A. INTRODUCTION

A.1. What are microRNAs?

A.1.1. General information

The past several years have witnessed a remarkable expansion in the recognized functions of RNA beyond the traditional trinity of ribosomal RNA, transfer RNA and messenger RNA. RNA-mediated gene regulation encompasses a wealth of mechanisms, including RNA editing, differential splicing, RNA*i* and gene silencing, heterochromatin formation, genetic recombination, and translational control by auto-regulatory mRNA (riboswitsches) and microRNAs (miRNA). Cells contain a variety of non-coding RNAs, including components of the machinery of gene expression, such as tRNAs and rRNAs, and regulatory RNAs that influence the expression of other genes, such as siRNAs, miRNAs, small nuclear RNAs (snRNAs), and small modulatory RNAs (smRNAs) [1,2]. Non-coding RNAs are impressively diverse and widespread, genomic tiling arrays have revealed that about half of all transcripts are non-coding [3]. One major class of small non-coding RNAs - the miRNAs – was first recognized in 2001 [4-6]. miRNA genes produce small transcripts of about 22-25 nucleotides (nt) in length that are thought to act primarily as antisense inhibitors of mRNA translation. As such, miRNAs can be thought of as a genetic filter that screens the transcriptional output of a cell and determines which mRNAs will be translated into protein. Thousands of miRNAs have been identified in worms, flies, plants, mammals and viruses; and many miRNAs are evolutionarily conserved [7]. With more then 200 members per species in higher eukaryotes, miRNAs are one of the largest gene families, accounting for ~1% of the genome [8]. Recent studies suggest that miRNAs regulate the expression of more than 30% of the protein-coding genes in humans and other organisms [9,10]. Functional analysis of individual miRNA genes has only begun, however, the available studies have already revealed that miRNAs play key roles in diverse regulatory pathways, including control of developmental timing, haematopoietic cell differentiation, apoptosis, cell proliferation and organ development (reviewed in [8,11]).

A.1.2. The discovery of miRNAs

The archetype miRNA gene, *lin-4*, was identified in *C. elegans* through a genetic screen for defects in the temporal control of post-embryonic development [12,13]. *C. elegans* has four larval stages (L1-L4) with distinct and characteristic cell lineages. Mutations in *lin-4* disrupt the regulation of larval development, causing the temporal misregulation of cell fate patterns. Specifically, cell fates characteristic for the first larval stage (L1) are reiterated at later developmental stages [12]. Opposite developmental phenotypes – omission of the L1 cell fates and

premature entry into the L2 stage – are observed in worms deficient for the *lin-14* gene [14]. *lin-14* encodes a nuclear protein, downregulation of which at the end of the L1 initiates the developmental progression to L2 [15,16]. The *lin-4* gene, on the other hand, was shown to produce a non-coding RNA. Ambros and co-workers identified two products of *lin-4*: a ~70 nt transcript with a predicted stem-loop structure, and a 22 nt RNA derived from the stem sequence [15]. The authors concluded that the 70 nt form is most likely a precursor, and that the 22 nt RNA is released by nucleolytic processing. Furthermore, the 22 nt *lin-4* RNA was found to have antisense complementarity to multiply sites in the 3'-untranslated region (UTR) of the *lin-14* gene [15,17]. These observations led to a model for developmental regulation of *lin-14* by *lin-4* in which *lin-4* negatively regulates LIN-14 protein expression through imperfect binding to the 3'UTR of *lin-14* mRNA [18-20]. Due to this negative regulation, *lin-4* controls cell fate transitions through early stages of larval development (L1/L2).

The discovery of *lin-4* and its function in target-specific translational repression opened a new posttranscriptional mechanism of gene regulation during development. In 2000, the second *C. elegans* miRNA, *let-7*, was discovered. Like *lin-4*, *let-7* mutants display perturbations in the timing of developmental events [21]. *let-7* encodes a 21 nt RNA that controls the developmental transition from the fourth larval stage to the adult stage. *let-7* binds to the 3' UTR of *lin-41* and hunchback-like gene (*hbl-1*) and inhibits their translation [21-25]. Like *lin-4, let-7* was shown to be produced from a characteristic ~70 nt hairpin RNA precursor [26]. *let-7* is phylogenetically conserved; homologs of the *let-7* gene were identified in the human, mouse, rat, chicken, mollusc and fly genomes. The expression of mature *let-7* is temporally regulated during *C. elegans* development, another feature that is widely conserved throughout the phyla [26]. Moreover, the sequence of the *let-7* target, *lin-41*, was shown to be conserved in zebrafish, fruit fly and mouse [24]. The 3'UTRs of *lin-41* homologs carry *let-7* complementary sites [26]. The conservation of both the *let-7* and *lin-41* sequences as well as their temporal expression patterns during development indicate that the new posttranscriptional regulatory mechanisms involving ~22 nt RNA molecules in *C. elegans* are evolutionarily ancient.

A.1.3. miRNA genes

The realization that *C. elegans* miRNA genes are evolutionarily conserved spurred efforts by several laboratories to clone and sequence small RNA molecules expressed in *C. elegans*, *D. melanogaster*, mouse and human cell lines. Initially, hundreds of unique ~22 mers were identified, including about 20 new genes in *D. melanogaster*, 30 in humans and 60 in *C. elegans*. All these new transcripts met the following criteria: they were ~22 nt endogenously expressed RNAs; they were predicted (using the MFOLD program [27]) to be produced from parental ~70 nt stem-loop precursors, and they were broadly conserved throughout the phyla [4-6]. This new class of small RNA was termed microRNA, or miRNA. Many, but not all, of the newly identified miRNAs have

developmentally regulated expression patterns, including cell and tissue specificity. It is therefore likely that miRNAs may have a wider variety of functions than just developmental timing. In addition to cloning strategies, numerous other miRNA genes were found in bioinformatics searches of different species, including mouse, rat, human, zebrafish, worm, flies, and plants. Together, cloning and bioinformatic identification have led to current estimates of over 800 miRNA genes in the human genome [28-37]. Griffiths-Jones and co-workers have established an online data bank in order to catalog miRNAs and facilitate the nomenclature of newly identified miRNA genes (http://www.sanger.ac.uk/Software/Rfam/mirna/) [38].

