Chapter 1: Introduction

The full potential of non-coding RNAs (ncRNAs) is only recently beginning to be realised (Mattick and Makunin, 2006). A prime example of this surge of interest in novel RNA functions is that of microRNAs (miRNAs). Not only has the number of annotated miRNAs increased rapidly, but it has become increasingly apparent that miRNA mediated post-transcriptional regulation is widespread and linked to many key biological processes, including cancer, development and embryonic stem cell self renewal and pluripotency (Gangaraju and Lin, 2009; Medina and Slack, 2008; Zhao and Srivastava, 2007). The identification of novel miRNAs has progressed so rapidly since the advent of this relatively new area of cellular research that there is very limited functional annotation of these miRNAs, bearing in mind the potential complexity of the regulatory networks within which each may participate. As a consequence, the aim of my research was to develop a system that would help to address this deficit in functional annotation through the derivation of a large number of experimentally supported candidate target messenger RNAs (mRNAs) for miRNAs expressed in mouse embryonic stem (ES) cells.

1.1 miRNAs

Metazoan miRNAs are ~21-22 nucleotide (nt) small RNA molecules which, as a general rule, guide a ribonucleoprotein complex (miRNP) to target mRNA molecules by partial complementarity between themselves and the mRNA molecule. The vast majority of mRNAs targeted by a miRNA are either degraded or translationally inhibited. This report will concentrate solely on the miRNAs of metazoans, as plant miRNAs have been hypothesized to

have evolved independently and therefore obey a related, but broadly non-applicable set of rules (Axtell and Bowman, 2008; Mallory and Bouche, 2008).

The related process of RNA interference (RNAi) is triggered by double stranded RNA (dsRNA) molecules, which are processed into small interfering RNAs (siRNAs); RNA molecules that are of approximately the same length as the miRNA. As a rule these are believed to guide ribonucleoprotein (RNP) complexes to mRNA targets, which they match with perfect complementarity. This leads to the cleavage of the target molecules at the point at which they bind. This mechanism of target regulation is not the same as that used by miRNAs in all but atypical circumstances.

Since the discovery of the first miRNA in 1993, the number of known miRNAs has rapidly increased, with the vast majority being identified since the turn of the century (Lee et al., 1993). Currently there are 695 human miRNAs and 488 mouse miRNAs registered in miRBase (Release 12) (Table 1.1A) (Griffiths-Jones et al., 2008). Originally believed to be exceptional, it is now clear that miRNA mediated post-transcriptional regulation has a major influence on both the RNA and protein expression profile of cells. Tens of miRNA species are expressed within every tissue (Landgraf et al., 2007) and each is expected to have hundreds of targets (Friedman et al., 2009). The network of miRNA-mediated control is complex, with elements of both combinatorial regulation by multiple miRNAs targeting the same molecule, and functional redundancy between different miRNA species. In addition feed forward and feedback loops have been recognized involving both miRNAs and proteins (Marson et al., 2008; Petrocca et al., 2008a).

In table 1.1B on the following page, I have included a list of definitions of miRNA related terms used through out this thesis for reference.

A

mmu-miR-92a-1

Name features	Purpose of feature
mmu-	Reference to the organism ('mmu-' = Mouse, 'has-' = human etc.)
92	Novel miRNAs are given a specific number
a	A letter appended to the miRNA number denotes closely related mature sequences, (eg. miR-92a and miR-92b)
-1	Identical mature sequences expressed from distinct loci are numbered sequentially
*****	If ~22nt RNA molecules are derived from opposite strands of a precursor hairpin in a cloning study and one form clearly predominates the less dominant form is appended with a '*'
-3p/5p	If ~22nt RNA molecules are derived from opposite strands of a precursor hairpin in a cloning study and neither form clearly predominates the forms are appended with a '-3p' or '-5p' depending upon which strand of the miRNA hairpin the miRNA originates

B

Term	Definition
miRNA	~21-22nt RNA molecules responsible for guiding a miRNP to partially complementary mRNA molecules leading to post-translational regulation of targets.
pri-miRNA	Primary transcripts from within which miRNAs are initially transcribed.
pre-miRNA	Folded RNA hairpins containing the miRNA molecule. Released from the primary transcript by Drosha cleavage, exported from the nucleus and further processed by Dicer to release the miRNA.
microprocessor	A protein complex containing Drosha and DGCR8 proteins responsible for releasing pre-miRNAs from pri-miRNA transcripts.
miRNP/RISC	Protein complex containing miRNAs (or siRNAs) responsible for orchestrating miRNA guided post-transcriptional gene regulation.
mirtron	A small subset of miRNAs transcribed within short intronic structures. The ends of the pre-miRNA are determined by splice events and hence these are processed in a microprocessor independent manner.

Table 1.1: The miRNA naming convention and definitions of relevant terms used throughout this thesis. A) This is an example of the miRNA naming convention used by miRBase. The names can be divided into 4 or 5 parts, each of which contains information concerning the miRNA's origins or limited information concerning miRNA:miRNA relationships. B) Definitions of terms used throughout.

1.1.1 MicroRNA processing pathway

Mature miRNAs are released from larger RNA molecules by two rounds of RNA cleavage. They are transcribed within primary miRNA (pri-miRNA) molecules, which fold to incorporate the miRNA within the stem of a hairpin structure. Initially the hairpin is released from this longer transcript by the RNase III family protein Drosha. This released hairpin is termed the precursor miRNA (pre-miRNA). Subsequently the loop is removed from the hairpin by the RNase III family protein Dicer and the miRNA is liberated (Fig.1.1). Next I shall discuss the biogenesis of miRNAs in more detail.

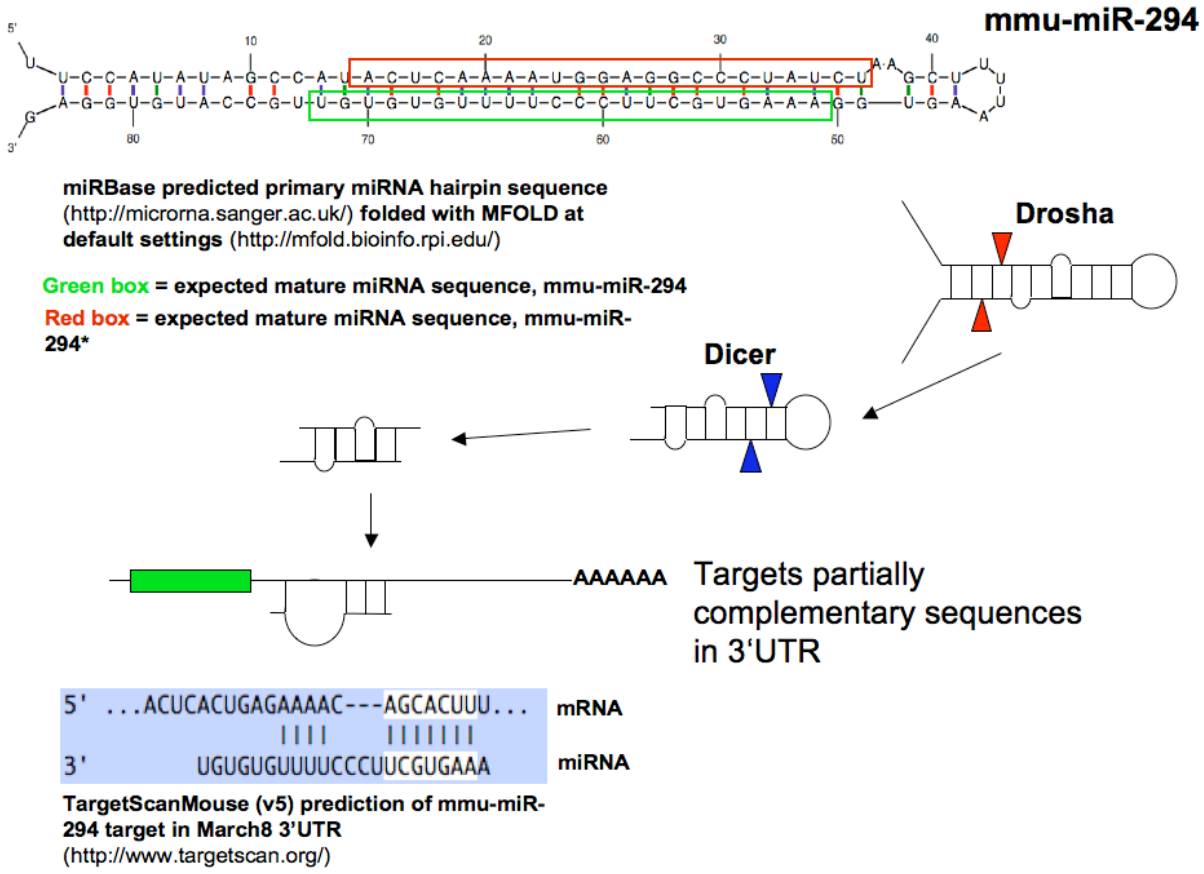

Fig.1.1: A simplified outline of the miRNA processing pathway, depicting the two cleavage steps required to liberate the mature miRNA. Sequence structures depict a representation of the local RNA sequence and structure into which the mature mmu-miR-294 miRNAs are embedded (http://microrna.sanger.ac.uk) (**top**) and a predicted target site interaction between miRNA and target sequence (http://www.targetscan.org) (**bottom**).

1.1.1.1 Pri-miRNAs

Pri-miRNAs can take on several forms. miRNAs can reside within introns or exons of independently transcribed ncRNA molecules or are transcribed along with the mRNAs of protein coding genes, embedded in their introns or untranslated regions (UTRs) (Fig.1.2) (Kim and Kim, 2007; Rodriguez et al., 2004). Multiple miRNAs can reside in clusters, transcribed together as a single unit, to be subsequently cleaved out and processed separately (Houbaviy et al., 2005).

1.1.1.1.1 Transcription of pri-miRNAs

In addition to miRNAs found within the introns of protein coding genes, independently transcribed non-coding pri-miRNA molecules are most commonly transcribed by polymerase II (pol II), although pol III regulated transcription cannot be excluded with upstream Alu repeat sequences seemingly capable of supporting miRNA expression from the human C19MC locus (Borchert et al., 2006). A range of these pri-miRNAs have been demonstrated to possess a 7-methyl guanosine cap and to be polyadenylated (Cai et al., 2004; Houbaviy et al., 2005; Lee et al., 2004a). In addition, α-amanitin, a pol II inhibitor, appears to reduce the levels of 7 pri-miRNAs within HeLa cells and pol II has been demonstrated to bind directly to the miR-23a-cluster promoter region (Lee et al., 2004a). Furthermore, the promoter region of hsa-miR-21, also identified in HeLa cells, is capable of supporting the transcription of functional mRNA transcripts which further suggests these promoter elements are capable of recruiting Pol II (Cai et al., 2004). The mmu-miR-290 cluster has been annotated in considerable detail. Houbaviy et al identified a pol II promoter region upstream of the miRNA cluster. This promoter region contained a TATA box conserved in *H. sapiens*, *M. musculus*, *B. Taurus* and *C. familiaris*, within 35 base pairs (bp) of the transcriptional start

site of the cluster (Houbaviy et al., 2005). More recently, through a comparison of regions found upstream of intergenic miRNAs in *C. elegans*, *H .sapiens*, *A thaliana* and *O. sativa* to a collection of pol III and pol II promoters, Zhou *et al.* demonstrated that the vast majority of miRNAs appear to possess pol II promoters, with 100% of regions upstream (2000bp) of *C. elegans* pre-miRNAs and 96.3% of regions upstream of *H. sapiens* pre-miRNAs predicted to contain possible or definitive pol II promoters. The remaining regions were found to contain either possible pol III promoters or random sequence (Zhou et al., 2007).

