 | minus | 28956 | 29154 | -0,6638 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>TACCGACCGTTTGGGGCTCATTA | 3p |
| Cluster_6920 | mature 1<br>star 1 | scaffold_68 | minus | 255497 | 255695 | -0,6683 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>TACCGACCGTTTGGGGCTCATT | 3p |
| Cluster_6923 | mature 1 | scaffold_90 | minus | 11904 | 12102 | -0,6271 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC | 3p |
| Cluster_6954 | star 1<br>mature 1 | scaffold_26 | plus | 1420750 | 1420919 | -0,6112 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>TACCGACCGTTTGGGGCCCATT | 3p<br>5p |
| Cluster_6963 | mature 1<br>star 1 | scaffold_95 | plus | 21818 | 22000 | -0,6694 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>ACCGACCGCTTGGGGCCCATT | 3p |
| Cluster_6974 | mature 1<br>star 1 | scaffold_1 | minus | 3120005 | 3120203 | -0,6683 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>TACCGACCGTTTGGGGCTCATT | 3p |
| Cluster_7014 | mature 1<br>star 1 | scaffold_11 | plus | 1036057 | 1036255 | -0,6246 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>TACCGACTGTTTGGGGCCCATT | 3p |
| Cluster_7046 | mature 1 | scaffold_4 | plus | 3520095 | 3520350 | -1,0555 | 0,00E+00 | TTGGCAGGCTCCGGAGCGGACT | 5p |
| Cluster_7179 | mature 1<br>mature 2 | scaffold_53 | plus | 485814 | 486048 | -0,5153 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>TGGGCCCCAAACGGTCGGTAAC | 3p<br>3p |
| Cluster_7216 | mature 1 | scaffold_74 | plus | 256121 | 256353 | -0,6953 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC | 3p |
| Cluster_7334 | mature 1<br>star 1<br>mature 2<br>star 2 | scaffold_27 | plus | 1728277 | 1728550 | -0,8777 | 0,00E+00 | TGCGGACGAGGTCGAGAAGAGC<br>TGCTCTTCTCGACCTCGTCCGC<br>GTGGGGAGCCGTCTGCACAACTC<br>TTGTGCAGACGGCTCCCCACC | 5p<br><br>3p |
| Cluster_7756 | mature 1<br>star 1 | scaffold_51 | minus | 270611 | 270827 | -0,624 | 0,00E+00 | TGGACACGATATCGTCTGGCT<br>AGCCAGACGATATCGTGTCCAT | 5p |
| Cluster_7958 | mature 1<br>mature 2 | scaffold_40 | minus | 430258 | 430543 | -0,7839 | 0,00E+00 | TTTCTTCATCGCTACCTCTGAG<br>TGACCTTCTTGCATCCCGCTGC | 3p |
| Cluster_7976 | mature 1<br>star 1<br>mature 3<br>mature 2 | scaffold_5 | minus | 3611556 | 3611879 | -0,8287 | 0,00E+00 | TTCCGACGGTGCTGGTCCCACG<br>TGGGACCAGCACCATCGGAAGA<br>ACGCATAGACATAGACATGT<br>CGGAAGAAACCTTTGCAGCATT | 5p |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Cluster_8081 | mature 1<br>mature 2 | scaffold_40 | plus | 430264 | 430542 | -0,8122 | 0,00E+00 | TTTCTTCATCGCTACCTCTGAG<br>TGACCTTCTTGCATCCCGCTGC | 3p |
| Cluster_8206 | mature 1<br>star 1 | scaffold_1 | plus | 10097709 | 10097959 | -0,806 | 0,00E+00 | TGGAAGCTGGGTCTGGAGGGC<br>TGCCCTCCAGACCCAGCTTC | 3p |
| Cluster_8210 | mature 1<br>star 1 | scaffold_321 | plus | 2280 | 2530 | -0,8072 | 0,00E+00 | TGGAAGCTGGGTCTGGAGGGCATG<br>TGCCCTCCAGACCCAGCTTC | 3p |
| Cluster_8211 | mature 1<br>star 1 | scaffold_321 | plus | 6541 | 6736 | -0,7694 | 0,00E+00 | TGGAAGCTGGGTCTGGAGGGCATG<br>TGCCCTCCAGACCCAGCTTCCAC | 3p |
| Cluster_8230 | mature 1<br>star 1<br>mature 2 | scaffold_321 | minus | 6541 | 6736 | -0,7148 | 0,00E+00 | TCCAGACCCAGCTTCCACTCCT<br>TGGAAGCTGGGTCTGGAGGGCATG<br>TACGTGACTCATCAATGAGCT | 5p |
| Cluster_8376 | mature 1<br>star 1 | scaffold_40 | minus | 552236 | 552619 | -0,5286 | 0,00E+00 | TAACGACGTCGTGAGTTTACT<br>TAAACTCATGACGTCGCTTTAAG | 3p |
| Cluster_8423 | mature 2<br>mature 1 | scaffold_1 | minus | 14115752 | 14115955 | -0,7583 | 0,00E+00 | TGGGCCCCGAACGGTCGGTAAC<br>TAACGGTCGGTTGGGTTAATGGG | 3p |
| Cluster_8819 | mature 1 | scaffold_15 | plus | 1962924 | 1963123 | -0,6535 | 0,00E+00 | TCAGAGGACACACATGGAGT | 5p |
| Cluster_9048 | mature 1<br>star 1 | scaffold_6 | minus | 1121305 | 1121548 | -0,6516 | 0,00E+00 | TCACAGCCAGCGGAGAGATCGT<br>TGCATTGCGATCTCTCCGCTGG | 3p |
| Cluster_9225 | mature 1<br>star 1 | scaffold_23 | minus | 948820 | 949127 | -0,8708 | 0,00E+00 | TTCCCGCTCTGACCCCATGCC<br>TGGCAAGGGGTCAGAACGGGA | 3p |
| Cluster_9481 | mature 1<br>star 1 | scaffold_23 | plus | 1464526 | 1464786 | -0,8326 | 0,00E+00 | TATTGTGGGCGGTGTACTGA<br>TCAGTACACCGCCCACAATAG | 5p |
| Cluster_10404 | mature 1 | scaffold_5 | plus | 3371149 | 3371227 | -0,4025 | 1,42E-03 | TGAACGTGTGCAAGTCCTGAGC | 5p |
| Cluster_10771 | mature 1 | scaffold_8 | plus | 972518 | 972804 | -0,9087 | 0,00E+00 | TAATATGGGTTGGAGTTGGGC | 5p |
| Cluster_10930 | mature 1<br>star 1 | scaffold_8 | minus | 972518 | 972805 | -0,9333 | 0,00E+00 | TAATATGGGTTGGAGTTGGGC<br>TGGACCCCAAACGGTCGGTAAC | 5p |
| Cluster_11020 | mature 1 | scaffold_49 | minus | 113928 | 114160 | -0,5077 | 4,62E-03 | TACCCAGGAATGTGTTGCTCA | 5p |
| Cluster_11070 | mature 1<br>star 1 | scaffold_84 | plus | 117102 | 117319 | -0,6399 | 0,00E+00 | TGGGCCCCAAACGGTCGGTAAC<br>TACCGACCGTTTGGGGCTCATT | 3p |
| Cluster_11121 | mature 1 | scaffold_72 | minus | 37065 | 37263 | -0,6472 | 0,00E+00 | TACCGACCGTTTGGGGCTCATT | 5p |
| Cluster_11340 | mature 1 | scaffold_1 | plus | 4501912 | 4502231 | -0,7084 | 0,00E+00 | TGAGGAGTGAGTCAGATATGG | 5p |
| Cluster_11349 | mature 1<br>star 1 | scaffold_15 | plus | 124912 | 125239 | -0,8183 | 0,00E+00 | TACCGACCGTTTGGGGCCCATT<br>TGGGCCCCAAACGGTCGGTAAC | 5p |
| Cluster_11517 | mature 1<br>star 1 | scaffold_2 | minus | 5039337 | 5039669 | -0,546 | 0,00E+00 | TGCGAGACCACTGACATGACA<br>TGTCATGTCAACGGTCTCGCA | 3p |
| Cluster_11821 | mature 1 | scaffold_31 | minus | 520637 | 520741 | -0,4886 | 7,41E-03 | TTCGTCAAGCTGCTCACGCTGC | 3p |
| Cluster_11900 | mature 1<br>mature 2 | scaffold_63 | plus | 278359 | 278625 | -0,7079 | 0,00E+00 | TATTTGTGGTTTCGGGAACGGG<br>TACCGGAAGACAGAAATAGACC | 3p |
| Cluster_12062 | mature 1 | scaffold_66 | plus | 60650 | 60753 | -0,6481 | 0,00E+00 | TTCTTATCACATGGTGGACAGG | 5p |
| Cluster_12125 | mature 1 | scaffold_40 | minus | 430127 | 430227 | -0,4079 | 5,26E-03 | TTGTCAAAGTCAAAGAGTGGG | 5p |
| Cluster_12239 | mature 1<br>star 1 | scaffold_15 | plus | 530690 | 530944 | -0,8133 | 0,00E+00 | TTCGTTGACGTGAATCCCCCT<br>TGAGGGGGATTCACGTCAACT | 3p |
| Cluster_12311 | mature 1 | scaffold_1 | plus | 4061756 | 4062016 | -0,6651 | 0,00E+00 | TACCGATCCTTCCTGGTGCTAG | 5p |
| Cluster_12414 | mature 1<br>star 1 | scaffold_23 | minus | 1437556 | 1437665 | -0,8191 | 0,00E+00 | TTGCCACCGTCGCCATCCCTGC<br>GGGGATGGTGATGGCGGCAGTG | 5p |
| Cluster_12489 | mature 1 | scaffold_2 | plus | 4800060 | 4800185 | -0,4968 | 3,24E-03 | TCGGCGATGAGGGATGTGGC | 5p |
| Cluster_12522 | mature 1 | scaffold_5 | plus | 2402794 | 2403117 | -0,8605 | 0,00E+00 | TGTAGAAGTGACCCTTGGAC | 3p |
| Cluster_12630 | mature 1 | scaffold_22 | plus | 1842233 | 1842611 | -0,4525 | 4,94E-15 | TAGTTTATAACATTGGGTCGCC | 5p |

# REFERENCES


Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207-221.

