Chapter 5: The derivation of ES cell miRNA candidate target lists in a background depleted of endogenous miRNA expression

5.1 Aim

The aim of this chapter is to use the *Dgcr8*$^{gt1/tm1}$ knock out cell line to identify mRNA targets of endogenous miRNAs in the context of the ES cell transcriptome. These cells have a broad depletion of endogenous miRNA levels and therefore derived targets will be identified in the absence of functional redundancy and the effects of co-regulation. In order to do this I optimize the reintroduction of miRNA mimics into the *Dgcr8*$^{gt1/tm1}$ cell line and subsequently introduce a set of ES cell expressed miRNAs, producing candidate target lists.

5.2 Introduction

The original large-scale analysis of miRNA target perturbations was conducted by Lim *et al.* (Lim et al., 2005). Following the reintroduction of miR-1 or miR-124 into HeLa cells the authors judged changes in the cells' expression profiles through the use of mRNA expression arrays. As a consequence they noted that by transfecting miRNAs not normally expressed in HeLa cells but highly expressed in the heart and the brain respectively, the mRNA profiles of the transfected cells were affected in such a way as to become more akin to the tissue in which the miRNAs are normally expressed. Although the expression changes witnessed in these experiments did significantly overlap with computational miRNA target predictions, it is logical that the system will be somewhat restricted to the identification of the *in vivo*

targets of those miRNAs that are incidentally also expressed in HeLa cells. This is a clear limitation if the intention was to better understand the significance of non-artificial miRNA-target interactions. Grimson *et al.* took advantage of this relatively simple cellular system to generate mRNA profiles from HeLa cells transfected with 11 different miRNA duplexes (Grimson et al., 2007). Again the miRNAs used were not HeLa cell specific and the experiment would thus suffer from the same limitations as those of the Lim study. However, while the limited expression of each miRNA's *in vivo* targets in HeLa cells will interfere with the identification of *in vivo* miRNA-target interactions these gene lists suit the purpose of better understanding the rules underlying miRNA target associations.

Mutant zebrafish embryos lacking both maternal and zygotic Dicer (*MZDicer*) activity display a severe morphological phenotype (Giraldez et al., 2005). The reintroduction of miR-430 into these embryos by microinjection went some way to rescuing the structural abnormalities in the embryonic brain. Subsequently, microarrays were used to identify the putative targets of the miR-430, through a comparison of the mRNA expression in embryos with no Dicer activity to either wild type embryos or embryos with miR-430 reintroduced, in which the expression of miR-430 targets will be down-regulated (Giraldez et al., 2006). The intersection of these two comparisons identified 328 genes down regulated (≥1.5 fold) in the presence of miR-430 with experimentally annotated 3'UTRs. Of these, 203 had sequences complementary to miR-430 seed sequence (Lewis et al., 2005) within their 3'UTR. Reporter constructs with the putative wild type and mutant target 3'UTRs fused downstream of a GFP reporter gene were used to validate predicted miRNA target UTRs, through injection into embryos of a wild type or *MZDicer* background. Of the 3'UTRs tested with a single 6mer (2-7) or 7mer target site, 5/7 were validated by these assays (>2-fold up regulation in *MZDicer*

background). An even greater proportion of targets were validated, if the 3'UTRs contained either a greater number of sites or an 8mer target site. Only ~25% of the proposed targets were conserved in both Tetradon and Fugu. This would mean that any predictive algorithm that required target conservation as predictive criteria would miss these targets.

Although miR-430 is very highly expressed in the early zebrafish embryo, in the *MZDicer* fish the lack of the usual spectrum of miRNAs expressed in the wild type embryos makes the role of individual miRNAs easier to discern. There will be no combinatorial regulation or functional redundancy in the system directed by other miRNAs. As a result the function of miR-430 is potentially even more clear and startling than if individual miR-430 duplexes were knocked out or over expressed in an otherwise wild type system.

Recently similar systems have been developed and used in mouse embryonic stem cells. *Dgcr8* knock out cells have been used to identify miRNAs responsible for the regulation of the cell cycle (Wang et al., 2008). Mouse embryonic stem cells with a disrupted miRNA-processing pathway accumulate in the G1 phase of the cell cycle (Murchison et al., 2005; Wang et al., 2007). Wang *et al.* performed a screen of mouse miRNAs to identify miRNAs that improved this proliferation phenotype (Wang et al., 2008), subsequently identifying members of the miR-290 cluster and miRNAs with similar sequences as important regulators of the p21 (*Cdkn1a*) CDK inhibitor, and central to cell cycle regulation. The knockout background was fundamental to the identification of p21 as a target of these miRNAs as both over-expression and knock down of individual miRNAs in wild-type cells had minimal effect on the ES cell proliferation. This demonstrates the advantages attributable to the removal of

functional redundancy and target saturation from the system prior to screening for miRNA targets.

Whereas Wang *et al.* initially relied on TargetScan predictions to identify p21 as a candidate miRNA target, Sinkkonen *et al.* took an alternative approach and extended the system used by Giraldez *et al.* (Giraldez et al., 2006) to an embryonic stem cell system (Sinkkonen et al., 2008). They compared the expression profile of *Dicer*$^{-/-}$ cells to heterozygous counterparts and then identified from among those genes up-regulated in the *Dicer* $^{-/-}$ cells mRNAs that are down-regulated upon the re-addition of the miR-290 cluster by electroporation. This transferal of the Giraldez approach to a simpler cellular system produced a list of 253 predicted targets of this cluster (including *Cdkn1a*). These genes provided the authors with a functional shortlist enriched with miRNA targets ultimately leading to the identification of *Rbl2* as a potential target of the miR-290 cluster, which could go some way to accounting for the lack of *de novo* DNA methylation at the Oct4 promoter upon differentiation. Simultaneously a second independent study reached the same conclusion concerning the miR-290 cluster orchestrating methylation control via the *Rbl2* gene (Benetti et al., 2008).

In this chapter I assess the expression profiles of the cell lines derived in Chapter 3 and use these profiles to investigate the broad roles of miRNAs in embryonic stem cells. I then explain the optimization of the re-addition of miRNAs into the *Dgcr8*$^{gt1/tm1}$ cell line. By using the expression profiles of the cell lines and expression data from *Dgcr8*$^{gt1/tm1}$ cells transfected with miRNA mimics, I was able to produce lists of genes enriched in miRNA targets for mmu-miR-25 and mmu-miR-291a-3p, using a system of intersecting gene sets similar to that used by both Giraldez *et al*. and Sinkkonen *et al*. (Giraldez et al., 2006; Sinkkonen et al.,

2008). Based on these gene lists I am able to construct hypotheses and functional relationships for future investigation.

5.3 Results

5.3.1 A comparison of growth conditions and their effect on cell phenotype

Initially I performed a comparison between the mRNA expression profiles of *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{gt2/tm1}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$ and *Dgcr8*$^{+/+}$ cells. RNA was prepared from each cell line in triplicate and the expression profiles were assessed by Illumina Mouse-6 V1.1 expression chips. However as has been mentioned previously it was necessary to alter the cell culture conditions during the course of the work presented in this thesis. The centrifugation step was initially included in the cell splitting protocol to remove the residual trypsin from the cells prior to plating to fresh plates, given the low split ratios used to divide the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells (see section 2.4.1.1). It also initially seemed that as short a period in trypsin as possible might help to limit the morphological phenotype I was seeing amongst the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells. However, the initial culture conditions used were causing a greater than optimal proportion of the *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$ and wild type cells to differentiate. It was judged that this could be the result of the mechanistic stress of the centrifugation step, used to remove the cells from the trypsin. Subsequently it was deemed more appropriate to culture the cells in the conditions most optimal for the propagation of the control cell lines, as any subsequent differences seen between the growth of the control and mutant cells could still be attributed to the depletion of DGCR8 (see section 2.4.1.2). As a result I regrew the cells in triplicate using the new set of growth conditions and repeated the arrays. It was considered important that this experiment be repeated in full as the results would not only provide the basis for judging likely miRNA

controlled genes, but comparison between the two array sets prepared with RNA grown under the two conditions could be used to judge whether the growth condition were having a fundamental effect on the cells and the results gathered under the two conditions.

In addition to this alteration in protocol between the two array sets, I prepared and labeled the RNA for analysis under the first set of conditions personally. The RNA prepared under the fresh growth conditions was labeled for the array by Dr Peter Ellis in the Sanger Institute microarray facility.

In order to determine the magnitude of the effect of the growth conditions on the expression data subsequently gathered by array, I performed cluster analysis based on the correlation in normalized expression between samples and arrays. Initially, if the samples are clustered based on the entire normalised expression set for each sample, the separation of the clustered structure appears to be based on growth conditions and the identity of the person who conducted the labeling of the samples (Fig.5.1A), (although there is evidence that some of the *Dgcr8*$^{gt1/tm1}$ miRNA transfected samples, grown under the new growth conditions also cluster with the older expression array samples (See section 5.3.3.4)). It is worth bearing in mind, however, that when considering the entire normalized expression set, a large proportion of the data contributing to the inter-sample correlation, will be derived from probes that vary by an insignificant amount, due to small changes attributable to the growth condition or array and experimental noise. As an alternative I refined the selection of probes used to cluster the arrays. First I removed probes from the set based on the detection score attributed to probes by the Illumina Beadstudio programme, which is a measure of the likelihood that the probe target is expressed above background (detection score < 0.05 in more than 5 of the samples).

Subsequently the probe section was further narrowed by using only those probes whose intensity varied significantly between the set of 6 combined *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ arrays and the six *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$ arrays, all grown without spinning the cells out of the trypsin (Section 2.4.1.2) (P-value < 0.05, LFC > $\log_2(1.2)$). The resultant probe set used to separate the samples contained 3489 probes. When only considering these probes the samples clustered as would be expected if the clustering was attributable to genotype instead of growth condition (Fig.5.1B). All of the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ samples cluster together irrespective of the culture method as do the control cells. Also of note, when the clustering is based on the intensity of this limited set of probes, the *Dgcr8*$^{gt1/tm1}$ cells transfected with miRNA mimics (Section 5.3.3.4) also cluster with the other *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ samples. By limiting the set of probes used to only those that change most significantly in a single growth condition, the selection of this probe set is not biased by those probes which change significantly between cells with differing genotypes grown under the alternative growth conditions. The choice of probes therefore does not assume that cells with the same genotypes grown under each condition will necessarily behave in the same manner. However, by limiting the probe set in this way, the clustering is now based on the expression changes evident within a set of the most biologically relevant genes, removing a large fraction of the noisy probes whose expression varied by small amounts.

A

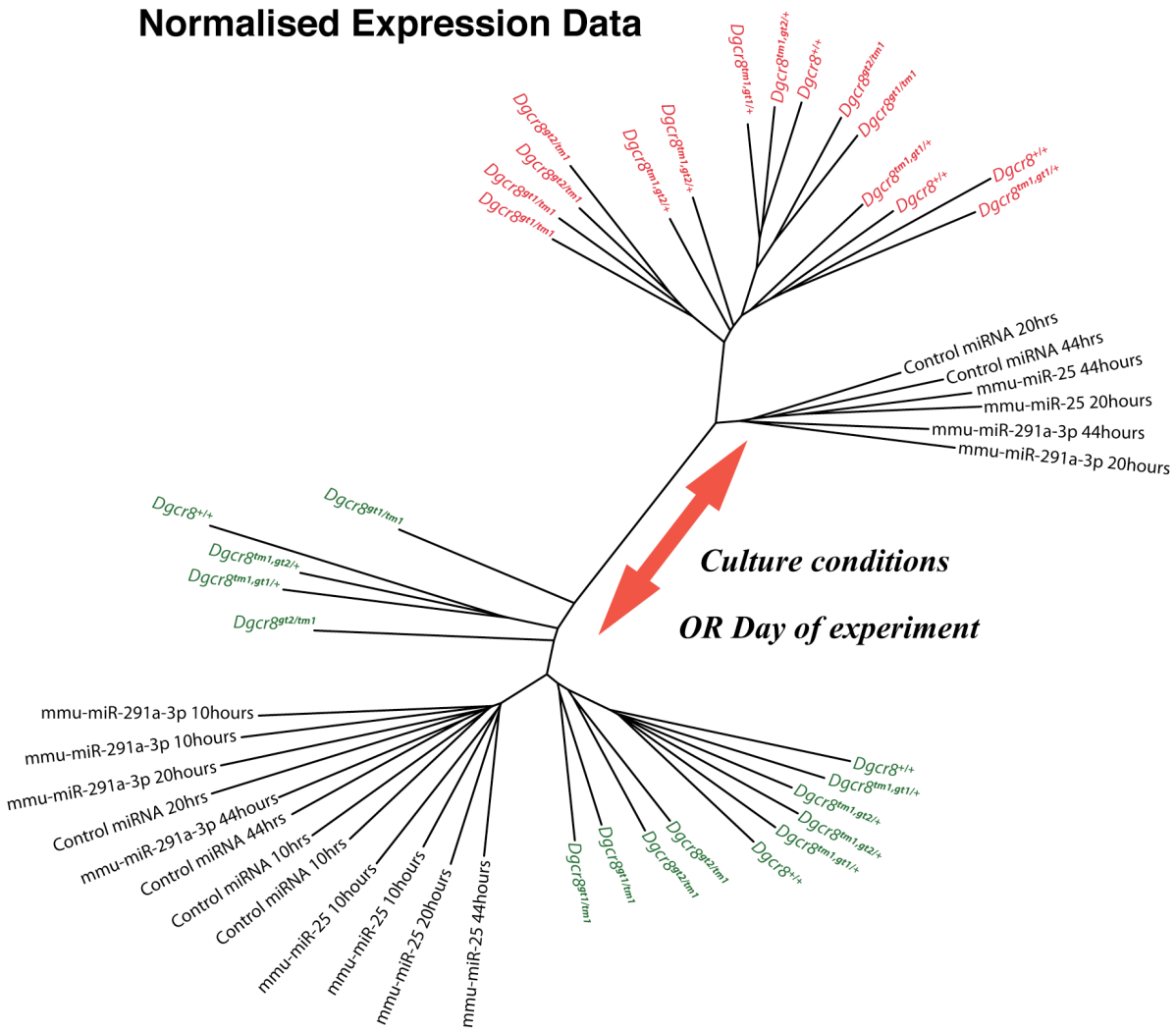


Fig.5.1A: A rootless tree depicting the degree of correlation between all of the mRNA array profiles on the basis of the complete set of normalized probe intensities. Spearman's correlation was used to determine the degree of similarity between the results of all of the expression arrays used in this study. In red are the arrays conducted upon samples cultured following the original culture method (2.4.1.1). In green are all the arrays conducted upon samples prepared with the more recent culture method (2.4.1.2). In black are the arrays conducted as part of the miRNA re-addition experiments (See Section 5.3.3).

B

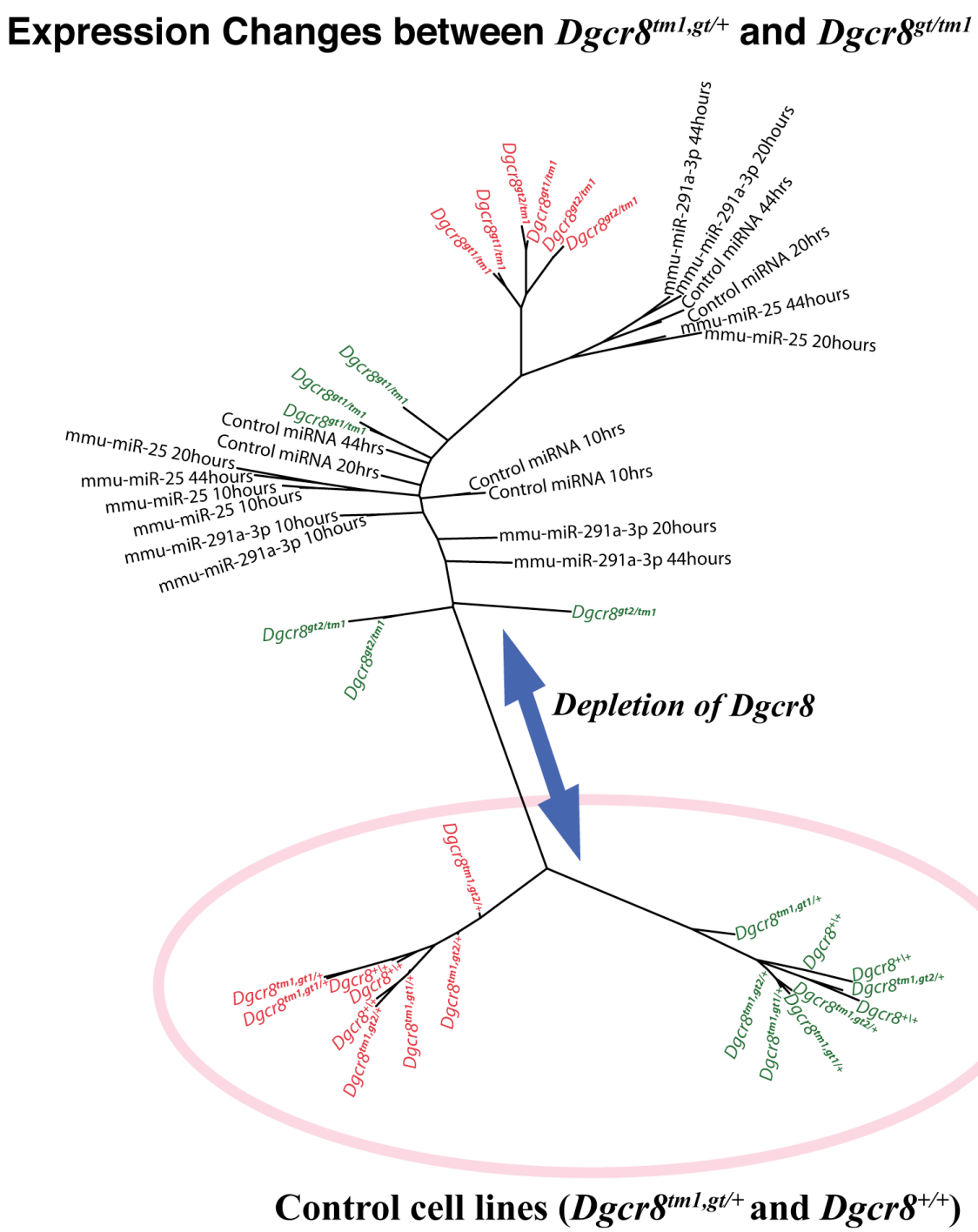


Fig.5.1B: A rootless tree depicting the degree of correlation between all of the mRNA array profiles on the basis of a refined set of probe intensities. The probes used to calculate the correlation between samples were selected as those which were registered as "present" in more than 5 of the array samples as judged by the

lumiBatch detection calls (*P*-value < 0.05) and which demonstrated significant differential expression between all the of the arrays for cells with two disrupted alleles compared to arrays with a singe disrupted allele. All the cells in the comparison were grown under the most recent culture conditions (2.4.1.2). Sample relationships were calculated by Spearman's correlation.

