Chapter 6: Discussion

The aim of this project was to develop a cell based system for the identification of miRNA targets on a large scale in a cellular system. The system was intended to resemble that of Giraldez *et al.*, which was used to investigate the role of miRNAs in zebrafish (Giraldez et al., 2006). The principle of the method used is that miRNAs introduced into a biological system will elicit expression changes in target mRNAs detectable by microarray. However, use of an *MZDicer* mutant fish allowed the derivation of candidate miRNA target lists in a cellular background that was depleted of endogenous miRNA expression. As a consequence the identification of candidate target mRNAs, through the reintroduction of duplexes by injection, would not be affected by target site saturation by endogenous miRNAs or complicated by the combinatorial relationships inherent in the miRNA mediated regulatory network. In addition, unlike the knockdown of individual miRNAs the system is not affected by the functional redundancy between miRNA family members that share the same seed regions, which are expected to target the same mRNAs.

6.1 Disruption of the *Dgcr8* locus to deplete mature miRNAs in mouse ES cells

To this end, using mouse ES cells with a gene trap inserted within a single allele at the *Dgcr8* locus, which I obtained from Bay Genomics (Nord et al., 2006; Stryke et al., 2003), I have disrupted the second allele with a targeted trap inserted by homologous recombination. Consequently I used the expression of marker genes on each of the traps, Northern blots and RT-PCR to confirm the configuration of these traps within the locus and in the case of the Northern Blot I demonstrated a dramatic

reduction of the expression of wild type *Dgcr8* in the *Dgcr8*$^{gt/tm1}$ cells. I also identified a substantial reduction in ES cell miRNA expression (Houbaviy et al., 2003) through the use of Northern blots, which confirmed the functional significance of the mutations at the *Dgcr8* locus. Western blots and immuno-staining confirmed that the *Dgcr8*$^{gt/tm1}$ cells were still expressing ES markers; transcription factors at the centre of the regulatory network which modulates pluripotency and self renewal.

The phenotypic effects of the disruption of the miRNA processing pathway in ES cells broadly resembles the phenotypes seen by others both published before and since this project was begun (Fukagawa et al., 2004; Murchison et al., 2005; Wang et al., 2007). Notably that *Dgcr8*$^{gt/tm1}$ cells appear to accumulate in the G1 phase of the cell cycle and are unable to form wild type EBs, although like the *Dgcr8* knock out cell line, they do appear to differentiate with a degree of success, forming cells with various, conspicuous morphologies. How this change in morphology is reflected with respect to the expression of markers of differentiation and the silencing of ES cell markers in the *Dgcr8*$^{gt/tm1}$ cells is a subject for future research.

6.2 The small RNA profile of *Dgcr8*$^{gt/tm1}$ and *Dgcr8*$^{tm1,gt/+}$ cells

I used the highly sensitive Illumina/Solexa RNA sequencing platform to sequence small RNA libraries to confirm the depletion of miRNA expression in the *Dgcr8*$^{gt/tm1}$ cells. This was an important step, since ES cells are pluripotent and capable of differentiation. As a consequence the disruption of *Dgcr8* may have caused the cells to change or adapt, altering their expression profile. In this case, an incomplete depletion of DGCR8 may have been sufficient to trigger cellular alterations without significantly depleting miRNA processing. As a result a limited miRNA profile by

Northern blot may have simply missed miRNAs still expressed or newly expressed in these cells. The Solexa/Illumina sequencing of the miRNA profiles of the *Dgcr8*$^{gt/tm1}$ and *Dgcr8*$^{tm1,gt/+}$ cells and the comparison of these profiles confirmed that this was not the case and demonstrated a significant depletion of the expression of the vast majority of miRNAs in these cells. In addition, by comparing the raw maximum mapped sequence depths across a range of ncRNA species annotated in Ensembl it was clear that these species seemed to be proportionally represented in both the *Dgcr8*$^{gt/tm1}$ and *Dgcr8*$^{tm1,gt/+}$ cells. This adds weight to the observation that DGCR8 does not appear to play a significant role in the processing of other non-coding RNA species, including rRNAs, for which Drosha has been implicated as a processing enzyme (Wu et al., 2000).