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., and Carrington, J.C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet *36*, 1282-1290.

Allo, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A.*, et al.* (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nat Struct Mol Biol *16*, 717-724.

Ameres, S.L., Hung, J.H., Xu, J., Weng, Z., and Zamore, P.D. (2011). Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two Drosophila Argonaute proteins. Rna *17*, 54-63.

Ameres, S.L., and Zamore, P.D. (2013). Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol *14*, 475-488.

Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J.C.*, et al.* (2012). Argonaute proteins couple chromatin silencing to alternative splicing. Nat Struct Mol Biol *19*, 998-1004.

Androulidaki, A., Iliopoulos, D., Arranz, A., Doxaki, C., Schworer, S., Zacharioudaki, V., Margioris, A.N., Tsichlis, P.N., and Tsatsanis, C. (2009). The kinase Akt1 controls macrophage response to lipopolysaccharide by regulating microRNAs. Immunity *31*, 220-231.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Ashe, A., Sapetschnig, A., Weick, E.M., Mitchell, J., Bagijn, M.P., Cording, A.C., Doebley, A.L., Goldstein, L.D., Lehrbach, N.J., Le Pen, J.*, et al.* (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. Cell *150*, 88-99.

Axtell, M.J. (2013). Classification and comparison of small RNAs from plants. Annu Rev Plant Biol *64*, 137-159.

Axtell, M.J., and Bartel, D.P. (2005). Antiquity of MicroRNAs and Their Targets in Land Plants. Plant Cell *17*, 1658-1673.

Axtell, M.J., and Bowman, J.L. (2008). Evolution of plant microRNAs and their targets. Trends Plant Sci *13*, 343-349.

Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P. (2006). A two-hit trigger for siRNA biogenesis in plants. Cell *127*, 565-577.

Axtell, M.J., Snyder, J.A., and Bartel, D.P. (2007). Common functions for diverse small RNAs of land plants. Plant Cell *19*, 1750-1769.

Babiarz, J.E., Hsu, R., Melton, C., Thomas, M., Ullian, E.M., and Blelloch, R. (2011). A role for noncanonical microRNAs in the mammalian brain revealed by phenotypic differences in Dgcr8 versus Dicer1 knockouts and small RNA sequencing. Rna *17*, 1489-1501.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev *22*, 2773-2785.

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. Science *290*, 972-977.

Barent, R.L., Nair, S.C., Carr, D.C., Ruan, Y., Rimerman, R.A., Fulton, J., Zhang, Y., and Smith, D.F. (1998). Analysis of FKBP51/FKBP52 chimeras and mutants for Hsp90 binding and association with progesterone receptor complexes. Mol Endocrinol *12*, 342-354.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S.*, et al.* (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol Cell *31*, 67-78.

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

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev *20*, 1885-1898.

Beitzinger, M., Peters, L., Zhu, J.Y., Kremmer, E., and Meister, G. (2007). Identification of Human microRNA Targets From Isolated Argonaute Protein Complexes. RNA Biol *4*.

Berezikov, E. (2011). Evolution of microRNA diversity and regulation in animals. Nat Rev Genet *12*, 846-860.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol Cell *28*, 328-336.

Berezikov, E., Thuemmler, F., van Laake, L.W., Kondova, I., Bontrop, R., Cuppen, E., and Plasterk, R.H. (2006). Diversity of microRNAs in human and chimpanzee brain. Nat Genet *38*, 1375-1377.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-Tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J.*, et al.* (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. Nat Methods *3*, 199-204.

Blow, M.J., Grocock, R.J., van Dongen, S., Enright, A.J., Dicks, E., Futreal, P.A., Wooster, R., and Stratton, M.R. (2006). RNA editing of human microRNAs. Genome Biol *7*, R27.

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

Bologna, N.G., Mateos, J.L., Bresso, E.G., and Palatnik, J.F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. Embo J *28*, 3646-3656.

Bonner, J.T. (2001). First Signals: The Evolution of Multicellular Development (Princeton, N.J.: Princeton University Press).

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell *123*, 1279-1291.

Bouhouche, K., Gout, J.F., Kapusta, A., Betermier, M., and Meyer, E. (2011). Functional specialization of Piwi proteins in Paramecium tetraurelia from post-transcriptional gene silencing to genome remodelling. Nucleic Acids Res *39*, 4249-4264.

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., and Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. Nucleic Acids Res *39*, 675-686.

Bramsen, J.B., Laursen, M.B., Damgaard, C.K., Lena, S.W., Babu, B.R., Wengel, J., and Kjems, J. (2007). Improved silencing properties using small internally segmented interfering RNAs. Nucleic Acids Res *35*, 5886-5897.

Bramsen, J.B., Laursen, M.B., Nielsen, A.F., Hansen, T.B., Bus, C., Langkjaer, N., Babu, B.R., Hojland, T., Abramov, M., Van Aerschot, A.*, et al.* (2009). A large-scale chemical modification screen identifies design rules to generate siRNAs with high activity, high stability and low toxicity. Nucleic Acids Res *37*, 2867-2881.

Bramsen, J.B., Pakula, M.M., Hansen, T.B., Bus, C., Langkjaer, N., Odadzic, D., Smicius, R., Wengel, S.L., Chattopadhyaya, J., Engels, J.W.*, et al.* (2010). A screen of chemical modifications identifies position-specific modification by UNA to most potently reduce siRNA off-target effects. Nucleic Acids Res *38*, 5761-5773.

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

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

Buckley, B.A., Burkhart, K.B., Gu, S.G., Spracklin, G., Kershner, A., Fritz, H., Kimble, J., Fire, A., and Kennedy, S. (2012). A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. Nature *489*, 447-451.

Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Nishibu, T., Ukekawa, R., Funakoshi, T., Kurokawa, T., Suzuki, H., Hayashizaki, Y.*, et al.* (2010). A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. Genome Res *20*, 1398-1410.

Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2011). Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biol *8*, 158-177.

Burton, N.O., Burkhart, K.B., and Kennedy, S. (2011). Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. Proc Natl Acad Sci U S A *108*, 19683-19688.

Busch, M., and Zernecke, A. (2012). microRNAs in the regulation of dendritic cell functions in inflammation and atherosclerosis. J Mol Med (Berl) *90*, 877-885.

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

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K.*, et al.* (2002). Frequent deletions and down-regulation of microRNA genes miR-15 and miR-16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A *99*, 15524-15529.

Campo-Paysaa, F., Semon, M., Cameron, R.A., Peterson, K.J., and Schubert, M. (2011). microRNA complements in deuterostomes: origin and evolution of microRNAs. Evolution & development *13*, 15-27.

Caplan, A.J., Cyr, D.M., and Douglas, M.G. (1993). Eukaryotic homologues of Escherichia coli dnaJ: a diverse protein family that functions with hsp70 stress proteins. Mol Biol Cell *4*, 555-563.

Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K.B., Montgomery, T.A., Nguyen, T., Cuperus, J.T., and Carrington, J.C. (2012). Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. Plant Cell *24*, 3613-3629.

Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Carrello, A., Ingley, E., Minchin, R.F., Tsai, S., and Ratajczak, T. (1999). The common tetratricopeptide repeat acceptor site for steroid receptor-associated immunophilins and hop is located in the dimerization domain of Hsp90. J Biol Chem *274*, 2682-2689.

Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2000). Gene silencing in worms and fungi. Nature *404*, 245.

Chalker, D.L., and Yao, M.C. (2011). DNA elimination in ciliates: transposon domestication and genome surveillance. Annu Rev Genet *45*, 227-246.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chaulk, S.G., Thede, G.L., Kent, O.A., Xu, Z., Gesner, E.M., Veldhoen, R.A., Khanna, S.K., Goping, I.S., MacMillan, A.M., Mendell, J.T.*, et al.* (2011). Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. RNA Biol *8*, 1105-1114.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol *18*, 1218-1226.

Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature.

Chen, H.M., Chen, L.T., Patel, K., Li, Y.H., Baulcombe, D.C., and Wu, S.H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Natl Acad Sci U S A *107*, 15269-15274.

Chen, H.M., Li, Y.H., and Wu, S.H. (2007). Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. Proc Natl Acad Sci U S A *104*, 3318-3323.

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

Chen, P.Y., Weinmann, L., Gaidatzis, D., Pei, Y., Zavolan, M., Tuschl, T., and Meister, G. (2008). Strand-specific 5'-O-methylation of siRNA duplexes controls guide strand selection and targeting specificity. Rna *14*, 263-274.

Chen, Y., Boland, A., Kuzuoglu-Ozturk, D., Bawankar, P., Loh, B., Chang, C.T., Weichenrieder, O., and Izaurralde, E. (2014). A DDX6-CNOT1 complex and W-binding pockets in CNOT9 reveal direct links between miRNA target recognition and silencing. Mol Cell *54*, 737-750.

Chen, Y., Siegel, F., Kipschull, S., Haas, B., Frohlich, H., Meister, G., and Pfeifer, A. (2013). miR-155 regulates differentiation of brown and beige adipocytes via a bistable circuit. Nature communications *4*, 1769.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Cheng, Q., Fowler, R., Tam, L.W., Edwards, L., and Miller, S.M. (2003). The role of GlsA in the evolution of asymmetric cell division in the green alga Volvox carteri. Dev Genes Evol *213*, 328-335.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E.*, et al.* (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev *24*, 992-1009.

Chitwood, D.H., Guo, M., Nogueira, F.T., and Timmermans, M.C. (2007). Establishing leaf polarity: the role of small RNAs and positional signals in the shoot apex. Development *134*, 813-823.

Christodoulou, F., Raible, F., Tomer, R., Simakov, O., Trachana, K., Klaus, S., Snyman, H., Hannon, G.J., Bork, P., and Arendt, D. (2010). Ancient animal microRNAs and the evolution of tissue identity. Nature *463*, 1084-1088.