1.1.1.1.2 Intronic miRNAs

Direct evidence that miRNAs may be expressed along with host mRNAs was demonstrated in a comparison of 90 human miRNAs to the NCBI expressed sequence tag (EST) database. Chimeric ESTs were found to contain both mRNA sequence and miRNA precursor sequence (Smalheiser, 2003). A recent and more comprehensive study (Kim and Kim, 2007) verified work performed earlier by Rodriguez et al. (Rodriguez et al., 2004), demonstrating that ~80% of the miRNAs that map to ESTs map to introns. Considered alongside all miRNAs (including those with no EST information or those which do not map to known genes), this intronic population accounts for ~50% of the miRNAs investigated in these two independent studies. Of the definitively intronic miRNAs from the Rodriguez *et al.* study, ~3/4 are within the introns of protein coding genes.

Where the intronic miRNAs appear on the same strand as the host gene, it is expected that the miRNAs will be transcribed along with the host transcript and subsequently processed. Microarrays have been employed to demonstrate correlation between the expression of miRNAs and their parent transcripts, again adding weight to the co-expression hypothesis

(Baskerville and Bartel, 2005). These results were replicated by deriving expression data for host transcripts by reverse transcriptase polymerase chain reaction (RT-PCR) and comparing these results to previously generated miRNA expression data (Rodriguez et al., 2004). Both intronic miRNAs tested in this way exhibited the same expression profile as their host transcript.

1.1.1.2 Cleavage of the primary miRNA transcript by the microprocessor protein complex

The initial miRNA processing step (release of pre-miRNAs from the pri-miRNA transcript) takes place within the nucleus. The characteristic 2 nt 3' overhangs and 5' monophosphate groups of mature miRNA duplexes prompted the identification of Drosha (RNASEN) as the enzyme that performs the initial restriction of the miRNA maturation process, as these features are also the byproduct of an RNase III cleavage reaction (Lee et al., 2003).

1.1.1.2.1 The microprocessor

Gregory *et al.* conducted an analysis of two Drosha containing protein complexes of differing sizes in HEK-293 cells (An apparent third complex containing only a smaller isoform of Drosha was not pursued further by this study) (Gregory et al., 2004). The authors found Drosha to be associating with multiple and varying proteins, including DEAD-box helicases, heterogeneous nuclear ribonucleoproteins and Ewing's sarcoma proteins. However, of the proteins tested, Drosha appeared to associate with DGCR8 (a dsRNA binding protein) alone in the smaller ~600kDa complex. This was also the complex which accounted for by far the largest proportion of pri-miRNA processing activity *in vitro*, processing pri-miRNAs into

pre-miRNAs in a site-specific manner. This protein complex constituting of Drosha and DGCR8 has been termed the microprocessor. The microprocessor-pri-miRNA processing activity could be replicated *in vitro* by combining recombinant Drosha and DGCR8, although alone, neither could process pri-miRNAs effectively and Drosha alone exhibited non-specific RNase activity. Examining the role of the microprocessor *in vivo*, siRNAs targeted to either Drosha or *DGCR8* were found to block miRNA processing at the pri-miRNA step. However it should be noted that *in vivo* depletion of three of the other components of the largest Drosha containing complex also had a small effect on mature miRNA levels although as this was much less significant than the effect seen with the smaller complex, the authors conclude it is "therefore more likely that the large Drosha-containing complex has a function in other RNA processing pathways" (Gregory et al., 2004).

A second study was published simultaneously, identifying the homologue of *DGCR8* (Pasha) as the partner of Drosha in *Drosophila* (*pasha*) and *C. elegans* (*pash-1*) (Denli et al., 2004). Again these two proteins co-immunoprecipitated in *Drosophila* S2 cells. Although PASHA co-precipitates with pri-miRNAs, the Drosha:PASHA interaction was unaffected by RNase treatment of the immunoprecipitates, implying a direct interaction between the two proteins. RNAi experiments targeting Pasha in both *Drosophila* and *C. elegans* lead to an accumulation of pri-miRNA and a depletion of mature miRNA as expected.

Fukuda *et al.* were only able to purify the larger of the two Drosha containing complexes, described above, from mouse cells, although they found DGCR8 within this complex (Fukuda et al., 2007). Drosha is thought to be involved in ribosomal RNA (rRNA) processing in addition to miRNA processing (Wu et al., 2000) and this complex was able to process both

miRNAs and rRNAs. This study implicated further proteins (DEAD-box RNA helicase subunits) in the processing of a subset of pri-miRNA sequences in addition to the minimal microprocessor. DGCR8 is not required for rRNA processing, however, as demonstrated in a *Dgcr8* mouse ES cell knock out experiment, in which no effect was seen on the levels of rRNAs caused by the removal of the functional protein (Wang et al., 2007).

Within this newly discovered microprocessor complex it appears that multiple copies of Drosha and DGCR8 interact. However, the enzymatic processing centre of Drosha is formed by intramolecular dimerisation of the two RNase sites within each Drosha molecule (Han et al., 2004).

This first excision defines one end of the mature miRNA sequence. This initial processing step proceeds co-transcriptionally; Drosha associating with the nascent strand in a DGCR8 dependent manner (Morlando et al., 2008).

1.1.1.2.2 Recognition and mechanism of pri-miRNA processing

Han *et al*. (Han et al., 2006) calculated the average local RNA structure of miRNAs in humans and flies. This structure consisted of an approximately 33bp stem (approximately 3 helical turns) with a terminal loop (for an example of a primary miRNA structure see Fig.1.1). The stem structure was flanked by single-stranded RNA segments. From observations discerned by the use of labeled transcripts containing various artificial mutations and structural alterations, and an immunopurified microprocessor complex, Han *et al.* proposed a model whereby DGCR8 binds firmly to the single stranded to double stranded junction of the pri-miRNA structure. Drosha binds to this anchoring complex transiently with

its RNase domains positioned approximately 11bp from the base of the stem; the site of RNase cleavage. Zeng et al. reached broadly similar conclusions with regard to microprocessor function. Again, using a series of *in vitro* assays in addition to transfected plasmid expression constructs they concurred that the microprocessor most efficiently processed a stem flanked at each end by lengths of single stranded RNA (Zeng and Cullen, 2005). However, in contrast to the other study, they proposed a mechanism whereby Drosha or a protein complex recognized the pre-miRNA hairpin loop and then mediated RNase cleavage ~22bp from that junction (Zeng et al., 2005). Han *et al*. attempted to reconcile some of the differences presented by these two conflicting models for microprocessor action by observing that the presence of a large terminal loop required in the Zeng hypothesis could be a consequence of the requirement for the single stranded RNA (ssRNA) at both ends of the duplex for efficient cleavage. Han *et al*. were unable to replicate the critical experiments of Zeng *et al*. where alterations to the position of the terminal loop of the pre-miRNA structure altered the site of cleavage accordingly.

1.1.1.2.3 Processing of intronic miRNAs

Until recently it had been assumed that miRNAs were processed from introns post-splice. However, Kim *et al.* have observed spliced ESTs that begin a few base pairs from a proposed Drosha cleavage site and contained sequences from the miRNA containing intron along with adjacent spliced exons from the parent transcript. This EST, which may be derived from a Drosha cleavage product, would imply that miRNAs are processed from unspliced introns within otherwise spliced mRNA molecules (Kim and Kim, 2007). Further experiments in HeLa cells confirmed the presence of partially spliced parent mRNA transcripts within which the miRNA containing intron appears to be spliced after other intronic sequences, implying

that the microprocessor may interfere with the splice machinery. Artificial expression constructs containing miRNAs within the introns of protein coding genes were used to demonstrate that splicing is not required for intronic miRNA processing. These constructs were also used to demonstrate that the presence of a miRNA within the intron of a gene did not significantly affect the levels of fully spliced mRNA (Kim and Kim, 2007). These observations imply that mRNAs cleaved by Drosha are subsequently spliced. The authors propose the "exon-tethering" model of Dye et al. as a possible explanation for this (Dye et al., 2006).

1.1.1.2.4 miRNAs within the 3'UTRs of protein coding genes

Intriguingly Rodriguez *et al.* identified 2 miRNAs that map within the exons of the 3'UTRs of protein coding genes, in a study of mouse and human miRNAs (ENSMUSG00000018171 and ENSG00000163430) (Rodriguez et al., 2004). It is worth noting that experiments conducted in Hek-293T cells with luciferase reporter genes harboring a pre-miRNA in their 3'UTR detected a large proportion of the truncated, Drosha processed, luciferase transcript in the cytoplasm and a much smaller fall in luciferase activity than expected. This prompted a hypothesis that the processed transcripts were still able to function as mRNAs despite the truncation event (Cai et al., 2004).

1.1.1.2.5 Co-transcriptional miRNA processing

Morlando *et al.* performed Drosha chromatin immunoprecipitation (ChIP) in HeLa cells. They discovered an enrichment of Drosha at the sites of expressed intronic and intergenic miRNAs that was both RNase sensitive and DGCR8 dependent (Morlando et al., 2008). The authors proceeded to investigate the process of co-transcriptional miRNA processing in

detail. In addition to copious other experiments, the authors used Northern blots to demonstrate the successful maturation of miRNAs expressed downstream of a poly-A site from highly unstable transcripts, generated by 3' transcription from unterminated ß-globin gene constructs. Furthermore, the authors demonstrated the recruitment of exonucleases to the sites of Drosha cleavage in HeLa cells. These nucleases seem to enhance splicing efficiency following Drosha cleavage as depletion of these proteins by RNAi led to a reduction of splicing efficiency. The effect of this depletion was itself reduced upon the depletion of Drosha.

