Chapter 1

Introduction

In the first part of the Introduction I describe the discovery of miRNAs and how this class of molecules came into the scientific limelight. I also describe findings that demonstrate functions and roles of miRNAs, and their importance for the brain and neurons. The second part of the Introduction states the aims of this thesis, and introduces the experimental paradigm, that have been followed during the course of the thesis project.

1.1 Significance of miRNAs

1.1.1 Discovery of miRNAs

Nucleic acids were discovered in the nineteenth century, and a function for RNA was described in the fifties of the twentieth century, when it was identified to be a messenger between DNA and protein. The perception of RNA as a passive ancillary carrier of genetic information changed dramatically during the following six decades. Discoveries, including introns capable of self-splicing by Thomas Cech, the RNase P that cleaves tRNA using RNA at its core by Sidney Altman, and ultimately the discovery of Venkatraman Ramakrishnan, Thomas Steitz and Ada Yonath that synthesis of all proteins in the ribosome is catalyzed solely by its RNA component shifted RNA to the center of biology. The hypothesis of a primordial RNA-world became generally accepted, and few scientists doubt the importance of RNA-dependent mechanisms for evolution and the existence of life's complexity, such as retrotransposition driven genome rearrangements and alternative splicing.

Contrary to the original functional paradigm of RNAs as DNA-protein intermediates, more recently discovered regulatory RNAs do not encode proteins, but have a separate

functional significance of their own. One type of regulatory RNAs, called miRNAs, is the subject of this thesis. The first miRNA was discovered in 1993 by Victor Ambros and Gary Ruvkun (Lee et al., 1993; Wightman et al., 1993) in *Caenorhabditis elegans* (*C. elegans*) as a post-transcriptional regulator of gene expression. The discovered miRNA, called *lin-4*, was found to decrease expression of its target gene, *lin-14*, through interaction of the *lin-4* miRNA with the transcript of *lin-14*, leading to the decrease in levels of LIN-14 protein. Since their discovery in 1993, miRNAs have been identified in all multicellular animals and plants and also in some unicellular plants (Grimson et al., 2008). The official repository of information concerning metazoan miRNAs is miRBase (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008) (http://www.mirbase.org/). According to the current release (miRBase Release 16), there are 667 and 1,049 known genes for miRNAs in the mouse and human genomes respectively.

Curiously, after their original discovery, miRNAs received relatively little attention from the scientific community. However, studies of a different type of non-coding RNA, that were also performed in *C. elegans*, eventually brought miRNAs to the attention of scientist world wide. In 1995, an injection of antisense RNA into *C. elegans*, was shown to repress expression of a gene to which it was complementary, a phenomenon that was later called RNA interference (RNAi) (Guo and Kemphues, 1995). Mysteriously, the injection of the sense sequence also induced RNAi. Three years later, Andrew Fire and Craig Mello found an explanation by showing that double-stranded RNA (dsRNA) was the effective trigger of the phenomenon. Soon after the role of dsRNA was established, RNAi was identified in *Drosophila* (Kennerdell and Carthew, 1998), successfully used to silence a gene in *Xenopus* (Oelgeschläger et al., 2000) and was described in mice (Svoboda et al., 2000; Wianny and Zernicka-Goetz, 2000). Moreover, RNAi explained the enigmatic phenomenon of post-transcriptional silencing of endogenous genes by clones of homologous sequences, which was reported in plants (Napoli et al., 1990; van der Krol et al., 1990; Smith et al., 1990; de Carvalho et al., 1992) and fungi (Romano and Macino, 1992; Cogoni et al., 1996). As RNAi is triggered by tiny amounts of dsRNA, it suggested the existence of mechanisms that can propagate and sustain the RNA-mediated gene silencing (as opposed to purely stoichiometric sense-antisense interactions) in all major branches of eukaryotic tree of life. Very soon after the discovery of dsRNA as the trigger of RNAi, several groups showed that in both plants (Hamilton and Baulcombe, 1999) and animals (Hammond et al., 2000; Zamore et al., 2000; Elbashir et al., 2001) the effective dsRNA was converted to short functional oligomers (21 to 25 nucleotids long), that were termed short interfering RNAs (siRNAs). The discovery of siRNAs, which were chemically identical

to several then known miRNAs, suggested that RNAi and miRNA induced regulation of gene expression were related processes, with miRNAs being an endogenous form of siRNAs (siRNAs and miRNA and other short RNAs are together referred to as sRNAs). As RNAi was demonstrated in a variety of species, these insights prompted the cataloging miRNAs and other endogenous sRNAs in various species. The discoveries of new sRNAs followed soon. In the year 2001, Tomas Tuschl's laboratory were able to clone dozens of miRNAs from *Drosophila* embryos and human cell cultures (Lagos-Quintana et al., 2001), which confirmed that a pool of sRNAs existed naturally in different animals. Finally, 8 years after their discovery, miRNAs came into the scientific limelight.

In the ensuing decade, miRNAs and other sRNAs remained at the frontier of research world-wide. Entirely new classes of sRNAs were discovered in metazoans, including Piwi-interacting RNAs (piRNAs) (Aravin et al., 2007), recently indentified promoter-associated short RNAs (PASRs) (Taft et al., 2009) and splice-site RNAs (spliRNAs) (Taft et al., 2010). Two excellent reviews give a comprehensive summary of information about siRNAs and piRNAs (Carthew and Sontheimer, 2009; Malone and Hannon, 2009). There are several important distinctions between plant and animal miRNAs, despite general principles being similar in both kingdoms (Voinnet, 2009). The biogenesis and function of animal[1] miRNAs is described in the next section.

1.1.2 Biogenesis and molecular mechanisms of miRNA function

The main role of miRNAs in the cell is the regulation of expression of genes (miRNA targets) at a post-transcriptional level. The absence of direct miRNA-DNA interactions (Sharp, 2009) and demonstration of co-localisations of miRNAs and mRNAs to the cytoplasmic compartments (Liu et al., 2005) supported this view. As in the case of siRNAs, a majority of published reports showed miRNAs to have an inhibitory effect on expression of their targets. RNAi, a cleavage of the mRNA triggered by siRNAs (see section 1.1.1), is catalyzed by an enzymatic complex to which siRNAs are bound and which they direct to the mRNA targets. This protein complex is located in the cytoplasm, and is known as RNA-induced silencing complex or RISC (Hammond et al., 2000). Soon after the discovery of RISC being a catalytic machine of RNAi, it was demonstrated that miRNAs were also associated with RISC (Hutvágner and Zamore, 2002). Complementarity between ≈

[1]The work of this thesis is focused on miRNAs in the mouse, therefore the introduction to miRNA biology in animals is also focused on the mouse. Genes, transcripts and protein names will, by default, refer to those in the mouse, unless specified otherwise. Conventions for the mouse notations will be used (gene names are in italic with the first letter capitalised, while products of the genes (transcripts and proteins) are in a regular font with the first letter capitalised), unless specified otherwise.

7 bases located at the 5′-end of miRNA (called the seed region) and an mRNA transcript was found to be a good predictor of the transcript being targeted by the miRNA (Lewis et al., 2003, 2005). Properties of the seed region and the target sites (called the seed matching sites) are discussed in detail in Introduction, section 1.2.1. Here it should be noted that the requirement of only a partial complementarity between a miRNA and an mRNA transcripts enables, in principle, a single miRNA to target hundreds of mRNA transcripts (Enright et al., 2003; Stark et al., 2003; Farh et al., 2005; Lim et al., 2005; Giraldez et al., 2006; Baek et al., 2008; Selbach et al., 2008).

Biogenesis of miRNAs

The ≈ 22nt miRNAs that are incorporated into RISC and act as its guide are sometimes referred to as mature miRNAs. Mature miRNAs originate from longer transcripts, called primary-miRNAs (pri-miRNAs), that are produced by Pol II (Polymerase II) transcription[1], and are capped and polyadenylated (Lee et al., 2004; Cai et al., 2004). Frequently, pri-miRNA transcripts give rise to more than one mature miRNA (i.e. they are polycistronic), or, as was shown for ≈ 40% of human miRNAs, pri-miRNA transcripts can also encode a protein sequence, in which case mature miRNA sequences are usually located within introns and are called intragenic (Kim et al., 2009). Within a pri-miRNA, the sequence of a mature miRNA is within a secondary structure, a hairpin (Winter et al., 2009; Kim et al., 2009). These hairpins are recognized in the nucleus by the enzyme-complex that is sometimes referred to as the Microprocessor, the principal component of which is an RNase III type endonuclease, called Drosha (Lee et al., 2002). Drosha introduces a cut in the stem of the hairpin within the pri-miRNA releasing a shorter hairpin (with the stem ≈ 33 nt), called the precursor-miRNA (pre-miRNA). Characteristically for the RNase III type endonucleases, when Drosha cuts the base of the dsRNA hairpin, it leaves a ≈ 2nt overhang of the 3′RNA-end, and a phosphate at the 5′-end (Basyuk et al., 2003; Lee et al., 2003). The Drosha cut is important, as it produces one end of the mature miRNA. Interestingly, in the case of a few intronic miRNAs, called mirtrons, their splicing produces pre-miRNAs directly, thus bypassing the Microprocessor (Babiarz et al., 2008).