Chung, W.J., Agius, P., Westholm, J.O., Chen, M., Okamura, K., Robine, N., Leslie, C.S., and Lai, E.C. (2011). Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans. Genome Res *21*, 286-300.

Chung, W.J., Okamura, K., Martin, R., and Lai, E.C. (2008). Endogenous RNA Interference Provides a Somatic Defense against Drosophila Transposons. Curr Biol.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D.*, et al.* (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694-1698.

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F.*, et al.* (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell *139*, 123-134.

Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E.*, et al.* (2011). MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biol *12*, R126.

Cognat, V., Pawlak, G., Duchene, A.M., Daujat, M., Gigant, A., Salinas, T., Michaud, M., Gutmann, B., Giege, P., Gobert, A.*, et al.* (2013). PlantRNA, a database for tRNAs of photosynthetic eukaryotes. Nucleic Acids Res *41*, D273-279.

Cogoni, C., and Macino, G. (1997). Isolation of quelling-defective (qde) mutants impaired in posttranscriptional transgene-induced gene silencing in Neurospora crassa. Proc Natl Acad Sci U S A *94*, 10233-10238.

Cogoni, C., and Macino, G. (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. Nature *399*, 166-169.

Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W., Green, P.J., Barton, G.J., and Hutvagner, G. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. Rna *15*, 2147-2160.

Coleman, A.W., and Maguire, M.J. (1982). A microspectrofluorometric analysis of nuclear and chloroplast DNA in Volvox. Dev Biol *94*, 441-450.

Conine, C.C., Batista, P.J., Gu, W., Claycomb, J.M., Chaves, D.A., Shirayama, M., and Mello, C.C. (2010). Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. Proc Natl Acad Sci U S A *107*, 3588-3593.

Conine, C.C., Moresco, J.J., Gu, W., Shirayama, M., Conte, D., Jr., Yates, J.R., 3rd, and Mello, C.C. (2013). Argonautes Promote Male Fertility and Provide a Paternal Memory of Germline Gene Expression in C. elegans. Cell *155*, 1532-1544.

Couvillion, M.T., Bounova, G., Purdom, E., Speed, T.P., and Collins, K. (2012). A Tetrahymena Piwi bound to mature tRNA 3' fragments activates the exonuclease Xrn2 for RNA processing in the nucleus. Mol Cell *48*, 509-520.

Couvillion, M.T., Lee, S.R., Hogstad, B., Malone, C.D., Tonkin, L.A., Sachidanandam, R., Hannon, G.J., and Collins, K. (2009). Sequence, biogenesis, and function of diverse small RNA classes bound to the Piwi family proteins of Tetrahymena thermophila. Genes Dev *23*, 2016-2032.

Couvillion, M.T., Sachidanandam, R., and Collins, K. (2010). A growth-essential Tetrahymena Piwi protein carries tRNA fragment cargo. Genes Dev *24*, 2742-2747.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.

Creasey, K.M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B.C., and Martienssen, R.A. (2014). miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. Nature *508*, 411-415.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Cummins, L.L., Owens, S.R., Risen, L.M., Lesnik, E.A., Freier, S.M., McGee, D., Guinosso, C.J., and Cook, P.D. (1995). Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. Nucleic Acids Res *23*, 2019-2024.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol *17*, 997-1003.

Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet *12*, 19-31.

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., and Hannon, G.J. (2009). Hierarchical rules for Argonaute loading in Drosophila. Mol Cell *36*, 445-456.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell *101*, 543-553.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R.*, et al.* (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol Cell *31*, 79-90.

De, N., Young, L., Lau, P.W., Meisner, N.C., Morrissey, D.V., and MacRae, I.J. (2013). Highly complementary target RNAs promote release of guide RNAs from human Argonaute2. Mol Cell *50*, 344-355.

Deng, W., and Lin, H. (2002). *miwi*, a murine homolog of *piwi*, encodes a cytoplasmic protein essential for spermatogenesis. Dev Cell *2*, 819-830.

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

Diederichs, S., and Haber, D.A. (2007). Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell *131*, 1097-1108.

Dong, Z., Han, M.H., and Fedoroff, N. (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc Natl Acad Sci U S A *105*, 9970-9975.

Dudda, J.C., Salaun, B., Ji, Y., Palmer, D.C., Monnot, G.C., Merck, E., Boudousquie, C., Utzschneider, D.T., Escobar, T.M., Perret, R.*, et al.* (2013). MicroRNA-155 is required for effector CD8+ T cell responses to virus infection and cancer. Immunity *38*, 742-753.

Dueck, A., Eichner, A., Sixt, M., and Meister, G. (2014). A miR-155-dependent microRNA hierarchy in dendritic cell maturation and macrophage activation. FEBS Lett *588*, 632-640.

Dueck, A., and Meister, G. (2014). Assembly and function of small RNA - argonaute protein complexes. Biol Chem *395*, 611-629.

Dueck, A., Ziegler, C., Eichner, A., Berezikov, E., and Meister, G. (2012). microRNAs associated with the different human Argonaute proteins. Nucleic Acids Res *40*, 9850-9862.

Duncan, L., Bouckaert, K., Yeh, F., and Kirk, D.L. (2002). kangaroo, a mobile element from Volvox carteri, is a member of a newly recognized third class of retrotransposons. Genetics *162*, 1617-1630.

Duncan, L., Nishii, I., Harryman, A., Buckley, S., Howard, A., Friedman, N.R., and Miller, S.M. (2007). The VARL gene family and the evolutionary origins of the master cell-type regulatory gene, regA, in Volvox carteri. J Mol Evol *65*, 1-11.

Duncan, L., Nishii, I., Howard, A., Kirk, D., and Miller, S.M. (2006). Orthologs and paralogs of regA, a master cell-type regulatory gene in Volvox carteri. Current genetics *50*, 61-72.

Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., and Voinnet, O. (2007). Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. Nat Genet *39*, 848-856.

Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C., and Voinnet, O. (2004). Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. Plant Cell *16*, 1235-1250.

Duran-Figueroa, N., and Vielle-Calzada, J.P. (2010). ARGONAUTE9-dependent silencing of transposable elements in pericentromeric regions of Arabidopsis. Plant signaling & behavior *5*, 1476-1479.

Eamens, A.L., Smith, N.A., Curtin, S.J., Wang, M.B., and Waterhouse, P.M. (2009). The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. Rna *15*, 2219-2235.

Eiring, A.M., Harb, J.G., Neviani, P., Garton, C., Oaks, J.J., Spizzo, R., Liu, S., Schwind, S., Santhanam, R., Hickey, C.J.*, et al.* (2010). miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. Cell *140*, 652-665.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100-110.

Ender, C., Krek, A., Friedlander, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Mol Cell *32*, 519-528.

Ertl, H., Mengele, R., Wenzl, S., Engel, J., and Sumper, M. (1989). The extracellular matrix of Volvox carteri: molecular structure of the cellular compartment. J Cell Biol *109*, 3493-3501.

Eun, C., Lorkovic, Z.J., Naumann, U., Long, Q., Havecker, E.R., Simon, S.A., Meyers, B.C., Matzke, A.J., and Matzke, M. (2011). AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in Arabidopsis thaliana. PLoS ONE *6*, e25730.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F.*, et al.* (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat Struct Mol Biol *18*, 1211-1217.

Faehnle, C.R., Elkayam, E., Haase, A.D., Hannon, G.J., and Joshua-Tor, L. (2013). The making of a slicer: activation of human Argonaute-1. Cell reports *3*, 1901-1909.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L.*, et al.* (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS ONE *2*, e219.

Fahlgren, N., Montgomery, T.A., Howell, M.D., Allen, E., Dvorak, S.K., Alexander, A.L., and Carrington, J.C. (2006). Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. Curr Biol *16*, 939-944.

Fang, W., Wang, X., Bracht, J.R., Nowacki, M., and Landweber, L.F. (2012). Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. Cell *151*, 1243-1255.

Fang, Y., and Spector, D.L. (2007). Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. Curr Biol *17*, 818-823.

Felippes, F.F., Schneeberger, K., Dezulian, T., Huson, D.H., and Weigel, D. (2008). Evolution of Arabidopsis thaliana microRNAs from random sequences. Rna *14*, 2455-2459.

Fernandez-Valverde, S.L., Taft, R.J., and Mattick, J.S. (2010). Dynamic isomiR regulation in Drosophila development. Rna *16*, 1881-1888.

Ferris, P., Olson, B.J., De Hoff, P.L., Douglass, S., Casero, D., Prochnik, S., Geng, S., Rai, R., Grimwood, J., Schmutz, J.*, et al.* (2010). Evolution of an expanded sex-determining locus in Volvox. Science *328*, 351-354.

Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. Nat Rev Genet *3*, 329-341.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J.*, et al.* (2014). Pfam: the protein families database. Nucleic Acids Res *42*, D222-230.

Floyd, S.K., and Bowman, J.L. (2004). Gene regulation: ancient microRNA target sequences in plants. Nature *428*, 485-486.

Fluiter, K., Mook, O.R., Vreijling, J., Langkjaer, N., Hojland, T., Wengel, J., and Baas, F. (2009). Filling the gap in LNA antisense oligo gapmers: the effects of unlocked nucleic acid (UNA) and 4'-C-hydroxymethyl-DNA modifications on RNase H recruitment and efficacy of an LNA gapmer. Mol Biosyst *5*, 838-843.

Flynt, A.S., Greimann, J.C., Chung, W.J., Lima, C.D., and Lai, E.C. (2010). MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol Cell *38*, 900-907.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA Maturation and Germ-Line Stem Cell Maintenance Requires Loquacious, a Double-Stranded RNA-Binding Domain Protein. PLoS Biol *3*, e236.

Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P., and d'Adda di Fagagna, F. (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature *488*, 231-235.