If the arrays grown under the older set of conditions are considered and the same expression comparison is conducted (e.g. $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ vs. $Dgcr8^{tm1,gt1/+}$, $Dgcr8^{tm1,gt2/+}$) 3738 probe intensities change significantly. The overlap between the two sets of significant probes accounts for 2697 probes in this set. Again it is apparent that the two sets of array experiments both identify similar gene sets as significantly altered following the depletion of functional DGCR8 irrespective of the culture conditions.

To further ensure that the cells cultured by the two methods behave in an approximately equivalent fashion, I decided to determine whether there was an enrichment of miRNA seed sequences within the 3'UTRs of genes up regulated upon the depletion of DGCR8, considering the cells grown by each method independently. Sylamer is a program that searches ordered lists of sequences for the enrichment of defined motifs (van Dongen et al., 2008). The program begins at one end of the sequence list and progresses through the list by considering increasingly large sets of sequences for an enrichment of a given motif above what would be considered usual given the number of motifs within the remainder of the sequences in the list. If provided with a list of UTRs ordered by their log fold change in expression following the removal of DGCR8, the program is able to search the list to identify miRNA seed sequences that are enriched within the UTRs of the genes whose expression has either up-regulated or down-regulated by the depletion.

For each set of culture conditions I again compared the 6 expression arrays for *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ against the six arrays for *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ (3 of each genotype) to produce two lists of probes; ordered by their log fold change following the disruption of both alleles of *Dgcr8*. These ordered lists of probes were converted into lists of associated 3' UTRs and analysed (by Dr. Cei Abreu-Goodger) using the Sylamer program for the enrichment of any miRNA seed sequence motifs within either the 3'UTRs of up-regulated or down-regulated genes. As can be seen in Fig.5.2, the most significantly enriched miRNA seed sequences in the genes up-regulated upon the disruption of both *Dgcr8* alleles are essentially the same for both culture conditions. This again implies that fundamentally the two sets of arrays are in agreement as to the most significant expression changes despite practical alterations to the experimental procedure. It is also of note that seed sequences that appear to be amongst the most enriched correspond with miRNAs among those most highly expressed within the *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ cells, judged by Solexa/Illumina sequencing (See Chapter 4, Fig.4.8), and whose expression was subsequently reduced in *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells. This enrichment of these miRNA specific seeds implies that a significant proportion of the genes up regulated following miRNA depletion are subject to ES cell miRNA mediated regulation, which is disrupted following miRNA depletion. There are two miRNAs for whom the target seed sequences appear to be significantly depleted within a portion of the ordered gene list (miR-879-7mer-2/m8 and miR-127-7mer-1A), with their line falling below 0 on the y-axis of the chart, although these *P*-values seem to be marginal. The relevance of the enriched seed sequences will be considered more thoroughly later in the chapter, where I will consider the changes in gene expression following DGCR8 depletion, by combining the array data for cells of differing genotypes, grown by either culture method.

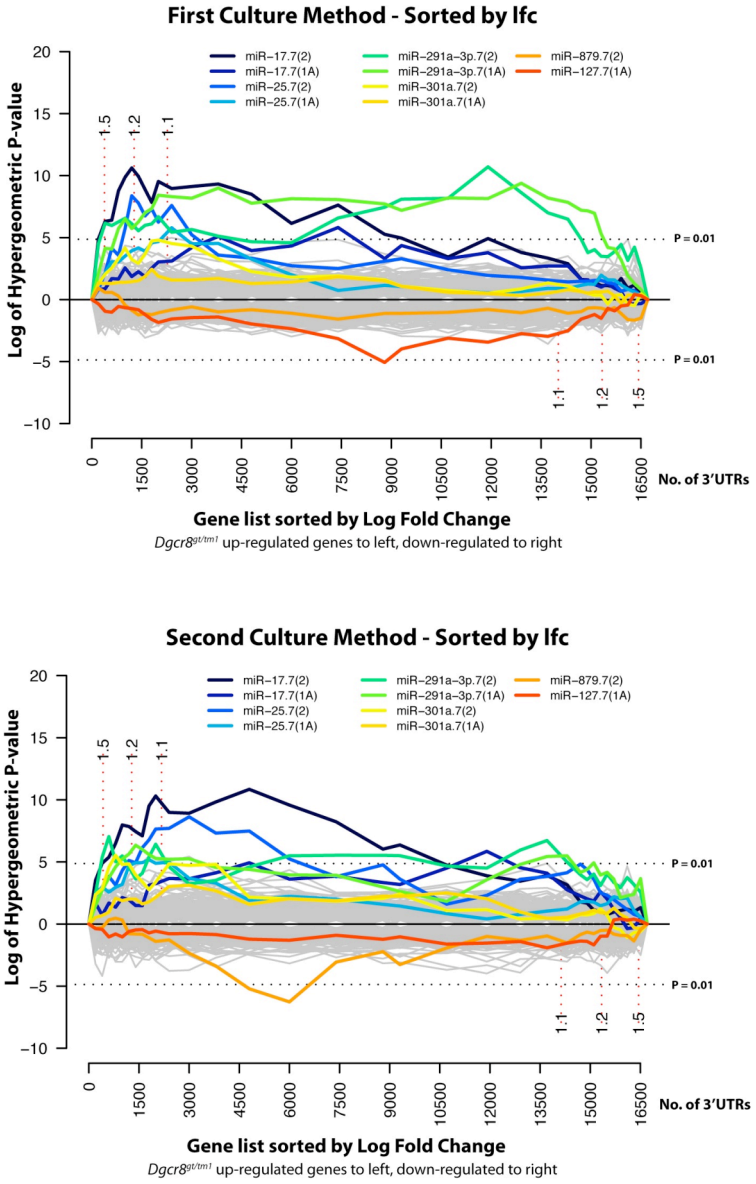


Fig 5.2: Sylamer plots to identify miRNA seed sequence enrichment or depletion within up-regulated or down-regulated genes following the depletion of DGCR8, conducted separately using arrays derived from cells cultured under the older culture conditions (2.4.1.1) (top) and the more recent culture methods (2.4.1.2) (bottom). Genes were ordered according to their log fold change following the depletion of DGCR8 (*Dgcr8*$^{gt/tm1}$ vs. *Dgcr8*$^{tm1,gt/+}$). These gene lists are plotted along the x-axis, with up-regulated genes on the left and down-regulated genes on the right. Subsequently the 3'UTRs of the ordered list were assessed in sets anchored at 0 on the x-axis, but of increasing size, for enrichment of all miRNA 7mer seed sequences (7mer-1A and 7mer-m8 (7mer-2) seeds based on mirBase v12 miRNA sequences) via a series of hypergeometric tests. The test results for each miRNA are represented by a line on the plot. miRNA target enrichment within a potion

of the list is associated with a either a peak to the left of the graph (if enriched in up-regulated genes) or trough to the right (if enriched in down-regulated genes). *P*-values cutoffs of 0.01 are represented as dotted lines stemming from the y-axis. On the Sylamer plot, red dotted lines represent the positions of various log-fold-change cutoffs. Sylamer analysis of the gene lists was conducted by Dr. Cei Abreu-Goodger.

5.3.2 Expression profiles of *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{gt2/tm1}$ and *Dgcr8*$^{+/+}$ cells

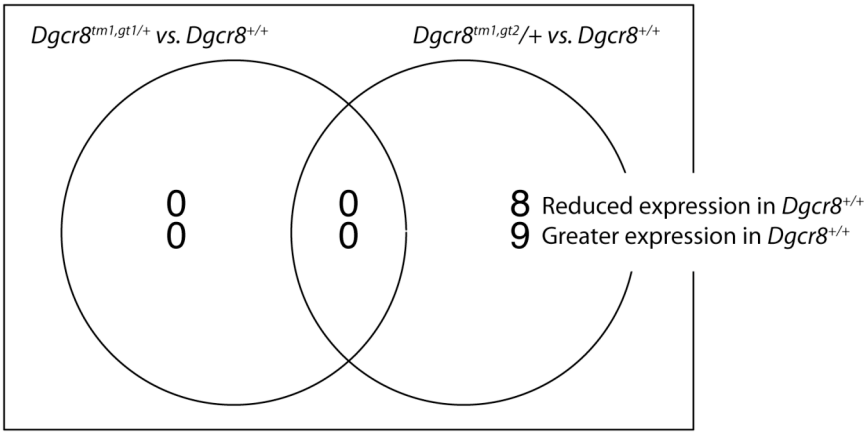
5.3.2.1 Comparison of the expression profiles of *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{gt2/tm1}$ and *Dgcr8*$^{+/+}$ cells

As the expression profiles for the cells cultured by the two methods seem broadly comparable, I decided to combine the array results for each genotype from each culture method into a single analysis. As a result I have 6 expression arrays corresponding to each genotype, which will increase the statistical confidence of any derived expression changes.

First I compared the expression of *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ cells to the *Dgcr8*$^{+/+}$ cells to ensure that the single disrupted allele wasn't having profound effects on the gene expression in the cells (Fig.5.3A). As can be seen, there are very few significant differences in the expression profiles of these cells. This suggests that *Dgcr8* isn't a limiting factor in the miRNA mediated gene regulation as there are few significant alterations to expression when it is partially depleted. Alternatively the partial reduction in functional DGCR8 within the cell could trigger an autoregulatory feed back loop to compensate with an increase in the stability of *Dgcr8* mRNA, as described by Han *et al.* (Han et al., 2009). Indeed they noted that although a 50% reduction in *Dgcr8* mRNA levels is expected in heterozygous *Dgcr8* knockout ES cells and MEFs, these cells appear to retain 90% of the wild type mRNA

expression and neither cell line demonstrated a significant reduction in DGCR8 or Drosha protein levels.

A



B

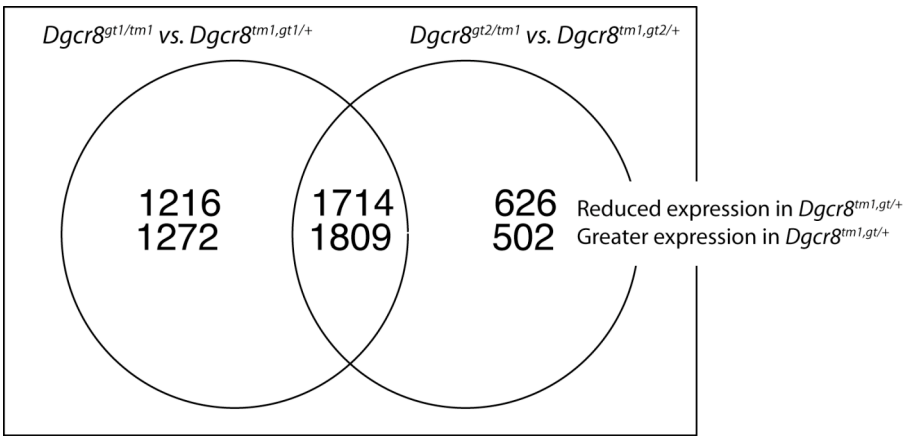


Fig.5.3: A Venn diagram depicting the number of probes registering significant expression changes between cell types. A) Comparisons between wild type cells (*Dgcr8*$^{+/+}$) and cells with a single disrupted allele (*Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$). B) Comparisons between cells with a single disrupted allele (*Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$) and their paired DGCR8 depleted cell lines, each with a trap within both alleles of *Dgcr8* (*Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$). Log fold change > $\log_2(1.1)$, *P*-value < 0.05.

A comparison between the expression profiles of the *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ cells and the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines identified a considerable overlap between the

genes with significantly altered expression when the two biological replicate comparisons were made separately (Fig.5.3B). Therefore subsequently I combined the expression array data for the independent replicate cell lines (*Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$) to give me data from 12 arrays corresponding to each broad genotype; *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$. By using more arrays I would increase the statistical robustness of any comparison and therefore be able to discern significant expression changes more accurately. Data from these two sets of 12 arrays were compared to determine the expression changes that result from the depletion of DGCR8, stemming from the insertion of a gene trap cassette into both alleles. A total of 6695 probes altered significantly between these two sets (3251 probes were up regulated in the *Dgcr8*$^{gt/tm1}$ cells compared to *Dgcr8*$^{tm1,gt/+}$ and 3444 probes were down regulated. *P*-value < 0.05, LFC > $\log_2(1.1)$ (Fig.5.4A)) . In addition the array probes for this joint analysis were arranged in order according to their LFC and differential expression related t-statistic that resulted from this comparison, and again Sylamer was used (Dr. Cei Abreu-Goodger) to determine if there were any miRNA seed enrichments associated with these expression changes as explained above (Fig.5.2). A similar set of seed sequences were found to be enriched in the genes up-regulated upon the insertion of a trap into both alleles of *Dgcr8* in this combined analysis as were found when the two growth conditions were considered separately (Fig.5.4B). However, when all the arrays are combined there are no longer any significant miRNA seed depletion signals seen below 0 on the y-axis. The list of miRNAs for whom the seed sequences are enriched amongst those genes up regulated closely resembles the list of the most highly expressed miRNAs in the *Dgcr8*$^{tm1,gt/+}$cells (Fig.5.4C, Fig4.8). This is to be expected as it would seem reasonable that these highly expressed miRNAs would be playing an influential role in post-transcriptional regulation in mouse ES cells. What is not necessarily as obvious is that this influential role

would encompass a broad set of target genes, to the extent at which a strong enrichment signal would be seen following the miRNA depletion, rather than influencing a small refined set of target genes. Also of note is the number of highly expressed miRNAs sharing the same enriched seed sequences (Fig.5.4C). In particular the miR-17 7mer seeds are shared by 4 of the top 17 most highly expressed miRNAs in the *Dgcr8*$^{tm1,gt/+}$cells. In this case the enrichment of the miR-17 seed could likely result from the deregulation of the targets of all of these miRNAs. Conversely all of these miRNAs are likely to overlap in their target lists.