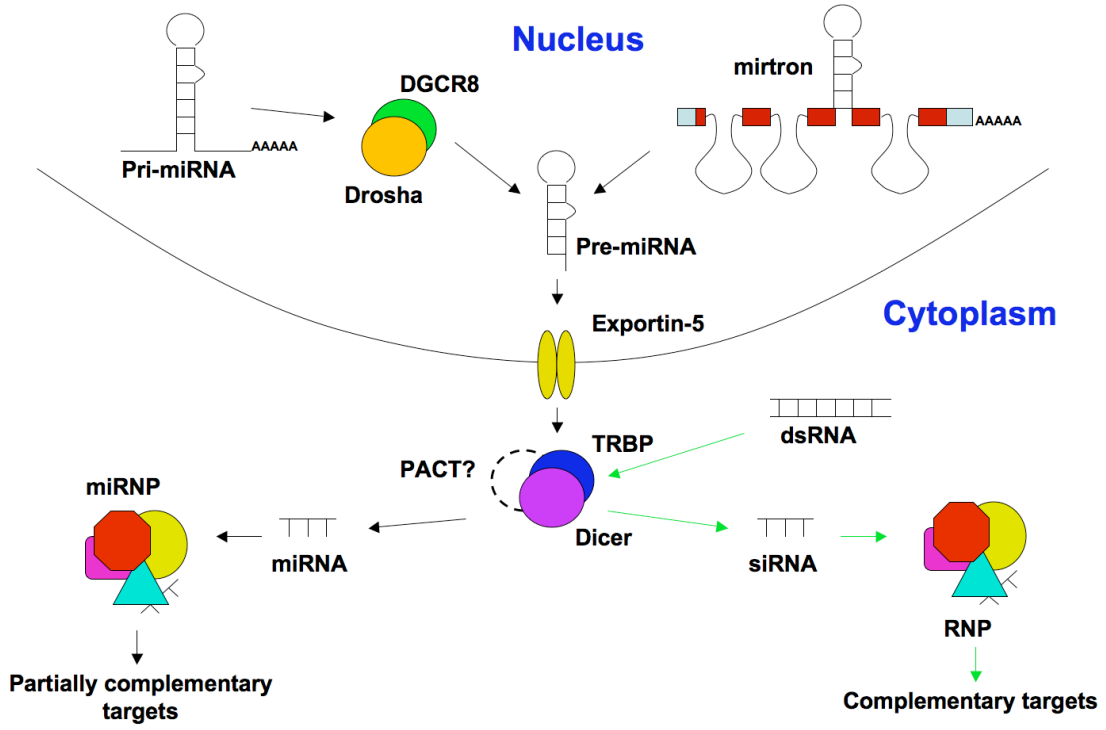

Fig.1.2: **Canonical microRNA processing pathway in vertebrates, with the introduction of mirtronic miRNAs at the pre-miRNA stage.** miRNAs are generally transcribed in longer pri-miRNA molecules, replete with secondary structure. Drosha (RNASEN) and DGCR8 operate in unison to liberate the pre-miRNA hairpin. This is exported from the nucleus to the cytoplasm by exportin 5 (XPO5). In the cytoplasm the pre-miRNA is further processed by Dicer (DICER1) with associated cofactors to release the mature miRNA (PACT (PRKRA) and TRBP (TARBP2)). One strand of the miRNA duplex is selected and incorporated into the miRNP, which it guides to target mRNA molecules to generally block translation or cause degradation via deadenylation. Mirtrons follow the same processing pathway for the most part, but are initially excised from the parent RNA molecule via a splice reaction (see section 1.1.1.6). The RNAi pathway overlaps with the miRNA processing

pathway at the Dicer cleavage stage. As research continues the distinction of these two pathways is becoming more murky and complicated by overlaps in components and function.

1.1.1.3 The fate of the precursor miRNA

Above I have discussed in detail the first cleavage event in the miRNA processing pathway required to release the pre-miRNA from the pri-miRNA transcript. This initial processing step takes place in the nucleus. The precursor hairpin (~70nt molecule) is then transported from the nucleus by exportin 5 (Yi et al., 2003; Zeng and Cullen, 2004) in cooperation with a Ran-GTP cofactor (Fig.1.2). Once in the cytoplasm the pre-miRNA is further processed by a second RNase III enzyme, Dicer, in order to liberate the mature miRNA from the larger RNA molecule.

1.1.1.3.1 Pre-miRNA processing

Dicer removes the loop from the end of the pre-miRNA, releasing the mature miRNA and once again leaving a 5' phosphate, 3' hydroxyl and a 2 nt 3' overhang, all of which are characteristic of miRNAs (Grishok et al., 2001; Ketting et al., 2001; Knight and Bass, 2001). It is worth noting that the miRNA processing pathways and the RNAi pathway converge at this point (Bernstein et al., 2001; Hutvagner et al., 2001) with Dicer responsible for the processing of dsRNAs in both pathways (Fig.1.2). These roles are performed by distinct Dicer orthologues in *Drosophila* (*Dcr-1* and *Dcr-2*) (Lee et al., 2004b), but in other animals from *C. elegans* (*dcr-1*) to *M. musculus* (*Dicer1*) and *H. sapiens* (*DICER1*) a single Dicer gene exists and this single protein performs both duties (Fig.1.2).

Like Drosha, Dicer has been shown to associate with dsRNA binding proteins (R2D2 and Loquacious in *Drosophila* and PACT (PRKRA) and TRBP (TARBP2) in human cell lines

and in the mouse (Chendrimada et al., 2005; Forstemann et al., 2005; Kok et al., 2007; Liu et al., 2006)) (Fig.1.2). However, unlike Drosha, which requires DGCR8 in order to cleave primary miRNAs *in vitro* (Han et al., 2004), Dicer is capable of cleaving both dsRNAs and pre-miRNAs *in vitro*, in the absence of a dsRNA binding partner (Chendrimada et al., 2005).

1.1.1.3.2 Dicer associated proteins

From this point on the understanding of the mechanism of miRNA function becomes more difficult to discern with considerable disagreement evident between a number of papers, some of which are at present very difficult to reconcile.

Within mammalian systems, Dicer has been found to interact *in vitro* and *in vivo* with PACT and TRBP in stable complexes. Chendrimada *et al.* found that TRBP bound Dicer and in turn allowed the association of Argonaute 2 (EIF2C2) (discussed later) to an siRNA associated complex (Chendrimada et al., 2005). They suggested that TRBP might therefore be involved in the initial stages of RNP/RISC (RNA-induced silencing complex) complex assembly (the miRNA and siRNA effector complexes). They also found that the depletion of TRBP in HEK-293 cells seemed to lead to the destabilisation of Dicer and a reduction of mature miRNA and siRNAs. Haase *et al.* did not demonstrate a fall in mature miRNA levels in cell lines depleted for TRBP (Haase et al., 2005), nor did they see a destabilisation of Dicer. They did however see that endogenous miRNAs had a reduced effect on reporter gene transcripts bearing perfectly complementary target sites in their UTRs following TRBP depletion, again implying a role for TRBP in RNP/RISC assembly.

Kok *et al*. found Dicer, PACT and TRBP to form trimeric complexes in HEK-293 cells and in mouse testicular tissue (Kok et al., 2007). Through a series of processing experiments both in artificial systems and in human cell lines the authors found that Dicer cleavage of dsRNAs and short hairpin RNAs (shRNAs) was improved by the presence of both PACT and TRBP. However, in contrast to Haase *et al.* they also found that the Hek-293 cells depleted in TRBP remained susceptible to siRNA transfection but not shRNA transfection, whereas Haase *et al.* found that siRNA effects were abolished upon the depletion of TRBP in a Hek293T-REx cell line. Ultimately, Kok *et al.* suggested that TRBP and PACT function at the stage of siRNA production. These differences are difficult to reconcile and further experiments are necessary to clarify the situation.

1.1.1.4 Regulation of the miRNA processing pathway

It is clear that miRNAs co-expressed in clusters are not always present within the cell in equal quantities in their mature form. This would suggest that there are mechanisms whereby either the processing of mRNAs to their mature forms is regulated or that they are degraded at different rates. One such mechanism has been described for the regulation of the let-7 miRNA family by Lin28a (*LIN28*) and Lin28b (*LIN28B*) in vertebrates. Two systems have been proposed for lin-28's interaction with the miRNA. Newman *et al.* found that Lin28 binds to specific sequences in the loop of the let-7 hairpin, subsequently blocking the Drosha cleavage reaction (Newman et al., 2008). Heo *et al.* found the block to be prior to Dicer cleavage of their target precursor. Again the authors found Lin-28 was able to bind the let-7 molecule but in this case the let-7 precursor appeared to acquire a uracil tail in a Lin-28 dependent manner and was more rapidly degraded than the standard precursor-miRNA (Heo et al., 2008).

1.1.1.5 miRNPs, the effector complexes of miRNA mediated regulation

Both siRNAs and miRNAs are incorporated into RNPs which they then guide to target sequences in mRNAs. In the case of miRNAs in the vast majority of cases these target sequences are partially complementary to the miRNA sequence in metazoans. It could be expected that the strand of the miRNA selected to guide the miRNP is the strand with the least thermodynamically stable 5' end, with 5' instability being a property of functional siRNAs (Khvorova et al., 2003). However, with the advent of high throughput sequencing it is becoming more apparent that the so called "star" or "passenger" strand also plays an important part in post-transcriptional regulation of target sequences. This is supported by evolutionary data suggesting a biological role for both miRNA strands (Okamura et al., 2008).

At the core of the miRNP are the RISC-associated argonaute proteins (reviewed in Hutvagner *et al.* (Hutvagner and Simard, 2008)). These proteins are thought to be important effectors of miRNA function. Tethering of Argonaute-like proteins to mRNAs in the absence of miRNAs is sufficient to initiate post-translational control of the target sequence (Pillai et al., 2004), while Argonaute-like proteins have also been demonstrated to possess the "Slicer" activity required to cleave target sequences with perfect complementarity to an siRNA or miRNA (Liu et al., 2004; Meister et al., 2004).

Humans and mice possess four of these proteins AGO1-4 (EIF2C1, EIF2C2, EIF2C3, and EIF2C4). The PIWI module (Mid domain) of these proteins binds the 5' end of the miRNA, while the PAZ domain is thought to recognize the 2 nt 3' ssRNA overhang, produced by the

RNase III processing of the miRNAs and siRNAs. The AGO proteins are incorporated within larger protein structures. A recent study of human AGO1 and AGO2 identified a large number of proteins with numerous associated functions in what appeared to be 3 complexes of differing molecular weight (Hock et al., 2007). Interestingly, all four AGO proteins have been purified alongside a similar selection of proteins which implies a degree of functional redundancy between each of the AGO containing complexes (Landthaler et al., 2008).

Further studies have endeavored to identify the miRNAs associated with each AGO in humans. Argonaute-like proteins are known to have diversified in their function, with only human AGO2 exhibiting "slicer" activity required to cleave target sequences that are perfectly complementary to a miRNA or siRNA (Liu et al., 2004; Meister et al., 2004). Disruption of the *Eif2c2* gene in the mouse also leads to severe abnormalities and an embryonic lethal phenotype arguing against total functional redundancy amongst mammalian AGO proteins. It is therefore potentially surprising that evidence supports miRNAs binding to AGO proteins indiscriminately (Liu et al., 2004; Meister et al., 2004). It should be noted however that more recent experiments conducted through the immuno-precipitation of endogenous AGO proteins and the pyrosequencing and 454 sequencing of associated miRNAs, suggest that although overall the miRNA population from AGO1, AGO2 and AGO3 complexes are broadly similar, there are some differences which could direct the three complexes to subtly different target populations (Azuma-Mukai et al., 2008; Ender et al., 2008).

1.1.1.6 Mirtrons and other exceptions to the canonical rules

The advent of high throughput sequencing has allowed the small RNA complement of organisms and cell lines to be characterized beyond anything that was achievable through low throughput cloning methods, identifying rare miRNAs beyond the sequencing depth of more conventional methods. In addition, as these new methods do not require sequence complementarity for miRNA detection, they are able to easily profile these relatively rare small RNA species previously not profiled by microarray based detection strategies.