The next step, pre-miRNA is exported from the nucleus to the cytoplasm via the Exportin 5 complex (Yi et al., 2003; Bohnsack et al., 2004). Upon export to the cytoplasm, pre-miRNAs are recognized and cleaved by another protein complex, a principle

[1]Pol III transcription was reported to produce pri-miRNAs in case of Alu-element derived miRNAs (Borchert et al., 2006), but such cases are rare.

component of which is an RNase type III enzyme, called Dicer (encoded by a single gene in the mouse, *Dicer1*) (Grishok et al., 2001; Hutvágner et al., 2001; Bernstein et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). Dicer introduces a cut into the pre-miRNA, removing the loop from the stem, and generates the second end of the mature miRNA. This end also has ≈ 2 nt overhang at the 3′-end and a phosphate at the 5′-end. The RNA, that is generated by Dicer, is a ≈ 22 nt double-stranded oligonucleotide, one strand of which is going to become a mature miRNA (called a guide strand). The duplex is thought to be loaded into RISC in a process dependant on the interaction of Dicer with the Argonaute component of RISC (Ago proteins (Höck and Meister, 2008; Joshua-Tor and Hannon, 2010)) and other ancillary proteins (Chendrimada et al., 2005; Maniataki and Mourelatos, 2005). However, miRNA duplexes introduced into a Dicer-null background have been reported to efficiently downregulate expression of the targets despite lacking the Dicer protein (Hanina et al., 2010). Therefore, at least for exogenously produced miRNA duplexes, RISC assembly can be independent of Dicer. To produce a functionally competent RISC one of the two strands of the duplex is degraded (passenger strand). The strand of the duplex with the weakest hydrogen bonds at its 5′-end is more likely to survive (Schwarz et al., 2003; Khvorova et al., 2003). Mechanism of degradation of the passenger strand is not enirely clear, as endonucleolytic activity of the Argonaute component is not essential for assembly of the miRNA RISC (Matranga et al., 2005; Leuschner et al., 2006). Recently, Dicer independent generation of one mature miRNA (miR-451) from pre-miRNA was reported (Cifuentes et al., 2010; Yang et al., 2010). However, as in the case of Microprocessor independent miRNAs, Dicer independent miRNAs are thought to be rare and currently only one such miRNA is known (Cifuentes et al., 2010; Yang et al., 2010).

Molecular mechanisms of miRNA function

Historically, the function of miRNAs was thought to be the inhibition of the translation of targeted mRNAs, while siRNAs were thought to trigger endonucleolytic cleavage of targeted mRNAs. However, the destabilization and subsequent degradation of target mRNA transcripts was subsequently observed (Lim et al., 2005; Giraldez et al., 2006). Subsequently, a significant correlation was detected between the regulatory effects of several miRNAs measured at mRNA and protein levels (Baek et al., 2008; Selbach et al., 2008). Moreover, according to recent reports, the inhibitory effect of several miRNAs on protein production is predominantly explained by the reduction in mRNA levels (Hendrickson et al., 2009; Guo et al., 2010). The ability of miRNAs to destabilize target mRNAs was

at the basis of the experimental paradigm in this thesis, where lowering of mRNA levels served as an indication of miRNA activity (see Introduction, section 1.2.2).

The regulatory function of miRNAs is not exerted by the miRNA itself. The same is true for siRNAs during RNAi. miRNAs bind the RNA induced silencing complex (RISC), which it guides to target transcripts (Fabian et al., 2010). The target recognition by miRNAs is observed in complementarity of miRNA sequences to the sequence of their target transcript. The rules of guiding itself (i.e. recognition of target transcripts by miRNAs) will be discussed in section 1.2.1, in conjunction with the use of these rules for the computational prediction of miRNA targets. Below is a brief overview of the mechanisms by which RISC destabilises the target transcript and/or represses its translation after it was brought to the transcript by the miRNA.

A protein of the Argonaute family is the component of RISC that directly binds the miRNA (Peters and Meister, 2007). There are four genes in the mammalian genome that encode members of Argonaute family (*Ago1*, *Ago2*, *Ago3* and *Ago4*), each of which can bind miRNAs (Azuma-Mukai et al., 2008; Landthaler et al., 2008). All four Ago proteins are thought to be functionally competent as RISC components, although only Ago2 possesses endonucleolytic activity, which catalyzes RNAi (Liu et al., 2004; Baillat and Shiekhattar, 2009). As a consequence, the main role of Ago proteins in animal miRNA mediated regulation is thought to be either in the repression of translation or in directing mRNA targets to components of a generic mRNA degradation machinery (Fabian et al., 2010). One of the proteins that binds to Ago and that is essential for both miRNA mediated mRNA destabilisation and translational repression is GW182 (Eulalio et al., 2008). In fact, both mRNA destabilisation and the repression of translation can be triggered by chemical tethering of GW182 to mRNA, in the absence of miRNAs and Ago (Pillai et al., 2004; Behm-Ansmant et al., 2006; Chekulaeva et al., 2009; Zipprich et al., 2009), which confirms that GW182 operates downstream of Ago. GW182 is thought to recruit the mRNA to the deadenylase complex, which triggers deadenylation of mRNA transcripts, followed by their decapping and subsequent degradation by exonucleases (Eulalio et al., 2009). The inhibition of translation occurs both during initiation and elongation stages. During initiation RISC can interfere with assembly of the initiation complex at the Cap of the mRNA (Pillai et al., 2005; Humphreys et al., 2005; Wang et al., 2006; Wakiyama et al., 2007; Thermann and Hentze, 2007; Mathonnet et al., 2007) or with assembly of the 80S ribosome (Wang et al., 2008a; Chendrimada et al., 2007). During the elongation stage RISC can cause ribosomes to stall and trigger them to drop-off the transcript (Gu et al., 2009; Olsen and Ambros, 1999; Nottrott et al., 2006; Maroney et al., 2006; Petersen et al.,

2006). More detailed information on molecular mechanism of miRNA mediated mRNA destabilisation and translational repression is described elsewhere (Fabian et al., 2010).

There are currently three reports of miRNAs activating gene expression. Two of these described miRNA targeting of mRNA transcripts through seed matching sites in 5′UTRs (which is unconventional for miRNA-mRNA interaction, see section 1.2.1). Subsequently the activation of translation of mRNA transcripts was observed (Ørom et al., 2008; Henke et al., 2008). In another report, the activation of translation of miRNA targets was observed in a human cell culture system upon the stress of serum withdrawal and cell-cycle arrest (Vasudevan et al., 2007). It is still unclear, however, how general and reproducible these observations are, and the activatory roles of miRNAs will not be further discussed in this Introduction[1].

1.1.3 miRNAs in brain development and neuronal function

Despite research of miRNA function outside of the *C. elegans* developmental paradigm commencing only a decade Ago, their significance for neuronal biology was already obvious by 2007, when this thesis project was designed. The evidence of the importance of miRNAs for brain development and neurogenesis came, primarily, from the ablation of components of the miRNA biogenesis pathways (e.g. *Dicer1* deletions) and simultaneous removal of nearly all miRNAs as a consequence. Further details of the roles of miRNAs in determining neuronal identity and neuronal function were revealed through experiments with individual miRNAs. Results of these experiments are discussed in the current section with the purpose of illustrating importance of miRNAs for brain development, the establishment of neuronal identity and neuronal function.

miRNAs in brain development

The depletion of all miRNAs by the disruption of miRNA biogenesis provided some of the first insights into the function of miRNAs in brain development. Dicer is an RNase III type enzyme that is indispensable for biogenesis of all but one known miRNA (Cifuentes et al.,

[1]Analysis of the results of experiments performed in this thesis project, where miRNAs were exogenously added to cultured neurons, were not consistent with activatory activity of overexpressed miRNAs (Chapters 4 and 5). However, a signature of activity of endogenous miRNAs in certain conditions, such as the stresses described in Chapter 6 (section 6.2.1) was consistent with the relief of miRNA mediated regulation during stresses, or, perhaps, a switch to an activatory mode. Before such conclusions can be drawn, additional experiments specifically designed to directly test this proposition are required to convincingly demonstrate that the modulation of miRNA regulation in primary neurons under stresses does take place (see Discussion, section 7.4).