6.2.1 DGCR8 independent miRNA processing

Finally, the Solexa/Illumina sequencing of the small RNA populations allowed me to identify a number of miRNAs that appear to be processed, at least in part, in a DGCR8 independent fashion; let-7a-1, let-7a-2, miR-98, miR-720, miR-689-2, miR-712, miR-1196, miR-712*, miR-1186, miR-805, miR-344-1 or 2, miR-320, miR-668, miR-877 and miR-484. These include a number of DGCR8 independent miRNAs proposed by Babiarz et al. (miR-320, miR-344, miR-668, miR-877 and miR-484) (Babiarz et al., 2008). Of these *miR-877* has been suggested as a mammalian mirtron through bioinformatics analysis (Berezikov et al., 2007). To this end both the results presented in this thesis and the work of Babiarz *et al.* confirm that this miRNA does not require DGCR8 to be processed. As our understanding of miRNAs and their underlying functional mechanisms progresses it is increasingly clear that a number of the canonical rules previously believed to apply to all miRNAs are prone to

exceptions. The microprocessor dependent release of pre-miRNAs from their pri-miRNAs is clearly one such rule and is deserving of further research.

6.3 Generating miRNA candidate target lists

In order to identify miRNA candidate target lists I selected probes by intersecting those up regulated upon DGCR8 depletion with those down regulated upon the re-addition of individual miRNAs to *Dgcr8*$^{gt1/tm1}$ cells. The intersection with the up regulated probe lists would constrain the candidate targets to those relevant to the ES cell system and remove those demonstrating a significant change due to the potential over-expression of the added miRNA. Conversely, by restricting the list to those down regulated upon the re-addition of specific miRNAs, the list would be limited to a large extent to specific miRNA primary targets. This would remove any genes that contain predicted but non-functional target sites whose expression is altered upon miRNA depletion due to secondary effects as a consequence of the expression and phenotype of the cells adapting to their new miRNA transcriptome.

6.3.1 The influence of miRNAs on the ES cell transcriptome

The comparison of the expression profiles of the *Dgcr8*$^{gt/tm1}$ and *Dgcr8*$^{tm1,gt/+}$ clearly revealed an enrichment of miRNA seed sequences within the 3'UTRs of genes up regulated upon the depletion of functional *Dgcr8*. This enrichment includes the seed sequences of miRNAs from amongst those most highly expressed in the *Dgcr8*$^{tm1,gt/+}$ cells but depleted in the *Dgcr8*$^{gt/tm1}$ cells, as judged by Sylamer analysis (van Dongen et al., 2008). Although this enrichment of miRNA seed sequences amongst the genes up regulated upon miRNA depletion implies that the miRNAs in these sets regulate a broad spectrum of mRNAs in wild type ES cells, it does not exclude other miRNAs

from playing equally important roles in ES cell transcriptome regulation. Indeed it seems likely that other miRNAs can regulate smaller sets of highly influential genes that may be more difficult to detect through miRNA seed sequence enrichment analysis.