These techniques identified mirtrons as a new miRNA species that were not processed in the same way as the majority of miRNAs (see above). First discovered in *C. elegans* and *D. melanogaster* (Okamura et al., 2007; Ruby et al., 2007), these miRNAs are expressed within the short introns of other genes. However, rather than be sliced from the intron by Drosha, the entire intron is spliced by the splicing machinery and the spliced lariat is debranched to form a miRNA-precursor in a fashion that is microprocessor independent. Subsequently these pre-miRNAs seem to re-enter the canonical miRNA processing machinery at the stage of nuclear export, as defined by a series of RNAi experiments in S2 cells (Okamura et al., 2007; Ruby et al., 2007).

Berezikov *et al.* extended the search for mirtrons to mammals, proposing 19 mirtrons and 46 mirtron candidates based on high throughput sequencing data, restricting the search to short introns in mammalian genomes (Berezikov et al., 2007). Mechanistic evidence for the existence of mammalian mirtrons was provided by small RNA sequence libraries from Dicer and *Dgcr8* knockout mouse embryonic stem cell lines (Babiarz et al., 2008). Although by far the majority of miRNAs demonstrated a canonical requirement for both proteins, mirtrons

and other DGCR8 independent miRNA species were identified. Further sequence evidence was provided for miR-877 which had been predicted by Berezikov *et al.* as a mirtron, and the libraries also allowed the authors to demonstrate that these mirtrons are indeed DGCR8 independent as expected (Fig.1.2). miR-702 and miR-1981 were also identified as mirtrons. These mirtrons appear to be longer than those seen in invertebrates but still fold into pre-miRNA-like hairpins. miR-1982 also had a mirtron like structure and the same enzymatic dependencies. However, this intron folded into a structure reminiscent of dme-mir-1017 identified by Ruby *et al.* in *Drosophila*, with a single stranded RNA tail at one end of the hairpin proposed to be released by splicing (Ruby et al., 2007). In order for this pre-miRNA to enter the canonical pathway, the authors hypothesise that this tail would be removed by an as yet unidentified nuclease.

In addition to mammalian mirtrons, a number of other hairpin, DGCR8-independent, Dicer-dependent miRNA precursors were identified. miR-320 produced the most abundant reads in this category. Unusually, it was not highly conserved beyond the hairpin pre-miRNA. This is in contrast to the majority of conserved miRNAs, as miRNAs must maintain the stem structure for a further helical turn (approximately) in order to provide the optimal substrates for the microprocessor. The authors also noted that the majority of the reads mapped to the 3' arm of this precursor, as would be expected if the 5' end did not possess the residual phosphate left by RNase III cleavage. Hence, fragments from this side of the hairpin would not be cloned successfully. miR-484 also fell into a category displaying similar features to miR-320. miR-1980 had a tailed hairpin-like structure reminiscent of miR-1982 (discussed above) although miR-1980 was not within an intron. Clearly there are a number of non-canonical pathways by which a small subset of miRNAs can be processed.

One of the most novel findings of this paper was the concept that miRNAs may also be processed from other non-coding RNAs under a certain set of circumstances. A stack of sequence reads resembling a distribution normally associated with a miRNA gene, mapped to an annotated transfer RNA (tRNA). This stack was again Dicer dependent but DGCR8 independent. The locus also seems to be capable of being processed as a tRNA and will fold into either a hairpin structure or a tRNA cloverleaf structure. As a tRNA, this locus is pol III transcribed.

As another example of DGCR8 independent processing of a non-coding RNA into a potentially functional miRNA, a human small nucleolar RNA (snoRNA) has recently been demonstrated to produce a functional miRNA (Ender et al., 2008). Although it seems to be a small proportion of this snoRNA that is converted into the miRNA, luciferase assays have been used to demonstrate that CDC2L6 may be an example of an endogenous target of this unconventional miRNA.

1.1.2 Mechanism of miRNA function

In the vast majority of cases, miRNAs guide the miRNP to transcripts with target sites partially complementary to the miRNA sequence, generally within a mRNA's 3'UTR. The miRNP is then responsible for the post-transcriptional regulation of the target transcript. The mechanisms by which metazoan miRNAs/miRNPs function are still poorly understood (For review see (Filipowicz et al., 2008)). This is in part due to a rapidly increasing plethora of exceptions to the general rules of miRNA function and partly due to conflicting hypotheses being proposed and experimentally supported by differing experimental evidence. Two broad

methods of post-transcriptional gene regulation are thought to predominate. The first is destabilization and degradation of the target transcripts, while the second is through translational inhibition. I will discuss each below together with a variety of exceptions.

Advances in the high throughput analysis of the cellular proteome has allowed an investigation of the relative contribution of these two mechanisms to the effect miRNAs have on cellular expression (Baek et al., 2008; Selbach et al., 2008). Using the new "pulsed stable isotope labeling with amino acids in cell culture" (pSILAC) method to differentially label protein samples in culture these papers show, through the over expression of miRNAs or disruption of endogenous miRNAs, that the majority of miRNA targets are either repressed at both the mRNA level and the translational level or that mRNA destabilisation seems to account for the changes. However, some targets were seen to be almost totally regulated at the translational level with no apparent change to mRNA levels.

1.1.2.1 Deadenylation: Destabilisation of targets through the removal of the poly-A tail

Of these two widely accepted mechanisms of miRNA action, target degradation is perhaps the most clearly understood. While investigating the role of miR-430 in the early development of the zebrafish, Giraldez *et al.* demonstrated that this miRNA triggered the deadenylation of its targets (Giraldez et al., 2006). Wu *et al.* noted the same phenomenon in a mammalian system (Wu et al., 2006). This deadenylation would be expected to precede miRNA triggered mRNA degradation. Indeed, it has been recently shown that the depletion of the deadenylation complex in *Drosophila* S2 cells leads to an enrichment of miRNA targets in these cells (Eulalio et al., 2009). The authors of this paper went on to predict that

"60% of transcripts up-regulated in AGO-1 depleted cells are normally degraded through deadenylation".

Further to this they demonstrated that the depletion of decapping activators (*Ge-1* and *me3IB*) also inhibited miRNA mediated degradation of targets, although there remains evidence that targets were still deadenylated despite their stability. This replicated earlier work by Eulalio *et al.* in 2007 (Eulalio et al., 2007) and implies that the deadenylation of miRNA targets is followed by the removal of their 5' cap. It is noted that the degradation of a miRNA target appears to be independent of miRNA induced inhibition of translation. Even with a background of a total block in the initiation of reporter mRNA translation, the reporter construct cotransfected with a miRNA for which it bears targets is degraded more effectively than a reporter transfected with no targeting miRNA (Eulalio et al., 2009; Giraldez et al., 2006; Wu et al., 2006).

The widespread destabilisation of miRNA targets has been utilised by the recent Sylamer program, which can identify enrichments of predicted miRNA targets within gene lists ordered depending upon mRNA expression changes following miRNA addition and disruption experiments (van Dongen et al., 2008). These clear enrichments are seen either amongst the up regulated or down regulated gene sets depending upon whether the miRNAs are being added to or removed from the system.

1.1.2.2 Translational inhibition

From very early in the study of miRNAs, translational inhibition has been recognised as a method by which miRNAs inhibit the expression of their target mRNAs (Wightman et al.,

1993). However, it remains far from clear how miRNAs achieve this. In particular, it is not obvious whether miRNAs inhibit translation at the initiation step or at a post initiation stage. Polysomal fractionation has been used to demonstrate blocks at both stages of translation. Inhibition of reporter mRNAs by let-7 has been demonstrated to cause a shift of target mRNA to the top of the sucrose gradient implying a block in the association of the target with ribosomes at the initiation of translation (Pillai et al., 2005). In contrast other studies have suggested that the inhibition of target mRNAs leads to no change in the polysomal association seen on these gradients, suggesting that the translational inhibition occurs post-initiation (Olsen and Ambros, 1999; Petersen et al., 2006).

A number of mechanisms have been suggested for the miRNA dependent regulation of translation at the stage of initiation. Wakiyama *et al.* noted that the polyadenylation of transcripts appears necessary for translational inhibition, implying that deadenylation may play a role in the inhibition (Wakiyama et al., 2007). It is known that the interaction between the poly-A tails of mRNAs and their caps enhance translation. However, contradictory results from subsequent studies refute this. Pillai *et al.* found that a poly-A tail is not required for translational inhibition (Pillai et al., 2005) and this was again corroborated by Wu *et al.* (Wu et al., 2006).

It also appears that the m(7)G-cap of the mRNA plays an important role in translational suppression. Pillai *et al.* found that mRNAs required a m(7)G-cap for repression with neither an internal ribosome entry site (IRES) nor tethered initiation factors acting as adequate substitutes (Pillai et al., 2005). Kiriakidou *et al.* noted that the AGO proteins possess a domain that resembles EIF4E, capable of binding the m(7)G-cap. They went on to

demonstrate that the disruption of this domain led to a loss of the block of initiation, implying that AGO proteins may disrupt the initiation complex (Kiriakidou et al., 2007). This finding was later extended to demonstrate that the block disrupts the recruitment of the 80S ribosome to the targeted transcripts (Mathonnet et al., 2007).

An alternative mechanism has been presented by Chendrimada *et al.*, who demonstrated an association between TRBP and EIF6. EIF6 associates with the 60S ribosomal associated factor and in doing so disrupts the assembly of the translationally competent 80S ribosome (Chendrimada et al., 2007).

miRNAs have also been shown to co-sediment with the polysome fraction on a sucrose gradient. By subsequently blocking mRNA translation with exogenous agents, it has been demonstrated that at least under specific circumstances miRNAs are associated with actively translated mRNAs (Maroney et al., 2006). In contrast to other publications, Petersen *et al.* found IRES dependent translation to remain susceptible to regulation by miRNAs/bulged-siRNAs. As mentioned above they also noted no change in polysomal occupancy on inhibition. Therefore, judging inhibition to occur post initiation, they then proceeded to inhibit translation initiation and noted more rapid dissociation of target mRNAs from polysomes than control mRNAs. They propose that miRNA-triggered premature release of target peptides from ribosomes is a cause of miRNA mediated translational inhibition (Petersen et al., 2006).

1.1.2.3 A role for the P-body

Processing bodies (P-bodies) are discrete cytoplasmic foci; a site of mRNA sequestration and degradation in the cytoplasm, (Reviewed (Eulalio et al., 2007a). It is becoming increasingly apparent that P-bodies are intricately involved in miRNA function. GW182 (a P-body associated protein) binds AGO proteins. Behm-Ansmant *et al.* revealed that GW182 is required for both translational inhibition and miRNA target degradation. They also noted that the CCR4:NOT1 deadenylation complex and DCP1 and DCP2 from the decapping complex, all of which are associated with the P-body, are required for target degradation (Behm-Ansmant et al., 2006). Given these associations and the localization of mRNAs to P-bodies in a miRNA dependent manner (Liu et al., 2005), it is surprising that the disruption of P-body integrity does not have a pronounced effect on miRNA dependent regulation of reporter constructs (Eulalio et al., 2007b).