2010). There is one copy of *Dicer1* gene in the mouse and human genomes and a stable mouse knockout (Dicer-null) was generated (Bernstein et al., 2003). Development of Dicer-null embryos did not proceed beyond 7.5 days of development, which is before formation of the body plan during gastrulation (Bernstein et al., 2003). Although this result is likely to mean that Dicer and miRNAs are essential for mouse development, it did not prove that Dicer and miRNAs are important for brain development *per se*, as developmental arrest occurred too early for this conclusion to be drawn. Ablation in mice of one of the four Ago encoding genes, *Ago2*, also lead to a severe developmental delay, however its onset was at a later time (E10.5). Interestingly, one of the most prominent developmental defects in Ago2-null mice was failure of the neural tube closure (Liu et al., 2004). Similar observations were made by Antonio Giraldez and colleagues in their experiments on *Danio rerio* (Giraldez et al., 2005). Giraldez created *D. rerio* Dicer-null zygotes that were lacking maternal Dicer (*MZdicer*). Interestingly, although *MZdicer* embryos were not viable, their development progressed further relative to the development of Dicer-null mouse embryos, and severe abnormalities in development of both neural and nonneural systems were uncovered. For example, formation of the neurocel and the midbrain-hindbrain boundary were severely undermined in *MZdicer* embryos. Importantly, it was possible to confirm that the phenotype of *MZdicer* embryos was triggered by the lack of mature miRNAs through rescue experiments. Indeed, injection of a dsRNA mimic of dre-miR-430, a highly abundant miRNA in early embryonic development, rescued many aspects of neural development (including formation of normal size brain ventricles and the midbrain-hindbrain boundary). This experiment showed that miRNAs were not only essential for early vertebrate development, but that they also played a significant role in development of the nervous system.

With improvement in gene targeting technologies, conditional knockout mice were created, which enables one to directly observe the consequences of disrupting miRNA biogenesis for mammalian brain development and for mature neurons. By deleting Dicer at a specific time and in a specific cell type it was possible to circumvent the requirement for Dicer in the early embryonic development of the mouse and study consequences of its loss for later developmental stages. Isolation of Dicer-null neural progenitors suggested that miRNAs are essential for commitment of neural progenitors to differentiation. Neural progenitors without detectably expressed miRNAs were obtained from the embryonic Dicer-null cerebral cortex and were shown to be incapable of differentiation (Andersson et al., 2010). Similarly, Dicer-null oligodenrocyte progenitors were also incapable of differentiation. This phenotype was partially rescued by ectopic expression of miR-219 and

miR-338 (Zhao et al., 2010). The requirement of Dicer and miRNAs for commitment of neural progenitors was in agreement with reports of miRNAs being essential for the differentiation of mouse embryonic stem (ES) cells (Kanellopoulou et al., 2005; Wang et al., 2007). In addition to the regulation of stem cells commitment, miRNAs were found to be important for the survival of differentiated cell types. Increased apoptosis was frequently reported upon deletion of Dicer in various cell types, including dopaminergic neurons (Kim et al., 2007), Purkinje cells (Schaefer et al., 2007), and forebrain neurons (Davis et al., 2008; Konopka et al., 2010; Hébert et al., 2010).

miRNAs in establishment of neuronal identity

Studying consequences of perturbation (downregulation and/or overexpression) of individual miRNAs provided further evidence of their functional significance in neuronal development, particularly in differentiation and acquisition of a cellular identity. The first evidence of the role of miRNAs in the establishment of differentiated cell types, including the neuronal cell type, came from an experiment on ectopic expression of two tissue specific miRNAs in HeLa cell culture (Lim et al., 2005). One of this miRNAs was miR-124, a miRNA highly conserved and highly expressed in the central nervous system (CNS) and specific to neurons (Lagos-Quintana et al., 2002; Landgraf et al., 2007; Cheng et al., 2009; Clark et al., 2010). Ectopic expression of miR-124 in HeLa cell culture[1] caused inhibition of a number of genes that were normally expressed at a low level in the brain (Lim et al., 2005). Therefore, expression of miR-124 transformed the gene expression of HeLa cells to be more like that of a neuron, which hinted at its role in the establishment of neuronal gene expression (neuronal state). Subsequently, this proposition was supported by the observation that introduction of miR-124 into dividing neural precursors caused them to cease division and undergo neuronal differentiation (Cheng et al., 2009). In addition to miR-124, other miRNAs, such as miR-9, miR-125b and miRNAs of let-7 family were shown to trigger premature differentiation upon overexpression in neuronal progenitors (Leucht et al., 2008; Le et al., 2009; Rybak et al., 2008).

Experiments with individual miRNAs enabled the identification of genes regulated by these miRNAs (i.e. miRNA targets). In some cases targets of miRNAs themselves were shown to inhibit neuronal differentiation (i.e. they had an anti-neuronal activity), and therefore these miRNAs themselves can be said to have a pro-neuronal role. Perhaps the most studied example of a pro-neuronal miRNA is miR-124, and its targets were iden-

[1]HeLa was derived from a cervical carcinoma (Scherer et al., 1953) and, hence, is non-neuronal.

tified in several distinct pathways important for neuronal differentiation. For example, miR-124 was found to participate in a double negative feedback loop that involves REST, a transcription factor that has a gate-keeper role in the acquisition of the neuronal state (Conaco et al., 2006; Visvanathan et al., 2007). REST binds RE-1 elements in promoter regions of many neuronal genes. In non-neuronal cells it recruits co-repressors to promoters of these genes and causes their transcriptional inhibition (Ballas and Mandel, 2005; Ballas et al., 2005). One of these co-repressors, SCP1, is directly targeted by miR-124 and, thus miR-124 counteracts the anti-neuronal activity of REST (Visvanathan et al., 2007). Interestingly, in non-neuronal cells, REST was shown to inhibit expression of miR-124 (Conaco et al., 2006; Visvanathan et al., 2007), an interaction that completes the double negative feedback loop. Apart from being involved in inhibition of the function of the REST/SCP1 pathway, miR-124 was also implicated in the inhibition of at least three other anti-neuronal pathways. Its targets include PTBP1, a global inhibitor of the neuron specific alternative splicing (Makeyev et al., 2007), BAF53a, a neural-progenitor specific chromatin remodeling factor (Yoo et al., 2009) and Sox9, a transcription factor important for proliferation of neural progenitors (Cheng et al., 2009). In addition to miR-124, over a dozen different miRNAs were also shown to be regulators of neuronal differentiation through the inhibition of genes with anti-neuronal activity. An in depth description of the function of miR-124 and other miRNAs in neuronal differentiation can be found in an excellent review by Xuekun Li and Peng Jin (Li and Jin, 2010).

miRNAs in neuronal function

The first discovered miRNA, lin-4, was identified because of its essential role in regulation of *C. elegans* development (Lee et al., 1993; Wightman et al., 1993). Subsequently, as was described above, other miRNAs were also shown to have an important role in regulation of organ development and cell differentiation. At the same time, miRNAs, including miRNAs important for brain and neuronal development, were being discovered in adult organisms. This posed a challenge to identify the functions of miRNAs in mature neurons.

A seminal study in this field was published by Gerhard Schratt and colleagues, which described a role of miR-134 in regulation of synaptic morphology that was dependent on neuronal activity (Schratt et al., 2006). This miRNA was found to inhibit expression of *Limk1*, a gene that was shown to regulate actin filament dynamics. At the morphological level, miR-134 acted to decrease dendritic spine size. This work proposed the biological importance of miR-134 as an inhibitor of synaptic plasticity because activity of miR-134 itself was relieved by BDNF, a major stimulant of synaptic growth and function. The role

of miR-134 in neuronal function was further explored in a recent work from the laboratory of Li-Huei Tsai (Gao et al., 2010). Overexpression of miR-134 in the mouse hippocampus was shown to impair performance of the animals in a context fear-conditioning task and to abrogate induction of CA1-CA3 long-term potentiation (LTP). Inhibition of LTP and memory was suggested to be linked to miR-134 targeting Creb1 transcript that encodes a transcription factor important for induction of long-term synaptic plasticity (Gao et al., 2010).

Activity of other miRNAs was also shown to modulate the function of mature neurons, and their activity was frequently described as inhibitory to genes upregulated in neuronal plasticity (Schratt, 2009). Surprisingly, this mode of inhibition of synaptic plasticity was also shown for miRNAs that were previously demonstrated to promote neuronal differentiation (i.e. pro-neuronal miRNAs). For example, when miR-124 was injected into the cultured neurons of *Aplysia californica*, it significantly reduced long term facilitation of synaptic transmission (Rajasethupathy et al., 2009). Perhaps the most striking evidence of the inhibitory effect of miRNAs to neuronal plasticity came from demonstration that the loss of all miRNAs leads to enhancement of learning and memory in mice (Konopka et al., 2010). Using a transgenic mouse line in which ablation of Dicer could be induced in the adult forebrain neurons, it was possible to significantly deplete the pool of miRNAs in the mature neurons of a living animal. All mice eventually died, presumably due to the massive neurodegeneration that was observed after 14 weeks from inducing the deletion of *Dicer*, which was consistent with previously reported elevated apoptosis in Dicer-null backgrounds (see above). There was a time-window, however, at 12 weeks from the time of deletion, when the pool of miRNAs was significantly depleted, while the onset of apoptosis had not yet commenced. At this time, the mutant mice had increased performance in four different learning and memory tests and also displayed an elevated post-tetanic CA1-CA3 synaptic potentiation (Konopka et al., 2010).