miRNA genes are non-randomly distributed within genomes. They can be grouped according to their location:

- miRNAs located in intergenic regions away from annotated genes, implying that they derive from autonomous transcription units. Intergenic miRNAs are transcribed from their own promoters as long primary transcripts, termed pri-miRNAs [39,40] *let-7, mir-23a, mir-27a, mir-24-2, mir-155, mir-21* are examples of such intergenic miRNAs.
- miRNAs located in intronic regions of protein-coding genes in the sense or antisense orientation [29,41]. Many of these miRNAs are co-expressed with their protein-coding host genes, implying that they generally derive from a common transcript. The location of some of these miRNAs is phylogenetically conserved, moreover, similarity in expression patterns with their host genes also was found to be conserved in different organisms [29,32,42]. Examples of such intronic miRNAs are *mir-186* and *mir-126. mir-186* is expressed in the human and mouse genomes; and in both cases is located in intron 8 of the pre-mRNA of the zinc finger protein 265 [29]. The *mir-126* gene is located in an intron of the EGFL7 gene in human, mouse and zebrafish [43]. Both EGFL7 and *mir-126* show similar cell specificity in human and zebrafish [42]. Another striking case of intronic conservation involves *mir-7*, which is located in the last intron of the heterogeneous nuclear ribonucleoprotein K (hnRNP K) gene in flies and mammals [32].
- Intronic miRNAs located within non-protein-coding transcription units, for example, leukemia associated *mir-15* and *mir-16*. These both miRNAs are located within the fourth intron of a transcribed region, DLEU2 on chromosome 13q14 [44].
- About 50% of the miRNA genes are clustered in the genome and possibly processed from long, polycistronic transcripts. Examples of such clusters are the *mir-23a~27a~24-2* genes on human chromosome 19 and the *let-7a-1~let-7f-1* clusters on human chromosomes 9 and 17 and mouse chromosome 13 [5,6,30]. The clusters may contain miRNAs from a single family, ensuring their co-expression [42] and suggesting that they may act together to regulate a set of target mRNAs. An example of such a cluster, containing six closely related miRNAs genes, is

the mouse *mir-290~mir-295* cluster or its human homolog the *mir-371~mir-373* cluster. Both are expressed specifically in embryonic stem (ES) cells and downregulated upon differentiation [34,45]. However, in some cases clusters contain unrelated miRNAs genes, which are also co-expressed. This suggests that, in the absence of sequence homology, these miRNAs may share functional relationships and coordinately regulate their various targets. For example, orthologs of *C. elegans let-7* and *lin-4* in flies and vertebrates (*let-7* and *mir-125* respectively) are clustered and temporally regulated during development (Figure A.1) [46-49]. In contrast, in *C. elegans lin-4* and *let-7* are unlinked and are expressed sequentially [21,50], this may be unique to the nematodes.

- Some of the miRNAs show interesting genomic locations. For example, *mir-155* was found in an exon of the non-coding RNA *bic* [41], and *mir-10*, which is located in the *Hox* gene cluster in insects, zebrafish, mouse and human [29]. miRNA genes are enriched at genomic fragile sites and in genomic regions associated with cancer [51]. This suggests that duplication, deletion and translocation of miRNA genes may frequently be involved in oncogenesis.

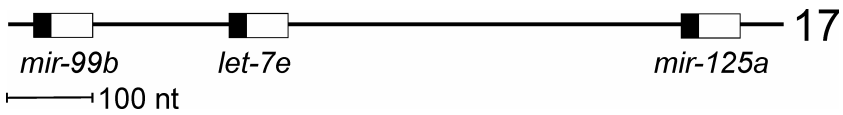

Figure A.1. Genomic organization of mouse's *let-7~mir-125* cluster. The ~70 nt stem-loop precursor structures are indicated as white boxes, the mature miRNA forms are shown within the precursors as black blocks, the chromosomal localization is indicated to the right.

A.1.4. miRNA expression

The application of modern approaches such as Northern blotting, microarrays and *in situ* hybridisation approaches were used to study miRNA-expression profiles in diverse organisms (reviewed in [2,8,43,52]). miRNAs show a wide variety of expression patterns. The expression of some miRNAs is developmentally regulated, while others are expressed constitutively. Some miRNAs display cell and tissue specificity in their expression, others are expressed ubiquitously. Examples of miRNA expression profiles in humans and mice are presented below:

- miRNAs with dynamic expression patterns during development: *let-7, mir-125b, mir-128, mir-266, mir-131* [48,53]
- Constitutively expressed miRNAs: *mir-16, mir-92* [54].
- Tissue specific miRNAs: heart specific *mir-1, mir-208*; liver specific *mir-122a*; spleen specific *mir-189, mir-212*; brain specific *mir-124*, *mir-9*; lung specific *mir-18, mir-19a* [28,54].
- Tissue enriched miRNAs: brain enriched *mir-125b*, *mir-128*; liver enriched *mir-152, mir-215*; heart enriched *mir-133, mir-206*; spleen enriched *mir-99a* [54].

- miRNAs with lineage specificity: granulocyte and macrophage specific *mir-223*; B-lymphoid cell specific *mir-181* [55]; mouse ES cell specific *mir-290~mir295* cluster [34]; neuron specific *mir-124*, *mir-128*, astrocyte specific *mir-23* [48].
- Ubiquitously expressed miRNAs: *mir-16, let-7a, let-7b, mir-26a, mir-30b, mir-30c* [54].
- Abnormal miRNA expression profiles in cancer: *mir-15* and *mir-16* are downregulated in chronic lymphocytic leukemias [44]; *mir-26a* and *mir-99a* are downregulated in lung cancer cell lines [51]; *mir-155* is upregulated in Burkitt lymphoma [56].