A

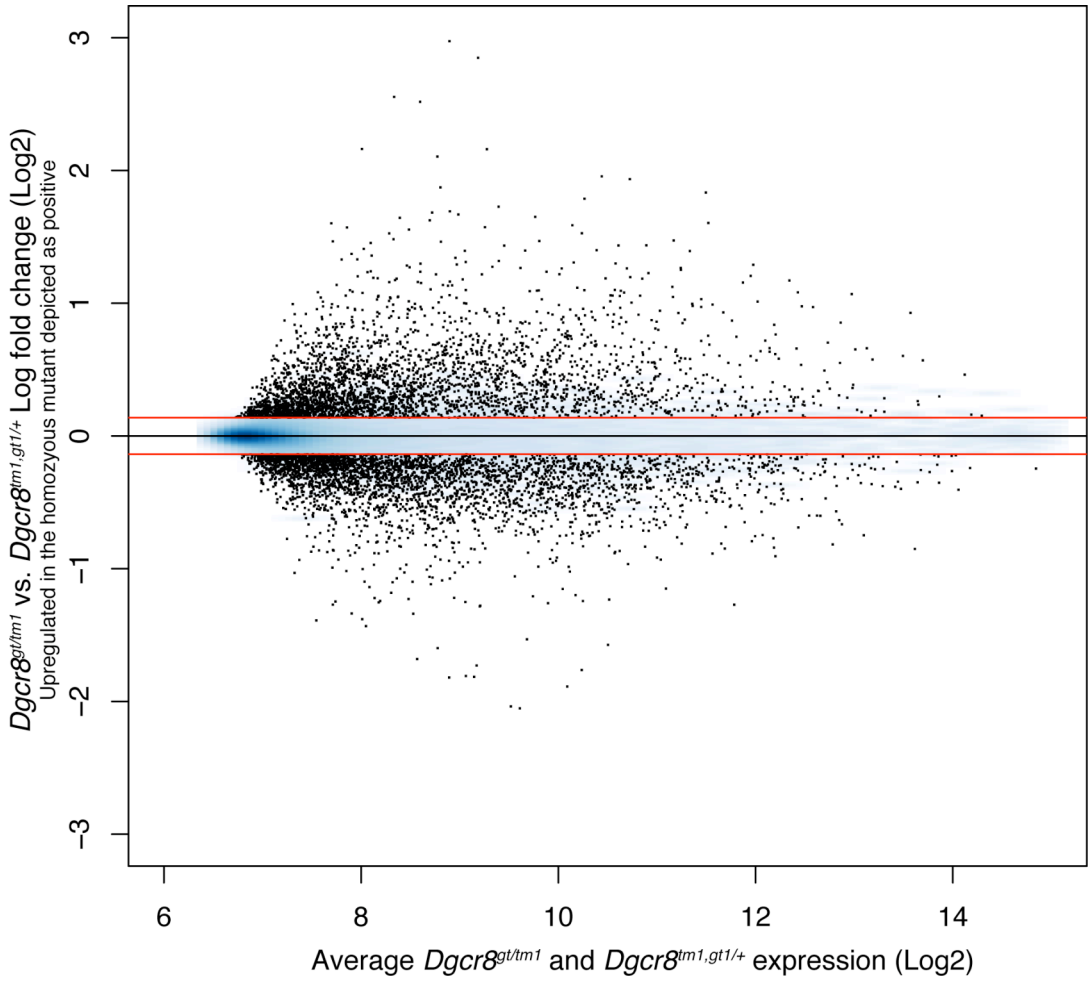


Fig.5.4: A) A plot depicting the average expression of each probe in the heterozygote and homozygous mutant cell lines against the log fold change of each probe following the depletion of DGCR8 from the ES cells. The red lines define a log fold cut off of log2(1.1). Probes deemed insignificant given an adjusted p-value cut off of 0.05 and a LFC cut off of log2(1.1) are plotted as a density gradient in blue. Probes whose expression alters significantly between the cell lines are plotted as black spots.

B

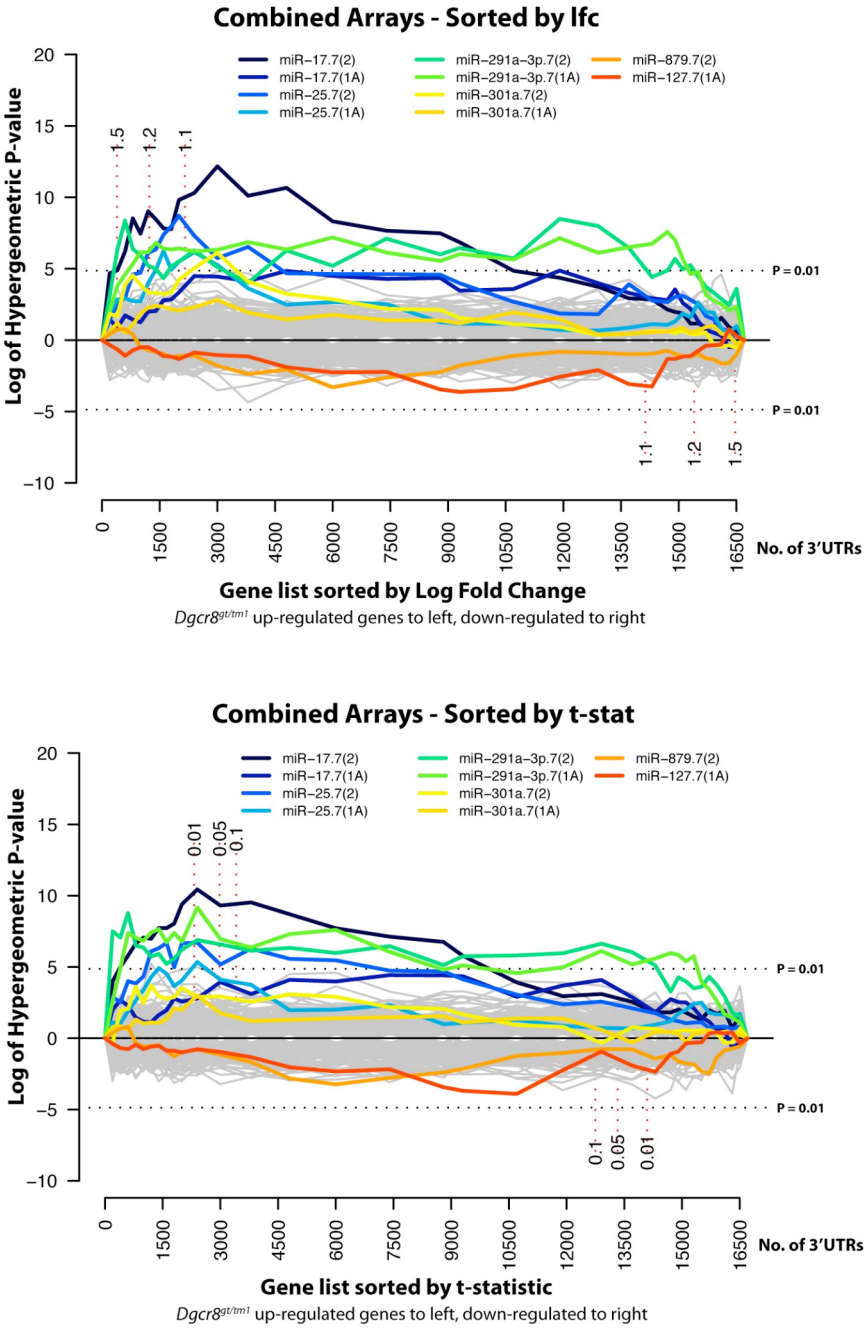


Fig.5.4: B) Sylamer plots to identify miRNA seed sequence enrichment or depletion within up-regulated or down-regulated genes following the depletion of DGCR8 based on expression data derived from cells grown under both the older and the most recent culture conditions. For a desciption of these plots see Fig.5.2. In this case the gene lists were ordered according to their log fold change (LFC) and t-statistic. Sylamer analysis of the gene lists was conducted by Dr. Cei Abreu-Goodger.

C

miRNA Seed	miRNA Target Seed Sequence	miRNAs with a common seed
miR-17.7(2)	GCACUUU	**miR-20a**.7(2), miR-20b.7(2), **miR-93**.7(2), **miR-106a**.7(2), **miR-106b**.7(2)
miR-17.7(1A)	CACUUUA	**miR-20a**.7(1A), miR-20b.7(1A), **miR-93**.7(1A), **miR-106a**.7(1A), **miR-106b**.7(1A)
miR-25.7(2)	GUGCAAU	miR-32.7(2), miR-92a.7(2), miR-92b.7(2), miR-363.7(2), miR-367.7(2)
miR-25.7(1A)	UGCAAUA	miR-32.7(1A), miR-92a.7(1A), miR-92b.7(1A), miR-363.7(1A), miR-367.7(1A)
miR-291a-3p.7(2)	AGCACUU	miR-294.7(2) **miR-295**.7(2), miR-302a.7(2), miR-302b.7(2), miR-302d.7(2)
miR-291a-3p.7(1A)	GCACUUA	miR-290-3p.7(1A), miR-291b-3p.7(1A), miR-292-3p.7(1A), miR-294.7(1A), **miR-295**.7(1A), miR-302a.7(1A), miR-302b.7(1A), miR-302d.7(1A), miR-467a.7(1A), miR-467b.7(2), miR-467c.7(1A), miR-467d.7(1A)
miR-301a.7(2)	UGCACUA	miR-130a.7(1A), miR-130b.7(1A), miR-301b.7(1A), miR-721.7(1A)
miR-301a.7(1A)	UUGCACU	miR-130a.7(2), miR-130b.7(2), miR-301b.7(2), miR-721.7(2)
miR-879.7(2)	AAGCCUC	
miR-127.7(1A)	GAUCCGA	

Fig.5.4: C) The miRNA seed sequences highlighted as enriched or depleted by Sylamer analysis in gene lists ordered by LFC or t-statistic following the disruption of both alleles of the *Dgcr8* locus. The right hand column lists miRNAs that share the same target seed, which is given in the second column. Highlighted in red are miRNAs with an average maximum read depth >10,000 between the *Dgcr8*$^{tm1,gt/+}$ cells (Fig.4.8).

5.3.2.2 Functional analysis of genes up regulated upon DGCR8 depletion

To gain an insight into the broad cellular functions that may be subject to miRNA regulation in ES cells, the group of genes that were up regulated upon the depletion of DGCR8 and miRNAs were subjected to GO analysis and KEGG pathway analysis (3251 probes, *P*-value < 0.05, LFC > $\log_2(1.1)$) (Table 5.1, Appendix B (CD)). Perhaps the most interesting trend is seen amongst those "Biological Process" GO terms over-represented amongst this set of genes. The most enriched terms include a large number of terms relating to "development".

This implies that miRNAs may play an important role in the dampening of developmental genes in pluripotent stem cells, perhaps stabilizing this undifferentiated and pluripotent state (Table 5.1).

Also of interest is the enrichment of 'ECM-receptor interaction' genes identified by the KEGG pathway analysis. Cellular interactions with the extra-cellular matrix are known to affect cellular differentiation and proliferation. miRNA involvement in the regulation of these pathways may be in an interesting avenue for future research (Table 5.1).

KEGG ID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
4512	6.50E-05	3.3	9.7	22	53	**ECM-receptor interaction**
4510	7.70E-04	2	24.1	39	131	**Focal adhesion**
5222	1.90E-03	2.3	11.9	22	65	**Small cell lung cancer**
10	2.30E-03	3.4	5.1	12	28	**Glycolysis / Gluconeogenesis**
280	3.20E-03	3.2	5.3	12	29	**Valine, leucine and isoleucine degradation**
5212	3.60E-03	2.2	11.8	21	64	**Pancreatic cancer**
720	4.00E-03	6.7	1.8	6	10	**Reductive carboxylate cycle (CO2 fixation)**
640	5.40E-03	3.7	3.7	9	20	**Propanoate metabolism**
4070	9.00E-03	2.3	8.8	16	48	**Phosphatidylinositol signaling system**

GO BP IDs	Pvalue	OddsRatio	ExpCount	Count	Size	Term
GO:0001525	9.60E-06	2.6	18.6	37	98	**angiogenesis**
GO:0016044	5.50E-05	2	31.1	52	164	**membrane organization and biogenesis**
GO:0009887	8.40E-05	1.7	60.6	88	321	**organ morphogenesis**
GO:0006897	9.40E-05	2.2	21.8	39	115	**endocytosis**
GO:0009790	2.00E-04	1.6	63.6	90	337	**embryonic development**
GO:0001568	2.10E-04	2	27.9	46	147	**blood vessel development**
GO:0045941	2.90E-04	1.7	45.9	68	242	**positive regulation of transcription**
GO:0031325	8.30E-04	1.6	56	78	295	**positive regulation of cellular metabolic process**
GO:0048518	9.00E-04	1.4	121.2	152	639	**positive regulation of biological process**
GO:0045944	9.20E-04	1.7	35.3	53	186	**positive regulation of transcription from RNA polymerase II promoter**
GO:0001889	9.30E-04	4	4.7	12	25	**liver development**
GO:0009891	1.10E-03	1.6	50.5	71	266	**positive regulation of biosynthetic process**
GO:0051254	1.10E-03	1.7	39.6	58	209	**positive regulation of RNA metabolic process**
GO:0008285	1.20E-03	2	19	32	100	**negative regulation of cell proliferation**
GO:0045595	1.50E-03	1.8	28.6	44	151	**regulation of cell differentiation**
GO:0032774	1.50E-03	1.3	216.8	254	1143	**RNA biosynthetic process**
GO:0010604	1.50E-03	1.5	53.7	74	283	**positive regulation of macromolecule metabolic process**
GO:0006732	1.90E-03	2.3	11.9	22	63	**coenzyme metabolic process**
GO:0045892	2.50E-03	1.7	30.2	45	159	**negative regulation of transcription, DNA-dependent**
GO:0048519	2.60E-03	1.3	126.9	155	669	**negative regulation of biological process**
GO:0045670	2.80E-03	6	2.3	7	12	**regulation of osteoclast differentiation**
GO:0051188	3.40E-03	2.4	9.5	18	50	**cofactor biosynthetic process**
GO:0043009	3.90E-03	1.6	41.7	58	220	**chordate embryonic development**
GO:0019219	4.10E-03	1.2	229.1	263	1208	**regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process**
GO:0007160	4.30E-03	3	5.5	12	29	**cell-matrix adhesion**
GO:0044262	5.00E-03	1.7	25.4	38	134	**cellular carbohydrate metabolic process**
GO:0048704	5.10E-03	2.8	6.3	13	33	**embryonic skeletal morphogenesis**
GO:0021575	5.10E-03	5	2.5	7	13	**hindbrain morphogenesis**
GO:0048731	5.60E-03	1.2	205.8	237	1085	**system development**
GO:0009880	6.10E-03	2.9	5.7	12	30	**embryonic pattern specification**
GO:0006790	6.10E-03	2.5	7.8	15	41	**sulfur metabolic process**
GO:0003002	6.30E-03	1.7	26.6	39	140	**regionalization**
GO:0010629	6.60E-03	1.6	35.9	50	189	**negative regulation of gene expression**
GO:0006733	6.90E-03	3.5	3.8	9	20	**oxidoreduction coenzyme metabolic process**
GO:0006766	7.40E-03	2.5	7.2	14	38	**vitamin metabolic process**
GO:0030888	7.80E-03	3.8	3.2	8	17	**regulation of B cell proliferation**
GO:0032318	7.80E-03	2.4	8	15	42	**regulation of Ras GTPase activity**
GO:0007157	8.60E-03	4.3	2.7	7	14	**heterophilic cell adhesion**

Table 5.1: GO term and KEGG pathway analysis of genes up regulated upon the depletion of DGCR8. Genes significantly up regulated (*P*-value < 0.05, LFC > $\log_2(1.1)$) between cells with a trap in each allele of *Dgcr8* when compared with cells with a single disrupted allele ($Dgcr8^{gt/tm1}$ vs. $Dgcr8^{tm1,gt/+}$) were subjected to GO and KEGG analysis for over and under-represented terms. Displayed above are the results of the over-represented "Biological Process" GO terms and over-represented KEGG pathway terms. Additional analyses are available in Appendix B (CD).

5.3.2.3 Identifying DGCR8 dependent alterations to the targets of the ES cell core transcriptional network

The mouse ES cell core transcriptional network (including Oct4, *Sox2*, *Nanog*, *Klf4* and c-Myc), plays a vital role in the maintenance of pluripotency and ES cell identity. Given the strongly enriched miRNA seed sequences identified amongst genes up regulated upon DGCR8 depletion (Fig.5.4B) and given the enrichment of "development" and "differentiation" GO terms amongst the significantly up regulated genes (Table.5.1) it seems reasonable to examine the potential influence of miRNA-coordinated regulation on the targets of this core transcriptional network. To achieve this I downloaded ChIP-Seq and ChIP-chip data for two studies that identified targets of transcription factors within this network (Chen et al., 2008; Kim et al., 2008b). There can be considerable variation between the targets predicted within each data set, for example Kim *et al.* predict 753 target genes for Oct4, while Chen *et al.* predict 6851 targets. Of these 420 are predicted by both studies. In order to ensure that there is the minimum noise associated with the data as possible I refined the target lists for each transcription factor to include only those target genes identified by both studies.

Subsequently, for these five transcription factors listed above, I compared the distribution of the fold changes of their target genes between the $Dgcr8^{tm1,gt/+}$ and $Dgcr8^{gt/tm1}$ cell lines, to the distribution of the other genes within the profile. All of the target distributions were significantly different to the background distribution (Fig.5.5). This suggests a significant perturbation of the targets of the core transcriptional network following the depletion of DGCR8 and miRNAs within ES cells. Interestingly, while the targets of the other 4

transcription factors all seem to be up regulated upon the removal of miRNAs from the system, the targets of c-Myc appear to be down regulated.