In summary, miRNA mediated regulation of gene expression was revealed to be important for both differentiation of neuronal progenitors and for the function of mature neurons. Several miRNAs, including pro-neuronal miRNAs, were suggested to inhibit neuronal plasticity. Both development and plasticity are characterized by widespread changes in gene expression, and miRNA regulation of these processes is consistent with their proposed function as buffers of differential gene expression (Wu et al., 2009b). In the role as buffers, miRNAs are well suited to contribute to canalisation of developmental programs (Hornstein and Shomron, 2006) and confer robustness to gene expression

networks (Herranz and Cohen, 2010). The robustness may be particularly important in the face of stresses to biological systems and miRNAs were proposed to be important for adequate stress responses (Leung and Sharp, 2010). This aspect of miRNA function will be important for interpretation of experimental results obtained during the course of this thesis project, therefore the role of miRNAs in stress responses is reviewed in the next section.

1.1.4 The paradox of miRNAs and role of miRNAs in stress responses

A feature of miRNAs, that is sometimes referred to as a paradox of miRNAs, is their high degree of evolutionary conservation and yet apparently non-essential role in cell and organism viability. Some miRNAs are highly conserved between deeply branching metazoans (Pasquinelli et al., 2000; Sempere et al., 2006). Moreover, the very origin of metazoan organ systems, such as the central nervous system, the sensory tissue, the musculature and the gut, was found to coincide with the origin of tissue specific expression of certain miRNAs (Christodoulou et al., 2010). In the light of deep evolutionary conservation, finding that a majority of miRNAs was not essential for viability came as a surprise. For example, simultaneous removal of a majority of miRNAs from the ES cells through deletion of the *Dgcr8* gene[1], which is a key factor in miRNA biogenesis did not trigger their death (Wang et al., 2007). The individual deletion of a majority of genes encoding miRNAs or entire miRNA families from the nematode worm *Caenorhabditis elegans* genome did not induce significant phenotypic abnormalities (Miska et al., 2007; Alvarez-Saavedra and Horvitz, 2010). According to one report, inhibition of a highly expressed, evolutionarily conserved neuron-specific miRNA, miR-124, during development of the chick spinal cord did not affect neurogenesis (Cao et al., 2007). Additionally, both inhibition and overexpression of miRNAs was found to induce only subtle changes in the abundance and translation of the target transcripts (Baek et al., 2008; Selbach et al., 2008). To reconcile the seemingly conflicting evolutionary conservation and dispensability, miRNAs were suggested to act as tuners and buffers of gene expression, rather than major regulatory switches (Wu et al., 2009b). In this role, the function of miRNAs would be to ensure robustness of gene expression programs. In agreement with this proposition, the significance of miRNAs was identified for processes that shift equilibria of gene expression programs. The

[1]Together with Drosha, Dgcr8 is an essential component in the Microprocessor complex (see section 1.1.2) (Han et al., 2004).

role of miRNA mediated regulation in some of these processes, such as development and plasticity, had already been discussed in section 1.1.3. Significance of miRNAs in other equilibrium shifting processes (induced mutations and stresses) is described below.

The function of miRNAs in conferring robustness to biological systems was revealed when nematodes with deletions of miRNA genes, but no deleterious phenotype, developed significant abnormalities in the context of perturbations to the transcriptome (Brenner et al., 2010). In nematodes with stable deletions of seemingly non-essential miRNAs, Brenner and colleagues knocked down five hub nodes of the gene expression network, each of which was a component of several major signaling pathways (e.g. EGF, Wnt, Notch and etc.). Knock down of transcripts encoding these hub proteins was expected to perturb gene expression equilibrium. Wild type nematode worms could sustain these perturbations without developing significant abnormalities, however nematodes with miRNA deletions developed notable defects of germline development and a significant proportion of the worms was sterile. Such combinatorial interaction was identified for four non-essential for viability miRNAs (out of eleven included in the analysis) and three out of five genes involved in signalling (Brenner et al., 2010). Interestingly, the nature of the signalling pathways is such that expression of their members vary in different contexts, which makes them good candidate targets of miRNA mediated buffering. Indeed, regulation of signalling pathways was noted to be a feature of miRNA targeting (Inui et al., 2010).

Cellular stresses are known to shift gene expression from the state of homeostatic equilibrium, and several miRNAs have been implicated in conferring robustness to both developmental programs and the homeostatic state in the face of stresses (Herranz and Cohen, 2010; Leung and Sharp, 2010). As an extension to the example discussed above, of non-essential *C. elegans* miRNAs acting as buffers of perturbation in gene expression, several non-essential miRNAs in other organisms have been identified as key regulators of stress responses. Perhaps the best studied example of a miRNA imparting robustness to a developmental program under stress is that of miR-7 in sensory organ development of the common fruit fly, *Drosophila melanogaster* (Li et al., 2009). Despite miR-7 being perfectly conserved between protostomes and deuterostomes (Sempere et al., 2006), mutant fruit flies that lacked miR-7 developed normally. However, if the mutant larvae were subjected to temperature fluctuations, the two key transcription factors, Yan and Atonal, were abnormally expressed in the eye and antennal cells. Additionally, a sensory organ precursors (SOP), called arista, failed to develop and SOPs for coeloconic sensillae either failed to develop or were patterned abnormally. In eye and SOP development, miR-7 is

thought to act as a buffer against perturbation in a gene expression network through a feed forward inhibition (both direct and indirect) of the two aforementioned transcription factors. Another miRNA in *D. melanogaster*, miR-14, was identified as protective against stress (Xu et al., 2003). Fruit flies with deletion of *miR-14* were viable, but had reduced lifespan and were significantly more susceptible to salt stress. Cell death was increased in the mutants under the stress, and at the same time several pro-apoptotic genes were found to be inhibited by miR-14, which suggested miR-14 role in suppressing the apoptosis during the stress. The importance of miRNAs for an appropriate stress response was also reported in vertebrates. In the zebrafish *Danio rerio*, miRNAs of the miR-8 family were found to be expressed in the skin and kidneys and to be required for adaptation to osmotic fluctuations (Flynt et al., 2009). The development of fish, with miR-8 knocked down, was indistinguishable from the wild type under normal conditions. However, if the fish were placed into a high osmomolarity buffer and then transferred into distilled water, oedema was observed in the mutants with significantly greater frequency than in wild type. In the mouse, knock out of the gene for miR-208 did not reduce viability, but disrupted normal stress response to thoracic aortic banding (van Rooij et al., 2007).

Summary of section 1.1

In this section I described the discovery of miRNAs and the current understanding of the functions and roles of miRNAs in the brain and neurons. The ability of a single miRNA to regulate the expression of dozens to hundreds of genes and the involvement of these targets in a spectrum of key developmental and neurological processes, suggests a role of great significance in the nervous systems and neurons in particular. In this regard, dispensability of many individual miRNAs is paradoxical. Recent research into miRNA function showed that miRNAs can act as buffers of plasticity in gene expression (for example, during neuronal plasticity and during stresses), which suggested an explanation of the apparent dispensability of many miRNAs in standard, stable laboratory environments.

Although the roles of miRNAs as buffers of gene expression programs has began to emerge from the current research, the full extent of the functional significance of miRNA is likely not yet fully appreciated. The primary reason for this is a traditional "single gene approach" that was taken in many of the previous studies into miRNA biology, where some of the miRNA targets are studied in detail, while the effect of miRNAs on hundreds of other potential targets was frequently overlooked. This set the scene for this thesis project, where I used methods of whole transcriptome profiling for the identification of

hundreds of miRNA targets with the goal to better understand the roles of miRNAs in neurons.

1.2 Thesis aims and an experimental paradigm

Individual miRNAs have been shown to regulate hundreds of target mRNA transcripts (Stark et al., 2003; Enright et al., 2003; Farh et al., 2005; Lim et al., 2005; Giraldez et al., 2006; Baek et al., 2008; Selbach et al., 2008). The functions of even the most studied miRNAs are generally perceived through a prism of only a handful of validated miRNA-mRNA interactions. The significance of miRNAs for the development of metazoans has been recognised since their discovery, however relatively little is known about their function in committed cell types. Numerous published studies described targets of miRNAs important for development of the nervous system as a whole and neurons in particular. However, the function of miRNAs in mature neurons is less well understood. This incomplete understanding of the roles of miRNAs in differentiated cell types, in conjunction with the paradoxical dispensability of many individual miRNAs for animal viability (Miska et al., 2007; Brenner et al., 2010), made the study of the roles of miRNAs in differentiated neurons relevant.