The diversity of miRNA expression patterns indicates that miRNAs may carry out a diverse spectrum of regulatory activities in animals. They can play evolutionary conserved developmental or physiological roles in many species. Furthermore, the finding that miRNA expression is regulated during tumorigenesis suggests that misregulation of miRNA expression and function may contribute to the disease process.

A.2. Biogenesis of miRNAs

miRNA biogenesis is a multi-step process as described in Figure A.2:

1. Synthesis of the primary miRNA transcript (pri-miRNA)
2. Processing of the nascent transcript to the ~70 nt stem-loop precursor miRNA (pre-miRNA)
3. Export of the pre-miRNA from the nucleus
4. Processing of the ~22 bp miRNA duplex from the ~70 nt pre-miRNA
5. Release of the ~22 nt mature miRNA
6. Assembly of the miRNA-induced silencing complex (miRISC)
7. miRNA binding to the 3'UTR of a target mRNA and regulation of translation
8. Transport of the mRNA/miRISC complex to P-bodies.

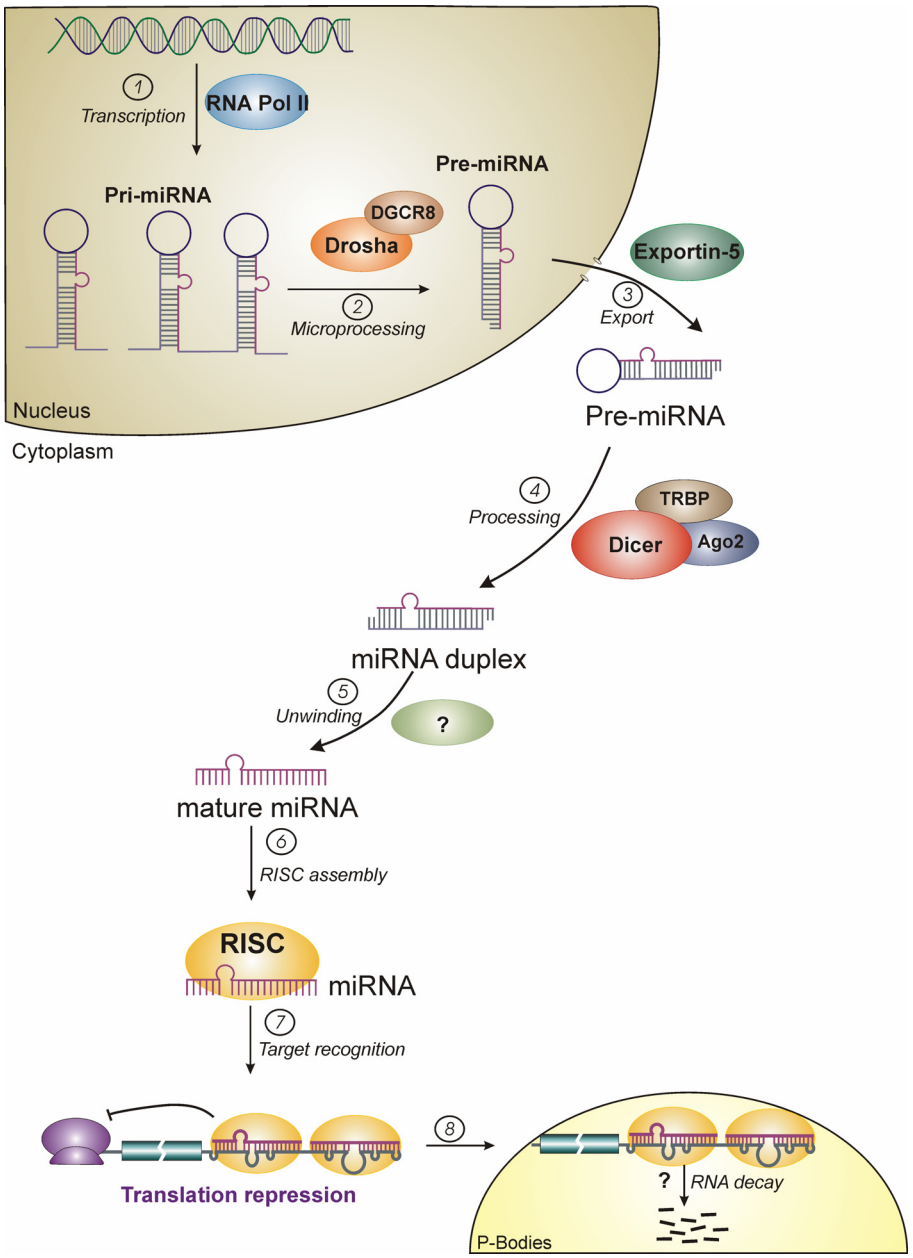

Figure A.2. Model of miRNA biogenesis process. See the text below for details. Figure is adapted from [52].

A.2.1. miRNA transcription

The small size and frequent location of miRNA genes in intergenic regions led to the question of which RNA polymerase is responsible for miRNA transcription. A second, related question was whether miRNA clusters are organized as multiple, independent transcriptional units or as longer, polycistronic transcripts. The first clue came with the demonstration of primary transcripts termed pri-miRNA that are much longer than the stem-loop precursors they contain [5,57]. RT-PCR and RNAse protection assays revealed that pri-miRNA may be several kilobases in length and may contain multiple miRNA precursors [57]. Next came the demonstration that RNA polymerase II (pol II) is responsible for pri-miRNA transcription (Figure A.2, Step 1) [39]. Initially, RNA polymerase III (pol III) was thought to mediate miRNA transcription because it is responsible for the production of

a range of small non-coding RNAs, including tRNAs, 5S rRNA and U6 snRNA. However, the following observations provide evidence that a majority of miRNA are transcribed by pol II, on equal terms with protein coding genes and some of the non-coding genes such as U1-U5 snRNAs:

- pri-miRNAs are quite long, longer that typical pol III transcripts.
- Many pri-miRNA transcripts carry the signature pol II features of a 5´ 7-methylguanosine cap and 3´ poly-adenosine tail [39,40].
- pri-miRNAs contain internal runs of more then four uridine residues, which would terminate pol III transcription.
- The expression of many miRNAs is developmentally regulated, which is more characteristic for pol II than for pol III dependent genes.
- miRNA transcription is sensitive to α-amanitin, a strong inhibitor of pol II activity. Moreover, pol II is physically associated with miRNA promoters, as has been shown by chromatin immunoprecipitation (ChIP) assay [39].
- Fusion of predicted promoter elements from miRNA genes to the green fluorescent protein (GFP) or to luciferase genes gives rise to reporter protein expression, suggesting that pri-miRNA transcripts are pol II transcripts [39,58,59].