1.1.2.4 Reconciling the different mechanisms for miRNA mediated post-transcriptional regulation

Reconciling these proposed differences in the miRNA functional mechanism is not easy. It is of note that the experiments were conducted in systems varying from *C. elegans* to human cell lines, and that for some of the findings cell extracts were used in addition to examining the system in a cellular environment and whole organisms. It is intriguing to think that the role of miRNAs and their mode of action may vary depending upon system and circumstance. Indeed Kong *et al.* has demonstrated that in HeLa cells, target mRNA constructs expressed from a simian virus 40 (SV40) promoter appear to be regulated at the initiation stage of translation, while mRNAs expressed from a thymidine kinase (TK)

promoter are repressed at a post-initiation stage. Surprisingly the promoter of the target gene appears to determine the form of translational inhibition (Kong et al., 2008).

1.1.2.5 Other miRNA mediated regulatory mechanisms

In addition, to the more widespread mechanisms of post-transcriptional gene silencing by miRNAs, there are what appear to be less common miRNA mechanisms. MiR-196 has been demonstrated to trigger the cleavage at a highly complementary site (with a single G:U wobble), in the 3'UTR of *Hoxb8* transcripts by the same mechanism used for siRNA directed mRNA cleavage (Yekta et al., 2004, Mansfield et al., 2004). In addition in mammalian systems, a cluster of maternally expressed miRNAs at the imprinted *antiPeg11* locus regulate a paternally expressed antisense transcript transcribed from a gene on the opposite strand (*Rtl1/Peg11*) (Davis et al., 2005). Again, these miRNAs appear to trigger the cleavage of their complementary mRNA target sequences. However, this method of targeting is massively outweighed by the more canonical targeting of partially complementary sites by miRNAs.

Under specific circumstances miRNAs can also cause the post-transcriptional up regulation of target transcripts (Vasudevan and Steitz, 2007). During cell cycle arrest three different miRNAs flipped their mode of regulation from suppression to activation.

In addition miRNAs have very recently been shown to activate (Place et al., 2008) and repress gene transcription in mammalian cells (Kim et al., 2008a). In both cases the miRNA appears to target each gene upstream of its start site. MiR-320 is encoded directly upstream

of the *POLR3D* gene and seems to direct Ago1 to the promoter and cause transcriptional silencing. MiR-320 expression also appears to correlate with H3K27me3 and EZH2 (a methyl transferase) at the promoter. In the case of transcriptional activation the miRNA does not match the target site with perfect complementarity, but miR-373 appears to cause the activation of E-cadherin (CDH1) and CSDC2 by binding the promoter region, although the authors report no defined mechanism by which this occurs.

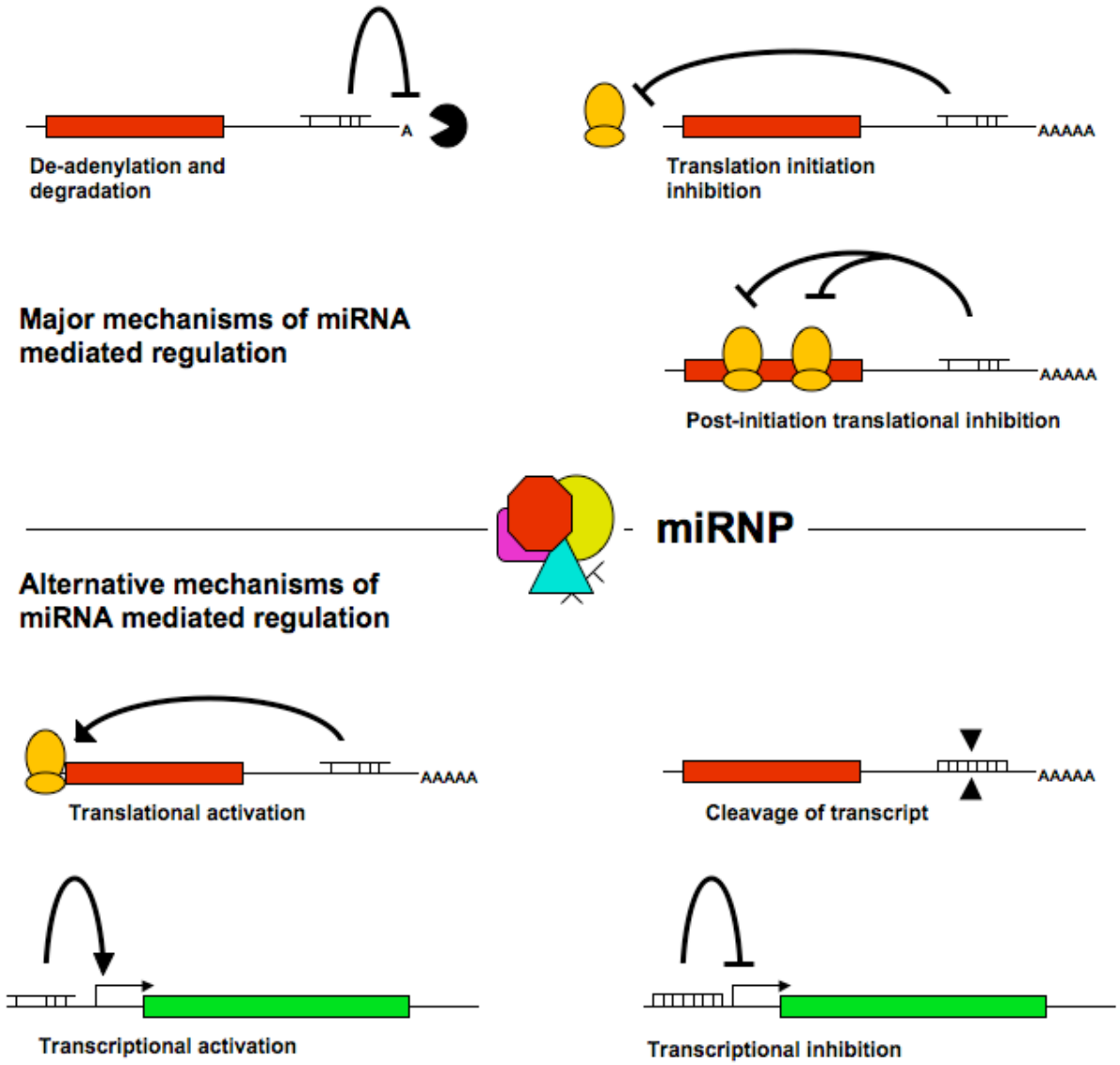

Fig.1.3: The mechanisms of miRNA mediated regulation. Top: Most common and widely researched miRNA/miRNP functions. **Bottom:** Alternative miRNA functions.

1.1.3 The rules of miRNA target recognition and target prediction algorithms

An increase in relevant data is leading to a better understanding of the rules which govern miRNA:mRNA target interactions and to rapid improvements in miRNA target prediction. In this section I intend to describe the progress in this field both before the inception of my studies and over the course of the last few years.

It was clear from some of the earliest miRNA targets identified experimentally that miRNAs bind targets of partial complementarity, with a preference for binding at the 5' end of the miRNA and within the 3'UTRs of mRNAs. MiRanda, one of the earliest miRNA target prediction algorithms, used filters based on these observations to select potential targets. Potential targets are identified by looking for complementarity across the length of the miRNA allowing G:U wobbles and a degree of mismatch. Preference is given to targets bound more completely at the 5'end. This initial matching process is followed by a series of filters which calculate both the thermal stability of the predicted targets and their evolutionary conservation across species. Multiple sites within the same UTR were summed to provide a list of the most confident predictions (Enright et al., 2003; John et al., 2004). A second method, TargetScan, used a signal/noise calculation based on the number of predicted targets for a true miRNA sequence divided by the number of targets predicted for shuffled sequence to define a set of rules for optimal target prediction (Lewis et al., 2003). This identified perfect complementarity to the "seed" sequences of each miRNAs (the region from bases 2-8 counted from the 5' end of the miRNA), (Fig.1.4) as one of the best predictors of miRNA targets. Better signal to noise ratios were achieved by a requirement for target site conservation. In addition to these criteria the program considered 3' complementarity, target

site free energy and the number of targets for each miRNA within a UTR as key criteria for target prediction. One perceived disadvantages of this method would be its propensity to miss targets with imperfect seed sequences, such as one of the let-7 target sites in the 3'UTR of the *lin-41* gene in *C. elegans* (Vella et al., 2004). At this early stage experimentally validated targets were rare so reporter assays were used to confirm target predictions. In brief, a segment from a 3'UTR containing a suspected target was cloned downstream of a luciferase open reading frame (ORF). An identical construct was also designed with point mutations within the target sites to disrupt miRNA binding. These target constructs were transfected into cells expressing the miRNAs of interest and the effect of the unmodified target site was monitored in relation to the mutated version to determine whether the relevant site is susceptible to miRNA targeting and induced post-transcriptional regulation (Lewis et al., 2003). Failing this, target prediction algorithms were assessed against the limited data available (John et al., 2004).

Studies using the reporter construct principle explained above and transfected into HeLa cells along with siRNA/miRNA duplexes, or in the presence of endogenously expressed miRNAs provided a further investigation of the properties of effective mRNA targets (Doench and Sharp, 2004). These experiments confirmed the importance of the 5' pairing of the miRNA to its target in leading to the down-regulation of the luciferase reporter, whereas pairing at the 3' end was deemed less important, although it was regarded as a modulating factor. G:U wobbles within the pairing appeared detrimental to miRNA control and miRNAs appeared to repress the reporter gene in a concentration dependent manner. Constructs with the target sites for multiple miRNAs/bulged-siRNAs in their 3' UTR also clearly demonstrated that miRNAs were able to control mRNAs in a combinatorial manner with multiple miRNA

regulating the same mRNA simultaneously by targeting different sites in the UTR. The fold repression of a construct with 2 sets of 2 target sites within it's 3' UTR, each set partially complementary to a different siRNA, increased from approximately 3 fold to 8 fold if both siRNAs were introduced together as opposed to one at a time.

Brennecke *et al.* extended this work, expressing green fluorescent protein (GFP) with target sites in its 3'UTR in an imaginal disc of *Drosophila*. miRNAs expressed within a region of the same disc were used to assess the effectiveness of each target site (Brennecke et al., 2005). The authors found that perfect complementarity within the region from 5' bases 2-8 of the miRNA was sufficient to confer repression of a target without complementarity within the 3' end of the miRNA. Complementarity between bases 1-8 of the miRNA and the target provided even greater repression of the target mRNA. They found no correlation with the pairing energy of the 5' bases. Targets with limited 5' pairing to the miRNA (miRNA bases 2-5) were functional but required substantial pairing at the 3'end of the miRNA. Again G:U bases within the seed appeared detrimental but potentially tolerable for more limited function. Ultimately the authors defined 3 types of target; 1) Canonical sites with substantial pairing at both the 5'and 3' ends of the miRNA; 2) Seed sites with 5' pairing but little 3' pairing; 3) 3' compensatory sites with at least 4 bases paired within the seed and strong pairing at the 3' end of the miRNA. The seed sequences of true miRNAs were found to be more conserved within 3' UTRs than random sequences, while regions adjacent to these sites were rarely conserved. In addition the authors estimated that there are probably only 1-20 3' compensatory sites per miRNA.

Transfection of miRNA duplexes for miR-1 and miR-124 into HeLa cells caused a reorganization of the cellular expression profile, as measured by microarray, to more closely resemble the profiles of those tissues within which these miRNAs are normally expressed (Lim et al., 2005). The transfection of miRNAs and the use of microarray technology allowed the experimental investigation of miRNA-target interactions on a large scale. Once again down-regulated transcripts demonstrated an enrichment for sequences complementary to bases 2-7 of the transfected miRNA, again revealing the significant contribution of this region of the miRNA to target selection.