Computational approaches for miRNA target analysis have been prevalent since 2003. Purely computational approaches have many pitfalls and a demonstrated over-prediction bias (Giraldez et al., 2006; Baek et al., 2008). In order to address this situation a number of approaches have been developed since. The three main methods are currently: mRNA profiling after miRNA perturbation, large-scale proteomic approaches after perturbation and direct sequencing of mRNA targets bound to RISC component enzymes. At the time of the design of this thesis project, only the first method of high throughput identification of putative miRNA targets had been established and received experimental validation. This method involves overexpression of miRNAs in cells by exogenous addition of miRNA mimics, and subsequent elucidation of miRNA targeting through microarray profiling of incurred changes in the transcriptome (Lim et al., 2005; Giraldez et al., 2006). This technique was suitable to study miRNAs in neurons, because neuronal gene expression can be studied in primary cultures (Valor et al., 2007), and these cultures can be efficiently transfected with miRNA mimics (Conaco et al., 2006). Additionally, the technology behind microarrays and protocols for their analysis are both mature and inexpensive. For these reasons I decided to assay miRNA targets in neuronal cultures using microarray profiling of differential expression triggered by transfection of miRNA mimics.

While the other two approaches were not available at the time of the design of this thesis project, it is still worth describing their strengths and weaknesses in comparison to the assay used. The first method is conceptually similar to the approach used during the thesis project. However, instead of profiling the effects of miRNA mimics at the transcript level, profiling is performed at the level of the proteome (Baek et al., 2008; Selbach et al., 2008). The advantage of this method is the possibility to identify miRNA targets that are not regulated by miRNAs at the level of mRNA stability. However, recent research shows that a majority of mammalian miRNA targets are indeed regulated predominantly at the level of mRNA stability (Guo et al., 2010). Additionally, these experiments are expensive and time-consuming.

The second method is based on sequencing of RNA co-precipitated with proteins of the RNA silencing machinery (Licatalosi et al., 2008). These methods, such as HITS-CLIP, par-CLIP and iCLIP are very promising for miRNA target research. Firstly, because the technique does not require perturbation of miRNA expression, it can potentially identify the *in vivo* targets of miRNAs in cognate tissue and organs. However, this is a new technology with great promise but many obstacles remain. Issues arise for a number of reasons including: availability and specificity of antibody used, technical variation and amplification biases from sequencing and finally poor capture of mRNA sequence compared to an abundance of miRNA sequence (Anton Enright, Eric Miska, Jernej Ule and Donál O'Carroll personal communication). Although this approach currently has many technical limitations that need to be addressed, it would appear, at least conceptually, to be the most promising approach for future work. A limited comparison between published miRNA targets defined by HITS-CLIP and by my own approach is given in Chapters 5 and 6.

Aims: To summarise, in this thesis I aimed to characterise the roles of miRNAs in neurons by identification and analysis of miRNA targets in primary neuronal cultures. The identification of targets was achieved through perturbation of levels of miRNA expression in the cultures, analysis of the incurred differential gene expression by microarrays and derivation of the lists of targets from the profiling data.

Below I will describe methods of chemical transfection as a way of inducing miRNA mediated perturbations of the transcriptome. I will also describe the seed enrichment analysis as an approach to derive putative direct targets from microarray profiling data. Finally, E17.5 mouse primary neuronal cultures will be introduced as a model system to study the biology of neurons.

1.2.1 Widespread effect of miRNAs on the transcriptome is guided by miRNA seed region

Complementarity between miRNA and mRNA sequences was suggested as a mechanism of targeting in the two articles reporting the discovery of the first known miRNA, lin-4, and its target, lin-14 mRNA (Lee et al., 1993; Wightman et al., 1993). Other features of target recognition by miRNAs were also noted in the original reports: only partial complementarity was required between miRNA and mRNA transcripts, and the sites in the mRNA that annealed to the miRNA were located in the 3′UTR of the transcript. Subsequently, another miRNA discovered in *C. elegans*, called let-7, was also shown to recognise the mRNA targets through basepairing with their 3′UTRs (Reinhart et al., 2000). These observations laid the foundation for the first computational sequence based miRNA target prediction algorithms (Stark et al., 2003; Enright et al., 2003; Lewis et al., 2003). These algorithms have one assumption in common: transcripts with one or more sites in their 3′UTRs, which are partially complementary to a miRNA, are more likely to be targeted by that miRNA than transcripts drawn at random. Around the time when the first target prediction algorithms were being developed, Eric Lai noted that the sequences between positions 2 to 8 at the 5′-end of several miRNAs were perfectly complementary to the 3′UTRs of a selection of post-transcriptionally regulated transcripts (Lai, 2002). Additionally, 5′-regions of fly and worm miRNAs were noted to be more evolutionarily conserved than the rest of the sequence (Lai, 2002; Lim et al., 2003). Based on this observation, and on analysis of the available validated miRNA-target pairs, Lewis and colleagues proposed that perfect complementarity in the positions from 2 to 8 at the 5′-end of miRNAs with the sequence of 3′UTRs was a key determinant in miRNA target recognition (Lewis et al., 2003, 2005). This important for target recognition region in the miRNA sequence received the name “miRNA seed region”, while the complementary sequence in the target sequence was called the seed matching site. Experiments with reporter constructs confirmed the importance of basepairing to the seed for targeting. Presence of a single seed matching site in a 3′UTR was shown to be sometimes sufficient for miRNA targeting of a transcript (Doench and Sharp, 2004; Lai et al., 2005), which further justified the use of seed matching sites for prediction of the targets.

Efforts from the laboratory of David Bartel greatly contributed to current understanding of the role of miRNA seed region in target recognition (Bartel, 2004, 2009). Through microarray profiling of changes in transcriptomes upon overexpression of miRNAs, Bartel and colleagues showed that complementarity of six to eight bases between a miRNA seed

region and 3′UTRs was frequently associated with destabilisation of these transcripts (Grimson et al., 2007). The efficiency of destabilisation was lower for transcripts that had sites with only six bases of complementarity than sites with seven and eight bases. Interestingly, sites with six bases of complementary to the seed region (positions 2 to 7) followed by an adenine nucleotide were significantly more effective at destabilisation than the six bases on their own. Sites with six or eight base complementarity will be referred to as 6(2)-type and 8(2)-type seed matching sites respectively, while the sites of seven base complementarity and six base complementarity followed by an adenine nucleotide will be referred to as 7(2)-type and 7(1A)-type seed matching sites.

On the basis of complementarity to the seed being sufficient to enable miRNA mediated regulation, individual miRNAs were predicted to potentially directly regulate expression of dozens to hundreds of genes (Brennecke et al., 2005; Grün et al., 2005; Lewis et al., 2005; Stark et al., 2005). Validation of a wide spread regulation of gene expression by miRNAs came from the whole transcriptome analysis of miRNA activity. Ectopic expression of miRNAs was found to induce wide spread changes in gene expression, where 3′UTRs of downregulated transcripts were significantly enriched in miRNA seed matching sites (Lim et al., 2005; Giraldez et al., 2006). Crucially, a significant fraction of the downregulated transcripts with the seed matching sites in their 3′UTRs were validated as direct miRNA targets using luciferase reporter assay (Lim et al., 2005). The validation of direct targeting served as a proof that a significant component of the widespread downregulation of gene expression was directly caused by miRNAs inhibiting their direct targets. The same conclusions were drawn from the reciprocal experiments, where individual miRNAs were removed from the system. In these experiments, 3′UTRs of transcripts upregulated were enriched in the seed matching sites (Krützfeldt et al., 2005; Rodriguez et al., 2007), which is consistent with the inhibitory role of miRNA. An elegant demonstration of the scale of direct miRNA mediated effects that can be identified through whole transcriptome profiling was made in the laboratories of Karen Steel and Anton Enright (Lewis et al., 2009). There, transcripts upregulated in a mouse mutant, which had a single substitution in the seed region of miR-96, were enriched in seed matching sites for the original miR-96 seed, while transcripts downregulated in the mutant were enriched in the sites complementary to the acquired miR-96 seed.

Regulation of gene expression by miRNAs on a large scale was also observed for endogenous miRNAs through profiling of transcription in specific tissues (Farh et al., 2005; Sood et al., 2006). Expression of some miRNAs was shown to be highly tissue specific (Wienholds et al., 2005; Landgraf et al., 2007). Therefore, given the multitude of targets

an individual miRNAs may have, the inhibitory activity of highly expressed and tissue specific miRNAs could shape gene expression of cognate tissues (Stark et al., 2005). Systematic analysis of the distribution of seed matching sites in transcriptomes of a wide range of tissues confirmed the broad impact of miRNA regulation on tissue specific gene expression (Farh et al., 2005; Sood et al., 2006). For example, the 3′UTRs of highly expressed in the liver transcripts were depleted of seed matching sites for liver specific miR-122, in the muscle – for muscle specific miR-1 and in the brain – for brain specific miR-124.

The biological reason behind the importance of basepairing between the seed region and the target sequence was understood when the crystal structure was solved for miRNA bound to a bacterial homologue of the metazoan Argonaute-component of RISC complex (Ago proteins) (Wang et al., 2008c). The edges of the bases in position 2 to 6 at the 5′-end of the miRNA were found to be readily available, and thought to nucleate annealing to the target. When the structure of the ternary complex of Ago-miRNA-target sequence was solved, it was found that bulges in positions 2 to 8 at the 5′-end of the miRNA were poorly accommodated (Wang et al., 2008b), which confirmed the importance of complementarity between the seed site and the target.