Frank, F., Hauver, J., Sonenberg, N., and Nagar, B. (2012). Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. Embo J *31*, 3588-3595.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

Frohn, A., Eberl, H.C., Stohr, J., Glasmacher, E., Rudel, S., Heissmeyer, V., Mann, M., and Meister, G. (2012). Dicer-dependent and -independent Argonaute2 Protein Interaction Networks in Mammalian Cells. Mol Cell Proteomics *11*, 1442-1456.

Fuhrmann, M., Oertel, W., and Hegemann, P. (1999). A synthetic gene coding for the green fluorescent protein (GFP) is a versatile reporter in Chlamydomonas reinhardtii. Plant J *19*, 353-361.

Fukunaga, R., Han, B.W., Hung, J.H., Xu, J., Weng, Z., and Zamore, P.D. (2012). Dicer partner proteins tune the length of mature miRNAs in flies and mammals. Cell *151*, 533-546.

Garcia, D. (2008). A miRacle in plant development: role of microRNAs in cell differentiation and patterning. Semin Cell Dev Biol *19*, 586-595.

Garcia, D., Garcia, S., Pontier, D., Marchais, A., Renou, J.P., Lagrange, T., and Voinnet, O. (2012). Ago Hook and RNA Helicase Motifs Underpin Dual Roles for SDE3 in Antiviral Defense and Silencing of Nonconserved Intergenic Regions. Mol Cell.

Garzon, R., Calin, G.A., and Croce, C.M. (2009). MicroRNAs in Cancer. Annu Rev Med *60*, 167-179.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P.D. (2010). Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. Rna *16*, 43-56.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Gonzalez-Gonzalez, E., Lopez-Casas, P.P., and del Mazo, J. (2008). The expression patterns of genes involved in the RNAi pathways are tissue-dependent and differ in the germ and somatic cells of mouse testis. Biochim Biophys Acta *1779*, 306-311.

Gorman, D.S., and Levine, R.P. (1965). Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of Chlamydomonas reinhardi. Proc Natl Acad Sci U S A *54*, 1665-1669.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell *123*, 631-640.

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

Grigg, S.P., Canales, C., Hay, A., and Tsiantis, M. (2005). SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. Nature *437*, 1022-1026.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23-34.

Grishok, A., Sinskey, J.L., and Sharp, P.A. (2005). Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. Genes Dev *19*, 683-696.

Grivna, S.T., Pyhtila, B., and Lin, H. (2006). MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. Proc Natl Acad Sci U S A *103*, 13415-13420.

Gu, S., Jin, L., Zhang, F., Huang, Y., Grimm, D., Rossi, J.J., and Kay, M.A. (2011). Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes. Proc Natl Acad Sci U S A *108*, 9208-9213.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., *et al.* (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell *36*, 231-244.

Guang, S., Bochner, A.F., Pavelec, D.M., Burkhart, K.B., Harding, S., Lachowiec, J., and Kennedy, S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. Science.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

Haas, E., and Sumper, M. (1991). The sexual inducer of Volvox carteri. Its large-scale production and secretion by Saccharomyces cerevisiae. FEBS Lett *294*, 282-284.

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

Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J.T., Nusbaum, J.D., Morozov, P., Ludwig, J., *et al.* (2011). RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. Rna *17*, 1697-1712.

Hagan, J.P., Piskounova, E., and Gregory, R.I. (2009). Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol *16*, 1021-1025.

Halic, M., and Moazed, D. (2010). Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. Cell *140*, 504-516.

Hallmann, A. (2011). Evolution of reproductive development in the volvocine algae. Sex Plant Reprod *24*, 97-112.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

Han, B.W., Hung, J.H., Weng, Z., Zamore, P.D., and Ameres, S.L. (2011). The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to Drosophila Argonaute1. Curr Biol *21*, 1878-1887.

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

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Han, M.H., Goud, S., Song, L., and Fedoroff, N. (2004b). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci U S A *101*, 1093-1098.

Hartig, J.V., Esslinger, S., Bottcher, R., Saito, K., and Forstemann, K. (2009). Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. Embo J *28*, 2932-2944.

Hartig, J.V., and Forstemann, K. (2011). Loqs-PD and R2D2 define independent pathways for RISC generation in Drosophila. Nucleic Acids Res *39*, 3836-3851.

Hartig, J.V., Tomari, Y., and Forstemann, K. (2007). piRNAs--the ancient hunters of genome invaders. Genes Dev *21*, 1707-1713.

Hauptmann, J., Dueck, A., Harlander, S., Pfaff, J., Merkl, R., and Meister, G. (2013). Turning catalytically inactive human Argonaute proteins into active slicer enzymes. Nat Struct Mol Biol *20*, 814-817.

Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., and Kay, M.A. (2010). Human tRNA-derived small RNAs in the global regulation of RNA silencing. Rna *16*, 673-695.

Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H., and Baulcombe, D.C. (2010). The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell *22*, 321-334.

Heimberg, A.M., Cowper-Sal-lari, R., Semon, M., Donoghue, P.C., and Peterson, K.J. (2010). microRNAs reveal the interrelationships of hagfish, lampreys, and gnathostomes and the nature of the ancestral vertebrate. Proc Natl Acad Sci U S A *107*, 19379-19383.

Heimberg, A.M., Sempere, L.F., Moy, V.N., Donoghue, P.C., and Peterson, K.J. (2008). MicroRNAs and the advent of vertebrate morphological complexity. Proc Natl Acad Sci U S A *105*, 2946-2950.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell *32*, 276-284.

Heo, I., Joo, C., Kim, Y.K., Ha, M., Yoon, M.J., Cho, J., Yeom, K.H., Han, J., and Kim, V.N. (2009). TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell *138*, 696-708.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Herron, M.D., Hackett, J.D., Aylward, F.O., and Michod, R.E. (2009). Triassic origin and early radiation of multicellular volvocine algae. Proc Natl Acad Sci U S A *106*, 3254-3258.

Hertel, J., Lindemeyer, M., Missal, K., Fried, C., Tanzer, A., Flamm, C., Hofacker, I.L., and Stadler, P.F. (2006). The expansion of the metazoan microRNA repertoire. BMC Genomics *7*, 25.

Ho, C.K., and Shuman, S. (2002). Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. Proc Natl Acad Sci U S A *99*, 12709-12714.

Ho, C.K., Wang, L.K., Lima, C.D., and Shuman, S. (2004). Structure and mechanism of RNA ligase. Cell, in press.

Hock, J., Weinmann, L., Ender, C., Rudel, S., Kremmer, E., Raabe, M., Urlaub, H., and Meister, G. (2007). Proteomic and functional analysis of Argonaute-containing mRNA-protein complexes in human cells. EMBO Rep *8*, 1052-1060.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

Houwing, S., Berezikov, E., and Ketting, R.F. (2008). Zili is required for germ cell differentiation and meiosis in zebrafish. Embo J *27*, 2702-2711.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B.*, et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

Huang, C.R., Burns, K.H., and Boeke, J.D. (2012). Active transposition in genomes. Annu Rev Genet *46*, 651-675.

Hunter, C., Sun, H., and Poethig, R.S. (2003). The Arabidopsis heterochronic gene ZIPPY is an ARGONAUTE family member. Curr Biol *13*, 1734-1739.

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

Huntzinger, E., Kuzuoglu-Ozturk, D., Braun, J.E., Eulalio, A., Wohlbold, L., and Izaurralde, E. (2012). The interactions of GW182 proteins with PABP and deadenylases are required for both translational repression and degradation of miRNA targets. Nucleic Acids Res.

Huskey, R.J., and Griffin, B.E. (1979). Genetic control of somatic cell differentiation in Volvox analysis of somatic regenerator mutants. Dev Biol *72*, 226-235.

Hutner, S.H. (1950). Anaerobic and aerobic growth of purple bacteria (Athiorhodaceae) in chemically defined media. Journal of general microbiology *4*, 286-293.

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

Iki, T., Yoshikawa, M., Nishikiori, M., Jaudal, M.C., Matsumoto-Yokoyama, E., Mitsuhara, I., Meshi, T., and Ishikawa, M. (2010). In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. Mol Cell *39*, 282-291.

Initiative, T.A.G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.

Iwakawa, H.O., and Tomari, Y. (2013). Molecular insights into microRNA-mediated translational repression in plants. Mol Cell *52*, 591-601.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell *39*, 292-299.

Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J., Johnson, J.M., Lim, L., Karpilow, J., Nichols, K.*, et al.* (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. Rna *12*, 1197-1205.

Jackson, A.L., and Linsley, P.S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. Nat Rev Drug Discov *9*, 57-67.

Jakobiak, T., Mages, W., Scharf, B., Babinger, P., Stark, K., and Schmitt, R. (2004). The bacterial paromomycin resistance gene, aphH, as a dominant selectable marker in Volvox carteri. Protist *155*, 381-393.

Jakymiw, A., Lian, S., Eystathioy, T., Li, S., Satoh, M., Hamel, J.C., Fritzler, M.J., and Chan, E.K. (2005). Disruption of GW bodies impairs mammalian RNA interference. Nat Cell Biol *7*, 1267-1274.

Janas, M.M., Wang, B., Harris, A.S., Aguiar, M., Shaffer, J.M., Subrahmanyam, Y.V., Behlke, M.A., Wucherpfennig, K.W., Gygi, S.P., Gagnon, E.*, et al.* (2012). Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins. Rna *18*, 2041-2055.

Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D., and Corey, D.R. (2006). Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. Nat Struct Mol Biol *13*, 787-792.

Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R.E., Kim, Y.J., Dinh, T.T., Liu, J., Cui, X., Zheng, B.*, et al.* (2011). ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis. PLoS Genet *7*, e1001358.

Jinek, M., and Doudna, J.A. (2009). A three-dimensional view of the molecular machinery of RNA interference. Nature *457*, 405-412.

Johnston, M., Geoffroy, M.C., Sobala, A., Hay, R., and Hutvagner, G. (2010). HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. Mol Biol Cell *21*, 1462-1469.

Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell *14*, 787-799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol *57*, 19-53.