With respect to the large gene set significantly up regulated following the depletion of *Dgcr8* from the ES cell system (3251 probes were up regulated, *P*-value < 0.05, LFC > $\log_2(1.1)$), it would appear that miRNAs, between them, do indeed play a broad regulatory role in ES cells, which is to be expected given the number of miRNA targets predicted by computational algorithms (Friedman et al., 2009; Griffiths-Jones et al., 2008). What is perhaps more surprising therefore is the maintained expression of ES cell specific markers in these mutant cells rather than the catastrophic disruption of the transcriptional network required to maintain ES cell identity. It does seem however that miRNAs do influence the targets of this core transcriptional network (including Oct4, *Sox2*, *Nanog*, *Klf4* and c-Myc), as determined by an examination of the differential expression of their targets. What is less clear is whether this altered expression results from miRNAs directly interacting with the transcriptional targets of the TF, with the mRNAs of the TFs themselves or with factors upstream of the TFs that may alter their expression. Given the profligate nature of miRNA:target relationships, it is reasonable to predict that all three may be of an influence. Although there was no apparent enrichment of miRNA seed sequences amongst the TF target gene mRNA 3'UTRs, this does not rule out miRNA target enrichments on a scale below the sensitivity of this method. In addition there were apparent changes in the expression of several of the TFs themselves as assessed by Illumina array. These changes would require confirmation by a more sensitive and

specific method, but potential miRNA regulation of these TFs clearly does warrant further research to aid the understanding of the regulation of ES cell pluripotency. Indeed such relationships have previously be implied in non-differentiated ES cells although, to my knowledge, they are yet to be confirmed (Singh et al., 2008).

6.3.2 The re-addition of miRNAs to the DGCR8 deficient ES cells and miRNA candidate target lists

Finally I optimized the protocol for the re-addition of miRNA mimics to the *Dgcr8*$^{gt1/tm1}$ cells and reintroduced miR-25 and miR-291a-3p miRNAs into the cells by transfection. Subsequently I used Illumina mRNA expression microarrays to judge which genes are down regulated in the presence of these miRNAs with reference to cells transfected with a control duplex. As described above I then used these expression changes to determine candidate target mRNAs by intersecting the probe list with those probes up regulated upon DGCR8 depletion. This intersection followed by the removal of probes identifying the same gene and probes not annotated with an Entrez ID left 40 miR-25 candidate target genes and 25 miR-291a-3p candidates. The miR-291a-3p target candidates and a target candidate list for the miR-290 cluster generated by Sinkkonen *et al.* had an overlap of 9 genes (9/40 of my candidates and 9/253 of the Sinkkonen list) (Sinkkonen et al., 2008). The limited overlap may be a consequence of fundamental differences between the experiments, the most critical of which includes their transfection of a cluster of miRNAs, compared to my transfection of a single miRNA. Although the miRNAs of Sinkkonen *et al.* do predominantly share similar seed sequences (excluding miR-290-5p), the differences in seeds combined with combinatorial regulation could result large differences in the mRNA target identified by their study. Despite this both studies identified *Cdkn1a* as

a target of these miRNAs; a target that has been confirmed in independent study (Wang et al., 2008).

All but 3 of the candidate genes identified in my study, for which I derived annotated Ensembl or Vega transcripts with an associated 3'UTR, contained at least a single 6mer seed sequence corresponding to the relevant transfected miRNA that may identify the miRNA target site responsible for miRNA induced transcript degradation (Lim et al., 2005). Of the remaining three genes that do not contain a seed in the 3'UTR, two of the selected transcripts contained relevant 6mer seed in their ORF, which again may be prognostic of target sites (Grimson et al., 2007; Tay et al., 2008). The final gene's transcript contained no obvious seed sequences in the ORF or the 3'UTR. This gene may have alternative non-annotated transcripts with differing exons and 3'UTRs not searched here. Another possibility is that this gene is not a true miRNA target, but a secondary target of upstream genes affected by miRNA regulation.

These lists should prove invaluable for guiding research as they are experimentally derived and relevant in an ES cell context. As such they contain a number of interesting target genes worthy of further research for the roles that they may play in stem cell physiology. While the goal of this research was to generate large, inclusive target list for miRNAs, it is important to bear in mind that some of the genes in these target candidate lists may not be true targets and may have been selected as a result of experimental noise. If definitive relationships and pathways are to be concluded it would be necessary to conduct additional experiments.