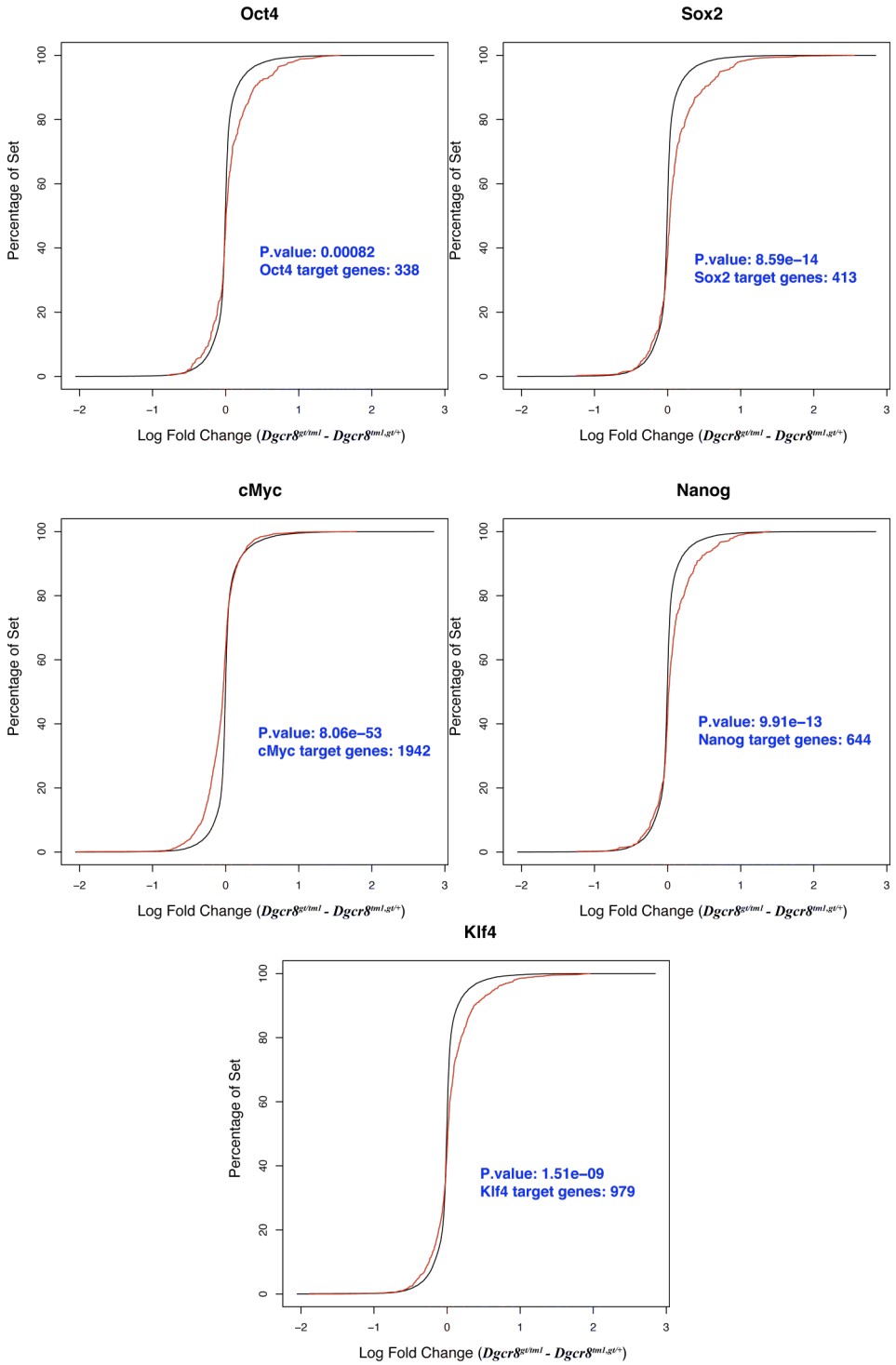


Fig.5.5: Plots to assess the change in the expression of the transcriptional targets of various core transcription factors relative to the other genes following the depletion of DGCR8 (*Dgcr8*$^{tm1,gt/+}$ vs.

***Dgcr8*$^{gt/tm1}$).** The transcriptional targets of each transcription factor were ordered according to their LFC as were the remainder of the genes on the array. These sets were plotted with the black line representing the background data and the red line representing the transcriptional targets. In each case the log fold change distributions of these two sets of genes were subjected to a Wilcoxon test to determine if they differed significantly, with the *P*-values given in blue.

There are three reasons why the target gene expression may have been broadly altered. Either the expression of the core transcription factor itself is affected by the loss of miRNAs or the targets themselves are broadly regulated by ES cell expressed miRNAs either directly or indirectly. To further address the first possibility, beyond the Western blots and immuno-staining presented in Chapter 3, I searched the probes whose expression altered significantly following the depletion of DGCR8, for the probes corresponding to these five transcription factors. Perhaps surprisingly, given my initial Western blots and immuno-staining (Fig.3.10 and Fig.3.11), probes corresponding to three of these TFs were indeed within these differentially expressed sets (Table 5.2). Microarrays are a notoriously noisy method, not suited to judging accurately expression changes for single genes and further confirmation by qRT-PCR or Western Blot would be necessary to confirm these alterations in gene expression.

Gene Symbol	Entrez IDs	Illumina Probe	Log Fold Change	Adjusted P-value
Klf4	16600	rpUHFdf15SFI5LRC1U	0.397	0.0011
Klf4	16600	35LRC1Xd1PNCJ05Ras	0.536	0.0012
Myc (c-Myc)	17869	T39XFZ0GEknZUDXsXY	-0.143	0.0029
Klf4	16600	0khLe85Huv0juQw.sQ	0.576	0.0123
Sox2	20674	ZSdkhcTgadlRcSTtcc	0.394	0.0198

Table 5.2: Probes associated with Oct4, *Sox2*, *Nanog*, *Klf4* and c-Myc that exhibit a significant expression change between *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ cell lines; (LFC > $\log_2(1.1)$, *P*-value < 0.05). Negative LFCs are associated with a reduced expression in *Dgcr8*$^{gt/tm1}$ cells, while positive LFCs are associated with increased expression upon DGCR8 depletion.

In addition, the c-Myc probe seemed to only detect an approximately 10% change in c-Myc transcript levels, while the *Sox2* probe was not significant at the 1% *P*-value cutoff. Despite this, it is intriguing to consider the possibility that some of the phenotypic effects of DGCR8 depletion may reflect the disruption of these core transcription factors. Indeed, a recent study in mouse ES cells noted that the transfection of miR-21 into the cells caused a reduction in cellular self-renewal. Based on computational predictions of miRNA target sites the authors proposed that the miRNA may target *Sox2* and *Nanog* directly (Singh et al., 2008). The Solexa/Illumina sequencing of the miRNA populations in $Dgcr8^{tm1,gt/+}$ cells revealed miR-21 as one of the most highly expressed miRNAs to show a significant reduction in expression upon the depletion of DGCR8 in $Dgcr8^{gt/tm1}$ cell lines. This reduction correlates with the potential increase in *Sox2* expression. If *Sox2* were indeed to be a miR-21 target this may be a route by which DGCR8 reduction can lead to the disruption of this important transcriptional network. However, this remains a hypothesis until further experimental evidence is presented to support the theory.

Three *Klf4* probes all seem to suggest significant changes in *Klf4* levels. TargetScanS (v.4.2) predicts two miR-25 targets in the 3'UTR of *Klf4*. miR-25 is one of the most significantly depleted miRNAs in the $Dgcr8^{gt/tm1}$ cells and if these were to be true miR-25 target sites, miRNAs could be directly manipulating this core transcription factor. It should be noted that *Klf4* wasn't within my miR-25 target-enriched list described later in this chapter (Section 5.3.3.4), so this target relationship again remains speculation.

Oct4, *Sox2*, *Nanog*, *Klf4* and to some extent c-Myc cooperate on their regulation of target genes (Chen et al., 2008; Kim et al., 2008b) and are also involved in the regulation of each other, so it would not be surprising if the disruption of a single member in the network would lead to alterations in the expression levels of the targets of other members of the system. Oct4, SOX2, NANOG and KLF4 bind to the promoters of genes that are either active or repressed in non-differentiated stem cells (Kim et al., 2008b; Liu et al., 2008b). c-Myc on the other hand has a much clearer tendency to bind to the promoters of genes that are active in undifferentiated stem cells (Kim et al., 2008b) and also seems to generally target a distinct set of genes with distinct functions (Kim et al., 2008b; Liu et al., 2008b). The trend for Oct4, SOX2, NANOG and KLF4 targets all to be up regulated in *Dgcr8*$^{gt/tm1}$ cells while c-Myc targets are down regulated is therefore not counterintuitive. In addition the down regulation of c-Myc, upon the depletion of DGCR8, implied by the microarray probes may conceivably account for a proportion of the repression of its target genes.

In order to address the possibility that the deregulation of these TF target gene sets following the depletion of DGCR8 may be due to an enrichment for the targets of ES cell expressed miRNAs amongst these genes (particularly amongst the up regulated gene sets) the genes from each set were mapped to Ensembl transcript IDs and the 3'UTRs of each set (annotated as described in section 2.9.8) were scanned for 7mer seed enrichments with Sylamer (Dr. Stijn van Dongen). No significant 7mer miRNA seed enrichments were identified in the target set. Consequently it is less likely that a broad deregulation of these target sets is due to a depletion of direct miRNA mediated target degradation and is more likely the result of changes at the level of transcription, although miRNA regulation of these sets not detectable by this method can not be ruled out.

5.3.3 Reintroducing miRNA mimics to *Dgcr8*$^{gt1/tm1}$ cells

In order to generate miRNA target lists I planned to reintroduce miRNA mimics into *Dgcr8* deficient ES cells and then interrogate the cells' expression profiles for miRNA dependent expression changes, with mRNA expression arrays. These ES cells do not express endogenous miRNAs and would not be saturated for miRNA mediated target regulation. They would also be devoid of any inter-miRNA redundancy caused by shared seed sequences and should be depleted of endogenous combinatorial miRNA mediated target regulation. This should provide a simplified system for understanding the roles of individual miRNAs. For these reasons I hoped that this system would allow me to generate accurate miRNA candidate target lists with a high degree of sensitivity.

5.3.3.1 Optimisation of the conditions for miRNA reintroduction

I decided to use the *Dgcr8*$^{gt1/tm1}$ cell line as the basis for the miRNA reintroduction experiments. In order to reintroduce miRNAs into the *Dgcr8*$^{gt1/tm1}$ cells they would have to be transfected as mature miRNA duplexes as it was difficult to identify commercially available miRNA mimics that could enter the miRNA processing pathway at any other stage and that could be guaranteed not to require a functional microprocessor in order to be processed to a mature form. To optimize the transfection conditions (Lipofectamine 2000 quantity, siRNA concentration and cell number) for the reintroduction of these miRNAs into the mutant ES cells, siRNAs were used that would have identifiable cellular effects when transfected successfully.

Initially the conditions for siRNA transfection were adapted from the experimental procedures presented in two papers for ES cell siRNA transfection and the recommended

Invitrogen protocol (http://tools.invitrogen.com/content/sfs/manuals/stealth_sirna_tsf_lf2k_man.pdf) (Chen et al., 2007; Takahashi and Yamanaka, 2006). These conditions were used to transfect *Dgcr8*$^{gt1/tm1}$ cells at various cellular densities and with various quantities of siGLO conjugated siRNAs and Lipofectamine 2000 (Section 2.12) (Data not shown). However, although it was apparent that the cells were successfully being transfected under these conditions it was very difficult to quantitate transfection efficiency as there was a relatively high background in each well of what appeared to be untransfected siGLO associated lipid complexes. The intense and punctate siGLO staining also made it more difficult to discern which cells, amongst those in groups, were successfully transfected and which just lay in close proximity to those transfected or an untransfected lipid complex.

Subsequently I altered the method of optimization. As an alternative I transfected the cells with an siRNA targeting *Kif11*, a gene encoding a motor protein. KIF11 depletion causes growth arrest and triggers apoptosis (Ambion). The effect on cell number and viability was measured through the use of an Alamar blue redox assay. This method was originally provided by Dr. Ian Sudbery and had been optimized for working with HeLa cells. Subsequently alterations were made to suit my purposes. In principle metabolically active cells cause a reduction of the Alamar Blue indictor resulting in a colour and fluorescence change. This change is easily quantifiable (Section 2.12) (Data not shown). However, after initial experiments, an alternative method was used as it became clear that this method would require further optimization if it were to be used to effectively quantitate the effectiveness of siRNA transfection, as there was an apparently non-linear relationship between the number of untransfected cells plated in a well and the subsequent Alamar Blue associated fluorescence reading. It was unclear if this non-linear relationship was caused by differential growth rates

at different cell densities or non-linear reduction of the Alamar blue as the cell numbers increased. In addition, due to the nature of mouse ES cells, with this cellular assay it is difficult to know exactly what is the cause of growth changes, whether it is cell cycle arrest or triggered differentiation.

Ultimately, a LacZ siRNA, which would target the β-geo fusion transcripts derived from the gene-trapped allele, was used to further optimise the transfection method. Successful transfection of this siRNA could be visualised by fixing and Xgal staining the cells (Section 2.11.1.2). Successfully transfected cells exhibited reduced blue staining (Fig.5.6) when compared to non-transfected cells or cells transfected with a control siRNA. It became clear from a time course experiment (Data not shown) that these siRNA transfections were very fast acting with clear differences in LacZ activity apparent ~21 hours after the transfection was begun. In addition, this method also revealed the importance of ensuring the cells are spread as evenly as possible across the well prior to transfection. Plating the cells and co-transfecting them caused cells to aggregate at the centre of the well due to the small volume of transfection medium. The aggregated cells seemed to be transfected far less efficiently than the more sparse populations. Subsequently it was necessary to plate the ES cells 3 hours prior to transfection. This allowed the cells time to settle and prevented them from aggregating at the centre of the well during the transfection.

Using this method transfection efficiency could be estimated by comparing the number of white to blue cells in representative photographs of the LacZ siRNA transfected population. An optimization of the cell numbers with this new method ($10x10^4$ cells to $40x10^4$ per 24 well) resulted in 'blue/total cell-number' ratios varying from ~20-36%.

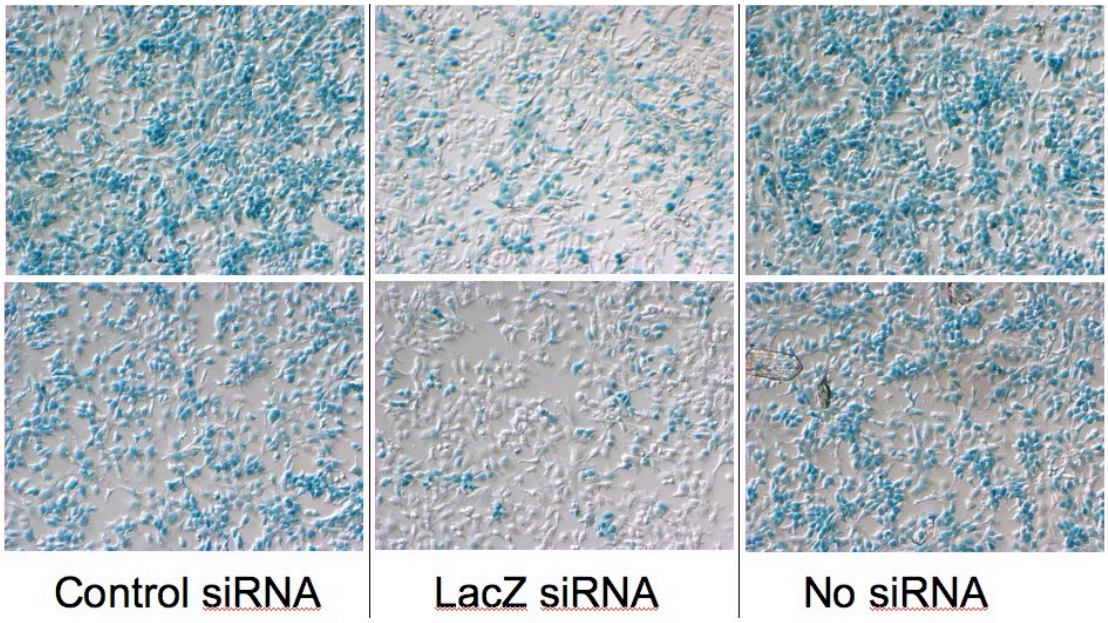


Fig.5.6: Xgal staining results of LacZ siRNA transfected *Dgcr8*$^{gt1/tm1}$ cells in the 6 well format to be used for miRNA mimic reintroduction experiments. Photos demonstrating the depletion of β-geo activity in cells transfected with a LacZ targeting siRNA. Images taken following Xgal staining which stains LacZ expressing cells blue. Taken at 10x relief contrast. Also depicted are cells transfected with a control siRNA or No siRNA. In 5 photos taken of the LacZ siRNA transfected cells, the proportion of blue cells to total cells varies between approximately 23% to 48% with an average ~34% staining positive for LacZ activity. By contrast, in 4 pictures of the control siRNA transfected cells, between 95-97% of the cells stained blue.

Finally, the method was then scaled up in order to transfect 6 well plates rather than 24 well plates to ensure sufficient RNA for the array analyses of the subsequent miRNA-transfected cells (for optimised method see section 2.11), (Fig.5.6).