Juang, B.T., Gu, C., Starnes, L., Palladino, F., Goga, A., Kennedy, S., and L'Etoile N, D. (2013). Endogenous Nuclear RNAi Mediates Behavioral Adaptation to Odor. Cell *154*, 1010-1022.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res *110*, 462-467.

Juvvuna, P.K., Khandelia, P., Lee, L.M., and Makeyev, E.V. (2012). Argonaute identity defines the length of mature mammalian microRNAs. Nucleic Acids Res.

Kamminga, L.M., van Wolfswinkel, J.C., Luteijn, M.J., Kaaij, L.J., Bagijn, M.P., Sapetschnig, A., Miska, E.A., Berezikov, E., and Ketting, R.F. (2012). Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of Caenorhabditis elegans. PLoS Genet *8*, e1002702.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol *5*, e57.

Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA unction. Dev Cell *4*, 205-217.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Jr., Zhu, J.K., Staskawicz, B.J., and Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci U S A *103*, 18002-18007.

Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhattar, R., and Nishikura, K. (2007a). RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep *8*, 763-769.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K. (2007b). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137-1140.

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

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

Kim, D.H., Villeneuve, L.M., Morris, K.V., and Rossi, J.J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat Struct Mol Biol *13*, 793-797.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Kirk, D.L. (1997). Volvox: A Search for the Molecular and Genetic Origins of Multicellularity and Cellular Differentiation (Cambridge: Cambridge University Press).

Kirk, D.L. (2005). A twelve-step program for evolving multicellularity and a division of labor. Bioessays *27*, 299-310.

Kirk, D.L., Baran, G.J., Harper, J.F., Huskey, R.J., Huson, K.S., and Zagris, N. (1987). Stage-specific hypermutability of the regA locus of Volvox, a gene regulating the germ-soma dichotomy. Cell *48*, 11-24.

Kirk, D.L., Birchem, R., and King, N. (1986). The extracellular matrix of Volvox: a comparative study and proposed system of nomenclature. J Cell Sci *80*, 207-231.

Kirk, D.L., Kaufman, M.R., Keeling, R.M., and Stamer, K.A. (1991). Genetic and cytological control of the asymmetric divisions that pattern the Volvox embryo. Dev Suppl *1*, 67-82.

Kirk, D.L., Viamontes, G.I., Green, K.J., and Bryant, J.L., Jr. (1982). In Developmental Order: Its Origin and Regulation, S. Subtelny, and P.B. Green, eds. (New York: Alan R. Liss), pp. 247-274.

Kirk, M.M., Ransick, A., McRae, S.E., and Kirk, D.L. (1993). The relationship between cell size and cell fate in Volvox carteri. J Cell Biol *123*, 191-208.

Kirk, M.M., Stark, K., Miller, S.M., Muller, W., Taillon, B.E., Gruber, H., Schmitt, R., and Kirk, D.L. (1999). regA, a Volvox gene that plays a central role in germ-soma differentiation, encodes a novel regulatory protein. Development *126*, 639-647.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet *11*, 597-610.

Kuchenbauer, F., Morin, R.D., Argiropoulos, B., Petriv, O.I., Griffith, M., Heuser, M., Yung, E., Piper, J., Delaney, A., Prabhu, A.L.*, et al.* (2008). In-depth characterization of the microRNA transcriptome in a leukemia progression model. Genome Res *18*, 1787-1797.

Kumar, A., and Bennetzen, J.L. (1999). Plant retrotransposons. Annu Rev Genet *33*, 479-532.

Kumar, R., Singh, S.K., Koshkin, A.A., Rajwanshi, V.K., Meldgaard, M., and Wengel, J. (1998). The first analogues of LNA (locked nucleic acids): phosphorothioate-LNA and 2'-thio-LNA. Bioorg Med Chem Lett *8*, 2219-2222.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W.*, et al.* (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. Development *131*, 839-849.

Kurihara, Y., Takashi, Y., and Watanabe, Y. (2006). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. Rna *12*, 206-212.

Kurth, H.M., and Mochizuki, K. (2009). 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in Tetrahymena. Rna *15*, 675-685.

Kuzuoglu-Ozturk, D., Huntzinger, E., Schmidt, S., and Izaurralde, E. (2012). The Caenorhabditis elegans GW182 protein AIN-1 interacts with PAB-1 and subunits of the PAN2-PAN3 and CCR4-NOT deadenylase complexes. Nucleic Acids Res *40*, 5651-5665.

Kwak, P.B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat Struct Mol Biol *19*, 145-151.

Ladewig, E., Okamura, K., Flynt, A.S., Westholm, J.O., and Lai, E.C. (2012). Discovery of hundreds of mirtrons in mouse and human small RNA data. Genome Res *22*, 1634-1645.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of Tissue-Specific MicroRNAs from Mouse. Curr Biol *12*, 735-739.

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

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr Biol *14*, 2162-2167.

Lanet, E., Delannoy, E., Sormani, R., Floris, M., Brodersen, P., Crete, P., Voinnet, O., and Robaglia, C. (2009). Biochemical evidence for translational repression by Arabidopsis microRNAs. Plant Cell *21*, 1762-1768.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Laursen, M.B., Pakula, M.M., Gao, S., Fluiter, K., Mook, O.R., Baas, F., Langklaer, N., Wengel, S.L., Wengel, J., Kjems, J.*, et al.* (2010). Utilization of unlocked nucleic acid (UNA) to enhance siRNA performance in vitro and in vivo. Mol Biosyst *6*, 862-870.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Lee, D.W., Pratt, R.J., McLaughlin, M., and Aramayo, R. (2003a). An argonaute-like protein is required for meiotic silencing. Genetics *164*, 821-828.

Lee, H.C., Chang, S.S., Choudhary, S., Aalto, A.P., Maiti, M., Bamford, D.H., and Liu, Y. (2009). qiRNA is a new type of small interfering RNA induced by DNA damage. Nature *459*, 274-277.

Lee, H.C., Li, L., Gu, W., Xue, Z., Crosthwaite, S.K., Pertsemlidis, A., Lewis, Z.A., Freitag, M., Selker, E.U., Mello, C.C.*, et al.* (2010). Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell *38*, 803-814.

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

Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R., and Kim, V.N. (2006). The role of PACT in the RNA silencing pathway. Embo J *25*, 522-532.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. Embo J *21*, 4663-4670.

Leuschner, P.J., Ameres, S.L., Kueng, S., and Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. EMBO Rep *7*, 314-320.

Levenshtein, V. (1965). BINARY CODES FOR CORRECTING DELETION INSERTION AND SUBSTITUTION ERRORS. Doklady Akademii Nauk Sssr *163*, 845-&.

Li, J., Wu, Y., and Qi, Y. (2014). MicroRNAs in a multicellular green alga Volvox carteri. Science China Life sciences *57*, 36-45.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol *15*, 1501-1507.

Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z.*, et al.* (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell *50*, 67-81.

Lima, W.F., Wu, H., Nichols, J.G., Sun, H., Murray, H.M., and Crooke, S.T. (2009). Binding and cleavage specificities of human Argonaute2. J Biol Chem *284*, 26017-26028.

Lin, X., Ruan, X., Anderson, M.G., McDowell, J.A., Kroeger, P.E., Fesik, S.W., and Shen, Y. (2005). siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. Nucleic Acids Res *33*, 4527-4535.

Lisch, D. (2013). How important are transposons for plant evolution? Nat Rev Genet *14*, 49-61.

Liu, G., Friggeri, A., Yang, Y., Park, Y.J., Tsuruta, Y., and Abraham, E. (2009). miR-147, a microRNA that is induced upon Toll-like receptor stimulation, regulates murine macrophage inflammatory responses. Proc Natl Acad Sci U S A *106*, 15819-15824.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R., and Hannon, G.J. (2005a). A role for the P-body component GW182 in microRNA function. Nat Cell Biol *7*, 1161-1166.

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

Liu, N., Abe, M., Sabin, L.R., Hendriks, G.J., Naqvi, A.S., Yu, Z., Cherry, S., and Bonini, N.M. (2011). The exoribonuclease Nibbler controls 3' end processing of microRNAs in Drosophila. Curr Biol *21*, 1888-1893.

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.

Lobbes, D., Rallapalli, G., Schmidt, D.D., Martin, C., and Clarke, J. (2006). SERRATE: a new player on the plant microRNA scene. EMBO Rep *7*, 1052-1058.

Loeb, G.B., Khan, A.A., Canner, D., Hiatt, J.B., Shendure, J., Darnell, R.B., Leslie, C.S., and Rudensky, A.Y. (2012). Transcriptome-wide miR-155 binding map reveals widespread noncanonical microRNA targeting. Mol Cell *48*, 760-770.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J*., et al.* (2006). MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. Genome Res *16*, 1276-1288.

Lu, J., Shen, Y., Wu, Q., Kumar, S., He, B., Shi, S., Carthew, R.W., Wang, S.M., and Wu, C.I. (2008). The birth and death of microRNA genes in Drosophila. Nat Genet *40*, 351-355.

Luciano, D.J., Mirsky, H., Vendetti, N.J., and Maas, S. (2004). RNA editing of a miRNA precursor. Rna *10*, 1174-1177.

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

Luteijn, M.J., van Bergeijk, P., Kaaij, L.J., Almeida, M.V., Roovers, E.F., Berezikov, E., and Ketting, R.F. (2012). Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. Embo J *31*, 3422-3430.

Lutz, M.B., Kukutsch, N., Ogilvie, A.L., Rossner, S., Koch, F., Romani, N., and Schuler, G. (1999). An advanced culture method for generating large quantities of highly pure dendritic cells from mouse bone marrow. Journal of immunological methods *223*, 77-92.

Machida, S., Chen, H.Y., and Adam Yuan, Y. (2011). Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. Nucleic Acids Res *39*, 7828-7836.

MacRae, I.J., Ma, E., Zhou, M., Robinson, C.V., and Doudna, J.A. (2008). In vitro reconstitution of the human RISC-loading complex. Proc Natl Acad Sci U S A *105*, 512-517.