In order to generate the optimal gene lists for further study it may be worth considering those genes identified as targets by multiple, independent methods. One approach would be to follow up those genes identified both here and by Sinkkonen *et al* (Sinkkonen et al., 2008). An alternative would be to use the experimentally identified gene lists presented in this thesis to refine the candidate genes that are predicted by target prediction programmes such as TargetScan (Friedman et al., 2009). 34/65 genes I identified as candidate miRNA targets were also predicted by TargetScan version 4.2. A final alternative would be to follow up those genes demonstrating the most significant expression changes upon miRNA removal and re-addition in preference to other targets. These shorter lists of genes would be a useful starting point for future work.

6.4 Future work

6.4.1 Expanding the system for the generation of candidate target lists

I have developed a system that is easily scalable in order to generate candidate target lists for a large number of miRNAs in a relatively short period of time. As such it would be interesting to generate target lists for a broader set of miRNAs. I would envisage these including miR-17, miR-20a, miR-93, miR-106a or miR-106b as the seed sequence of these miRNAs was also clearly enriched amongst the 3'UTRs of genes up regulated in *Dgcr8*$^{gt/tm1}$ cells when compared to *Dgcr8*$^{tm1,gt/+}$ cells. Whether more than one of these miRNAs should be included in a further set of experiments would be interesting to consider. As they share seed sequences, they are likely to share the majority of their targets (Friedman et al., 2009). Consequently it is difficult to know whether the methods presented here would be sufficiently sensitive to discern

whether members of the resulting candidate targets are likely to be specific to single miRNA. It is possible that through the expansion of the study in this way, features that discern targets beyond the seed sequence could be identified in order to discriminate between the targets of miRNAs with shared seeds. Other miRNAs worthy of transfection would include those most highly expressed in the *Dgcr8*$^{tm1,gt/+}$ cells. This large set of miRNAs could also be expected to regulate genes important for maintaining the undifferentiated state of ES cells. Finally it would be interesting to co-transfect a number of these miRNAs with differing seed sequences. This would allow an investigation of the extent and significance of coordinate regulation of miRNA targets by multiple miRNAs.

Other refinements could be adopted to improve the candidate gene lists. Further transfection replicates of miR-25 and miR-291a-3p would be expected to increase the statistical significance of any replicated expression changes, potentially improving the sensitivity of the method. Furthermore, repeating the transfections in the independently derived *Dgcr8*$^{gt2/tm1}$ cell line would again add a further refinement to the prediction lists. Ultimately an approximation of the number of false positives present in each of these lists could be gleaned through reporter gene assays the principle of which has been demonstrated in many published studies (Giraldez et al., 2006; Lewis et al., 2003). In brief, 3'UTRs containing predicted target sites would be inserted into a plasmid downstream of a reporter gene ORF. This plasmid is then co-transfected into HeLa cells with a miRNA mimic. The reporter gene expression is then compared to that of a similar experiment but in which the predicted target site contains point mutations within the target seed region. A significant increase in

reporter gene expression in cells transfected with the mutated plasmid can be used as an indication of the presence of a true miRNA target site.

An adaptation of the pSILAC method used to assess expression changes at the level of the proteome would allow me to assess the contribution of translational inhibition to target regulation for each transfected miRNA in a system with all of the advantages associated with the depletion of endogenous miRNAs (Baek et al., 2008; Selbach et al., 2008). The use of this method in the place of, or in addition to the Illumina microarray platform could prove an interesting addition to this study and it would be fascinating to determine whether the targets of particular miRNAs are more prone to translational inhibition than the targets of others and whether miRNA sequence or other factors may influence this.