5.3.3.2 Demonstration of Oct4 expression in transfected cells

In order to ensure that the transfection of the *Dgcr8*$^{gt1/tm1}$ cells with short RNAs wasn't causing the depletion of the ES cell marker expressing cells in the culture (either through cell death or by triggering differentiation) I tested the expression of Oct4, a factor essential in order to maintain the ES cells pluripotent state (Niwa et al., 2000), in transfected cells. I repeated the LacZ siRNA transfection in a chamber slide format, scaling the reagents to suit

the different size culture. Alongside cells transfected with the LacZ siRNA, I transfected an equal number of wells with the control siRNA and cultured four wells without transfection as negative controls. Slides were stained with Oct4 and LacZ primary antibodies, which were bound by Alexa 594 and 488 secondary antibodies respectively. 2 LacZ siRNA and 2 Control siRNA transfected wells were stained with these antibodies to replicate results. Simultaneously non-transfected cells were stained with combinations of these antibodies as controls. As a further control *Dgcr8*$^{tm1,gt1/+}$ cells (grown at a different density and under different conditions) were also stained with either both primary and secondary antibodies or simply secondary antibodies. *Dgcr8*$^{tm1,gt1/+}$ cells do not exhibit LacZ activity (Fig.3.7) and so would act as negative LacZ controls. Again it is important to bear in mind that the Alexa 488 and EGFP expressed from the targeted trap could potentially interfere with each other's signal. However after the cell staining experiments in Chapter 3 and further fluorescence microscopy experiments (data not shown), the EGFP appears to be expressed at very low levels in these cells and was not expected to interfere with the results.

All of the staining-control cells gave the expected background fluorescence for each of the antibodies used. Although there is diffuse fluorescence it is at a far lower level than that seen in the wells with both primary and secondary antibodies included in the staining. Only *Dgcr8*$^{gt1/tm1}$ cells known to express LacZ (β-geo) fusion protein gave a clear, nuclear-staining Alexa 488 signal when stained with both primary and secondary antibodies (Fig.5.7). This coincides with the nuclear localization of the LacZ activity seen in the Xgal staining experiments in Chapter 3. However, it is clear that the *Dgcr8*$^{tm1,gt1/+}$ cells also fluoresce when probed with the LacZ antibody, although this is a cytoplasmic stain rather than a nuclear stain. This does not coincide with Xgal staining activity seen in Chapter 3 for this cell line.

This stain is not seen if only the secondary antibodies are included in the staining procedure, so seems to be a non-specific binding caused by the primary antibody rather than EGFP expressed from the targeted trap cassette interfering with the interpretation of the results.

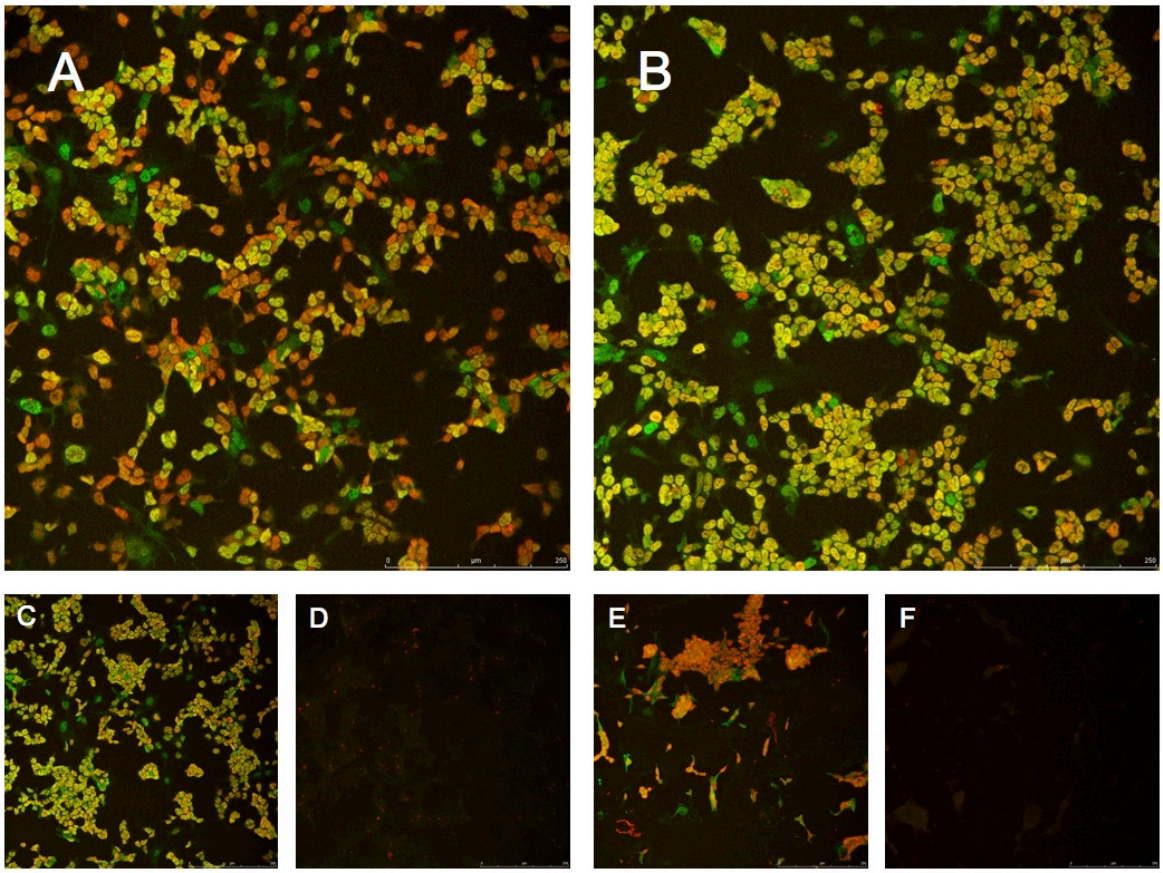


Fig.5.7: Immuno-staining of transfected *Dgcr8*$^{gt1/tm1}$ cells with LacZ and Oct4 specific antibodies to demonstrate siRNA transfection does not influence Oct4 expression. LacZ stained green. Oct4 stained red. A) *Dgcr8*$^{gt1/tm1}$ cells transfected with LacZ specific siRNA. B) *Dgcr8*$^{gt1/tm1}$ cells transfected with control siRNA. C) Non-transfected *Dgcr8*$^{gt1/tm1}$ cells. D) Non-transfected *Dgcr8*$^{gt1/tm1}$ cells with secondary antibodies only E) Non-transfected *Dgcr8*$^{tm1,gt1/+}$ cells F) Non-transfected *Dgcr8*$^{tm1,gt1/+}$ cells with secondary antibodies only. A,B,C and E – 10 sections through stained cells – Maximum projection. D and F – 5 sections through visible fluorescence – Maximum projection.

Transfection with the siRNA targeting the LacZ (β-geo) fusion transcript causes a subsequent reduction in nuclear 488 staining in a large proportion of the *Dgcr8*$^{gt1/tm1}$ cells when

compared to those cells subjected to the control transfection, which is expected (Fig 5.7A). However, there is still clear Oct4, 594 staining in the affected cells. I conclude that the successful transfection of pluripotent ES cells expressing marker proteins does not cause subsequent differentiation in the timeframe within which the cells are likely to be lysed and they are not depleted through cell death caused by lipid transfection.

5.3.3.3 Selecting miRNA mimics to reintroduce into *Dgcr8*$^{gt1/tm1}$ cells

mmu-miR-291a-3p and mmu-miR-25 miRNA mimics were selected as the two miRNA-mimics I would transfect to optimize and prove the system for the identification of miRNA targets. Mmu-miR-291a is a member of the miR-290 miRNA cluster. The functional study of this cluster is popular in the field. It is known to be highly expressed in mouse ES cells (Fig.4.8) and contains miRNAs with seed sequences common among ES cell expressed miRNAs (Fig.5.4C) (Calabrese et al., 2007). As a consequence of these studies miR-291a-3p has been associated with a number of target genes. These can therefore be used as positive controls for my target selection. Functional studies of the role of mmu-miR-25 are scarcer in ES cells. This makes the generation of mmu-miR-25 target lists an interesting prospect. Mmu-miR-25 was a miRNA highly expressed in the control cell lines as judged by Solexa sequencing (Fig.4.8) and exhibited a large and significant expression change when both alleles of *Dgcr8* are disrupted, identified when *Dgcr8*$^{tm1,gt/+}$ miRNA expression was compared to that of *Dgcr8*$^{gt/tm1}$ cells. The miR-291a-3p and miR-25 seed sequences were also among the most enriched seed sequences in the 3'UTRs of genes up regulated in *Dgcr8*$^{gt/tm1}$ cells when compared to *Dgcr8*$^{tm1,gt/+}$ cells (Fig.5.4B) adding to the evidence that these are important miRNAs in ES cells.

5.3.3.4 Transfection time course to identify the optimal time point for cell lysis

In order to identify mRNAs specifically degraded as a consequence of the presence of miRNAs, miRNA mimics would be transfected in parallel to control mimics (Dharmacon). The changes in mRNA expression resulting from the miRNA transfection would then be judged by a comparison to the mRNAs expression profile of the control transfected cells as estimated by Illumina expression microarray.

To determine the optimal period post-miRNA transfection for which to leave the cells prior to RNA lysis, I conducted a time series, lysing the cells 10 hours, 20 hours and 44 hours after the miRNA transfection was begun. It seemed reasonable to hypothesise that this period should be kept to a minimum since ES cells transfected with siRNAs seem to react rapidly to the transfection (Takada et al., 2005b). Assuming that a system devoid of competing small RNAs may respond still more rapidly, the response may be very quick and fairly brief if the siRNAs are cleared with little competition. It is also reasonable to assume that the shorter the delay between transfection and lysis the more likely that the genes that are affected are primary miRNA targets rather than downstream effects. However, it would be necessary to leave the cells for a sufficient time to allow the delivered miRNAs to enter the system and affect mRNA levels.

Both miR-25 and miR-291a-3p mimics were transfected as part of this time series. All transfections were duplicated (although both 10 hour replicates were conducted upon the same day, while the 20 hour and 44 hour replicates were conducted on two separate

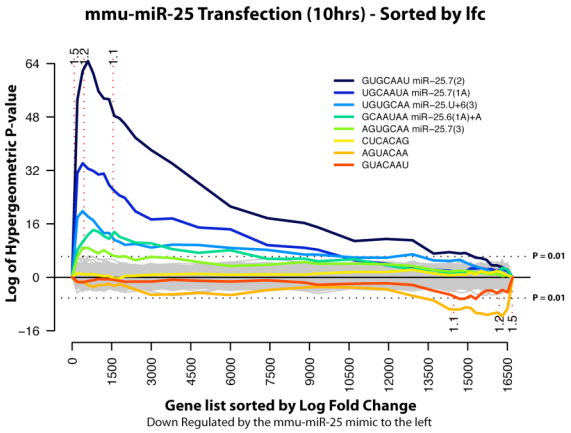
occasions). As described above, for each replicate a set of cells was also transfected with a control miRNA to control the derived expression data for the effects of the transfection itself.

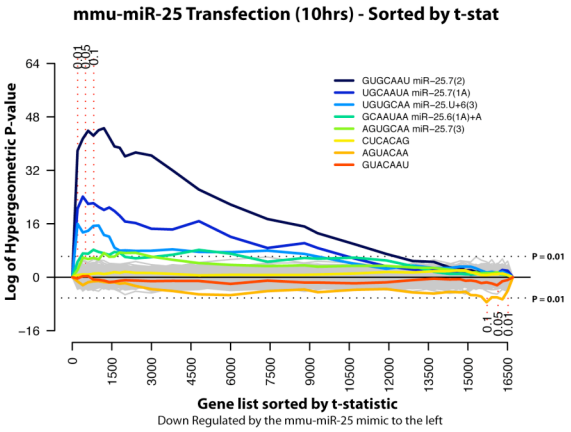
At each time point the Illumina mRNA expression microarray expression profiles for each miRNA were compared to the microarray expression profiles for the control miRNA. Once again, following this comparison, for each miRNA and time point the probes from the array were ordered according to their LFC and their associated t-statistic. Sylamer was used to examine these lists for the enrichment of miRNA seed sequences (Fig.5.8) (Dr. Cei Abreu-Goodger) in mRNAs that were up or down regulated. As expected the most significant seed sequences, seen amongst the genes down regulated by either miR-25 or miR-291a-3p relative to the control, were those for the miRNA that was transfected in each case. These enrichments were particularly significant at the 10 hour time point and subsequently became less significant as time increased. This loss of seed of enrichment amongst the down-regulated genes is particularly obvious in the case of miR291a-3p. At 44 hours the down-regulated genes do not appear to be particularly enriched for the miR-291a-3p seeds. This is despite the fact that at 44 hours there are approximately the same number of probes demonstrating a log fold change of at least '1.1' as was seen at 10 hours (Fig.5.8). It follows that the loss of enrichment is not entirely caused by a return to normal/control expression levels, but may perhaps be compounded by changes in the expression of indirect targets, that do not contain miRNA seed sequences, which dilute the enrichment signal.

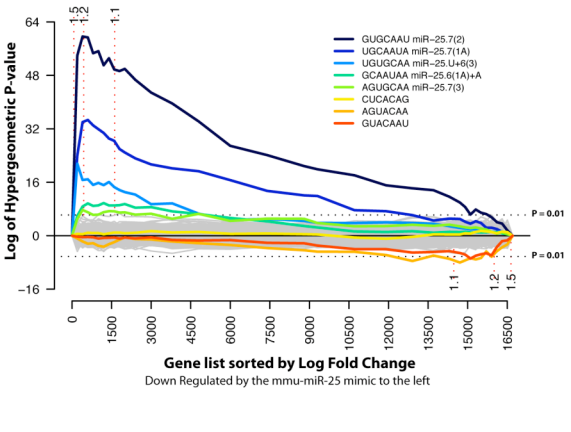
At the other end of the x-axis there are also enriched seed signals. Unsurprisingly these seed sequences correspond to the first 8 bases of the control miRNA used (Based on cel-miR-239 - UUGUACUACACAAAAGUACUG) or the 6mer with an extra U at the 3' end of the target

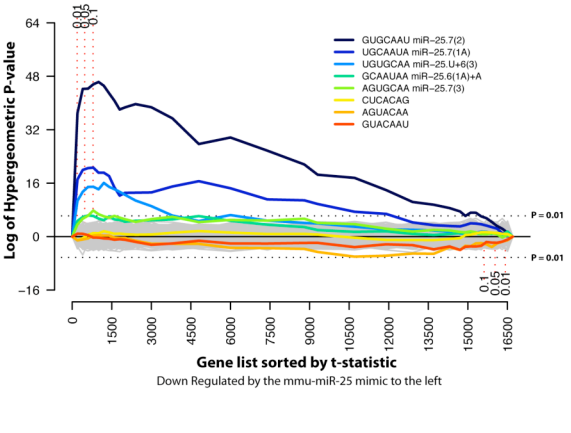
site. It is worth noting that the current miRBase annotation and the Dharmacon annotation for this miRNA differ, although as the miRNA negative controls are selected for having the minimal sequence similarity with other mouse miRNAs this is not important. Although the control mimic does appear to be affecting the expression of certain genes the enrichment signal is far weaker than that seen in the case of the transfected endogenous miRNA. This is probably because the targets and *C. elegans* miRNA mimics will not have co-evolved to optimize target selection, so that the signal will be due to random off target effects.