Maher, C., Stein, L., and Ware, D. (2006). Evolution of Arabidopsis microRNA families through duplication events. Genome Res *16*, 510-519.

Majlessi, M., Nelson, N.C., and Becker, M.M. (1998). Advantages of 2'-O-methyl oligoribonucleotide probes for detecting RNA targets. Nucleic Acids Res *26*, 2224-2229.

Mallory, A.C., Hinze, A., Tucker, M.R., Bouche, N., Gasciolli, V., Elmayan, T., Lauressergues, D., Jauvion, V., Vaucheret, H., and Laux, T. (2009). Redundant and specific roles of the ARGONAUTE proteins AGO1 and ZLL in development and small RNA-directed gene silencing. PLoS Genet *5*, e1000646.

Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. Genes Dev *19*, 2979-2990.

Marques, J.T., Kim, K., Wu, P.H., Alleyne, T.M., Jafari, N., and Carthew, R.W. (2010). Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. Nat Struct Mol Biol *17*, 24-30.

Martinez, N.J., Chang, H.M., Borrajo Jde, R., and Gregory, R.I. (2013). The co-chaperones Fkbp4/5 control Argonaute2 expression and facilitate RISC assembly. Rna *19*, 1583-1593.

Mateos, J.L., Bologna, N.G., Chorostecki, U., and Palatnik, J.F. (2010). Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. Curr Biol *20*, 49-54.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol *8*, 209-220.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E., and Filipowicz, W. (2014). Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. Mol Cell *54*, 751-765.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

Maute, R.L., Schneider, C., Sumazin, P., Holmes, A., Califano, A., Basso, K., and Dalla-Favera, R. (2013). tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A *110*, 1404-1409.

McGuffin, L.J., Bryson, K., and Jones, D.T. (2000). The PSIPRED protein structure prediction server. Bioinformatics *16*, 404-405.

Meissner, M., Stark, K., Cresnar, B., Kirk, D.L., and Schmitt, R. (1999). Volvox germline-specific genes that are putative targets of RegA repression encode chloroplast proteins. Current genetics *36*, 363-370.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nat Rev Genet *14*, 447-459.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 Mediates RNA Cleavage Targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. Curr Biol *15*, 2149-2155.

Meyers, B.C., Axtell, M.J., Bartel, B., Bartel, D.P., Baulcombe, D., Bowman, J.L., Cao, X., Carrington, J.C., Chen, X., Green, P.J.*, et al.* (2008). Criteria for Annotation of Plant MicroRNAs. The Plant Cell Online *20*, 3186-3190.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat Protoc *8*, 1551-1566.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C.*, et al.* (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell *133*, 116-127.

Michalik, K.M., Bottcher, R., and Forstemann, K. (2012). A small RNA response at DNA ends in Drosophila. Nucleic Acids Res *40*, 9596-9603.

Miller, S.M., and Kirk, D.L. (1999). glsA, a Volvox gene required for asymmetric division and germ cell specification, encodes a chaperone-like protein. Development *126*, 649-658.

Miller, S.M., Schmitt, R., and Kirk, D.L. (1993). Jordan, an active Volvox transposable element similar to higher plant transposons. Plant Cell *5*, 1125-1138.

Miyoshi, K., Miyoshi, T., Hartig, J.V., Siomi, H., and Siomi, M.C. (2010a). Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila. Rna *16*, 506-515.

Miyoshi, T., Takeuchi, A., Siomi, H., and Siomi, M.C. (2010b). A direct role for Hsp90 in pre-RISC formation in Drosophila. Nat Struct Mol Biol *17*, 1024-1026.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in Tetrahymena. Cell *110*, 689-699.

Mochizuki, K., and Gorovsky, M.A. (2004). Conjugation-specific small RNAs in Tetrahymena have predicted properties of scan (scn) RNAs involved in genome rearrangement. Genes Dev *18*, 2068-2073.

Molnar, A., Schwach, F., Studholme, D.J., Thuenemann, E.C., and Baulcombe, D.C. (2007). miRNAs control gene expression in the single-cell alga Chlamydomonas reinhardtii. Nature *447*, 1126-1129.

Montes, R.A., de Fatima Rosas-Cardenas, F., De Paoli, E., Accerbi, M., Rymarquis, L.A., Mahalingam, G., Marsch-Martinez, N., Meyers, B.C., Green, P.J., and de Folter, S. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nature communications *5*, 3722.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008a). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell *133*, 128-141.

Montgomery, T.A., Yoo, S.J., Fahlgren, N., Gilbert, S.D., Howell, M.D., Sullivan, C.M., Alexander, A., Nguyen, G., Allen, E., Ahn, J.H.*, et al.* (2008b). AGO1-miR173 complex initiates phased siRNA formation in plants. Proc Natl Acad Sci U S A *105*, 20055-20062.

Moore, K.J., Sheedy, F.J., and Fisher, E.A. (2013). Macrophages in atherosclerosis: a dynamic balance. Nature reviews Immunology *13*, 709-721.

Moore, K.J., and Tabas, I. (2011). Macrophages in the pathogenesis of atherosclerosis. Cell *145*, 341-355.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M.*, et al.* (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res *18*, 610-621.

Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. (2008). PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. Proc Natl Acad Sci U S A *105*, 3145-3150.

Murray, P.J., and Wynn, T.A. (2011). Protective and pathogenic functions of macrophage subsets. Nature reviews Immunology *11*, 723-737.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368-374.

Nazari-Jahantigh, M., Wei, Y., Noels, H., Akhtar, S., Zhou, Z., Koenen, R.R., Heyll, K., Gremse, F., Kiessling, F., Grommes, J., *et al.* (2012a). MicroRNA-155 promotes atherosclerosis by repressing Bcl6 in macrophages. J Clin Invest *122*, 4190-4202.

Nazari-Jahantigh, M., Wei, Y., and Schober, A. (2012b). The role of microRNAs in arterial remodelling. Thrombosis and haemostasis *107*, 611-618.

Nematollahi, G., Kianianmomeni, A., and Hallmann, A. (2006). Quantitative analysis of cell-type specific gene expression in the green alga Volvox carteri. BMC Genomics *7*, 321.

Neupert, J., Karcher, D., and Bock, R. (2009). Generation of Chlamydomonas strains that efficiently express nuclear transgenes. Plant J *57*, 1140-1150.

Nishii, I., and Miller, S.M. (2010). Volvox: simple steps to developmental complexity? Curr Opin Plant Biol *13*, 646-653.

Nishii, I., Ogihara, S., and Kirk, D.L. (2003). A kinesin, invA, plays an essential role in volvox morphogenesis. Cell *113*, 743-753.

Nishikura, K. (2009). Functions and Regulation of RNA Editing by ADAR Deaminases. Annu Rev Biochem.

Noland, C.L., and Doudna, J.A. (2013). Multiple sensors ensure guide strand selection in human RNAi pathways. Rna *19*, 639-648.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I. (2004). RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. Nat Genet *36*, 1174-1180.

O'Carroll, D., Mecklenbrauker, I., Das, P.P., Santana, A., Koenig, U., Enright, A.J., Miska, E.A., and Tarakhovsky, A. (2007). A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. Genes Dev *21*, 1999-2004.

O'Connell, R.M., Chaudhuri, A.A., Rao, D.S., and Baltimore, D. (2009). Inositol phosphatase SHIP1 is a primary target of miR-155. Proc Natl Acad Sci U S A *106*, 7113-7118.

O'Connell, R.M., Kahn, D., Gibson, W.S., Round, J.L., Scholz, R.L., Chaudhuri, A.A., Kahn, M.E., Rao, D.S., and Baltimore, D. (2010). MicroRNA-155 promotes autoimmune inflammation by enhancing inflammatory T cell development. Immunity *33*, 607-619.

O'Connell, R.M., Taganov, K.D., Boldin, M.P., Cheng, G., and Baltimore, D. (2007). MicroRNA-155 is induced during the macrophage inflammatory response. Proc Natl Acad Sci U S A *104*, 1604-1609.

Obika, S., Nanbu, D., Hari, Y., Morio, K.-i., In, Y., Ishida, T., and Imanishi, T. (1997). Synthesis of 2′-O,4′-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3, -endo sugar puckering. Tetrahedron Letters *38*, 8735-8738.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in Drosophila. Cell *130*, 89-100.

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev *18*, 1655-1666.

Okamura, K., Robine, N., Liu, Y., Liu, Q., and Lai, E.C. (2011). R2D2 organizes small regulatory RNA pathways in Drosophila. Mol Cell Biol *31*, 884-896.

Olmedo-Monfil, V., Duran-Figueroa, N., Arteaga-Vazquez, M., Demesa-Arevalo, E., Autran, D., Grimanelli, D., Slotkin, R.K., Martienssen, R.A., and Vielle-Calzada, J.P. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature *464*, 628-632.

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D.K., and Aravin, A.A. (2013). Bacterial argonaute samples the transcriptome to identify foreign DNA. Mol Cell *51*, 594-605.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613-622.

Pall, G.S., and Hamilton, A.J. (2008). Improved northern blot method for enhanced detection of small RNA. Nat Protoc *3*, 1077-1084.

Pantano, L., Estivill, X., and Marti, E. (2010). SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res *38*, e34.

Pare, J.M., LaPointe, P., and Hobman, T.C. (2013). Hsp90 cochaperones p23 and FKBP4 physically interact with hAgo2 and activate RNA interference-mediated silencing in mammalian cells. Mol Biol Cell *24*, 2303-2310.

Pare, J.M., Tahbaz, N., Lopez-Orozco, J., LaPointe, P., Lasko, P., and Hobman, T.C. (2009). Hsp90 regulates the function of argonaute 2 and its recruitment to stress granules and P-bodies. Mol Biol Cell *20*, 3273-3284.

Park, J.K., Liu, X., Strauss, T.J., McKearin, D.M., and Liu, Q. (2007). The miRNA pathway intrinsically controls self-renewal of Drosophila germline stem cells. Curr Biol *17*, 533-538.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A *102*, 3691-3696.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr Biol *12*, 1484-1495.