Despite these observations, a minor fraction of pri-miRNAs can be transcribed by other RNA polymerases (e.g. pol III). Thus, ectopic expression of some miRNAs from a pol III promoter efficiently produces mature miRNAs, which function *in vivo* [55].

A.2.2. miRNA maturation

The maturation of miRNAs occurs in steps and is compartmentalized. The first step is cleavage of the long pri-miRNA transcript in the nucleus by the class II RNAse III enzyme Drosha. This cleavage releases the ~70 nt stem-loop precursor miRNA [60] (Figure A.2, Step 2). Drosha is a ~160 kDa nuclear protein, containing two RNAse III catalytic domains (RIIIDa and RIIIDb), and a double-stranded RNA-binding domain (dsRBD) in the C-terminal half. The amino-terminal segment of the protein contains proline- and serine/arginine-rich regions of unknown functions (Figure A.3) [61].

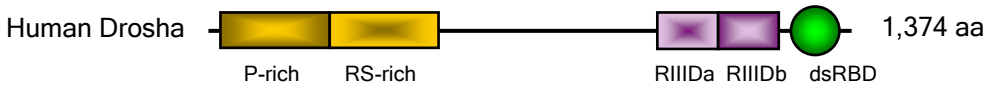

Figure A.3. Domain organization of human Drosha protein. Prolin-rich (P-rich) and serin-argenine-rich (RS-rich) N-terminal domains are shown in yellow. The RNAse III domains (RIIIDa and RIIIDb) are shown in violet, The dsRBD is green.

Drosha is a conserved protein; homologs can be found in human, mouse, *D. melanogaster* and *C. elegans* genomes [62], but not in plants. Drosha does not work in isolation, but forms a large complex of ~650 kDa with the DiGeorge syndrome critical region gene 8 (DGCR8) protein in mammals [61] or with the Pasha protein in *D. melanogaster* and *C. elegans* [63]. This complex was termed the microprocessing complex. DGCR8 is a ~120 kDa protein, containing dsRNA binding domains. The role of DGCR8 is unclear, but it is believed to assist Drosha in substrate recognition and in choosing the correct position on the pri-miRNA [61,63-65]. The Drosha complex must be able to recognize all pri-miRNAs despite their lack of sequence homology. It has been shown that the common tertiary structure of pri-miRNAs determines the substrate specificity. The efficiency of Drosha cleavage is dependent on the double-stranded stem structure, which is about 30 bp long, terminal loop size (≥ 10 nt) and flanking sequences [60,66,67]. After binding of the Drosha/DGCR8 microprocessing complex to the pri-miRNA, the two Drosha RNAse III domains form an intramolecular dimer and introduce cuts on both strands of the stem approximately 22 nt from the stem/loop junction [67]. RIIIDa cleaves the 3' strand, while RIIIDb makes a cut at the 5' strand of the pri-miRNA [61]. The flanking fragments are thought to be degraded. The released precursor miRNA has features typical of a RNAse III endonucleolytic cleavage: a 5' phosphate and a ~2 nt 3' overhang [60,68].

During the next step of miRNA maturation, the pre-miRNA is transported from the nucleus to the cytoplasm (Figure A.2 Step 3). Pre-miRNA export is mediated by the nuclear transport receptor Exportin 5 in a Ran-GTP dependent manner [69,70]. Exportin 5 is a member of the karyopherin family of nuclear/cytoplasmic factors. For nuclear cargo binding the karyopherins require the GTP-bound form of the Ran-GTPase. After nuclear export, cytoplasmic hydrolysis of Ran-GTP to Ran-GDP induces release of the cargo. Although Exportin 5 can also export tRNA and adenoviral VA1 non-coding RNA, pre-miRNAs are the main cargo. Pre-miRNAs have the characteristic structures required for recognition and export by Exportin 5: a ~2 nt 3' overhang and a RNA stem longer than 16 bp [71].

In the cytoplasm the ~70 nt pre-miRNA is processed to the mature ~22 nt miRNA by the class III RNAse III enzyme Dicer (Figure A.2, Step 4). Dicer is a conserved protein, which can be found in plants, insects, worms and mammals. Some organisms contain multiple Dicer homologs. For example, *D. melanogaster* has two: Dicer 1, which is required for miRNA processing and Dicer 2, which is responsible for siRNA generation [72]. Dicer is a large protein of about 200 kDa containing two catalytic RNAse III domains (RIIIDa and RIIIDb), a dsRBD, an Asp-Glu-Ala-Asp-box (DEAD-box) ATP-dependent RNA helicase domain, a Piwi-Argonaute-Zwilli (PAZ) domain and a domain of unknown function DUF283 (Figure A.4) [73].

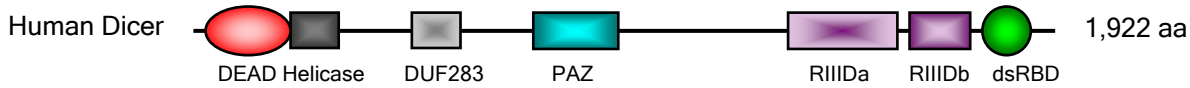

Figure A.4. Domain organization of human Dicer. The DEAD-box is shown in red, The RNA-helicase in dark grey, The DUF283 domain in light grey, The PAZ domain is shown in blue, The RIIID domains in violet, The dsRBD in green.