6.4.2 Novel miRNAs and siRNAs expressed in mouse ES cells

The Solexa/Illumina sequenced small RNA libraries generated in this study would also allow me to further investigate the small RNA populations of both wild type and DGCR8 depleted mouse ES cells in ways that are beyond the scope of this thesis. Deep sequencing of small RNA populations have recently allowed the identification of both novel miRNAs in mouse ES cells (Calabrese et al., 2007) and endogenously expressed siRNAs in mouse ES cells (Babiarz et al., 2008) and in mouse oocytes (Tam et al., 2008; Watanabe et al., 2008). In order to investigate the expression of these novel short RNAs more completely, the sequence libraries derived from this study should be mapped against the complete mouse genome. Regions within the genome in which short sequence reads are found to accumulate but which are not annotated as a known miRNA locus could be investigated further in order to

determine a mechanism by which the may originate. The local secondary structure of a transcript from the region could be used to identify whether the reads may originate from a canonical miRNA hairpin. In addition genomic features such as inverted repeat elements, diverging or converging gene structures or pseudogenes mapped at alternative loci all potentially suggest the generation of dsRNA molecules that may form Dicer substrates for the production of siRNAs that would map to this location.

6.4.3 The roles of DGCR8 independent miRNAs

Additional experiments could be conducted to ascribe roles to the DGCR8 independent miRNAs. Transfection of the *Dgcr8*$^{gt1/tm1}$ cells with Locked Nucleic Acid (LNA) antagomirs, that bind specific miRNAs and block their function, could be used to block selected miRNAs (Chan et al., 2005). As this experiment would be performed in a cellular background depleted of DGCR8 it would benefit form the same associated advantages concerning a reduction of functional redundancy. Once again microarrays could be used to identify transcripts that are up regulated following this functional block to compile candidate target lists for these miRNAs.

6.4.4 Prospects for improved pri-miRNA annotation

Currently intergenic miRNA primary transcripts are relatively poorly annotated. Although experiments have been conducted to specifically annotate a small set of these miRNAs (for example (Cai et al., 2004; Houbaviy et al., 2005; Lee et al., 2004)) and a recent effort has been made to compile relevant annotation data to annotate the structure of a much broader set of these pri-miRNAs (Saini et al., 2008), there is still requirement for further experimental evidence to support these annotations. One of the side effects of a depletion of functional DGCR8 is expected to be an accumulation

of unprocessed pri-miRNA transcripts (Wang et al., 2007). Previously Illumina/Solexa sequencing of fragmented poly-A selected transcripts from mouse tissues has identified mouse primary transcripts (Mortazavi et al., 2008). By using this technique to sequence the poly-A transcripts of the *Dgcr8*$^{gt/tm1}$ cells it may be possible to derive an RNA-seq transcriptome profile enriched for pri-miRNA sequences to aid their annotation.

6.4.5 Extending the investigation of the cellular phenotype upon mature miRNA depletion

There are several improvements and extensions to the phenotypic profiling of the *Dgcr8*$^{gt/tm1}$ cells that I would wish to include as part of my future studies. As described in Chapter 3 I wish to determine a growth curve for both the *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ cells prior to further cell cycle analysis. This would allow me to ensure that all of the cell samples subsequently processed for cell cycle profiling are in the exponential phase of their growth curve. It would also provide an additional comparison of the growth rate of the cells to supplement the cycle profile. I also wish to further my analysis of the potential of the *Dgcr8*$^{gt/tm1}$ cells to differentiate. Initially I would like to assess the expression of ES cell markers and differentiation markers in these cells following EB differentiation through qRT-PCR or immunostaining. Wang *et al.* noted that *Dgcr8* knock out ES cells were unable to silence ES cell markers upon the induction of differentiation (Wang et al., 2007). Sinkkonen *et al.* noticed a similar failure to repress Oct4 expression upon the induction of differentiation in Dicer knockout ES cells (Sinkkonen et al., 2008). They later ascribed this inability to silence Oct4 to a failure to methylate the Oct4 promoter during differentiation, a process indirectly modulated by the miR-290 cluster. I would like to confirm a similar

phenotype in my *Dgcr8*$^{gt1/tm1}$ cells. This will allow me to determine if my *Dgcr8*$^{gt1/tm1}$ cells are comparable to knockouts generated in other studies in this respect. In addition it may be informative to generate chimeric embryos from the *Dgcr8*$^{gt1/tm1}$ cells through injection into blastocysts. The *Dgcr8*$^{gt1/tm1}$ cells express lacZ and should therefore be traceable in the mouse embryo making it relatively easy to determine to which tissues, if any, the cells are able to contribute.