Peters, L., and Meister, G. (2007). Argonaute proteins: mediators of RNA silencing. Mol Cell *26*, 611-623.

Peterson, K.J., Dietrich, M.R., and McPeek, M.A. (2009). MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. Bioessays *31*, 736-747.

Petri, S., Dueck, A., Lehmann, G., Putz, N., Rudel, S., Kremmer, E., and Meister, G. (2011). Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects. Rna *17*, 737-749.

Pfaff, J., and Meister, G. (2013). Argonaute and GW182 proteins: an effective alliance in gene silencing. Biochem Soc Trans *41*, 855-860.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M.*, et al.* (2005). Identification of microRNAs of the herpesvirus family. Nat Methods *2*, 269-276.

Pillai, R.S., Artus, C.G., and Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. Rna *10*, 1518-1525.

Piriyapongsa, J., and Jordan, I.K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. Rna *14*, 814-821.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev *19*, 2030-2040.

Poulter, R.T., and Goodwin, T.J. (2005). DIRS-1 and the other tyrosine recombinase retrotransposons. Cytogenet Genome Res *110*, 575-588.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K.*, et al.* (2010). Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science *329*, 223-226.

Provasoli, L., and Pintner, I.J. (1959). Artificial media for freshwater algae: problems and suggestions. In The Ecology of Alga C.A. Tyron, and R.T. Hartman, eds. (Pittsburgh, PA: Pymatuning Laboratory of Field Biology, University of Pittsburgh), pp. 84-96.

Qi, J., Qiao, Y., Wang, P., Li, S., Zhao, W., and Gao, C. (2012). microRNA-210 negatively regulates LPS-induced production of proinflammatory cytokines by targeting NF-kappaB1 in murine macrophages. FEBS Lett *586*, 1201-1207.

Qi, Y., Denli, A.M., and Hannon, G.J. (2005). Biochemical specialization within Arabidopsis RNA silencing pathways. Mol Cell *19*, 421-428.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glockner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res *41*, D590-596.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev *20*, 3407-3425.

Ramalingam, P., Palanichamy, J.K., Singh, A., Das, P., Bhagat, M., Kassab, M.A., Sinha, S., and Chattopadhyay, P. (2014). Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. Rna *20*, 76-87.

Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621-629.

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

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

Remmert, M., Biegert, A., Hauser, A., and Soding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods *9*, 173-175.

Rhoades, M., Reinhart, B., Lim, L., Burge, C., Bartel, B., and Bartel, D. (2002). Prediction of plant microRNA targets. Cell *110*, 513.

Rissland, O.S., Mikulasova, A., and Norbury, C.J. (2007). Efficient RNA polyuridylation by noncanonical poly(A) polymerases. Mol Cell Biol *27*, 3612-3624.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. Genome Res *14*, 1902-1910.

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

Rohl, A., Rohrberg, J., and Buchner, J. (2013). The chaperone Hsp90: changing partners for demanding clients. Trends Biochem Sci *38*, 253-262.

Ron, M., Alandete Saez, M., Eshed Williams, L., Fletcher, J.C., and McCormick, S. (2010). Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. Genes Dev *24*, 1010-1021.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193-1207.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007a). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83-86.

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

Rudel, S., Flatley, A., Weinmann, L., Kremmer, E., and Meister, G. (2008). A multifunctional human Argonaute2-specific monoclonal antibody. Rna *14*, 1244-1253.

Rutherford, M.S., and Schook, L.B. (1992). Differential immunocompetence of macrophages derived using macrophage or granulocyte-macrophage colony-stimulating factor. Journal of leukocyte biology *51*, 69-76.

Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev *20*, 2214-2222.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev *21*, 1603-1608.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular Cloning, 2nd edn (Plainview, NY: Cold Spring Harbor Laboratory Press).

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037-1040.

Schroda, M. (2006). RNA silencing in Chlamydomonas: mechanisms and tools. Current genetics *49*, 69-84.

Schurmann, N., Trabuco, L.G., Bender, C., Russell, R.B., and Grimm, D. (2013). Molecular dissection of human Argonaute proteins by DNA shuffling. Nat Struct Mol Biol *20*, 818-826.

Schuster, P., Fontana, W., Stadler, P.F., and Hofacker, I.L. (1994). From sequences to shapes and back: a case study in RNA secondary structures. Proceedings Biological sciences / The Royal Society *255*, 279-284.

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

Seitz, H., Ghildiyal, M., and Zamore, P.D. (2008). Argonaute loading improves the 5' precision of both MicroRNAs and their miRNA* strands in flies. Curr Biol *18*, 147-151.

Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2013). The C. elegans CSR-1 Argonaute Pathway Counteracts Epigenetic Silencing to Promote Germline Gene Expression. Dev Cell.

Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D.C., van der Oost, J., Patel, D.J., and Wang, Y. (2014). Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. Proc Natl Acad Sci U S A *111*, 652-657.

Shi, C., and Pamer, E.G. (2011). Monocyte recruitment during infection and inflammation. Nature reviews Immunology *11*, 762-774.

Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell *150*, 65-77.

Shu, L., and Hu, Z. (2012). Characterization and differential expression of microRNAs elicited by sulfur deprivation in Chlamydomonas reinhardtii. BMC Genomics *13*, 108.

Sievers, F., and Higgins, D.G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol *1079*, 105-116.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology *7*, 539.

Sigova, A., Rhind, N., and Zamore, P.D. (2004). A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in Schizosaccharomyces pombe. Genes Dev *18*, 2359-2367.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.

Smardon, A., Spoerke, J., Stacey, S., Klein, M., Mackin, N., and Maine, E. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. Curr Biol *10*, 169-178.

Smibert, P., Yang, J.S., Azzam, G., Liu, J.L., and Lai, E.C. (2013). Homeostatic control of Argonaute stability by microRNA availability. Nat Struct Mol Biol *20*, 789-795.

Smith, M.R., Willmann, M.R., Wu, G., Berardini, T.Z., Moller, B., Weijers, D., and Poethig, R.S. (2009). Cyclophilin 40 is required for microRNA activity in Arabidopsis. Proc Natl Acad Sci U S A *106*, 5424-5429.

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

Song, L., Axtell, M.J., and Fedoroff, N.V. (2010). RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. Curr Biol *20*, 37-41.

Squadrito, M.L., Etzrodt, M., De Palma, M., and Pittet, M.J. (2013). MicroRNA-mediated control of macrophages and its implications for cancer. Trends in immunology *34*, 350-359.

Stalder, L., Heusermann, W., Sokol, L., Trojer, D., Wirz, J., Hean, J., Fritzsche, A., Aeschimann, F., Pfanzagl, V., Basselet, P.*, et al.* (2013). The rough endoplasmatic reticulum is a central nucleation site of siRNA-mediated RNA silencing. Embo J *32*, 1115-1127.

Starega-Roslan, J., Krol, J., Koscianska, E., Kozlowski, P., Szlachcic, W.J., Sobczak, K., and Krzyzosiak, W.J. (2011). Structural basis of microRNA length variety. Nucleic Acids Res *39*, 257-268.

Starr, R.C. (1970). Control of differentiation in Volvox. Symp Soc Dev Biol *29*, 59-100.

Starr, R.C., and Jaenicke, L. (1974). Purification and characterization of the hormone initiating sexual morphogenesis in Volvox carteri f. nagariensis Iyengar. Proc Natl Acad Sci U S A *71*, 1050-1054.

Steiner, F.A., Hoogstrate, S.W., Okihara, K.L., Thijssen, K.L., Ketting, R.F., Plasterk, R.H., and Sijen, T. (2007). Structural features of small RNA precursors determine Argonaute loading in Caenorhabditis elegans. Nat Struct Mol Biol *14*, 927-933.

Stocks, M.B., Moxon, S., Mapleson, D., Woolfenden, H.C., Mohorianu, I., Folkes, L., Schwach, F., Dalmay, T., and Moulton, V. (2012). The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics *28*, 2059-2061.

Svoboda, P. (2007). Off-targeting and other non-specific effects of RNAi experiments in mammalian cells. Curr Opin Mol Ther *9*, 248-257.

Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.*, et al.* (2014). DNA-guided DNA interference by a prokaryotic Argonaute. Nature *507*, 258-261.

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

Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009). Small RNAs derived from snoRNAs. Rna *15*, 1233-1240.

Tahbaz, N., Carmichael, J.B., and Hobman, T.C. (2001). GERp95 belongs to a family of signal-transducing proteins and requires Hsp90 activity for stability and Golgi localization. J Biol Chem *276*, 43294-43299.

Tahbaz, N., Kolb, F.A., Zhang, H., Jaronczyk, K., Filipowicz, W., and Hobman, T.C. (2004). Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. EMBO Rep *5*, 189-194.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant Cell Physiol *49*, 493-500.

Taliaferro, J.M., Aspden, J.L., Bradley, T., Marwha, D., Blanchette, M., and Rio, D.C. (2013). Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression. Genes Dev *27*, 378-389.

Talmor-Neiman, M., Stav, R., Klipcan, L., Buxdorf, K., Baulcombe, D.C., and Arazi, T. (2006). Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. Plant J *48*, 511-521.

Tam, L.W., and Kirk, D.L. (1991). Identification of cell-type-specific genes of Volvox carteri and characterization of their expression during the asexual life cycle. Dev Biol *145*, 51-66.

Tam, W. (2001). Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. Gene *274*, 157-167.

Tang, R., Li, L., Zhu, D., Hou, D., Cao, T., Gu, H., Zhang, J., Chen, J., Zhang, C.Y., and Zen, K. (2012). Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: evidence for a microRNA hierarchy system. Cell Res *22*, 504-515.

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

Tijsterman, M., Ketting, R.F., Okihara, K.L., and Plasterk, R.H. (2002). RNA helicase MUT-14-dependent silencing triggered in *C. elegans* by short antisense RNAs. Science *295*, 694-697.

Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the argonautes. Nat Chem Biol *3*, 36-43.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299-308.