Traditionally, functional seed matching sites were identified in 3′UTRs of the target transcripts (Lee et al., 1993; Wightman et al., 1993). Later 3′UTRs of miRNA targets were experimentally found to be significantly enriched in seed matching sites for that miRNA (Lim et al., 2005; Giraldez et al., 2006). This allowed the use of statistics of enrichment of seed matching sites in 3′UTRs of transcripts to discern miRNA mediated effects on gene expression and the compilation of lists of putative direct targets in this thesis (Chapter 5). However, using RNA precipitation and new generation sequencing (i.e. a method which is not reliant on seed enrichment statistics to identify putative direct targets) showed that a significant fraction of target sites are located outside of the 3′UTRs of putative targets (Chi et al., 2009). For example, in a recent study, of all sites in putative direct miRNA targets in embryonic stem cells $\approx 30\%$ were found to be located in coding region (Leung et al., 2011). Nevertheless, the greatest enrichment of the seed matching sites (per length of a sequence) in miRNA perturbation experiments was routinely observed in 3′UTRs of transcripts responding to the perturbations (Cei Goodger-Abreu, personal communication). Therefore, for the purpose of this thesis, miRNA targeting was considered in a traditional way and only 3′UTR sequences were used for seed matching site based compilation of putative miRNA targets (Chapter 5).

Two main conclusions can be drawn from the results of the experiments and analysis described in this section. Firstly, a single miRNA can directly inhibit the expression of dozens to hundreds of genes, a property that shapes expression of the whole transcriptome. Secondly, profiling of differential gene expression upon perturbation of individual miRNAs, in combination with the search for seed matching sites in 3′UTRs, enables identification of direct targets of miRNAs. These principles were used in this thesis to study function of miRNAs, which is described in more detail in the next section.

1.2.2 Transfections of miRNA mimics enables identification of direct miRNA targets

Cationic lipid transfections were developed over two decades ago as a method of introducing DNA molecules into cultured eukaryotic cells (Felgner et al., 1987). Efficiency of cationic lipid transfection was higher than that of some of the more traditional methods, including calcium phosphate transfections, and it became a popular tool that is now widely used for transfections of both DNA and RNA molecules into cells in culture. Transfections of mammalian cell cultures with miRNA mimics (double-stranded RNA, with one of the strands being equivalent to mature miRNA), was successfully used to overexpress miRNAs (Lim et al., 2005; Conaco et al., 2006; Baek et al., 2008; Selbach et al., 2008), and transfection with miRNA inhibitors (single-stranded chemically modified RNA complementary to mature miRNAs) was used to inhibit activity of endogenous miRNAs (Krützfeldt et al., 2005; Conaco et al., 2006; Selbach et al., 2008). Mechanism for the transfection of DNA was previously studied and reported in the literature (Zabner et al., 1995; Xu and Szoka, 1996), and the same mechanisms were suggested for transfection of RNA (Schroeder et al., 2010). Cationic lipids were suggested to form aggregates with nucleic acids that can be engulfed by cells through endocytosis (Zabner et al., 1995). After the nucleic acid/lipid complex is internalized, cationic lipids interact with endosomal membranes, which destabilises the aggregates and leads to the release of DNA from cationic lipids into the cytoplasm (Xu and Szoka, 1996). In the case of DNA transfections, the limiting step in efficiency of transfection was proposed to be the transfer of DNA from the cytoplasm into the nucleus (Zabner et al., 1995). This step was not considered to be an obstacle for miRNA perturbation experiments, because the biological activity of miRNA mimics and inhibitors takes place in the cytoplasm. Similarly high efficiency was expected in experiments with miRNA mimics and inhibitors, because efficiency of DNA

entry into the cytoplasm was shown to be high ($\approx$ 72.3% after 24 h incubation (Zabner et al., 1995)).

Stable knock out mice have been used in miRNA research (Rodriguez et al., 2007; Elia et al., 2009; Xin et al., 2009), however transfection of miRNAs had several advantages for studying miRNAs in mature neurons. The first advantage is the relative simplicity with which transfection experiments can be performed. In comparison to creation of knockout lines, transfections can be performed faster, thus, even within time constraints of the thesis projects, it was possible to study the functions of several miRNAs. The second advantage is that transfections are acute, meaning that the development of the transfected cells is no different from that of control cells prior to the experiment itself. The latter aspect was particularly important in this project on miRNAs in differentiated neurons, because miRNAs are involved in neuronal differentiation (see section 1.1.3). Because of this, analysis of mature neurons in stable knockout mutant lines would be problematic due to their development being affected prior to the experiment. This problem could have in principle been circumvented through creation of conditional knockout lines, however creation of such mice was not logistically possible within the time constraints of the project.

An important factor arguing for the use of transfections to study miRNAs was a report of successful detection of miRNA targets with this methodology in a seminal work by Lee Lim and colleagues (Lim et al., 2005). Two miRNAs, miR-1 and miR-124, were transfected into HeLa cell culture. Sets of genes significantly downregulated in the two experiments ($P < 0.001$) encoded transcripts that were enriched in seed matching sites for miR-1 and miR-124 ($P < 7.0e-27$ and $P < 1.1e-54$). This enrichment suggested that downregulation of a significant fraction of genes was directly caused by the inhibitory activity of the transfected miRNAs, and the downregulated transcripts with seed matching sites comprised lists of putative direct targets of the two miRNAs. This proposition was supported through a validation experiment: using a luciferase reporter system (Lewis et al., 2003), direct inhibition was confirmed for six out of ten selected targets. Further confirmation that targets identified in the transfection experiment by Lim and colleagues came from a different study conducted on primary neuronal cultures (Conaco et al., 2006). In that work, a mimic for miR-124 was transfected into primary neuronal cultures, and 17 out of 17 genes, which were inferred by Lim et al. as direct targets of miR-124, were significantly downregulated in neuronal cultures. Moreover, upon transfection of neuronal cultures with an inhibitor for miR-124, ten out of the 17 genes were significantly upregulated (Conaco et al., 2006). The experiments with miR-124 inhibition in neuronal

cultures showed that over half of genes, which were identified by transfection of mimics as putatively direct targets, were likely to have been the innate miR-124 targets in primary neurons.

Because of the logistical advantages and reported effectiveness of transfections of miRNA mimics for identification of direct miRNA targets, the transfection of miRNA mimics in primary neuronal cultures was chosen as the basis of the experimental approach in this thesis. A transfection protocol using a cationic lipid reagent, DharmaFECT 3, was optimised for the transfection of siRNAs into primary neuronal cultures by my colleague Dr. Erik MacLaren (Maclaren et al., 2011). This protocol was used for transfecting miRNA mimics and inhibitors (Methods, section 2.5), because miRNA mimics and inhibitors have a similar length to siRNAs.

1.2.3 The use of seed enrichment analysis of direct miRNA effects and identification of targets

Before the list of genes downregulated in transfection experiments could be used for the identification of putative direct miRNA targets, it was necessary to confirm that overexpressed miRNAs were likely to be the direct cause of differential gene expression. miRNAs can guide inhibitory RISC to the targets through complementarity of short sequences (six to eight nucleotides (Lewis et al., 2003, 2005)). Sequences that are complementary to a miRNA seed region can be encountered by chance in 3′UTRs, because the seed region is short. Therefore, even if a transfection of a miRNA mimic failed to induce miRNA mediated inhibition, it was possible to falsely identify downregulated transcripts with the seed matching sites as direct miRNA targets. To avoid such false positive results, it was important to develop a method that could discriminate between experiments where transfection of miRNA mimics failed to elicit miRNA mediated inhibition. A significant hypergeometric overrepresentation of seed matching sites for a particular miRNA in 3′UTRs of downregulated transcripts suggests that their downregulation was likely to be directly caused by the miRNA. Not all of the transcripts that are downregulated during miRNA overexpression experiments and bearing seed matching sites for the miRNA are real miRNA targets, however over 60% of these putative direct targets have been experimentally validated in previous studies (Lim et al., 2005; Giraldez et al., 2006). Furthermore, significant biases in the distribution of seed matchings sites have been observed in tissue specific expression profiles (Farh et al., 2005; Sood et al., 2006). These biases represent the depletion of the seed matching sites for highly expressed endogenous miR-

NAs from the 3′UTRs of highly expressed mRNA transcripts (Farh et al., 2005; Sood et al., 2006). The correlation of natural depletion signals with cognate miRNAs provided additional evidence for shifts in seed distribution as indicative of the biological activity of the miRNAs.