The DEAD-Box RNA helicase domain may have two functions: unwinding of the miRNA duplex and remodeling of the miRNA/miRISC complex [74]. The PAZ domain is a conserved RNA binding domain, unique to Dicer and Argonaute proteins, that recognizes the 3' overhang structure of pre-miRNA and preferentially interacts with 3' end. This suggests that the PAZ domain may be responsible for the recognition of substrates that have been generated by Drosha preprocessing [75]. Moreover, through this interaction the PAZ domain may position the site of Dicer cleavage on the stem of the pre-miRNA [76]. Like Drosha, Dicer forms a catalytic center by intramolecular dimerization of the RNAse III domains and cuts the dsRNA stem ~20 bp from its terminus, leaving the 5' phosphate and the ~2 nt 3' overhang. RIIIDa cleaves the 3' strand, and RIIIDb the 5' strand of the precursor [77]. The cleavage product, a ~22 nt dsRNA (miRNA:miRNA* duplex), is a transient intermediate. It is unwound by an unknown helicase, one of the strands, termed miRNA*, is degraded (Figure A.2, Step 5) and the other strand is incorporated into the miRISC (Figure A.2, Step 6). Which strand to select as the mature miRNA is determined by the relative thermodynamic stability of the two ends of the duplex. The arm with the less stable 5' end (G:U pair versus G:C pair) usually survives [78,79].

Recently it has been shown that Dicer works in a complex with two other proteins: the double-stranded RNA binding protein TRBP in humans (Loquacious in *D. melanogaster*), and the Argonaute 2 protein (Ago2) [80-83]. The functions of these two proteins in the Dicer complex are unclear. Presumably, TRBP assists substrate binding by Dicer, and Ago2 helps to release the cleavage product from the Dicer complex and to load it on to the miRISC.

A.2.3. miRNA incorporation into the miRISC

Following processing, the mature miRNA incorporates into the miRISC (Figure A.2, Step 6). RISC is a multi-turnover enzyme complex, carrying out target cleavage in the RNA*i* pathway but also translational repression in the miRNA pathway [84]. In addition, components of the RISC also participate in transcriptional silencing and heterochromatin formation [85,86].

The precise determination of RISC structure is still ongoing as additional components are continually being identified. The core components of RISC are Argonaute family proteins, also called PPD proteins because they all contain PAZ and PIWI domains. The number of PPD protein orthologs varies from 1 in *S. pombe* to 24 in *C. elegans* [87]. Mammals contain four Argonaute homologs: Ago1 to Ago4. Ago2 is unique in that it has RNAse activity (so-called slicer activity),

the function of the remaining Ago proteins in the context of RISC is less clear. Interaction of the PIWI domain with Dicer inhibits the Dicer ribonuclease activity [88]. Furthermore, the Ago proteins may engage the miRNA via the single-stranded RNA binding PAZ domain [75]. Taken together, Ago proteins may stimulate miRNA release from the Dicer complex and control the transfer of the miRNA to the miRISC.

In addition to Ago proteins, several other proteins were co-purified with RISC. However, it is not clear if they are also core components of the complex or accessory proteins that provide functional specificity for the RISCs. Gemin3 (a DEAD-box putative RNA helicase) and Gemin4 proteins were shown to coimmunoprecipitate with human Ago proteins and miRNAs [30]. Vasa Intronic Gene (VIG) and Fragile-X-related Mental Retardation Protein (FMRP) – two RNA-binding proteins – were also shown to be associated with RISC [89,90]. VIG is a conserved protein of unknown function containing a cluster of arginine and glycine residues (RGG-box, an RNA binding motif). FMRP is a conserved protein that contains several RNA binding domains including two K homology (KH) domains and a RGG-box. FMRP forms a complex with mRNA (mRNP, messenger ribonucleoprotein). This mRNP complex associates with polyribosomes to suppress protein translation [91]. The association of FMRP with the miRISC strongly suggests that FMRP could regulate translation through the miRNA pathway.

Recently two novel factors were found to co-purify with the RISC: the putative DEAD-box RNA helicase MOV10 and the RNA-recognition motif-containing protein TNRC6B/KIAA1093 [92]. Mammalian MOV10 is similar to Armitage in *D. melanogaster* and might be required for RISC maturation and assembly of siRNA into functional RISC [93]. TNRC6B is a poorly characterized protein with homology to the GW182 protein, a component of processing bodies (P-bodies, also termed GW-bodies) [94].

Several groups have analyzed the localization of miRISC components and miRNA in mammalian cells. It has been shown that Ago proteins, MOV10 and TNRC6B are all localized to cytoplasmic P-bodies [95-97]. Furthermore, miRNA binding to target mRNAs results in their association with P-bodies. P-bodies are small, discrete cytoplasmic foci that serve as centers for mRNA decay. P-bodies have been observed in animal and yeast cells [98]. The core components of P-bodies are GW182 and Dcp1 (decapping) proteins. Both proteins are involved in the post-transcriptional regulation of gene expression by sequestering specific mRNAs. Liu and co-workers have shown that Ago proteins physically interact with GW182 [96]. Moreover, GW182 also has a functional role in miRNA-mediated silencing: repression of GW182 leads to downregulation of miRNA functions [96]. However, the question still remains whether P-bodies are storage sites for the miRNA-RISC complex, or whether miRNA-dependent translational repression takes place in P-bodies.

A.3. Mechanism of miRNA function

Once incorporated into the miRISC, miRNAs interact with specific binding sites in the 3'UTR of target mRNAs and downregulate their expression, either by translational repression or by mRNA degradation (reviewed in [8]). However, the molecular details of miRNA-mediated suppression are still being investigated.