6.4.6 An investigation into the role of miRNAs in ES cell adhesion

Considering the enrichment of "ECM-receptor interaction" and "focal adhesion" GO terms among those genes significantly up regulated upon the depletion of DGCR8, which include several integrins, and in light of the morphological phenotype and the inability of the homozygous mutant cell lines to assemble into EBs it would be interesting to explore the adhesive properties of the *Dgcr8*$^{gt1/tm1}$ cells further. Initially, *Dgcr8*$^{gt/tm1}$ and *Dgcr8*$^{tm1,gt/+}$ cells could be plated onto culture plates coated with different ECM substrates in a fashion similar to that used by Ohnishi *et al.* (Ohnishi et al. 2008). This could then be used to identify substrates upon which the cells differentially adhere through fixing attached cells and staining them with crystal violet. Subsequent extraction of the dye and the measurement of its absorbance would provide a reading proportional to the number of the attached cells (Ohnishi et al. 2008). A comparison between the adhesive properties of *Dgcr8*$^{gt/tm1}$ cells transfected with a set of ES cell miRNAs and control miRNAs would also provide an assay that could be used to identify miRNAs that may rectify any adhesion phenotype in a screen similar to that used by Wang *et al.* which assigned miRNAs with roles associated with the regulation of the ES cell cycle (Wang et al. 2008). Once miRNA candidates have been selected, arrays could be used to formulate target lists and

ascribe a more explicit role to the miRNAs through the functional investigation of the potential targets.

6.4.7 Use of a conditional mutation in an alternative system

There is a risk when using constitutively mutated cell lines that compensatory mutations or changes in expression may accumulate that are secondary to the initial mutation. This has been suggested as the cause of the recovery of Dicer knock out cells from an initial growth deficit following the mutation of the Dicer locus (Murchison et al., 2005). One method by which irreversible secondary effects could be identified would be to attempt to rescue the cell lines through the re-expression of *Dgcr8* in the mutated cells. This could be achieved through either the transfection of a transiently expressed cDNA or through the insertion of this cDNA at a random or targeted locus.

An alternative approach would be to use the Bay Genomics gene trapped cell lines as a basis from which to derive a conditional knock out cell line. One option would be to flank essential exons of the second allele at the *Dgcr8* locus with loxP sites. The loxP system would allow the excision of the intervening DNA upon the expression of Cre recombinase. Expression of a Cre protein fused to a mutant oestrogen hormone binding domain allows the activity of the Cre recombinase to be controlled through the addition of an oestrogen analogue to the cell culture medium (Hameyer et al., 2007; Vooijs et al., 2001). In this way a mutation can be triggered within the second allele of *Dgcr8* upon demand and the subsequent alterations to the cell phenotype can be monitored over time to discriminate the initial effects of miRNA depletion from the effects of compensatory changes.

6.5 Conclusion

In this thesis I have described a method with which to conduct large-scale analyses of miRNA targeting in mouse ES cells. In addition I have used this methods to derive target candidate lists for 2 of the miRNAs highly expressed in these ES cells. It is becoming increasingly well understood that miRNAs play an essential role in modulating ES cell pluripotency and differentiation (Singh et al., 2008; Sinkkonen et al., 2008; Tay et al., 2008; Wang et al., 2007). Therefore, although it is not feasible to discern the intricate details of the individual miRNA-target relationships within the time-scale of my doctoral thesis, I strongly believe that both the candidate target lists and the underlying method for their generation should be of interest and assistance to a large body of researchers. Therefore I hope that the publication of the data will make it widely available to the scientific community and aid an understanding of the function of miRNAs both in mouse ES cells and in general.