The release of the chicken genome led to an update of TargetScan; TargetScanS (Lewis et al., 2005). Using the same signal to noise ratio as before to judge effectiveness, by extending required conservation to include 5 genomes, including the chicken genome, the algorithm was stripped back to predict targets based solely on the conservation of seed sequences (miRNA bases 2-7) (Fig.1.4). The signal to noise ratios were improved by requiring a Watson-Crick match between base 8 of the miRNA and the target (7mer-m8 seed) or by requiring the target base opposite the first base of the miRNA to be an A (7mer-t1A). Requiring both conditions to be met improved the signal to noise ratio still further (8mer). Imperfect seeds increased the associated noise. The authors also found a relatively faint signal for miRNA targets existing within the ORF of genes.

At a similar time a new algorithm was released; PicTar. This attempted to account for the synergistic and combinatorial effects of multiple targets within the same 3' UTR mentioned in Doench *et al.* (Doench and Sharp, 2004) to predict genes most likely to be under miRNA control. The program identifies miRNA base 1-7 and 2-8, 7nt complementary sites in 3'

UTRs and then calculates miRNA:target 3' complementarity. It subsequently filters sites for free energy of the association and by requiring "anchors" in multiple UTR alignments and uses this to make the initial target calls. The method subsequently provides a PicTar score for multiple targets within the same UTR.

At the time of the inception of my PhD, Giraldez *et al.* derived zebrafish embryos lacking both maternal and zygotic Dicer (*dicer1*) function (Giraldez et al., 2005). Expression changes were judged by array and initially genes up-regulated upon miRNA removal were derived from these embryos and compared to transcripts whose expression was altered upon the re-addition of miR-430 by microinjection (Giraldez et al., 2006). The intersection of genes whose expression was up-regulated in the Dicer mutant when compared to the two alternative conditions were searched for an enrichment of miR-430 seed sequences in their 3' UTRs. Of the 328 genes in the intersected region with annotated 3' UTRs, there was a significant enrichment for the miR-430 seed sequence.

At this point a comparison was made between the targets predicted by different methods (Sethupathy et al., 2006). Using sets of experimentally verified targets, the authors of this paper tested each algorithm for its ability to identify targets within this set. Interestingly miRanda, TargetScanS and PicTar algorithms could only identify approximately 45-50% of the targets and roughly 2/3 of conserved targets, with miRanda making roughly 7000-8000 more predictions than PicTar or TargetScanS. Also notable was that although PicTar and TargetScanS appeared to share a large proportion of their predictions, when predictions for these three methods were overlapped, the intersection covered only 40% of conserved targets.

The programmes were clearly making substantially different predictions, suggesting that a number of rules for target prediction remained to be found.

Further use of Luciferase reporter assays demonstrated that nonconserved 7 or 8nt miRNA "seed" matches could affect the translation of reporter genes in the same way as TargetScanS predicted conserved targets (Farh et al., 2005). This suggested that a whole potentially important class of miRNAs was being missed by demanding conservation of targets as a factor for their prediction.

Subsequently, miRanda has been updated for use in miRBase (miRBase Targets) (Griffiths-Jones et al., 2008). Within miRBase Targets the miRanda algorithm is used to calculate scores for a particular miRNA's predicted target sites based on complementarity, weighted for increased significance for complementarity at the 5' end of the miRNA. These scores are then incorporated into a *P*-value calculation along with the number of additional sites for a specific miRNA in the relevant 3'UTR. Conservation is also considered within the *P*-value calculation. Targets are no longer judged according to their thermostability or the requirement to be conserved, although generally targets with a higher degree of conservation will be attributed with a more significant *P*-value.

Currently, high throughput target identification is beginning to have an impact on elucidating miRNA targeting rules, filling the void left by the slow pace of miRNA target identification. Grimson *et al.* analysed data from 11 miRNA over expression experiments in HeLa cells (Grimson et al., 2007). Following transfection the RNA of these cells was purified and arrays were used to assess mRNA degradation on a large scale. They confirmed a multiplicative

effect of multiple targets within the same 3'UTR although intervals of 8-40nt produced an even stronger repressive effect. Pairing energy was seen to be a bad indicator of the efficacy of 3' pairing. Instead 4 base contiguous pairing starting at miRNA bases 13 to 16 had the greatest effect on down regulation. Again the authors concluded that canonical sites described by Brennecke *et al.* are rare (Brennecke et al., 2005). Functional sites were generally found in regions rich in AU bases. Interestingly, although 8mer sites were seen to have no detectable effect on expression when found in 5' UTRs, they appeared to have a mild effect on expression when found in the ORF of a gene. This dampened effect seen with 8mers within an ORF extended 15nt into the 3'UTR. In addition targets near the ends of the 3'UTR were seen to have a greater effect on expression than those in the middle. These additional findings were used to calculate a "context score" which was applied to both conserved and nonconserved 7mer and 8mer seed sites in TargetScanS. This score was seen to be a good predictor of the efficacy of miRNA target sites when used in concert with luciferase target site confirmation assays described above.

TargetScanS has recently been updated again, using a new method to normalize the speed of evolution between different sets of UTRs and therefore removing the requirement to separate conserved and non-conserved sites into independent groups (Friedman et al., 2009). This new method allows a much more sensitive analysis of sites and increases the proportion of predicted target genes for a combined set of all human miRNAs to approximately 60% of human protein coding genes at a "conservation cutoff of 1.0". They note that the order of signal to background ratios for the conservation of seed sequences reflects the target site efficacy, placing target "seed" sequences in an effective order of 8mer > 7mer-m8 > 7mer-tA1 > 6mer > offset 6mer (complementary to bases 3-8 of the miRNA sequence) (Fig.1.4). At

the same cutoff as above, of the non-complementary seed sequences only 8mer seeds with a bulge appeared to be under effective selection suggesting non-seed targets are rare. Compensatory sites also appeared to be rare with an estimated 4.5 such sites conserved for each miRNA family. An estimated 4.9% of all preferentially conserved target sites were supplemented by conserved 3' pairing.

It is worth noting that the significance of the 5' complementarity between miRNAs and their targets implies that miRNA families, which share a high degree of sequence similarity, are likely to target the same mRNA target sites and consequently to be redundant in function to some extent, if co-expressed.

In addition to these most widely used algorithms many further methods exist for target prediction. The "Probability of interaction by target accessibility" algorithm (PITA) is one of the most original (Kertesz et al., 2007). Instead of using conservation to select functional targets, PITA uses a filter based on the thermodynamic stability of secondary structures at the target site to filter targets into those likely to be effective and those that are not. Furthermore, they found a tendency for miRNA seed sequences to be positioned in thermodynamically accessible regions of the UTR.

Recently, functional miRNA targets sites situated within the coding regions of several genes have been identified in the mouse (Tay et al., 2008). Multiple targets have been found in the ORF of Oct4 (*Pou5f1*), *Sox2* and *Nanog* genes. In some cases these target sites cross exon boundaries, are not conserved or have incomplete base pairing within the seed region. All of

these factors combined suggest that there are still likely to be many functional target sites not identified by prediction methods for a variety of reasons.

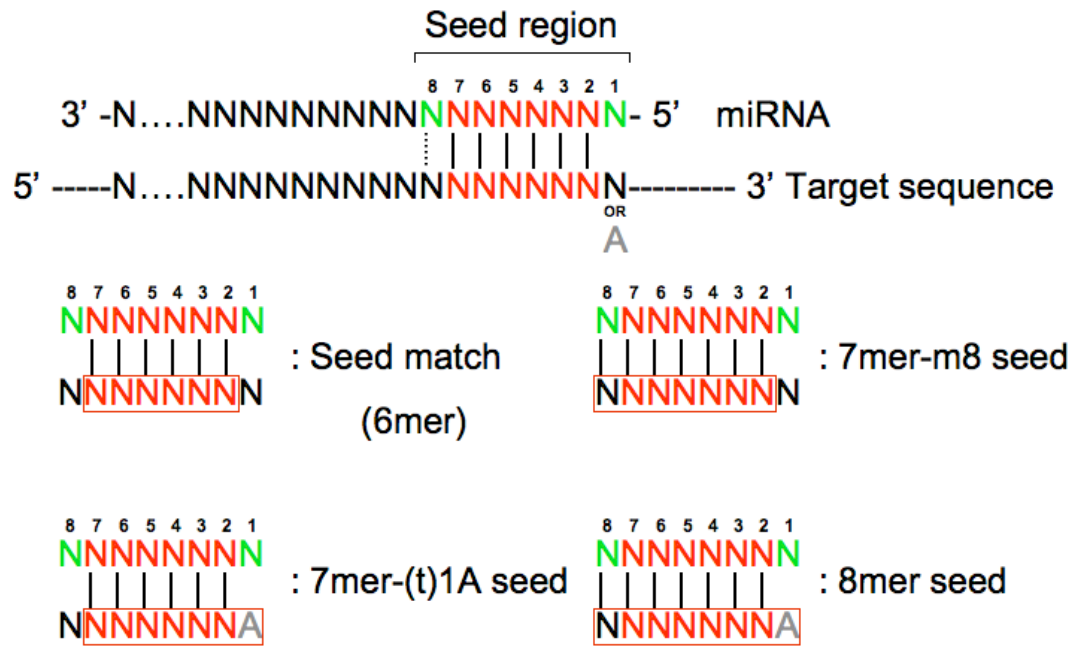

Fig.1.4: The miRNA seed region and the definitions of –mer target sites as given by http://www.targetscan.org. The red box outlines the target bases within the mRNA referred to in each case. Black lines indicate complementary bases. Red bases indicate the core 6mer. Numbered bases are counted from the 5' end of the miRNA.

1.1.4 Endogenous siRNAs

Also of interest as a closely related small RNA family is the recent discovery of endogenous siRNAs in mammals. Until very recently it had been assumed that endogenous siRNAs were restricted to those species that possess an RNA-dependent-RNA-polymerase, an enzyme not found in mammals, capable of generating long dsRNA substrates for Dicer, which would in turn generate siRNAs perfectly complementary to their targets. As a result it was assumed that RNAi in mammals was only targeted by dsRNA supplied from external sources. This theory was compounded by the presence of an interferon response in vertebrates that would be triggered by long dsRNAs present in the cells. However, the work of Babiarz *et al.*

(Babiarz et al., 2008) alongside the work of other teams (Tam et al., 2008; Watanabe et al., 2008) have now identified endogenously derived siRNAs from repeat sequences and/or pseudogenes in mouse ES cells and oocytes. These siRNAs appear to be functional. In the case of Barbiarz *et al.* they are demonstrated to be derived in a Dgcr8-independent, Dicer-dependent manner. Tam *et al.* speculate that these siRNAs will operate by a conventional RNAi mechanism, leading to the degradation of their targets, while Watanabe *et al.* demonstrate that Dicer (DICER1) and AGO2 (EIF2C2) (slicer competent AGO) both appear to be required for the endogenous siRNA driven regulation of transposons and pseudogenes, through examining the expression levels of predicted targets in knockout oocytes. It is also worth noting that both oocytes and embryonic stem cells are devoid of the dsRNA dependent interferon response.