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

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

Ueki, N., and Nishii, I. (2008). Idaten is a new cold-inducible transposon of Volvox carteri that can be used for tagging developmentally important genes. Genetics *180*, 1343-1353.

Ueki, N., and Nishii, I. (2009). Controlled enlargement of the glycoprotein vesicle surrounding a volvox embryo requires the InvB nucleotide-sugar transporter and is required for normal morphogenesis. Plant Cell *21*, 1166-1181.

Urbich, C., Kuehbacher, A., and Dimmeler, S. (2008). Role of microRNAs in vascular diseases, inflammation, and angiogenesis. Cardiovascular research *79*, 581-588.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

Vaish, N., Chen, F., Seth, S., Fosnaugh, K., Liu, Y., Adami, R., Brown, T., Chen, Y., Harvie, P., Johns, R.*, et al.* (2010). Improved specificity of gene silencing by siRNAs containing unlocked nucleobase analogs. Nucleic Acids Res.

Valdmanis, P.N., Gu, S., Schuermann, N., Sethupathy, P., Grimm, D., and Kay, M.A. (2012). Expression determinants of mammalian argonaute proteins in mediating gene silencing. Nucleic Acids Res *40*, 3704-3713.

Valen, E., Preker, P., Andersen, P.R., Zhao, X., Chen, Y., Ender, C., Dueck, A., Meister, G., Sandelin, A., and Jensen, T.H. (2011). Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. Nat Struct Mol Biol *18*, 1075-1082.

Varallyay, E., Valoczi, A., Agyi, A., Burgyan, J., and Havelda, Z. (2010). Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. Embo J *29*, 3507-3519.

Vasale, J.J., Gu, W., Thivierge, C., Batista, P.J., Claycomb, J.M., Youngman, E.M., Duchaine, T.F., Mello, C.C., and Conte, D., Jr. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. Proc Natl Acad Sci U S A *107*, 3582-3587.

Vastenhouw, N.L., Fischer, S.E., Robert, V.J., Thijssen, K.L., Fraser, A.G., Kamath, R.S., Ahringer, J., and Plasterk, R.H. (2003). A genome-wide screen identifies 27 genes involved in transposon silencing in C. elegans. Curr Biol *13*, 1311-1316.

Vaucheret, H., Vazquez, F., Crete, P., and Bartel, D.P. (2004). The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev *18*, 1187-1197.

Vazquez, F., Blevins, T., Ailhas, J., Boller, T., and Meins, F., Jr. (2008). Evolution of Arabidopsis MIR genes generates novel microRNA classes. Nucleic Acids Res *36*, 6429-6438.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004a). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol *14*, 346-351.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004b). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell *16*, 69-79.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Vigorito, E., Perks, K.L., Abreu-Goodger, C., Bunting, S., Xiang, Z., Kohlhaas, S., Das, P.P., Miska, E.A., Rodriguez, A., Bradley, A*., et al.* (2007). microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells. Immunity *27*, 847-859.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell *136*, 669-687.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Muller, P., Mathews, D.H., and Zuker, M. (1994). Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding. Proc Natl Acad Sci U S A *91*, 9218-9222.

Wang, B., Li, S., Qi, H.H., Chowdhury, D., Shi, Y., and Novina, C.D. (2009). Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. Nat Struct Mol Biol *16*, 1259-1266.

Wang, D., Zhang, Z., O'Loughlin, E., Lee, T., Houel, S., O'Carroll, D., Tarakhovsky, A., Ahn, N.G., and Yi, R. (2012). Quantitative functions of Argonaute proteins in mammalian development. Genes Dev *26*, 693-704.

Wang, G., and Reinke, V. (2008). A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. Curr Biol *18*, 861-867.

Warf, M.B., Johnson, W.E., and Bass, B.L. (2011). Improved annotation of C. elegans microRNAs by deep sequencing reveals structures associated with processing by Drosha and Dicer. Rna *17*, 563-577.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.G., and Qi, Y. (2012). A role for small RNAs in DNA double-strand break repair. Cell *149*, 101-112.

Weinmann, L., Hock, J., Ivacevic, T., Ohrt, T., Mutze, J., Schwille, P., Kremmer, E., Benes, V., Urlaub, H., and Meister, G. (2009). Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. Cell *136*, 496-507.

Werner, S., Wollmann, H., Schneeberger, K., and Weigel, D. (2010). Structure determinants for accurate processing of miR172a in Arabidopsis thaliana. Curr Biol *20*, 42-48.

Westholm, J.O., and Lai, E.C. (2011). Mirtrons: microRNA biogenesis via splicing. Biochimie *93*, 1897-1904.

Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S., and Peterson, K.J. (2009). The deep evolution of metazoan microRNAs. Evolution & development *11*, 50-68.

Wu, H., Neilson, J.R., Kumar, P., Manocha, M., Shankar, P., Sharp, P.A., and Manjunath, N. (2007). miRNA profiling of naive, effector and memory CD8 T cells. PLoS ONE *2*, e1020.

Wu, L., Fan, J., and Belasco, J.G. (2008). Importance of translation and nonnucleolytic ago proteins for on-target RNA interference. Curr Biol *18*, 1327-1332.

Wyman, S.K., Knouf, E.C., Parkin, R.K., Fritz, B.R., Lin, D.W., Dennis, L.M., Krouse, M.A., Webster, P.J., and Tewari, M. (2011). Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome Res *21*, 1450-1461.

Xie, M., Li, M., Vilborg, A., Lee, N., Shu, M.D., Yartseva, V., Sestan, N., and Steitz, J.A. (2013). Mammalian 5'-Capped MicroRNA Precursors that Generate a Single MicroRNA. Cell *155*, 1568-1580.

Xue, Z., Yuan, H., Guo, J., and Liu, Y. (2012). Reconstitution of an Argonaute-dependent small RNA biogenesis pathway reveals a handover mechanism involving the RNA exosome and the exonuclease QIP. Mol Cell *46*, 299-310.

Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010a). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A *107*, 15163-15168.

Yang, S.W., Chen, H.Y., Yang, J., Machida, S., Chua, N.H., and Yuan, Y.A. (2010b). Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. Structure *18*, 594-605.

Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Res *34*, 667-675.

Ye, X., Paroo, Z., and Liu, Q. (2007). Functional anatomy of the Drosophila microRNA-generating enzyme. J Biol Chem *282*, 28373-28378.

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

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell *127*, 747-757.

Yoda, M., Cifuentes, D., Izumi, N., Sakaguchi, Y., Suzuki, T., Giraldez, A.J., and Tomari, Y. (2013). Poly(A)-Specific Ribonuclease Mediates 3'-End Trimming of Argonaute2-Cleaved Precursor MicroRNAs. Cell reports.

Yoda, M., Kawamata, T., Paroo, Z., Ye, X., Iwasaki, S., Liu, Q., and Tomari, Y. (2010). ATP-dependent human RISC assembly pathways. Nat Struct Mol Biol *17*, 17-23.

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev *19*, 2164-2175.

Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C.*, et al.* (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci U S A *105*, 10073-10078.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932-935.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage. Mol Cell *19*, 405-419.

Zamudio, J.R., Kelly, T.J., and Sharp, P.A. (2014). Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. Cell *156*, 920-934.

Zernecke, A. (2012). MicroRNAs in the regulation of immune cell functions--implications for atherosclerotic vascular disease. Thrombosis and haemostasis *107*, 626-633.

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., and Anderson, T.A. (2006). Conservation and divergence of plant microRNA genes. Plant J *46*, 243-259.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.J., and Qi, Y. (2007). A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. Genes Dev *21*, 1190-1203.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. Embo J *26*, 1691-1701.

Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. (2009). Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. Rna *15*, 1886-1895.

Zhu, J.Y., Pfuhl, T., Motsch, N., Barth, S., Nicholls, J., Grasser, F., and Meister, G. (2009). Identification of novel Epstein-Barr virus microRNA genes from nasopharyngeal carcinomas. J Virol *83*, 3333-3341.

Zhu, J.Y., Strehle, M., Frohn, A., Kremmer, E., Hofig, K.P., Meister, G., and Adler, H. (2010). Identification and analysis of expression of novel microRNAs of murine gammaherpesvirus 68. J Virol *84*, 10266-10275.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.

Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res *9*, 133-148.

# ACKNOWLEDGEMENTS

This thesis could not have been completed without the help and support of many people.

MANY THANKS…

… to my supervisor Prof. Gunter Meister for all the inspiring scientific discussions, his positive attitude and encouragement. For always keeping his office door open (literally in Munich and figuratively in Regensburg) and last but not least expert conversations about last night's Bayern game.

… to all members of my thesis examination committee.

… to all my collaborators. Dr. Elisabeth Kremmer, for providing highly specific antibodies crucial for this work; Dr. Maurits Evers, Dr. Eugene Berezikov and Prof. Rainer Merkl for bioinformatics support; Prof. Torben Heick Jensen and Prof. Friedrich Grässer for great collaborations.

… to my colleagues in Munich as well as in Regensburg. To Sabine Rüdel, Christine Ender and especially Julia Stöhr for sharing their knowledge and friendship over the years. To Sebastian Petri, Daniele Hasler, Melina Musri and Daniel Schraivogel for always having a sympathetic ear and helpful advice and support.

… to my fellow student in Tübingen, fellow PhD student in Munich and very good friend Katja Grgur-Fleck for nice coffee breaks at the MPI. Her positiveness and our shared enthusiasm for fantasy books, cooking, baking, BBC series and a good deal more made the years in Munich and Regensburg much more fun.

… to my brilliant little brother for his never-failing optimism and his belief in me. For long conversations over the phone and being patient with me when explaining his PhD project to me (something involving math…). For being an enthusiastic soccer fan and cheering for the world's best team with me in the Allianz Arena.

… to my parents for always being there when I need them and for supporting me in everything I do. If I would write down everything that comes to my mind, I would need many more pages…

… to Mathias for helping me through tough times and sharing the good times, for liking me as I am and being my soulmate.