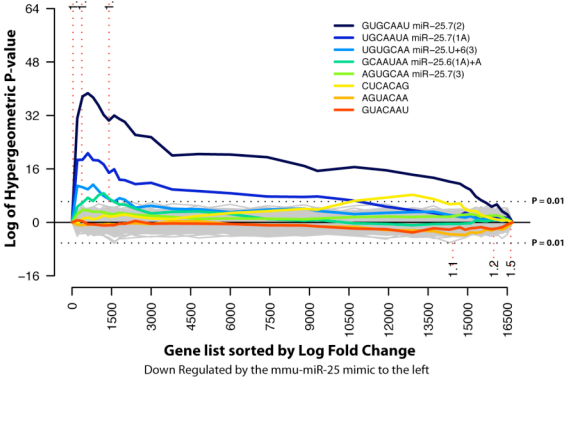
A

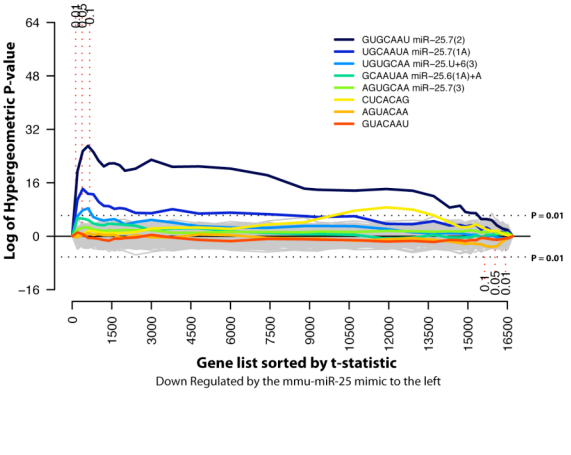

mmu-miR-25 Transfection (10hrs) - Sorted by lfc
Log of Hypergeometric P-value
GUGCAAU miR−25.7(2)
UGCAAUA miR−25.7(1A)
UGUGCAA miR−25.U+6(3)
GCAAUAA miR−25.6(1A)+A
AGUGCAA miR−25.7(3)
CUCACAG
AGUACAA
GUACAAU
P = 0.01
Gene list sorted by Log Fold Change
Down Regulated by the mmu-miR-25 mimic to the left


mmu-miR-25 Transfection (10hrs) - Sorted by t-stat
Log of Hypergeometric P-value
GUGCAAU miR−25.7(2)
UGCAAUA miR−25.7(1A)
UGUGCAA miR−25.U+6(3)
GCAAUAA miR−25.6(1A)+A
AGUGCAA miR−25.7(3)
CUCACAG
AGUACAA
GUACAAU
P = 0.01
Gene list sorted by t-statistic
Down Regulated by the mmu-miR-25 mimic to the left


mmu-miR-25 Transfection (20hrs) - Sorted by lfc
Log of Hypergeometric P-value
GUGCAAU miR−25.7(2)
UGCAAUA miR−25.7(1A)
UGUGCAA miR−25.U+6(3)
GCAAUAA miR−25.6(1A)+A
AGUGCAA miR−25.7(3)
CUCACAG
AGUACAA
GUACAAU
P = 0.01
Gene list sorted by Log Fold Change
Down Regulated by the mmu-miR-25 mimic to the left


mmu-miR-25 Transfection (20hrs) - Sorted by t-stat
Log of Hypergeometric P-value
GUGCAAU miR−25.7(2)
UGCAAUA miR−25.7(1A)
UGUGCAA miR−25.U+6(3)
GCAAUAA miR−25.6(1A)+A
AGUGCAA miR−25.7(3)
CUCACAG
AGUACAA
GUACAAU
P = 0.01
Gene list sorted by t-statistic
Down Regulated by the mmu-miR-25 mimic to the left


mmu-miR-25 Transfection (44hrs) - Sorted by lfc
Log of Hypergeometric P-value
GUGCAAU miR−25.7(2)
UGCAAUA miR−25.7(1A)
UGUGCAA miR−25.U+6(3)
GCAAUAA miR−25.6(1A)+A
AGUGCAA miR−25.7(3)
CUCACAG
AGUACAA
GUACAAU
P = 0.01
Gene list sorted by Log Fold Change
Down Regulated by the mmu-miR-25 mimic to the left


mmu-miR-25 Transfection (44hrs) - Sorted by t-stat
Log of Hypergeometric P-value
GUGCAAU miR−25.7(2)
UGCAAUA miR−25.7(1A)
UGUGCAA miR−25.U+6(3)
GCAAUAA miR−25.6(1A)+A
AGUGCAA miR−25.7(3)
CUCACAG
AGUACAA
GUACAAU
P = 0.01
Gene list sorted by t-statistic
Down Regulated by the mmu-miR-25 mimic to the left

B

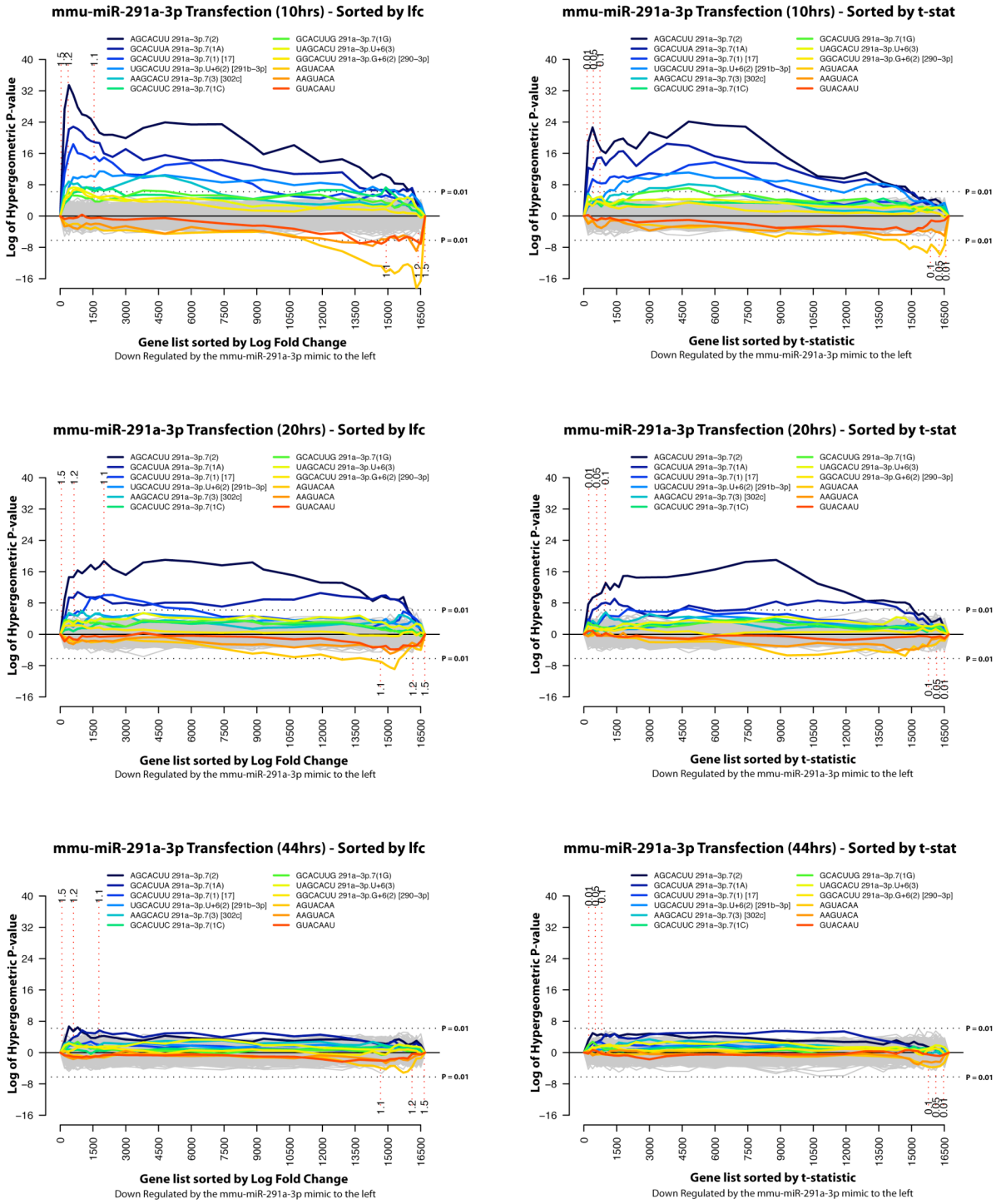


Fig.5.8: Sylamer plots querying gene lists ordered according to LFC and t-statistic following the transfection of DGCR8 depleted cells (*Dgcr8*$^{gt1/tm1}$) with miRNA mimics and lysed according to a time series. Two time series were conducted to find the optimal time point post-transfection at which to lyse cells and purify their RNA. Cells were transfected with mmu-miR-25 mimic (A) or mmu-miR-291a-3p mimic (B) and lysed after 10 hours, 20 hours or 44 hours. Subsequently differential gene expression was determined

relative to cells transfected with a control miRNA mimic. This differential expression was used to order the gene list according to LFC or t-statistic and Sylamer was used to query these gene lists for the enrichment of any 7mer sequences in any regions of these gene lists (down-regulated by miRNA mimic relative to control to the left of axis, up regulated to the right). See Fig.5.2 for the principle of Sylamer analysis and Sylamer plots. Significantly enriched 7mer seeds are represented by a coloured line on the plots. Sylamer analysis of these gene lists was conducted by Dr. Cei Abreu-Goodger.

5.3.3.5 miR-25 and miR-291a-3p potential target lists

I used the comparisons between the expression profiles of miRNA transfected cells and control transfected cells at the 10 hour time point as the basis for identifying relevant targets for miR-25 and miR-291a-3p. Working on the principle that the enrichment of miRNA seed sequences among those genes down regulated by the miRNA transfection was a consequence of the miRNA mimic causing the degradation of miRNA sensitive targets, I tried to select *P*-value and LFC cut offs for these comparisons that would provide a gene set most enriched for these down regulated seed bearing genes in each case. To do this I referred to the Sylamer plots of the transfection time course and selected *P*-value and LFC limits that encompassed the gene set with the greatest *P*-value for the enrichment of miR-25 or miR-291a-3p seed sequences. At the same time I bore in mind the importance of maintaining stringent criteria for the identification of significantly affected targets. As a result I selected probes that were down regulated in the presence of miRNA mimics as opposed to a control mimic transfection with a *P*-value cut off of 0.1 and a LFC cut off of $\log_2(1.2)$. This identified 83 probes demonstrating down-regulation in the case of mu-miR-25 and 31 probes in the case of mmu-miR-291a-3p.

These probes that reacted to the presence of each miRNA were further filtered by intersection with those probes that were identified as up regulated in the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ vs.

Dgcr8$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ expression analysis (see section 5.3.2.1 above). This requirement for the probes to be in both sets in order to be considered as targets will tend to remove any probes that may have changed due to the over expression of the transfected miRNA and may not be relevant under normal cellular conditions (as they remain relatively unchanged when miRNAs are depleted in the stem cell system). On the other hand, the transfection experiment allows the probes that change upon a broad miRNA depletion to be subsetted into those that are controlled by each stem cell expressed miRNA. Of the original 83 and 31 probes identifying down regulated genes upon miRNA transfection, 47 miR-25 affected probes and 28 miR-291a-3p probes remained after this filtering step. These correspond to 40 independent genes with annotated Entrez IDs in the case of miR-25 and 25 genes with annotated Entrez IDs in the case of miR-291a-3p. These probes and their associated annotation were amalgamated into HTML tables (Table 5.3/Appendix C (CD)) in order to aid the interpretation of these results. Some of the probes did not correspond to an annotated gene and transcript, but the probe ID is included for completeness.

Where possible the 3'UTRs of annotated transcripts associated with each of the microarray probes within the set of predicted targets was searched for the presence of target seed sequences that may account for the targeted degradation of the transcript by the transfected miRNA . Of the miR-25 potential targets, 35 out of 38 genes with annotated Havana or Ensembl transcripts and all but ENSMUST00000057037 of the miR-291a-3p annotated target transcripts contain at least one of the respective miRNA target seed 6mers within its 3'UTR. ENSMUST00000057037 itself has no associated 3'UTR and thus seed counts could not be conducted. The presence of seed sequences that may identify active miRNA target

sites reinforces the suggestion that these genes may be primary miRNA targets rather than secondary effects of miRNA mimic transfection.

There could be several reasons why the *49331047G18Rik*, *Gli2* and *Rgma* transcripts are within the miR-25 predicted target set and yet do not appear to have a seed sequence within their 3'UTR. Although both *Gli2* and *49331047G18Rik* only possess a single Ensembl annotated transcript there may be additional non-annotated alternative splices, with alternative 3'UTRs that may contain miRNA target sites not found here. *Rgma* has four transcripts annotated within Ensembl (v52). Differences in the 3'UTRs annotated for these transcripts will obviously affect seed sequences identified within these 3'UTRs. Seeds within alternative 3'UTRs could account for miRNA induced expression changes. Alternatively there may be miRNA targets within the ORFs that may be susceptible to miRNA targeting. Indeed a simple text search of the coding sequences for the transcripts for these genes annotated in Table 5.3 identified 4 GTGCAA seed sequences within the coding region of *Rgma* and 2 independent 6mer seeds (GTGCAA and GCAATA) in the *Gli2* ORF. A further possibility could be that these genes are not in fact direct targets of the miR-25 miRNA and are secondary targets affected downstream of other targets within or outside the list.

In contrast to the pervasive presence of seed sequences in the 3'UTRs of genes in these sets, TargetScan and the miRBase Targets version of the miRanda algorithm were not comprehensive in their prediction of these genes as targets of these miRNAs (Table 5.3). The databases were queried through the target genes' EntrezIDs. In the case of miRBase targets these were mapped via Ensembl IDs using the database's own tables (Dr. Stijn van Dongen). Of the probes with associated Entrez ID annotation, TargetScanS v4.2 identified 15/40 of the

miR-25 target gene set as miR-25 targets and 19/25 of the miR-291a-3p genes as targets. MiRBase Targets identified 9/40 of the miR-25 targets and 2/25 of the miR-291a-3p targets. Obviously the fact that these methods do not identify a significant proportion of my candidate targets is a cause for concern if these computational predictions are to be used as a basis for hypothesising further experimental investigation. As a consequence, these lists of candidate targets are a very useful supplement to these predictive methods. In addition, the list is a refinement on the broad and inclusive target lists derived from these computational predictions, which do not consider co-expression as a necessity for target prediction. Although it could be useful to use these candidate lists in parallel to produce lists of the most likely candidates predicted by all of the methods, before planning further research, it is also worth bearing in mind that the lists provided here are based on experimental data rather than computational prediction and as such should be more reliable when considering the role of miRNAs in ES cells.

Table 5.3: Tables summarizing the genes that are the suspected targets of mmu-miR-25 and mmu-miR-291a-3p as determined by transfection analysis. Information concerning each of the genes that was identified as a predicted target of mmu-miR-25 or mmu-miR-291a was summarised within HTML tables (See below and Appendix C (CD)). These tables contain the probe IDs that successfully fulfilled the criteria required to be associated with a miRNA target. The "Gene Symbols", "Gene Name" and "Entrez ID" for each probe were annotated according to the *lumiMouseAll* Bioconductor library. However, where these annotations were not available, "Gene Symbols" and "Entrez ID" were annotated according to the annotations of Dr. Cei Abreu-Goodger (derived from Ensembl, Nov 2008) and marked with an asterisk. Also included are the LFCs and *P*-values associated with each array experiment used to define the set, associated GO terms and KEGG pathway information, relevant miRNA seed counts within the 3'UTRs of associated transcripts and TargetScan (v4.2) and miRBase Target predictions for each gene for the respective miRNA.

Mmu-miR-25 Probable MiRNA Targets

Illumina Probe IDs	Gene Symbols	Gene Name	Entrez ID	Log Fold Change upon Dgcr8 Depletion	P-value of change upon Dgcr8 Depletion	Log Fold Change upon miR-25 Reintroduction	P-value of change upon miR-25 Reintroduction	GO Terms	KEGG Pathways	Transcript used for 3'UTR seed counts	6mer(1A)-GCAATA	6mer(3)-GTGCAA	6mer(2)-TGCAAT	7mer(2)(m8)-GTGCAAT	7mer(1A)-TGCAATA	8mer(1A)-GTGCAATA	Target Scan Conserved Target Context Score	Target Scan Non-conserved Target Context Score	miRBase Targets P-value
iXuHlOb1FlCWdz.rd4	Tmem184b	transmembrane protein 184b	223693	1.23	1.0e-18	0.43	6.4e-05			ENSMUST00000074991	1	1	1	1	1	1	-0.1188	Not Found	Not Found
Bp5e4GiPe018Rk3.L4	Adam23	a disintegrin and metallopeptidase domain 23	23792	1.26	4.2e-16	0.71	1.2e-05	cell adhesion(GO:0007155) | proteolysis(GO:0006508) | integrin-mediated signaling pathway(GO:0007229)		OTTMUST00000004901	2	2	5	2	2	1	-0.3949	Not Found	Not Found
lkuSKIB557us.i4.5I	Plekhm1	pleckstrin homology domain containing, family M (with RUN domain) member 1	353047	0.94	5.5e-16	0.98	3.2e-05	intracellular signaling cascade(GO:0007242)		OTTMUST00000005762	2	2	2	1	2	1	-0.1376	-0.3862	Not Found
Ej5ewT_J98eO7nrqEo	Whsc1l1	Wolf-Hirschhorn syndrome candidate 1-like 1 (human)	234135	0.66	3.9e-15	0.56	2.0e-05	transcription(GO:0006350) | regulation of transcription, DNA-dependent(GO:0006355) | chromatin modification(GO:0016568)		ENSMUST00000084026	1	3	3	1	1	0	Not Found	Not Found	Not Found
0Xul8aOr_34e.kHjnk uLp8e6gnXujLOheU7o	Rab8b	RAB8B, member RAS oncogene family	235442	0.61	4.7e-15	0.89	7.3e-08	transport(GO:0006810) | small GTPase mediated signal transduction(GO:0007264) | protein transport(GO:0015031)		ENSMUST00000041139	1	2	1	1	1	1	Not Found	-0.342	Not Found
oKJsk.kySo_5ezce7U	Gcnt2	glucosaminyl (N-acetyl) transferase 2, I-branching enzyme	14538	0.76	1.1e-13	0.45	1.2e-04		Glycosphingolipid biosynthesis - neo-lactoseries(00602) | Glycan structures - biosynthesis 2(01031)	ENSMUST00000110191	1	0	2	0	1	0	Not Found	-0.046	Not Found
3uKIKTKCuIcniH3tOk	Bicd2	bicaudal D homolog 2 (Drosophila)	76895	0.63	9.8e-12	0.39	1.6e-04	microtubule-based movement(GO:0007018)		ENSMUST00000048544	0	1	2	0	0	0	Not Found	Not Found	Not Found
EeMSH_Qure_2ep4910	4931407G18Rik	RIKEN cDNA 4931407G18 gene	70977	0.75	9.2e-11	0.57	3.3e-05			OTTMUST00000052375	0	0	0	0	0	0	Not Found	Not Found	Not Found
um2NuXFnc0Od13BQqg				0.55	1.2e-10	0.77	6.5e-06												
ldWImLHEDmUk16IL30	Pcaf *		18519*	0.55	1.5e-10	0.48	1.8e-04			ENSMUST00000000724	2	1	1	1	1	1	-0.5192	Not Found	Not Found
cuzncl0lPHj0yUv0Xw	Ccnc	cyclin C	51813	0.40	1.6e-10	0.93	1.2e-09	transcription(GO:0006350) | regulation of transcription, DNA-dependent(GO:0006355)		OTTMUST00000010608	1	2	1	1	1	1	-0.4342	Not Found	Not Found
c0rHdgr7fnpXuuvlDo	Plekhg3	pleckstrin homology domain containing, family G (with RhoGef domain) member 3	263406	0.62	1.0e-09	0.48	1.1e-04	intracellular signaling cascade(GO:0007242) | regulation of Rho protein signal transduction(GO:0035023)		ENSMUST00000101279	1	0	0	0	0	0	Not Found	Not Found	Not Found
Hg.3.j_oj70SgstdV0	Adam19 *		11492*	0.65	1.6e-09	0.58	1.2e-04			OTTMUST00000012216	3	1	3	1	3	1	-0.3185	-0.0716	Not Found
WfXy3tUW5N58XTpnv4	Rnf38	ring finger protein 38	73469	0.62	2.2e-09	0.94	3.2e-05			OTTMUST00000015974	2	3	3	2	1	1	-0.3052	Not Found	Not Found
0Xgh4_noK_vkoCFolU	Acbd4	acyl-Coenzyme A binding domain containing 4	67131	0.43	3.9e-09	0.46	6.7e-05	biological_process(GO:0008150)		OTTMUST00000006883	0	1	1	1	0	0	Not Found	-0.2064	0.0318833
KmNSu8OCCeiikq5ag	4930519N13Rik	RIKEN cDNA 4930519N13 gene	78177	0.45	4.8e-09	0.40	1.2e-04			OTTMUST00000037530	0	1	1	1	0	0	Not Found	-0.2048	0.00937581
ZZw4u1Ikd1fin5rUKI	Leprot	leptin receptor overlapping transcript	230514	0.47	4.9e-09	0.87	9.3e-07			OTTMUST00000019180	0	2	1	1	0	0	Not Found	Not Found	0.0246051
ohJKUK40unAvgCq_v0	Ric8b	resistance to inhibitors of cholinesterase 8 homolog B (C. elegans)	237422	0.47	6.7e-09	0.70	1.7e-05			ENSMUST00000038523	0	1	1	1	0	0	Not Found	-0.1647	Not Found