To identify signals of direct miRNA effects on gene expression profiles, it is possible to select genes based on a differential expression cutoff (e.g. $P < 0.05$), and to analyse composition of their 3′UTRs. Such an approach was used in the Lim et al. study, where significant enrichment of the seed matching sites for overexpressed miRNAs was detected in 3′UTRs of downregulated genes at the cutoff $P < 0.001$ (Lim et al., 2005). Although a simple hypergeometric test can produce useful information, there are several crucial drawbacks to this approach that prevent it from being used as the sole tool for discovery of miRNA mediated effects. First, 3′UTR length and composition biases cannot be easily accounted for with this traditional approach. Second, an arbitrary selection of the differential expression cutoff can artificially increase or decrease the size of the selection of putative direct miRNA targets. In order to account for length and composition biases and not to rely on arbitrary differential expression cutoffs, my colleagues, Stijn van Dongen and Cei Abreu-Goodger in the Enright laboratory, developed a method of nucleotide word enrichment analysis, called Sylamer (van Dongen et al., 2008).

Sylamer works by identifying occurrence biases of nucleotide words (one to 15 bases long) in a sorted list of sequences. For example, Sylamer can estimate enrichment or depletion of seed matching sites in a list of 3′UTRs of all transcripts that were detected by microarray transcriptome profiling, when this list is ranked from most downregulated to most upregulated. Assessment of the enrichment of a particular nucleotide word of a given length is done by calculation of hypergeometric enrichment P-value of that word in samples of sequences (or bins) from the list. Sampling of sequences from the ordered list is done from the most downregulated to the most upregulated and the size of a leading bin is iteratively incremented (i.e. at each step the leading bin includes all previously sampled sequences plus a certain number of new sequences). For example, if the size of the increment is 100 sequences, then the first bin includes 0 sequences, the second bin includes 100 sequences, the third bin includes 200 sequences and so on until the leading bin includes all sequences from the list. At each step, Sylamer calculates enrichment P-value of a particular nucleotide word by comparing its occurrence in the leading bin to its occurrence in all sequences of the list. Importantly, Sylamer operates on counts of nucleotide words of a given length per bin and per whole list, and the values of words-per-3′UTR are never a part of the equation, which automatically excludes the possibility

of a length bias. Additionally, correction for composition biases was incorporated into Sylamer, where hypergeometric statistics of a word of a given length can be adjusted to account for biases in underlying distributions of the related words of a shorter length (van Dongen et al., 2008).

The output of Sylamer can be visualised by plotting lines that correspond to occurrence biases of words of a given length in all bins from a sorted gene list. For example, Figures 1.1a and 1.1b show a cartoon diagram of a mock Sylamer line representing an occurrence bias of a single word. The x-axes of these plots corresponds to sorted 3′UTRs sequences. The y-axes values greater than 0 indicate enrichment and values less than 0 indicate depletion of the words in the bins when compared to the whole list. A shortcut to interpreting Sylamer plots is to consider the direction of the slope of the lines as indication of enrichment or depletion (shown as dashed arrows, Figures 1.1a and 1.1b). As shown in the cartoons, where the line goes upward there is an enrichment of the word in the underlying 3′UTRs, while if the line slopes downward there is a depletion.

In the analysis of the data presented in this thesis, the Sylamer plots illustrate occurrence biases of 876 distinct seven nucleotide seed matching sites[1], (7(2) and 7(1A) types) complementary to the seed regions of all known mature mouse miRNAs[2], which is 591 mature miRNAs according to miRBase Release 14 (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008) (Figure 1.1c and 1.1d). Two types of seven nucleotide seed matching sites (7(2) and 7(1A)-seed matching sites) were used here as predictors of miRNA targeting (Introduction, section 1.2.1). Therefore, in this thesis Sylamer analysis was always performed for nucleotide words of length seven. In the Sylamer plots, distribution of all words is shown as grey lines, apart from two to four manually selected words, which are shown in colors. As in the example cartoon diagrams, the x-axes of these plots correspond to sorted 3′UTRs sequences of genes, with one 3′UTR being selected per gene (Methods, section 2.7), and genes are usually sorted from most downregulated to most upregulated. The y-axes correspond to hypergeometric P-value of the words in each of the bins. The

[1] 6-mer and 8-mer seed matching sites can also be used as predictors of miRNA targeting, however 7-mers were reported to have a better predictive power for evolutionarily conserved miRNA targets (Bartel, 2009). In Sylamer analysis of distribution of nucleotide words and subsequent prediction of putative direct miRNA targets, peaks in occurrence biases of 7-mer sites (7(2) and 7(1A)-types) were seen to provide a higher signal to noise ratio for prediction of direct targets, than 6-mer and 8-mer sites (Cei Goodger-Abreu, personal communication).

[2] Three seed matching sites, *CAATAAA*, *TATTTAT* and *TCAATAA* that are similar to the *AATAAA*, a polyadenylation signal in 3′UTRs (Connelly and Manley, 1988), had to be excluded from Sylamer analyses, as biases in distribution of these words could not be attributed to miRNA mediated effects.

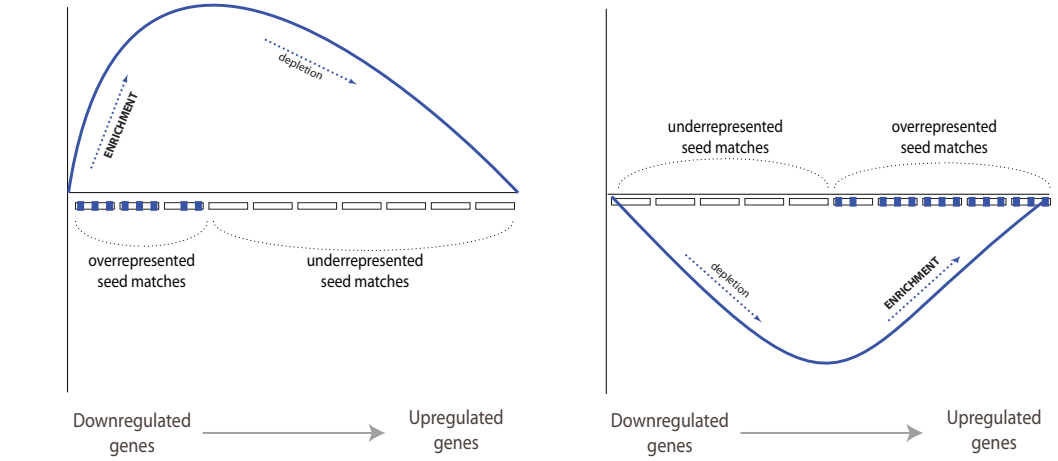
y-axes values above 0 indicate enrichment (equivalent to $-\log_{10}$ of the P-value) and values below 0 indicate depletion (equivalent to $\log_{10}$ of the P-value).

Visualisation of the results allows one to view the enrichment curve for a word or several words of interest in the context of the enrichment curves of all microRNA seed matching sites of the same length. This context defines a natural background, and a significant result should in all cases stand out among it. For example, the blue lines in Figures 1.1c and 1.1d show that occurrence bias of the word *GTGCCTT* is more significant than that of the background. This ability to instantly evaluate how unique the shifts in distribution of the words are, can qualitatively prove or disprove that selected miRNAs are specifically involved in the regulation of the gene expression patterns that are reflected in sorted 3′UTRs.

Sylamer plots can also help to select a threshold of differential expression that is most relevant to observed distributions, a choice that otherwise is arbitrary. For example, in Figure 1.1c genes are sorted by t-statistics from the most downregulated to the most upregulated upon overexpression of miR-124 (Chapter 5, section 5.1.2). Vertical lines show cutoffs of P-value of differential expression of 0.01 and 0.05 on both sides of the list (the x-axes). The steady rise of the colored line reflecting enrichment of a seed matching site (*GTGCCTT*, blue) peaks near P-value cutoff 0.01 at the left side of the plot (where downregulated genes are). Therefore, from this plot it follows that the most appropriate way to compile a list of candidate direct miR-124 targets is to select genes that were downregulated with differential expression P-value below 0.01, and subselect among them the genes that bare miR-124 seed matching sites in their 3′UTRs. Without this test, a researcher may be tempted to select a more relaxed cutoff (0.05) or a more stringent cutoff (0.001), which would artificially increase or decrease the selected list of candidate targets.