1.2 Embryonic stem cells

Mouse ES cells are cells derived from the inner cell mass of the blastocyst (Kaufman et al., 1983; Martin, 1981). They are pluripotent, (able to differentiate into all somatic lineages and, if introduced into pre-implantation embryos, colonise all foetal lineages including the germ cells in addition to some extra-embryonic tissues) and they are capable of self-renewal (Beddington and Robertson, 1989; Bradley et al., 1984).

1.2.1 Transcriptional networks for maintaining the stem cell state

At the centre of the transcriptional network responsible for maintaining the pluripotency of stem cells reside a set of transcription factors, that include Oct4 (Nichols et al., 1998; Niwa et al., 2000), *Sox2* (Avilion et al., 2003; Kopp et al., 2008) and *Nanog* (Mitsui et al., 2003). The perturbation of any of the factors given above has been demonstrated to have a profound

effect on the stability of the undifferentiated state. Expression of these genes has subsequently been used as a marker for embryonic stem cells.

The molecular factors underlying pluripotency have been further elucidated through the generation of induced pluripotent stem (iPS) cells by the introduction of artificially expressed gene combinations into differentiated somatic cell types. Initially the factors required for this reprogramming were Oct4, *Sox2*, *Klf4* and c-Myc (*Myc*), introduced to the cells using retroviruses (Takahashi and Yamanaka, 2006). The exact role that each of these genes plays has yet to be fully understood, as have the intricacies of the process. Subsequent experiments have reduced the number of factors required to derive iPS cells. Nakagawa *et al.* were able to derive iPS cells without c-Myc, a known oncogene (Nakagawa et al., 2008) while Huangfu *et al*. derived human iPS cells using Oct4 and *Sox2* and the histone deacetylase inhibitor, volproic acid (Huangfu et al., 2008). Given the role of *Nanog* in the maintenance of pluripotency in embryonic stem cells and its ability to improve the transfer of pluripotency to cells in fusion experiments (Silva et al., 2006), it is perhaps surprising that *Nanog* was not one of the required factors. As techniques and methods have been improved, iPS cells have been generated that are broadly comparable to ES cells and that are capable of producing germline competent chimeras to contribute to further generations (Okita et al., 2007).

In addition to the endogenous transcriptional network required to maintain ES cells in an undifferentiated state, exogenous signals are also necessary (reviewed in (Okita and Yamanaka, 2006)). Traditionally mouse ES cells have been maintained in culture in the presence of leukaemia inhibitory factor (LIF) and serum or bone morphogenetic protein (BMP) (Ying et al., 2003). The BMP (or serum) induction of inhibitor-of-differentiation

genes and LIF induction of STAT3 signaling pathways ensures that the cells are unable to differentiate. Recently it has been demonstrated, however, that these two pathways are not necessarily required to maintain stem cell identity, but may instead dampen the effects of exogenous, pro-differentiation stimuli (Ying et al., 2008). These culture conditions can be substituted by the disruption of the mitogen-activiated protein kinase (MAPK) pathway and the glycogen synthase kinase 3 (GSK3) pathway.

Attempts are being made to integrate both the transcriptional networks for self-renewal and pluripotency and the effectors of important cell signaling pathways into a unified system through ChIP and affinity purification experiments. Wang *et al.* used biotinylated proteins to purify the interacting partners from the pluripotency network, identifying interactions between a large number of proteins with known involvement in differentiation or the integrity of the inner cell mass (Wang et al., 2006). They later followed this work with a ChIP-Chip survey of the promoter occupancy of the factors identified by Takahashi *et al.* (Takahashi and Yamanaka, 2006) and a selection of previously identified interacting partners; 9 proteins in total (*Nanog*, Oct4, *Sox2*, *Klf4*, *Myc*, Dax1 (*Nr0b1*), Nac1 (*Nacc1*), *Zfp281*, Rex1 (*Zfp42*)) (Kim et al., 2008b). This work demonstrated the apparent complexity of the coordinated regulation of the network's targets. Approximately 800 gene promoters were bound by 4 or more of the proteins tested while approximately 50% were bound by only a single factor. MYC and Rex1 appeared to bind a different set of targets from the other transcription factors (TFs) tested. 96% of MYC promoters had a H3K4me3 signature that implies the chromatin region is open and active. This complements data that suggested that MYC targets have a generally greater expression in ES cells than genes with promoters bound by the other transcription factors. The targets of the other genes included proteins both expressed and

depleted in ES cells. Interestingly, the gene sets with a greater number of this subset of transcription factors associated with their promoters seem to have greater expression in ES cells.

Additional, independent, ChIP studies have been conducted repeating a number of these experiments using slightly different protocols. Chen *et al.* performed a series of ChIP-sequencing (ChIP-Seq) assessments for a similar set of genes (*Nanog*, Oct4, *Sox2*, *Klf4*, c-Myc (*Myc*), n-Myc (*Mycn*), *Esrrb*, *E2f1*, *Zfx*, *Smad1*, *Stat3*, *Tcfcp2l1*, *Ctcf*) (Chen et al., 2008). The authors identified sites bound by combinations of these TFs and once again they found groups of factors whose binding correlated into clusters, with c-Myc tending to bind with different factors to NANOG, SOX2 and Oct4. They noticed p300 (EP300) histone acetylase, known to associate with enhancer elements, tended to bind sites with 3-6 other TFs from the NANOG, SOX2, Oct4 group. They proceeded to find that 60% of genes that are up regulated in ES cells, when compared to differentiated cells tend to be associated with target sites enriched for NANOG, Oct4, SOX2, SMAD1, STAT3, c-Myc and n-Myc binding.

1.2.2 ES cell cycle

Mouse embryonic stem cells have a drastically shortened G1 phase of their cell cycle, compared to most somatic cells, which allow them to proliferate rapidly. This is the result of constitutively active cyclinE:CDK2, low cyclinD and CDK4 activity and permanently hyperphosphorylated Rb protein. Within somatic cells mitogen signaling induces the activity of CyclinD:CDK in early G1 phase. The increase in cyclinD:CDK activity ultimately leads to the hyperphosphorylation of Rb. Rb hyperphosphorylation allows CyclinE:CDK2 to become active and hence the cell cycle can proceed into S phase. Mouse ES cells essentially remove

the requirement for G1 mitogen control. Mitogens can also trigger the differentiation of ES cells, so this omission allows the cell cycle to continue without this risk of mitogen induced differentiation (reviewed in (Orford and Scadden, 2008)).

1.2.3 miRNAs and mouse ES Cells

1.2.3.1 The role of miRNAs in stem cells – Perturbing the processing pathway

In 2003 an attempt was made to breed a mouse with a homozygous, null Dicer gene, in order to further explore the role of miRNAs in development (Bernstein et al., 2003). However, homozygous mutants displayed an embryonic lethal phenotype. Development appears severely disrupted prior to embryonic day 7.5 (E7.5) with loss of embryonic Oct4 staining implying the loss of the stem cell population in early development. Subsequently the role of miRNAs in stem cells has become a focus for a number of laboratories.

Early experiments with hybrid DT40 chicken cells containing human chromosome 21 seemed to support the notion that Dicer (DICER1) dependent siRNAs derived from peri-centromeric transcripts played a role in centromeric structure, heterochromatin formation and cell division (Fukagawa et al., 2004). This work is complemented by later experiments in Dicer deficient mouse embryonic stem cells.

As expected, the removal of Dicer from the mouse ES cells seemed to lead to a loss of mature miRNAs and an accumulation of pre-miRNA transcripts (Kanellopoulou et al., 2005). However, it also seemed to lead to an increase in dsRNA species derived from centromeric

satellite repeats, which the authors suggest may be processed into short RNA species in a Dicer dependent manner. This correlates with changes in methylation and histone modification profiles. These findings remain controversial, however, in particular the concept of Dicer playing a major and direct role in chromatin modification via an RNA mediated process. Indeed, these findings were not replicated by a later study that found no changes in centromeric satellite associated DNA methylation or histone modification in Dicer deficient mouse ES cells (Murchison et al., 2005). However, the potential for a Dicer mediated role in the maintenance of heterochromatin, makes it more difficult to interpret the cellular functions of miRNAs in ES cells from the results of these knockout studies in isolation.

More recently, Wang *et al.* generated a *Dgcr8* conditional knockout mouse embryonic stem cell line (Wang et al., 2007). As DGCR8 plays no part in the generation of siRNAs from dsRNA substrates, DGCR8 was considered to be a more likely candidate for the generation of a miRNA specific phenotype. The *Dgcr8* null genotype also proved to be embryonic lethal and the *Dgcr8* mutants replicated a number of other phenotypes seen in Dicer mutant stem cells. All three mutant stem cell lines had a reduced rate of proliferation with both Murchison *et al.* and Wang *et al.* detecting an increase the number of cells in the G1 phase of the cell cycle in the absence of miRNAs (Murchison et al., 2005; Wang et al., 2007).

Furthermore, the loss of miRNAs had a profound effect on the differentiation potential of the stem cells. Dicer deficient cells were unable to form teratomas when injected subcutaneously into immuno-deficient mice (Kanellopoulou et al., 2005). Embryoid bodies formed from these cells grew for 8-10 days and then arrested and none of the markers of differentiation tested were expressed in these Dicer knockout embryoid bodies. In contrast, *Dgcr8* knockout

cells appeared to have a less profoundly compromised differentiation potential (Wang et al., 2007). They successfully formed teratomas following subcutaneous injection, although they appeared largely undifferentiated in structure. They also expressed a number of the markers of various cell lineages upon embryoid body induced differentiation and embryoid body growth did not arrest. The authors do note however that the mutant stem cells appear unable to silence pluripotency markers in the course of differentiation, with a larger proportion of mutant cells reverting to ES cell like growth following a period of induced differentiation than seen with control cells.

The differences evident between the phenotypes of these various knock out cell lines may in part be explained by molecular functions specific to either DGCR8 or Dicer. Pyrosequencing of the small RNA fraction of Dicer knockout cell lines and wild-type ES cells failed to identify the population of Dicer dependent centromeric heterochromatin associated siRNAs hypothesized by Kanellopoulou *et al* (Calabrese et al., 2007). However, as discussed in the sections "Mirtrons and other exceptions to the canonical rules" (see section 1.1.1.6) and "Endogenous siRNAs" (see section 1.1.4), a number of Dicer specific small RNAs have been identified by Illumina and 454 sequencing that may account for some of these phenotypic discrepancies (Babiarz et al., 2008).

1.2.3.2 miRNA expression in stem cells

Initially, ES cell miRNA expression profiles were constructed by cloning small RNAs and subsequently sequencing them. This process identified large numbers of previously unannotated miRNAs and revealed a number of miRNAs that seemed to be present specifically within mouse ES cells (Houbaviy et al., 2003). Perhaps the most notable, novel

miRNAs were found within a cluster on Chromosome 7 (the miR-290 cluster). A number of the miRNAs that have been subsequently ascribed to this cluster share a common seed sequence (7mer-1A; GCACTTA; mmu-miR-291a-3p, -291b-3p, -292-3p, -294, -295). By sharing the same seed it is expected that there can be a considerable overlap between the targets of these miRNAs. In 2004 a similar study was conducted in human ES cells (Suh et al., 2004). This study identified an apparently ES cell specific cluster (has-miR-371, -372, -373), that is orthologous to the mmu-miR-290 cluster. The human ES cells also expressed an orthologous miR-302 cluster that had been sequenced in the mouse ES cells. miR-302 shares a seed sequence with the mmu-miR-290 cluster.