W3K5fkivgpSUZR59RQ				0.38	7.4e-09	0.51	1.2e-05													
E8S57gidW_K_urHtf4imb8_k2n0aNqiOX2H8	Necap1	NECAP endocytosis associated 1	67602	0.67	1.1e-08	1.27	6.4e-07	transport(GO:0006810) | endocytosis(GO:0006897) | protein transport(GO:0015031)		ENSMUST00000032477	1	1	1	1	1	1	-0.4297	Not Found	Not Found
9nJyRQvr4SPv_45_V4	Wwp2	WW domain containing E3 ubiquitin protein ligase 2	66894	0.31	1.5e-08	0.62	2.0e-04	protein modification process(GO:0006464)	Ubiquitin mediated proteolysis(04120) | Dentatorubropallidoluysian atrophy (DRPLA)(05050)								-0.4071	Not Found	Not Found
6B6Xp7uTy0O1T6.53s	D16Ertd472e	DNA segment, Chr 16, ERATO Doi 472, expressed	67102	0.46	8.0e-08	0.70	4.5e-06			ENSMUST00000023571	2	1	0	0	0	0	Not Found	Not Found	Not Found
oronqodncvXoVAxvxE	Lmbr1l	limb region 1 like	74775	0.29	2.7e-07	0.75	2.3e-06	endocytosis(GO:0006897)		OTTMUST00000058993	2	1	1	1	1	1	-0.3165	Not Found	0.00104278
lW8eG7B8kk8i4bfS_s	Snap29	synaptosomal-associated protein 29	67474	0.49	2.9e-07	0.66	1.1e-04	transport(GO:0006810) | protein transport(GO:0015031)	SNARE interactions in vesicular transport(04130)	OTTMUST00000057883	1	1	1	1	0	0	Not Found	Not Found	Not Found
HQhFkIdI9XSOFUmVf4cSdHs4EaE_eHrDgud0	Gnaq	guanine nucleotide binding protein, alpha q polypeptide	14682	0.33	3.2e-07	0.98	2.6e-06	skeletal development(GO:0001501) | regulation of action potential(GO:0001508) | signal transduction(GO:0007165) | protein amino acid ADP-ribosylation(GO:0006471) | G-protein coupled receptor protein signaling pathway(GO:0007186) | G-protein signaling, adenylate cyclase activating pathway(GO:0007189) | glutamate signaling pathway(GO:0007215) | heart development(GO:0007507) | behavior(GO:0007610) | post-embryonic development(GO:0009791) | neuron remodeling(GO:0016322) | forebrain neuron development(GO:0021884) | embryonic digit morphogenesis(GO:0042733) | regulation of melanocyte differentiation(GO:0045634) | pigmentation during development(GO:0048066) | dopamine receptor, phospholipase C activating pathway(GO:0060158)	Calcium signaling pathway(04020) | Gap junction(04540) | Long-term potentiation(04720) | Long-term depression(04730) | GnRH signaling pathway(04912) | Melanogenesis(04916)	ENSMUST00000025541	0	2	2	2	0	0	Not Found	Not Found	Not Found
0uTv0tSkLxPa_7kMgo	Osbpl2	oxysterol binding protein-like 2	228983	0.37	4.3e-07	0.64	4.3e-06	transport(GO:0006810) | lipid transport(GO:0006869) | steroid metabolic process(GO:0008202)		OTTMUST00000038983	1	1	1	1	1	1	Not Found	-0.2365	0.0127902
Bh9F3khPV5KzHUS_UA	Brms1l *		52592*	0.35	9.5e-07	0.58	4.6e-05			ENSMUST00000012220	0	1	1	1	0	0	-0.1796	Not Found	Not Found
								negative regulation of transcription from RNA polymerase II promoter(GO:0000122) | skeletal development(GO:0001501) | kidney development(GO:0001822) | osteoblast development(GO:0002076) | pattern specification process(GO:0007389) | axon guidance(GO:0007411) | ventral midline development(GO:0007418) | hindgut morphogenesis(GO:0007442) | heart development(GO:0007507) | positive regulation of cell proliferation(GO:0008284) | dorsal/ventral pattern formation(GO:0009953) | proximal/distal pattern										

lenfuu6HT5eEyFcn1Q	Gli2	GLI-Kruppel family member GLI2	14633	0.68	1.5e-06	0.47	6.3e-05	formation(GO:0009954) | floor plate formation(GO:0021508) | spinal cord dorsal/ventral patterning(GO:0021513) | ventral spinal cord development(GO:0021517) | cerebellar cortex morphogenesis(GO:0021696) | smoothened signaling pathway involved in ventral spinal cord interneuron specification(GO:0021775) | smoothened signaling pathway involved in spinal cord motor neuron cell fate specification(GO:0021776) | dorsoventral neural tube patterning(GO:0021904) | smoothened signaling pathway in regulation of granule cell precursor cell proliferation(GO:0021938) | spinal cord ventral commissure morphogenesis(GO:0021965) | pituitary gland development(GO:0021983) | cell differentiation(GO:0030154) | lung development(GO:0030324) | mammary gland development(GO:0030879) | hindbrain development(GO:0030902) | tube development(GO:0035295) | odontogenesis of dentine-containing tooth(GO:0042475) | positive regulation of neuron differentiation(GO:0045666) | positive regulation of transcription(GO:0045941) | positive regulation of transcription from RNA polymerase II promoter(GO:0045944) | embryonic gut development(GO:0048566) | developmental growth(GO:0048589) | anatomical structure formation(GO:0048646) | neuron development(GO:0048666) | branching morphogenesis of a tube(GO:0048754) | anatomical structure development(GO:0048856) | notochord regression(GO:0060032)	Hedgehog signaling pathway(04340) | Basal cell carcinoma(05217)	ENSMUST00000062483	0	0	0	0	0	0	Not Found	Not Found	Not Found
fuf5P1KK.Vzv6sA78k	Znrf2	zinc and ring finger 2	387524	0.29	3.3e-06	0.90	6.2e-05	biological_process(GO:0008150)		OTTMUST00000056945	1	1	1	1	1	1	Not Found	-0.2721	Not Found
HLkRZF_ztElS698r3s	Ttc39b	tetratricopeptide repeat domain 39B	69863	0.51	6.4e-06	0.68	1.8e-04			OTTMUST00000000091	2	3	3	1	2	0	Not Found	Not Found	Not Found
TbIXcNVESkpfvK1MuI	Dbt	dihydrolipoamide branched chain transacylase E2	13171	0.24	2.2e-05	0.75	2.4e-06	metabolic process(GO:0008152)	Valine, leucine and isoleucine degradation(00280)	OTTMUST00000066974	0	1	1	1	0	0	Not Found	-0.2365	0.000120217
ECyrniJ06TKt568y6I	Dcbld1	discoidin, CUB and LCCL domain containing 1	66686	0.45	5.2e-05	1.04	4.5e-05										Not Found	Not Found	0.00284702
xQMgji9qDrurjHh1Kk	Sfxn3	sideroflexin 3	94280	0.49	6.6e-05	0.45	7.6e-05	transport(GO:0006810) | ion transport(GO:0006811) | cation transport(GO:0006812) | iron ion transport(GO:0006826)		ENSMUST00000062213	0	0	2	0	0	0	Not Found	Not Found	Not Found
fi_CR.P_vu.jRvefoI	Rgma	RGM domain family, member A	244058	0.26	7.3e-05	0.41	3.4e-05	neural tube closure(GO:0001843) | positive regulation of protein amino acid phosphorylation(GO:0001934) |		OTTMUST00000077776	0	0	0	0	0	0	Not Found	Not Found	Not Found

								regulation of BMP signaling pathway(GO:0030510)												
Eeuq_mu5hHURiKdEM8	Ndufa5	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5	68202	0.24	8.0e-05	1.40	7.4e-08	respiratory electron transport chain(GO:0022904) | transport(GO:0006810)	Oxidative phosphorylation(00190)	OTTMUST00000066417	0	1	1	1	0	0	Not Found	-0.3131	0.00396649
6X9TfVN8x7_f_OKfqo				0.42	1.1e-04	0.50	5.6e-06													
ry1e1qOt947dcQ5riQ	Ppm1b *		19043*	0.22	4.3e-04	0.60	1.4e-04			ENSMUST00000112304	1	0	0	0	0	0	Not Found	Not Found	Not Found
oUuo767U67hPAhSgl4	Hus1	Hus1 homolog (S. pombe)	15574	0.22	4.6e-04	0.87	6.2e-07	DNA damage checkpoint(GO:0000077) | double-strand break repair via homologous recombination(GO:0000724) | DNA repair(GO:0006281) | protein amino acid phosphorylation(GO:0006468) | response to DNA damage stimulus(GO:0006974) | embryonic development(GO:0009790) | negative regulation of DNA replication(GO:0008156) | response to UV(GO:0009411) | G1/S transition checkpoint(GO:0031575)		OTTMUST00000011503	1	0	2	0	1	0	Not Found	Not Found	Not Found
ld.9cvPeXILBVKCLck	Asah1	N-acylsphingosine amidohydrolase 1	11886	0.22	5.7e-04	0.60	1.4e-05		Sphingolipid metabolism(00600)	OTTMUST00000054076	1	1	1	1	0	0	Not Found	-0.1342	0.0236953
3zcTwolFuqW9i4Vkjs	Pols	polymerase (DNA directed) sigma	210106	0.22	1.4e-03	0.60	4.5e-05	DNA replication(GO:0006260) | cell cycle(GO:0007049) | mitosis(GO:0007067) | cell division(GO:0051301)		OTTMUST00000081019	2	1	1	1	1	1	-0.4701	Not Found	Not Found
9EKV3r68p3OAfM.lf0				0.17	1.7e-03	0.75	8.5e-05													
re7f7od_kU_3ejq6mA	Prkar1a	protein kinase, cAMP dependent regulatory, type I, alpha	19084	0.16	2.9e-03	0.99	7.6e-09	mesoderm formation(GO:0001707) | regulation of protein amino acid phosphorylation(GO:0001932) | signal transduction(GO:0007165) | protein amino acid phosphorylation(GO:0006468) | organ morphogenesis(GO:0009887) | cell proliferation(GO:0008283)	Apoptosis(04210) | Insulin signaling pathway(04910)	OTTMUST00000007044	0	2	1	1	0	0	Not Found	-0.2854	Not Found
Kl70hUQdU3V_lEg7Hw	Spryd4	SPRY domain containing 4	66701	0.17	5.0e-03	1.25	7.7e-10			ENSMUST00000061995	1	1	1	1	1	1	-0.4078	Not Found	Not Found
TJDtEB5LLt2heVnRSc	Fbxw7 *		50754*	0.28	6.4e-03	0.75	2.6e-06			OTTMUST00000054321	1	3	2	2	1	1	-0.5395	Not Found	Not Found

Mmu-miR-291a-3p Probable MiRNA Targets

Illumina Probe IDs	Gene Symbols	Gene Name	Entrez ID	Log Fold Change upon Dgcr8 Depletion	P-value of change upon Dgcr8 Depletion	Log Fold Change upon miR-291a-3p Reintroduction	P-value of change upon miR-291a-3p Reintroduction	GO Terms	KEGG Pathways	Transcript used for 3'UTR seed counts	6mer(3)-AGCACT	6mer(1A)-CACTTA	6mer(2)-GCACTT	7mer(2)(m8)-AGCACTT	7mer(1A)-GCACTTA	8mer(1A)-AGCACTTA	Target Scan Conserved Target Context Score	Target Scan Non-conserved Target Context Score	miRBase Targets P-value
Nbi0eo.u_ToKS39e_w	Hp1bp3	heterochromatin protein 1, binding protein 3	15441	1.47	2.9e-21	0.65	6.4e-05	nucleosome assembly(GO:0006334)		OTTMUST00000023193	3	3	3	1	3	1	Not Found	-0.2342	Not Found
HKTHleDlfk2_w4k118	Dctn4	dynactin 4	67665	1.10	5.8e-21	0.92	1.0e-05			ENSMUST00000025505	1	0	4	1	0	0	Not Found	-0.208	Not Found
xXFXjz.RJjJIk16QSo	Arhgef18	rho/rac guanine nucleotide exchange factor (GEF) 18	102098	1.20	2.6e-19	0.56	1.9e-05	actin cytoskeleton organization and biogenesis(GO:0030036) | small GTPase mediated signal transduction(GO:0007264) | regulation of cell shape(GO:0008360) | regulation of Rho protein signal transduction(GO:0035023)		OTTMUST00000076126	2	0	2	1	0	0	-0.1761	Not Found	Not Found
Zsu87Je6Dg9d9Vei1I	Chic1 *		12212*	1.24	2.2e-18	0.67	1.0e-05			OTTMUST00000044104	4	5	5	1	1	0	Not Found	-0.1875	Not Found
3uBv_Ub24UCMUp76E4	Tmem87b	transmembrane protein 87B	72477	0.67	1.3e-17	0.71	1.2e-05			OTTMUST00000036395	3	0	3	1	0	0	Not Found	-0.1248	Not Found
6SVCgVf5yvkUr3.d5E	Mgat4a	mannoside acetylglucosaminyltransferase 4, isoenzyme A	269181	0.66	1.6e-17	0.48	2.5e-05	carbohydrate metabolic process(GO:0005975)	N-Glycan biosynthesis(00510) | Glycan structures - biosynthesis 1(01030)								Not Found	-0.4217	Not Found
KFiKoeh3SIfxL7deS8	Tbcel	tubulin folding cofactor E-like	272589	0.79	1.4e-15	0.66	1.4e-06	protein modification process(GO:0006464)		ENSMUST00000066148	2	1	3	1	0	0	Not Found	Not Found	Not Found
Ej5ewT_J98eO7nrqEo	Whsc1l1	Wolf-Hirschhorn syndrome candidate 1-like 1 (human)	234135	0.66	3.9e-15	0.54	3.1e-05	transcription(GO:0006350) | regulation of transcription, DNA-dependent(GO:0006355) | chromatin modification(GO:0016568)		ENSMUST00000084026	7	2	7	5	2	1	Not Found	-0.1427	0.00112896
rKeJNVhH4qQBK_U9ck	Gnb5	guanine nucleotide binding protein (G protein), beta 5	14697	0.54	7.0e-15	0.69	2.1e-06	signal transduction(GO:0007165) | G-protein coupled receptor protein signaling pathway(GO:0007186)		ENSMUST00000076889	1	0	2	1	0	0	-0.1571	Not Found	Not Found
BXu_5I6pSBRyHJJf14	D030056L22Rik	RIKEN cDNA D030056L22 gene	225995	0.62	9.7e-15	0.50	8.7e-06			ENSMUST00000062753	1	0	2	1	0	0	Not Found	-0.2362	Not Found
QXS1Ku6d7erks5LTug				0.73	3.1e-14	0.76	1.3e-05												
Holi8e.uF76nVr.e.E	Tor1b	torsin family 1, member B	30934	0.56	3.9e-14	0.52	2.9e-05	biological_process(GO:0008150) | chaperone cofactor-dependent protein folding(GO:0051085)		OTTMUST00000027995	3	1	2	2	0	0	Not Found	-0.1008	Not Found
oyM1RN1xu0tf7Xiges	Il7	interleukin 7	16196	0.67	3.0e-12	0.74	1.0e-06	T cell lineage commitment(GO:0002360) | anti-apoptosis(GO:0006916) | immune response(GO:0006955) | regulation of gene expression(GO:0010468) | positive regulation of B cell proliferation(GO:0030890) | negative regulation of apoptosis(GO:0043066) | negative regulation of catalytic activity(GO:0043086) | bone resorption(GO:0045453) | positive regulation of T cell differentiation(GO:0045582) | positive regulation of organ growth(GO:0046622) | homeostasis of number of cells within a tissue(GO:0048873)	Cytokine-cytokine receptor interaction(04060) | Jak-STAT signaling pathway(04630) | Hematopoietic cell lineage(04640)								Not Found	-0.2369	Not Found
Nf_4d307578y3pW6oQ	Bat5	HLA-B associated transcript 5	193742	0.65	4.6e-12	0.52	2.8e-06			ENSMUSESTT00000055410	2	1	2	2	1	1	Not Found	-0.3073	0.00226352
								regulation of cell cycle(GO:0051726) | apoptosis(GO:0006915) | transcription(GO:0006350) | regulation of transcription, DNA-dependent(GO:0006355) |	Cell cycle(04110) | Pancreatic cancer(05212) | Glioma(05214) |									