A.3.1. Translational repression or mRNA cleavage by miRNA

Currently, there are two models for target mRNA regulation by miRNAs. The first model proposes that miRNA with nearly-perfect complementarity to target sites direct mRNA cleavage, like siRNA, that is then followed by mRNA degradation. According to the second model, partial complementarity between the miRNA and the mRNA results in translational repression without mRNA cleavage. In plants, miRNAs generally have near perfect complementarity to their mRNA targets [99,100]. In the animal kingdom, only a few instances of near-perfect complementarity to known miRNAs have been found. Therefore, the first model with mRNA cleavage is considered to be the primary pathway for plants, and translational repression is thought to be more prevalent in animals. However, there are some exceptions. In several cases miRNAs can guide mRNA cleavage in animals. For example, the 3'UTR of the *HOXB8* mRNA has some near-perfect sites for *mir-196a*. *mir-196a* triggers the cleavage of the *HOXB8* mRNA during mouse embryonic development [101,102]. In addition, in several experiments it has been shown that miRNAs can direct the cleavage of artificial target mRNAs designed to have perfect complementarity [84,103]. Conversely, siRNAs can mediate translational repression of partially complementary targets [104]. These observations suggest that the degree of complementarity between the miRNA and its binding sites determines the mechanism of regulation (translational repression or mRNA degradation).

Several mechanistic studies of translational repression have led to different conclusions. In the original studies of the *lin-4/lin14* interaction in *C. elegans* evidence was presented that repression occurs after translational initiation. The *lin-14* mRNA was found to stably associate with polysomes, suggesting either an arrest of translational elongation (ribosomal stalling), or co-translational peptide degradation as mechanisms [18]. In contrast, Pillai and colleagues have presented evidence for inhibition of translational initiation by *let-7* in human cells. They found that: 1) mRNA targets accumulated in P-bodies in a miRNA-dependent manner; 2) mRNAs in P-bodies were not engaged in translation; and 3) miRNAs prevented the loading of mRNAs to polysomes. In addition, alteration of translation initiation, for example by tethering the translation factors eIF-4E or eIF-4G directly to a mRNA, can make the mRNA resistant to miRNA–induced repression [105]. A further unresolved question relates to the targeting of the miRNA/mRNA complex to P-bodies. It is unclear

whether P-bodies are the site of translational inhibition by the miRNA machinery, or if P-bodies represent a storage area for repressed mRNA (reviewed in [106]).

A.3.2. Prediction and validation of miRNA targets

miRNA/target mRNA interactions have been found to share common features, allowing several groups to develop bioinformatic approaches to predict and identify potential miRNA target genes in mammals and flies [107-109]. All these approaches were based on the analysis of base-pairing between miRNAs and experimentally validated targets such as *let-7* and *lin-41* in *C. elegans*. The algorithms for identifying the targets rely on both thermodynamics-based modeling of miRNA:mRNA interactions and comparative sequence analysis and include the following criteria:

- The location of miRNA complementary elements in the 3'UTR of the mRNA.
- The concentration of nearly-perfect Watson-Crick complementarity in the 5' end of the miRNA ("seed" or "core" region which is believed to be most critical for target recognition).
- The free energy of folding (ΔG) for the miRNA:mRNA duplexes.
- The phylogenetic conservation of binding sites between orthologous genes.

Many of the prediction databases for miRNA targets are available on the internet, for example, TargetScan software (http://genes.mit.edu/targetscan), miRanda (http://www.microrna.org/), PicTar (http://pictar.bio.nyu.edu) and RNA22 (http://cbcsrv.watson.ibm.com/rna22.html).

As a next step in target identification the computationally predicted targets should be experimentally validated, for example, by fusing the UTR from the predicted targets to a reporter construct and analyzing the correlation of reporter and miRNA expression. The combination of *in silico* and *in vitro* analysis has led to the proposals that a single miRNA can target hundreds of mRNAs in different cell and tissue types; that more than one miRNA may interact with one message; and that miRNAs may regulate a very broad diversity of biological processes [107,109].

A.3.3. Principles of miRNA target recognition

How exactly miRNAs recognize their targets remains unclear, but there are some rules which seem to be important and common for target recognition and interaction processes. Cohen's group has evaluated minimal requirements for a functional miRNA:mRNA duplex [110]. They characterized two categories of miRNA target sites based primarily on the strength of 5´-base pairing to the miRNA. The first category consisted of "5' dominant" sites, which required extensive base-pairing to the 5' end of the miRNA. The minimal length of the 5' "seed" sufficient to regulate the target was 7-8 nt. This type of target had two subtypes: "canonical" and "seed" sites. "Canonical" sites paired at the 5' as well as 3' ends; "seed" sites did not require any 3' pairing. miRNA binding sites

of the second category were termed "3' compensatory" sites. The 5' "seed" of this group of sites is 4-6 nt and may contain single nucleotide bulges, G:U pairing or single mismatches. The weaker 5' binding required compensatory pairing to the 3' end of the miRNA in order to confer target regulation. Currently identified miRNA target sites are almost all canonical.

Moreover, Vella et al. have shown that sequences surrounding binding sites may also be important for target recognition and regulation. Of the six *let-7* complementary sites (LCS) found in the 3' UTR of the *C. elegans lin-41* mRNA only two (LCS1 and LCS2) were necessary and sufficient for regulation of the target mRNA. In addition, they provided evidence that a 27 nt sequence between LCS1 and LCS2 is necessary for target downregulation *in vivo.* This 27 nt spacer may be important for the secondary structure of the 3' UTR or required for binding additional RNA or protein factors [25].

A.3.4. miRNAs and siRNAs: What's the difference?

No formal connection between miRNA and siRNA pathways had been made until 2001, when Dicer, the enzyme that processes long double-strand RNAs (dsRNA) into siRNAs [73], was shown to convert ~70 nt stem-loop pre-miRNAs into mature miRNAs [87,111,112]. The link between RNAi and miRNA metabolism was further established by the characterization of the miRISC. The miRISC and RNAi effector complex (siRISC) were shown to overlap in both function and protein composition (reviewed in [74]) Nevertheless, a distinction between these two classes of small non-coding RNAs can be made with regard to their origin, evolutionary conservation, mechanism of target gene silencing and the type of genes that they regulate (Table A.1, reviewed in [8]).