(a) Enrichment in downregulated genes

(b) Enrichment in upregulated genes



(c) Enrichment in downregulated genes

(d) Enrichment in upregulated genes

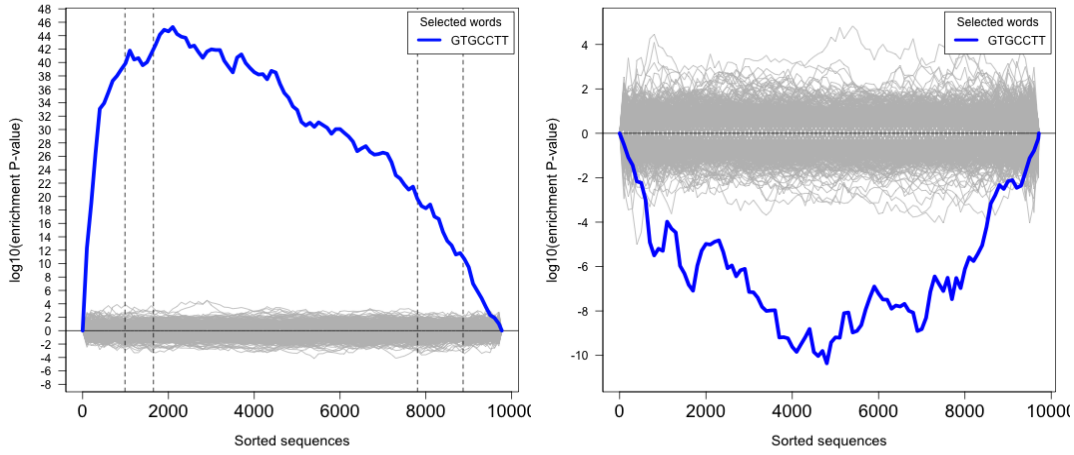


Figure 1.1: How to interpret Sylamer plots. *[The legend is on the next page]*

Figure 1.1: How to interpret Sylamer plots. *[The figure is on the previous page]*
The thick blue line in the cartoon Sylamer plots (Figures 1.1a and 1.1b) corresponds to the bias in the occurrence of one nucleotide word, which is represented by the small blue dashes in the ordered list of 3′UTR sequences (shown by the rectangles aligned under the x-axis). The 3′UTRs are ordered according to the change in expression of the corresponding genes: on the left side of the list there are downregulated genes, while on the right side there are upregulated genes. In Figure 1.1a the word is overrepresented in the left part of the list, where the 3′UTRs correspond to the downregulated genes. In Figure 1.1b the word is overrepresented in the right part of the list, where the 3′UTRs correspond to the upregulated genes. The y-axis qualitatively describes these biases: points on the line with the y-coordinate greater than 0 represent enrichment of the word in 3′UTRs preceding that point in the list, while the values less than 0 represent depletion of the word in the preceding 3′UTRs. The shortcut to understanding the Sylamer plots is to consider the direction of a slope of the line: the line that is going upward corresponds to the enrichment of the word in the underlying 3′UTRs, while the line going downward, corresponds to the depletion.
The lines in the Figures 1.1c and 1.1d correspond to the biases in occurrence of hundreds of different nucleotide words in thousands of ordered 3′UTR sequences (the x-axis). The 3′UTRs are ordered according to differential expression of the corresponding genes: on the left side of the list there are downregulated genes, while on the right side there are upregulated genes. The values on the y-axis correspond to the hypergeometric statistic for the enrichment/depletion of a word in the sequences that precede a point on the line in comparison to all sequences. The values greater than 0 correspond to the enrichment, while the values below zero correspond to the depletion. The occurrence of the majority of the words does not have significant biases (the grey lines). However, the distribution of one word ($GTGCCTT$), represented by the blue line, is empirically different from distribution of all other words. The blue line in the Figure 1.1c corresponds to the enrichment of the word in the 3′UTRs of the downregulated genes. In the Figure 1.1d the blue line shows enrichment in the 3′UTRs corresponding to the upregulated genes. The vertical dashed lines in Figure 1.1c show the P-value cutoffs (0.01 and 0.05 on the two ends of the x-axis) for the fold change t-statistic of the genes in the ordered lists. The dashed lines facilitate selection of the cutoff for identification of the lists of miRNA targets (see text), for example in Figure 1.1c the 0.01 cutoff on the left side corresponds almost exactly to the peak of the initial enrichment of the nucleotide word.

1.2.4 Primary neuronal cultures as a model system to study neuronal biology

In section 1.2.2 transfection of miRNA mimics was described as an efficient way to induce miRNA mediated changes in gene expression that can reveal the direct targets of miRNAs. Chemical transfections are usually performed on cells in culture, therefore to study the functions of miRNAs in neurons using transfections it was necessary to grow neurons in culture. Methods for culturing neurons were being actively developed for over a century (Nelson, 1975; Dichter, 1978). The cultures became a popular model for neuronal development and function, because they offered the researchers easy access to live cells, and made available nearly identical replicates for dose-response and timecourse experiments. One of these methods, known as dissociated primary neuronal culture, was used in this thesis to model neuronal growth and function and study the roles of miRNAs in differentiated neurons.

Dissociated primary neuronal cultures can be obtained from the brains of prenatal, neonatal and adult rats and mice (Brewer et al., 1993; Ahlemeyer and Baumgart-Vogt, 2005; Brewer, 1997; Brewer and Torricelli, 2007). Obtaining cultures from adult brains is more technically challenging, because dissociation of the brain cells is complicated by established adhesion between cell bodies and entanglement of mature neurites. Additionally, viability of cultured adult neurons was noted to be lower than that of prenatal and neonatal neurons and to require a constant supply of trophic factors (Brewer, 1997). Of the prenatal and neonatal brains, the former were selected as the source of neurons for this thesis project, as the use of embryonic material allowed the growth of more pure neuronal populations. The reason for this is due to the wave of embryonic neurogenesis preceding the wave of gliogenesis (Götz and Huttner, 2005; Freeman, 2010). Consequently it is possible to time the dissection of the embryonic brains in order to maximise proportion and number of neurons in the starting material. For example, it was shown that in primary cultures plated from the prenatal rat hippocampus (E18) less than 0.5% of cells were glial (Brewer et al., 1993), while in primary cultures from a neonatal mouse hippocampus the proportion of glial cells was approximately 7% (Ahlemeyer and Baumgart-Vogt, 2005), despite nearly identical isolation and culturing protocols.

In this thesis, in order to maximise the neuronal content of the plating material for primary neuronal cultures, mouse hippocampal and whole forebrain cultures were plated from E17.5 mouse embryonic brains, which approximately coincided with the end of embryonic neurogenesis (Götz and Huttner, 2005). To further enrich cultures for neurons,

the plated cells were cultivated in B27 supplemented Neurobasal (Methods, section 2.1), which is the media specifically developed to enhance neuronal and inhibit glial survival and growth (Brewer et al., 1993). Additionally, particular attention was paid to the control of the concentration of glutamine in the media, as it can deaminate to glutamate, an excitotoxic amino acid, and also because glial growth was suggested to be enhanced at higher glutamine concentrations (Brewer et al., 1993).

The experiments of this thesis project were designed based on the assumption that primary cultures, plated and incubated in conditions maximising neuronal content and survival, were a good model system to study neuronal biology. This assumption was supported by a body of existing evidence, where the cultures, similar to the ones used in this thesis, were shown to be a suitable system to study both neuronal physiology and gene expression. For example, some of the early studies on dissociated primary neuronal cultures showed that morphological characteristics of the cultured neurons were similar to the neurons in the brain (Dichter, 1978; Kriegstein and Dichter, 1983). Later it was demonstrated that neurons in the cultures were electro-physiologically active and developed functional synaptic connections (Bading et al., 1995; Hardingham et al., 2001). Moreover, it was also shown that it was possible to modify the strength of these connections, which meant that primary neuronal cultures may be used to study neuronal plasticity (Arnold et al., 2005).

Apart from studying the morphology and physiology of neurons, primary neuronal cultures can be a model system of choice to study the gene expression program of growing and functioning neurons. An advantage, that primary neuronal cultures provide, is the ease with which neuronal gene expression can be profiled: Because the cultures are enriched in neurons, neuronal gene expression can be studied simply by profiling the extract of total RNA. For example, in a study conducted in Seth Grant's laboratory, by profiling gene expression during development of cultured primary neurons, it was possible to identify hundreds of genes whose upregulation in cultures preceded the morphological and electrophysiological appearance of the synapses (Valor et al., 2007).

And finally, transfection of cultures with miRNA mimics and inhibitors, was shown to be suitable for the identification of miRNA targets (Conaco et al., 2006). Transfection of primary cortical cultures with the miR-124 mimic was shown to downregulate genes, while transfection of the inhibitor upregulates genes that were previously identified and validated as targets of miR-124 (Conaco et al., 2006).

Summary of section 1.2

Advances in our understanding of the mechanisms of miRNA target recognition, in conjunction with whole transcriptome analysis of gene expression, has enabled researchers to conduct experiments, which can identify the whole spectrum of putative direct miRNA targets at the same time. This approach is based on perturbation of expression of individual miRNAs followed by analysis of the sequences of 3′UTRs of differential expressed genes. 3′UTRs of transcripts targeted by a miRNA were shown be enriched in seed matching sites for that miRNA. Therefore identification of such transcripts, which also respond to perturbation of the miRNA expression levels, allows the compilation of lists of putative direct miRNA targets. Such an approach has been previously used, and lists of identified putative targets were significantly enriched in the validated direct miRNA targets (Lim et al., 2005; Giraldez et al., 2006). In this thesis I will describe the use of these experimental and computational techniques to identify putative direct targets of several miRNAs in primary neuronal cultures. Subsequent analysis of the lists of these targets allowed me to suggest explanations for some of the previous observation made in the published literature, and formulate testable hypotheses of miRNA functions in neurons.