As the number of annotated miRNA sequences was expanded, microarrays were designed to profile cells for the expression of known miRNAs (Laurent et al., 2008), as well as quantitative RT-PCR (qRT-PCR) assays that could profile miRNAs from a single ES cell (Tang et al., 2006). The latter could prove an important strategy for the dissection of self-renewal as ES cells form notoriously heterogeneous culture populations, containing small populations of spontaneously differentiated cells and cells expressing maker proteins at different levels (reviewed in (Silva and Smith, 2008)).

Ultimately, high throughput sequencing techniques have been employed to ascertain miRNA expression with a degree of sensitivity and specificity, unobtainable with the aforementioned techniques. The depth at which these new technologies allow the miRNA expression pool to be sampled allows miRNAs expressed at a low level to be detected and annotated. Pyrosequencing of the miRNA population from mouse ES cells attributes the majority (70-76%) of miRNA expression to 6 loci in the genome, some of which are home to clusters of

miRNAs (Calabrese et al., 2007). These include the mmu-miR-290 cluster (the most highly expressed cluster) and a cluster containing mmu-miR-467a and its paralogues. Once again mmu-miR-467a shares the same 7mer-1A seed as mmu-miR-291a-3p.

By contrast, human ES cells seem to express has-miR-302a and its paralogues at a far greater level than the has-miR-371 cluster, as measured by Solexa sequencing (Morin et al., 2008). Although this study identified many isomers of the canonical (miRBase annotated) mature miRNA sequences, with a variety of 5' and 3' extensions that may alter the seed sequence, this predominance of miR-302 over miR-371 to miR-373 is replicated by further studies, measured by pyrosequencing (Bar et al., 2008). It is intriguing that the predominant miRNAs expressed in human and mouse ES cells are different, while maintaining a common miRNA seed sequence, perhaps underlying a degree of redundancy in function between the two miRNA families. It is also worth bearing in mind that there are known phenotypic differences between mouse ES cells and human ES cells and any comparisons made between the two systems should be made with due caution (Discussed in (Tesar et al., 2007)).

A ChIP-Seq study has investigated the transcriptional control of miRNAs in mouse ES cells and the association of Oct4, SOX2, NANOG and TCF3 (a further TF) at miRNA promoters and correlated this with miRNA expression (Marson et al., 2008). It seems that as with protein coding genes, these transcription factors control miRNAs that are both activated and repressed in ES cells. In this way a putative and simple series of networks have been constructed, demonstrating the roles of miRNA in both coherent and incoherent feed-forward control of ES cell protein expression, fine tuning protein expression and poising the cells for differentiation.

1.2.3.3 The role of miRNAs in stem cells

Even if the recent prediction that 60% of human genes are likely to be targeted by miRNAs is considered to be an over-estimate (Friedman et al., 2009), when considering the ever-increasing number of annotated miRNAs the number of functional targets annotated for miRNAs remains tiny.

In mouse ES cells, miRNAs modulate the activity of DNA methyltransferases. Independent studies reported that miRNAs from the miR-290 cluster (mmu-miR-291-3p, -292-3p, -294 and -295) post-transcriptionally regulate the expression of *Rbl2* which in turn down-regulates *Dnmt3a* and *Dnmt3b* expression (Benetti et al., 2008; Sinkkonen et al., 2008). Dicer deficient mouse ES cells have decreased levels of these methyltransferases (in addition to *Dnmt1*) and exhibit global hypomethylation, substantial telomere length changes and increases in telomeric recombination (Benetti et al., 2008). The introduction of the miRNAs listed above into these cells led to decreases in *Rbl2* levels and increases in *Dnmt3a* and *Dnmt3b*. This regulation of *Dnmt3a* and *Dnmt3b* expression, via *Rbl2*, by the miR-290 cluster also has an important role in the silencing of the Oct4 promoter upon ES cell differentiation (Sinkkonen et al., 2008). Dicer deficient mouse ES cells are unable to efficiently silence the Oct4 promoter by methylation, upon differentiation. Interestingly, the transfection of the miR-290 cluster upon differentiation rescued this phenotype. In order to identify targets in this study, the authors transfected the Dicer knock out cells, deficient in mature miRNAs, with miRNA mimics and then assessed the resultant changes in mRNA levels by array. This system allowed the authors to identify miRNA targets in a system with no interference from functionally redundant miRNA families.

As implied by the results of the miRNA processing pathway disruption experiments described above, miRNAs also contribute to the regulation of the cell cycle. Wan *et al.* performed a screen of 266 mouse miRNAs; transfecting them into *Dgcr8* knock out mouse ES cells and examining the cell cycle for changes (Wang et al., 2008). Once again they identified mmu-miR-291-3p, -292-3p, -294, -295 and miRNAs sharing their seed sequence (including miR-302a and homologues) as increasing the proliferative rate of the mutant ES cells. Further investigation of the mechanism by which these miRNAs achieved this revealed p21 (*Cdkn1a*), a cyclinE-CDK2 inhibitor as a target of mmu-miR-291-3p, -292-3p, -294, -295 and miR-302d.

A further intriguing result published by Lin *et al.* concerned the transfection of cancerous cell lines with a retrovirus expressing the miR-302 cluster (Lin et al., 2008a). This was sufficient to convert these cell lines into an apparently pluripotent state, combined with the expression of various human ES cell markers, including Oct4. This implies that miRNAs may have a significant role to play at the very centre of the regulatory network which controls pluripotency.

1.2.3.4 The role of miRNAs in stem cells – Lessons from cancer

A large number of the miRNAs that are highly expressed in mouse and human ES cells have also been demonstrated to play a role in the pathogenesis of various forms of cancer. Intensive studies conducted in cancer and the identification of multiple targets of these miRNAs provides a resource when considering miRNA roles in stem cells.

Hsa-miR-373 (an orthologue of the mmu-miR-290 cluster of miRNAs) has been demonstrated to contribute to the migratory potential of cancer cell lines in part through the target gene *CD44* (Huang et al., 2008). In addition, a screen for miRNAs that can confer a degree of resistance to oncogenic-induced senescence identified both has-miR-372 and has-miR-373 as candidates (Voorhoeve et al., 2006). These miRNAs target *LATS2*, which is known to influence G1 to S phase transition in the cell cycle.

The miR-17-92 cluster was renamed oncomiR-1 after it was found to possess oncogenic potential (He et al., 2005). The cluster has been demonstrated to play a vital role in development, with its deletion in mice causing death at birth (Ventura et al., 2008). This cluster belongs to a set of three highly conserved clusters present in the mouse genome (including mmu-miR-106b-25 and mmu-miR-106a-363). Deletion of either of the other two clusters had no visible phenotype. However, when both mmu-miR-17-92 and mmu-miR-106b-25 were deleted, the embryos died before E15. This more severe phenotype suggests that the mmu-miR-17-92 cluster and the mmu-miR-106b-25 are to some extent redundant in their functions. This functional relationship is discussed in a review by Petrocca *et al.* (Petrocca et al., 2008b). These two clusters cooperate in the regulation of the TGFβ signaling pathway. MiR-106b, -93, -17 and -20a down regulate p21, while miR-25 and miR-92a-1 target Bim (a pro-apoptotic gene (*Bcl2l11*)). Both clusters are under *Myc* and *E2f1* regulation and both clusters (miR-106b, -93, -17 and -20a) create a feedback loop to down-regulate *E2f1*.

A further ES cell expressed miRNA, miR-21, down regulates a number of tumour suppressor genes and is over expressed in a wide variety of cancers. It has recently been found to target

the tumour suppressor gene *PDCD4* (Lu et al., 2008), in addition to previously described targets *PTEN* and *TPM1*. Contrary to this observation, identifying a miRNA acting as an oncogene, Liu *et al.* identified CCND1, *CCND3*, *CCNE1* and *CDK6* as miR-16 targets (Liu et al., 2008a). The authors used a HepG2 cell line depleted for Drosha (*RNASEN*) to investigate the extent of miRNA post-transcriptional control and to search expression profiles for genes mis-regulated in cancers that could be miRNA targets. As would be expected miR-16 over expression caused A549 cells (human lung carcinoma) to accumulate in G1 phase.

1.3 Aims of this project

While there have been constant and rapid advances in the field of miRNA target identification, there remains a large discrepancy between the number of known miRNA targets and the number of annotated miRNAs in miRBase. As a consequence there is a need for simple and effective methods by which to generate large numbers of experimentally supported miRNA-target interactions. Such data could then be used to both optimize miRNA target prediction algorithms and to directly annotate miRNAs with associated functional information, upon which to build hypotheses for further experimentation.

One novel approach developed as this project was being conceived was adopted by Giraldez *et al.* in zebrafish maternal-zygotic Dicer (*MZDicer)* mutant embryos (Giraldez et al., 2006). In order to eliminate the maternally contributed *dicer* activity from the early zebrafish embryo, Giraldez *et al.* conducted a germ line replacement experiment, whereby they depleted the germ cells from a wild type fish and reconstitute the germ line with *dicer1*$^{-/-}$ cells. Hence the offspring of a cross between these fish would no longer exhibit any endogenous Dicer activity. These *MZDicer* embryos allowed for the identification of a large

number of putative miRNA targets through the ectopic expression of a miRNA in conditions where endogenous targets are expressed. Target-miRNA associations were judged by monitoring the down-regulation of mRNA by expression array, following the injection of miRNAs into the mutant embryos. By drawing an intersection between those genes down-regulated when miRNAs are added and mRNAs up-regulated upon the removal of Dicer, a list of gene transcripts predicted to be enriched for specific miRNA targets was derived. The prediction of these putative targets would not be restricted by repression of other miRNAs working in concert with the miRNA of interest or by miRNAs sharing seeds with the introduced miRNA, which may mask its targets through functional redundancy and potential saturation of target sites. In this way the predicted target set would be expected to be "cleaner" than sets derived from over expressed miRNAs in unrelated systems (Lim et al., 2005), or target sets derived from studies in which a single miRNA has been disrupted, regardless of the aforementioned redundancy.

We decided to attempt a similar approach in a cell-based system in mammals, providing an easily manipulated foundation for the generation of miRNA target lists in a mammalian context. The aims were as follows:

1) To develop cell lines with a depletion of endogenous miRNA expression through the disruption of the *Dgcr8* locus using ES cell gene traps and homologous recombination.
2) To assess the expression of miRNAs in these mutant cells and control cell lines and so identify miRNAs expressed endogenously in a wild type and heterozygous background and to search for DGCR8 independent miRNA expression.

3) To optimise transfection conditions for the reintroduction of miRNA mimics into these cells.

4) To reintroduce selected miRNAs expressed in control cell lines into the miRNA depleted cells and monitor mRNA expression via microarrays.

5) To determine gene lists enriched for miRNA targets by comparing genes up-regulated upon removal of miRNAs from the system to genes down-regulated upon miRNA reintroduction.