Table A.1. Differences between miRNA and siRNA

miRNAs	siRNAs
miRNAs are found in various metazoan organisms including plants, worms, flies, and mammalians, but not in fungi or bacteria	endogenous siRNAs can be found in fungi, but not in flies, worms or mammalians
miRNAs are derived from endogenous genomic loci distinct from other genes	siRNAs are often derived from transposons, heterochromatic DNA, mRNAs, viruses or from artificially introduced dsRNAs
miRNAs are processed from transcripts forming stem-loop structures	mature siRNAs are derived from long double-stranded RNAs (~500 bp)
A single miRNA:miRNA* duplex is generated from the stem-loop precursor	one long dsRNA is processed by Dicer into many ~22nt siRNA duplexes
miRNA sequences are conserved in related organisms	endogenous siRNA sequences are rarely conserved
miRNAs specify "hetero-silencing" of their targets, regulating different genes	siRNAs specify the silencing of the same locus from which they originate ("auto-silencing")
Usually, miRNAs bind to the target 3'UTRs with imperfect complementarity at multiple sites and perform translational repression of target mRNAs	siRNAs form a perfect duplex with only one binding site of their targets and direct the cleavage of the target mRNAs at the site of complementarity

In order to understand the true biological functions of both classes of small non-coding RNAs it is necessary to focus on the details that distinguish the modes of siRNA and miRNA action *in vivo*.

A.4. Functional characterization of miRNAs

Hundreds of miRNAs have been cloned and described, and thousands of targets predicted, but experimentally validated functions have been described for only a few miRNA-mRNA regulatory pairs. miRNAs have a diverse spectrum of functions including developmental timing, cell proliferation and apoptosis, cell and tissue fate decisions, maintenance of tissue identity, and either tumor suppression or oncogenic activity. The significance of miRNAs in development has been most convincingly demonstrated in plants, where the RNAi and miRNA pathways are closely intertwined. For example, 15 of the first 17 plant miRNAs to be identified were found to target developmentally regulated transcription factors [113]. The initial difficulties in target gene prediction, and the dearth of genetic tools, have impeded similar investigations in animals. Nevertheless, in several instances the involvement of miRNAs in fundamental cellular processes has been established using forward (loss- and gain-of-function genetic screens) and reverse (miRNA knockdown or overexpression) genetics approaches. The known biological functions of miRNAs are reviewed in [11,43,114] and listed in Table A.2

Table A.2. Summary of miRNA biological functions.

miRNAs	Target(s)	Functions	References
C. elegans			
lin-4	*lin-14, lin-28*	Regulation of early developmental timing	[20,115]
let-7	*lin-41, hbl-1, ras*	Regulation of late developmental timing	[21,23]
lsy-6	*cog-1*	Determination of left/right neuronal asymmetry	[59]
mir-273	*die-1*	Determination of left/right neuronal asymmetry	[116]
D. melanogaster			
bantam	*hid*	Suppression of apoptosis, induction of cell proliferation	[117]
mir-14	*Drice ?*	Suppression of apoptosis, regulation of fat metabolism	[118]
mir-7	*Notch targets*	Regulation of Notch signalling pathway	[119]
Danio rerio			
mir-430	*?*	Regulation of brain morphogenesis	[120]
Mus musculus			
let-7	*lin-41*	Limb development	[121]
mir-125	*lin-28*	Regulation of neural differentiation	[122]
mir-196	*HoxB8*	Regulation of developmental patterning, limb development	[102,123]
mir-181	*?*	Promotion of haematopoietic cell specification towards the B-cell lineage	[55]
mir-181	*HoxA11*	Establishment of myoblast differentiated phenotype	[124]
mir-133	*?*	Control of skeletal muscle proliferation and differentiation	[125]

mir-1	*Hand2*	Regulation of cardiomyocyte differentiation and proliferation, skeletal muscle formation	[126]
mir-9	*STAT3*	Determination of neural fates in ES cell differentiation	[127]
mir-375	*Mtpn*	Suppression of insulin secretion	[128]
mir-122	*Tnp2*	Role in mouse spermatogenesis	[129]
Rat			
mir-134	*LimK1*	Regulation of dendritic spine development	[130]
Homo sapiens			
mir-143	*Erk5 ?*	Promotion of adipocyte differentiation	[131]
mir-32	*Retrovirus PFV-1*	Provision of antiviral defence	[132]
mir 16	several	AU-rich element-mediated mRNA instability	[133]
Cancer			
let-7	RAS	Downregulated in lung carcinoma, tumor suppression activity	[134]
mir-15, mir-16	?	Downregulated in B-cell chronic lymphocyte leukemia	[44]
mir-143,145	?	Downregulated in colon cancer tissue	[135]
mir-155	?	Upregulated in Burkitt's lymphoma	[56]
mir-17-92	?	Upregulated in B-cell lymphoma, promotion of oncogenesis	[136]
mir-17-92	E2F1	Regulation of c-Myc mediated cellular proliferation, tumor suppression activity	[137]

Gene abbreviations: *hbl-1* – hunchback like gene; *cog-1* – connection of gonad defective; *die-1* – dorsal intercalation and elongation defect; *hid* – head involution defective; *hoxB8* – homeobox B8; *hoxA11* – homeobox A11; *Hand2* - heart and neural crest derivatives expressed transcript 2; *STAT3* – signal transducer and activator of transcription 3; Mtpn – Myotropfin; *Tnp2* – Transition protein 2; *LimK1* – Lim-domain-containing protein kinase 1; *Erk 5* - extracellular signal–related kinase; Retrovirus PFV-1 - retrovirus primate foamy virus type 1.

A.5. The role of miRNA in nervous system and neuronal development

At the beginning of this thesis, we were attracted to the hypothesis that miRNAs play an important role in the generation of the enormous cellular diversity and the elaborate architecture that characterize the mammalian CNS. Over 80 miRNA genes have been detected in the mammalian brain, some of which are temporally regulated during development [28,33,34,37,48,53]. Evidence for the functional significance of neural miRNAs in gene regulation is beginning to accumulate. Genetic studies in *C. elegans* point to a significant role for miRNAs in the developmental regulation of neural gene expression. One of the functions of the *lin-14* gene is to control remodeling in the nervous system. Normally, during the first larval stage motor neurons innervate ventral body wall muscle and receive synaptic input dorsally. At a defined stage in development these neural cells reverse their polarity and innervate dorsal muscle. By controlling *lin-14* expression, which represses this reorganization, *lin-4* is responsible for the timing of this dramatic synaptic remodeling event [138]. The mammalian paralog of *lin-4 (mir-125)* has been cloned, and is expressed in mouse brain [28]. Recently, *mir-125* has been shown to downregulate mammalian *lin-28* during neural differentiation of embryonal carcinoma (EC) cells, suggesting a role for *mir-125* in mammalian neural differentiation [122]. Another role of *lin-4* is the regulation of the

hunchback gene, in cooperation with *let-7*. Hunchback is one of several transcription factors which determine the cell fate of neural progenitors in the *D. melanogaster's* CNS [22,23,139].

Furthermore, left/right asymmetry in the patterning of neuronal gene expression in *C. elegans* is established by inverse and sequential expression of two miRNAs, *lsy-6* and *mir-273* [59,116] *lsy-6* is expressed exclusively in left sensory neurons (ASEL) and inhibits the expression of the Nkx-type homeobox gene *cog-1*. Cog-1 is a repressor of the characteristic chemoreceptor in ASEL neurons, the guanylyl cyclase receptor (GCY-7). Right side sensory neurons (ASER) specifically express the *mir-273* miRNA, a repressor of *die-1*. Die-1 is a transcription factor needed for l*sy-6* transcription. Therefore, *mir-273* leads to both downregulation of l*sy-6* expression in ASER and also to upregulation of the GCY-5 chemoreceptor in ASER.

In mammals, a number of studies have used neural differentiation of ES and P19 EC cells to model the expression and function of brain-specific miRNAs during neurogenesis [48,54]. Krichevsky et al. have shown that overexpression of *mir-124a*, *mir-9*, *mir-125b* and *mir-22* in neural progenitor cells reduced differentiation to astroglial-like cells, while knockdown of *mir-9* increased the number of astroglia and reduced the number of neurons generated in culture [127]. Phosphorylated signal transducer and activator of transcription 3 (STAT3) has been shown to inhibit the neural terminal differentiation and selectively enhance differentiation of glial-like cells. *mir-9* is able to regulate the phosphorylation state of STAT3 [127]. Thus *mir-9* may regulate the balance between neurogenesis and gliogenesis in this cell model. A similar role has been demonstrated for the small non-coding RNA *NSRE*, which promotes neural specification by inactivating the NRSF/REST transcription factor. NRSF/REST is a repressor of neuron-specific gene expression; forced overexpression of the *NSRE* RNA in neural stem cells promotes neural differentiation at the expense of the astrocytic lineage [140].

miRNAs may also regulate synaptic development. Schratt and co-workers have shown that brain-specific *mir-134* negatively regulates dendritic spine volume in hippocampal neurons by controlling Lim-kinase 1(*LimK1*)-mediated spine development. *mir-134* can inhibit *LimK1* mRNA translation locally in dendrites in response to extracellular stimuli such as brain-derived neurotrophic factors (BDNF) that restrict the growth of dendritic spines [130].

A second line of evidence suggesting a functional role for miRNAs in mammalian neural development comes from studies of miRNA biogenesis. Several protein components of the miRISC are critical for CNS development. The Gemin3 and Gemin4 proteins are known to associate with the Survival of Motor Neurons protein, a protein that is defective or absent in the neurodegenerative syndrome Spinal Muscular Atrophy [30]. The RNA binding protein, FMRP, that is absent or mutated in patients with Fragile X Mental Retardation Syndrome, has been shown to play an important role in synaptic plasticity by regulation of mRNA transport and local protein synthesis at synapses. The role of FMRP in synaptic plasticity is supported by the observation of abnormal dendritic spines in

the brain of FMR1-knockout mice and in patients with Fragile X syndrome (reviewed in [141]). *Armitage*, a *D. melanogaster* homolog of the *MOV10* gene, was shown to be required in classical long term potentiation [142]. Targeted mutations in additional miRISC genes have been reported. The phenotypes are strongly supportive of vital developmental functions for miRNA in general, and in neural development in particular. Thus, Ago1 mutants are embryonic lethal, with defects in the CNS and PNS [143]; Ago2 mutants are also embryonic lethal and show failure of neural tube closure and neural tube deformity [144]. Dicer mutant mouse embryos die during gastrulation and have stem cell loss [145]. In zebrafish, maternal-zygotic Dicer mutants lacking all miRNAs show abnormal morphogenesis including defects in brain formation and neural differentiation. Interestingly, injection of *mir-430* rescued several defects in neural development, suggesting a role for this miRNA in zebrafish brain morphogenesis [120].

In summary, all these findings confirmed by the data presented below, implicate miRNAs as potentially widespread and significant regulators of neural gene expression during development and regeneration.

A.6. Aim of the thesis

The principal goal of this thesis entails an investigation into the role of miRNAs, in particular the *let-7* family, in the control of developmental timing in the mammalian nervous system and in the specification of neural cell fate during stem cell differentiation.

In pursuit of this goal, the following experimental program was adopted:

- Analysis of the expression kinetics of set of neural miRNAs in the developing brain, in primary neurons and astrocytes, and during neural differentiation of EC and ES cells.
- Comparison of miRNA processing pathway regulation in undifferentiated and differentiated stem cells.
- Determination of miRISC localization in EC and ES cells as well as in primary neurons.
- Development of an experimental system for the functional analysis of miRNAs and for target gene validation.
- Visualization of miRNA expression in primary cells and in P19 EC cells.
- Analysis of miRNA function: is miRNA expression important for EC cell differentiation?