KeR.j_.XBL3vV15.j8	E2f1	E2F transcription factor 1	13555	0.73	5.0e-12	0.49	3.2e-05	cell cycle(GO:0007049) | forebrain development(GO:0030900) | regulation of transcription(GO:0045449) | positive regulation of transcription, DNA-dependent(GO:0045893) | positive regulation of transcription from RNA polymerase II promoter(GO:0045944)	Prostate cancer(05215) | Melanoma(05218) | Bladder cancer(05219) | Chronic myeloid leukemia(05220) | Small cell lung cancer(05222) | Non-small cell lung cancer(05223)	OTTMUST00000037809	1	0	2	0	0	0	Not Found	Not Found	Not Found
3uKIKTKCulcniH3tOk	Bicd2	bicaudal D homolog 2 (Drosophila)	76895	0.63	9.8e-12	0.49	2.0e-05	microtubule-based movement(GO:0007018)		ENSMUST00000048544	1	2	3	1	0	0	Not Found	Not Found	Not Found
6YOkcXyoT3LPn3Pv6k	Pfn2	profilin 2	18645	1.00	6.8e-11	0.72	7.1e-05	actin cytoskeleton organization and biogenesis(GO:0030036) | cytoskeleton organization and biogenesis(GO:0007010)	Regulation of actin cytoskeleton(04810)	OTTMUST00000065997	1	0	4	1	0	0	-0.1369	Not Found	Not Found
QPvVV.Lwb_seLc3E4I	Cep170	centrosomal protein 170	545389	0.56	6.0e-10	0.57	1.1e-05			ENSMUST00000057037							Not Found	Not Found	Not Found
Nn_v556__f6DXtLhVc				0.43	1.8e-09	0.79	4.2e-06												
xT8NPGkLaPjFXE9.8	Entpd7	ectonucleoside triphosphate diphosphohydrolase 7	93685	0.58	3.8e-09	0.65	3.5e-06	ribonucleoside diphosphate catabolic process(GO:0009191) | ribonucleoside triphosphate catabolic process(GO:0009203)	Starch and sucrose metabolism(00500) | Folate biosynthesis(00790)	OTTMUST00000065276	1	1	1	1	1	1	Not Found	Not Found	Not Found
9o66jF5AB7up_TI4c4	Rab11a	RAB11a, member RAS oncogene family	53869	0.46	1.6e-08	0.52	4.7e-06	transport(GO:0006810) | small GTPase mediated signal transduction(GO:0007264) | protein transport(GO:0015031)									-0.3858	Not Found	Not Found
QKBEi8HyKHlp5q5K1I	Ifi35	interferon-induced protein 35	70110	0.28	4.7e-08	0.39	2.1e-05			OTTMUST00000005867	1	2	1	1	1	1	Not Found	Not Found	Not Found
xekoCDuV.HExea3S_c	Tnfaip1	tumor necrosis factor, alpha-induced protein 1 (endothelial)	21927	0.34	2.6e-07	0.90	6.0e-05	potassium ion transport(GO:0006813) | embryonic development(GO:0009790)		OTTMUST00000000462	2	0	2	2	0	0	-0.2121	Not Found	Not Found
o6cIXSzdUskffej_D0	4933434E20Rik *		99650*	0.51	8.2e-06	0.65	1.1e-06			ENSMUST00000068798	2	0	2	1	0	0	Not Found	-0.2088	Not Found
lq_eOfCErKvp0e8E5I	Zbtb41	zinc finger and BTB domain containing 41 homolog	226470	0.20	3.3e-05	0.68	4.5e-05	transcription(GO:0006350) | regulation of transcription, DNA-dependent(GO:0006355)		OTTMUST00000051928	4	5	5	3	2	2	-0.2492	Not Found	Not Found
6X9TfVN8x7_f_OKfqo				0.42	1.1e-04	0.38	6.5e-05												
ltnt_R3rt4ml4l9Vq4	Cdkn1a	cyclin-dependent kinase inhibitor 1A (P21)	12575	0.68	5.0e-04	0.94	2.6e-05	regulation of cell cycle(GO:0051726) | response to DNA damage stimulus(GO:0006974) | cell cycle(GO:0007049) | cell cycle arrest(GO:0007050) | negative regulation of cell proliferation(GO:0008285) | response to UV(GO:0009411) | positive regulation of B cell proliferation(GO:0030890) | negative regulation of apoptosis(GO:0043066) | positive regulation of non-apoptotic programmed cell death(GO:0043071) | negative regulation of cyclin-dependent protein kinase activity(GO:0045736)	ErbB signaling pathway(04012) | Cell cycle(04110) | p53 signaling pathway(04115) | Glioma(05214) | Prostate cancer(05215) | Melanoma(05218) | Bladder cancer(05219) | Chronic myeloid leukemia(05220)	OTTMUST00000078425	2	1	3	2	1	1	Not Found	-0.3089	Not Found
ud77jUnSmu2p_f.geM	Gtf3c2	general transcription factor IIIC, polypeptide 2, beta	71752	0.18	1.8e-03	0.58	5.5e-06	transcription(GO:0006350)		OTTMUST00000053730	4	1	2	2	1	1	Not Found	-0.1006	Not Found

5.3.3.6 Comparison of genes in the miR-291a-3p candidate target list to a previously published experimentally predicted miR-290 cluster target list

Sinkkonen *et al.* recently conducted a similar study in Dicer-deficient mouse ES cells (Sinkkonen et al., 2008). They compared the transcriptional profiles of *Dicer1*$^{-/-}$ cells to *Dicer1*$^{+/-}$ cells and then searched the 3'UTR sequences of the up regulated transcripts for the enrichment of 7mer motifs. They identified a list of 7mer sequences very similar to the list of enriched seeds that were identified in the up-regulated transcripts from the *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ comparison with Sylamer (Fig.5.4C). They found 7mer motifs GCACUUU, AGCACUU, GCACUUA, UGCACUU and AAGCACU to be enriched, which correspond to miR-17 (2) (and miR-291a-3p (bases 1-7)), miR-291a-3p (2), miR-291a-3p (1A), miR-291b-3p (2) and miR-302b (2) respectively. GCACUUU, AGCACUU and GCACUUA were all amongst the top motifs enriched within my comparison, although there was an additional enrichment of the miR-25 seeds and a miR-301a seed.

Next they transfected the *Dicer1*$^{-/-}$ cells with miRNAs of miR-290 cluster (mmu-miR-290-5p, 291a-3p, 292-3p, 293, 294 and 295) all of which, except mmu-miR-290-5p, share similar seed sequences. They therefore conducted an experiment like that described in this chapter. Upon comparison of the expression profiles of the miRNA transfected cells to those of the control transfections Sinkkonen *et al.* again found an enrichment of miR-291a-3p seed sequences, this time amongst those genes down regulated, in agreement with my results described above. They then intersected those genes up regulated in the knockout cells with those down regulated upon the reintroduction of the miRNAs and filtered this list for genes that contain the hexamer

GCACUU seed within their 3'UTR. The remaining list of predicted targets contained 253 genes.

A comparison between my list of miR-291a-3p predicted targets and the Sinkkonen *et al.* miR-290 cluster target predictions reveals an overlap of 9 genes (*Hp1bp3*, *Dctn4*, *Tbcel*, *D030056L22Rik*, *Il7*, *Pfn2*, *Tnfaip1*, *Zbtb41* and *Cdkn1a*). Significantly, both lists predict *Cdkn1a* to be a target of these miRNAs. *Cdkn1a* has indeed been recently demonstrated to be a miR-291a-3p target (Wang et al., 2008).

There could be many reasons for these discrepancies in both the number and identity of predicted targets. First of all there are several major differences between these experiments. The most important is that I transfected single miRNAs in order to generate my predicted target lists as opposed to a set of miRNAs. Although many of the miRNAs transfected by Sinkkonen *et al.* do share similar seed regions, mmu-miR-293-3p has a 2nt shift in its seed and mmu-miR-290-5p has an entirely different seed sequence. These two alterations could be expected to result in the targeting of entirely different mRNA transcripts. As miRNAs can co-regulate genes, combinatorial regulation could cause increased repression and degradation of some targets beyond that which would be expected if they were to be targeted by a single miRNA complementary to a single seed target site within the same UTR. Furthermore Sinkkonen *et al.* harvested the RNA from their cells 24 hours post-transfection as opposed to the 10 hours for which I cultured my cells. Assuming that both cell lines react with the same efficiency to the transfected duplexes and given the results of my optimization experiments, it could be that this period of 24 hours results in the

enrichment of their target lists with a greater number of secondary targets than lists generated by the method described here.

However, it is worth noting that Sinkkonen *et al.* identified potential targets with *P*-value cutoffs of 0.001, as opposed to 0.1 – 0.05 used by my analyses. The more stringent *P*-value criteria were in part made possible by the use of triplicate, rather than duplicate transfections. It is possible that further replicates of miR-291a-3p transfections would allow me to expand my miRNA target lists, as *P*-values could be expected to become more significant. Sinkkonen *et al.* also worked without LFC criteria, which may have expanded their lists of deregulated genes.

There may have also been differences in the transfection efficiencies between the two experiments. The greater over expression of a miRNA duplex within a cell could lead to far greater perturbations of target genes and may deregulate genes with far from optimal target sites in their transcripts, although to some extent off target effects would be remedied by the intersection of the array experiment results. On the other hand, a lower transfection efficiency could mean that the signal from down regulated transcripts may be dilute within the RNA pool by the transcripts from non-transfected cells whose RNA levels remain unaltered.

Finally, as described in Chapter 3, ES cells of different origins and cultured for different periods of time can accumulate genetic differences and these themselves may affect gene expression and hence predicted target lists. In future it may be informative to repeat the transfection experiments described in this chapter with the

Dgcr8$^{gt2/tm1}$ cell line in order to replicate results in an independent gene trap background that has been cultured in parallel.

5.3.3.7 Candidate targets of particular interest

The annotated GO terms and KEGG pathways amongst these predicted target gene sets contain some interesting trends. There are 4 genes associated with the cell cycle (*Pols* and *Hus1* for miR-25; *E2f1* and *Cdkn1a* for miR-291a-3p). The implication is that although the miR-290 cluster's regulation of *Cdkn1a* (p21) has recently been demonstrated as a fundamental cause of the G1/S phase transitional block following DGCR8 depletion, I would concur with the authors of that study that the deregulation of this transition is unlikely to be caused solely by a single gene (Wang et al., 2008).

Hus1, a potential miR-25 target, is involved in the recognition of DNA damage and the activation of a chain of reactions that can culminate in the phosphorylation of p53 (*Trp53*) and the subsequent transcriptional up-regulation of p21 and a block at the G1/S phase transition (Niida and Nakanishi, 2006). However, while miRNA mediated degradation of the transcripts of these two proteins in wild type cells might be expected to maintain cell cycling, *E2f1*, which encodes a transcription factor associated with the positive regulation of the G1 to S transition is also within the list of predicted miR-291a-3p targets (Cam and Dynlacht, 2003;Trimarchi and Lees, 2002). This at least appears to be a demonstration of the same miRNAs bearing a responsibility for the tightening of the regulation of a range of factors involved in the same cellular process, rather than as a molecular switch required to trigger a transition from one cellular state to another. It is worth bearing in mind, however, that the cell cycle involves highly complex interactions between a large number of proteins.

Obviously in such a background the influence of a miRNA and its targets can be affected by many outside factors, and as a consequence a gene list is insufficient to prescribe the effect of changes in miRNA and target levels.

As a further demonstration of the complexity involved in discerning the extent of miRNA control in G1/S phase transition, p21 has also been shown to be a transcriptional target of *Klf4* (Chen et al., 2003). As described above this marker is up regulated in the *Dgcr8* deficient cells. Whether this is a direct result of the removal of miRNA induced regulation of *Klf4* or a consequence of downstream cellular changes is uncertain, but it is clear that there are further pathways that impinge on the regulation of the cell cycle that will be influenced by miRNAs.

In addition to its role in cell cycle regulation, E2F1 can also trigger apoptosis. The recruitment of E2F1 to the promoter of p73 (*Trp73*), a pro-apoptotic gene, is enhanced by its acetylation. E2F1 stability is also enhanced by genotoxic stress induced acetylation (Ianari et al., 2004). This increase in stability is PCAF (KAT2B) dependent. PCAF is a histone acetyl-transferase also involved in transcriptional activation and also appears in the miR-25 predicted target lists presented here (Marmorstein, 2001).

The miR-106b-25 miRNA cluster is regulated by E2F1 (as they are embedded within the intron of the *Mcm7* mRNA). As a result the miR-290 cluster potentially plays a role in the regulation of the miR-106b-25 cluster via E2F1, adding to the connectivity of the miRNA network. Members of the latter cluster are also known to auto-regulate their own expression by targeting the *E2f1* mRNA and have been demonstrated to

target the p21 mRNA (Petrocca et al., 2008a; Petrocca et al., 2008b). The prospect of miR-291a-3p and miR-106b and miR-93 sharing targets is not that unusual as the first 8 bases of miR-291a-3p correspond to bases 2-9 of the other miRNAs. If these targets are indeed subsequently demonstrated to be true *in vivo* miRNA targets it would therefore appear that the regulation of *E2f1* by miRNAs occurs through both direct and indirect mechanisms all of which are tightly coordinated.

A homozygous deletion of *ADAM23*, a member of the ADAM family of proteins, which regulates cell-cell and cell-extracellular matrix interactions, has been identified in a gastric cancer cell line (Takada et al., 2005a). Subsequently a broader down regulation of *ADAM23* was seen in both gastric cancer cell lines and gastric primary tumours, although these were not necessarily associated with genomic deletions. Methylation of the gene's CpG island was attributed with a role in this down-regulation. However, a recent study has demonstrated miR-25 to be up regulated in gastric primary tumours (Petrocca et al., 2008b). *ADAM23* was a member of the miR-25 candidate target list generated in my study and given the number of miR-25 seed sequences in the 3'UTR of the gene it is intriguing to speculate that miR-25 may play a role in the down-regulation of this gene in gastric cancer. In the same study the reintroduction of *ADAM23* into gastric cancer cells seemed to suppress their growth in a colony formation assay.

A number of other tumour suppressor genes also appear within these target gene lists. Most notable is the *Fbxw7* gene that may be a target of miR-25. FBXW7 forms a part of the SCF type ubiquitin ligase complex, which targets a number of oncogenes (including c-Myc and cyclin E) for degradation (Welcker and Clurman, 2008) via the

26S proteasome. It is intriguing to speculate that the up regulation of *Fbxw7* in *Dgcr8*$^{gt/tm1}$ indirectly cause the disruption of c-Myc and in turn lead to the disruption of c-Myc regulated targets, as seen earlier in this chapter.

Wwp2 is another ubiquitin ligase. It seems to be ubiquitously expressed but is expressed at higher levels in undifferentiated ES cells. It seems that Oct-4 may be among the targets of *Wwp2* and ectopic expression of Oct4 and *Wwp2* along with Oct4 sensitive expression constructs demonstrates that *Wwp2* negatively regulates Oct4 transcriptional activity in ES cells (Xu et al., 2004). Although Oct4 levels do not seem to change significantly in cells depleted in DGCR8 it is plausible that more subtle changes could lead to a general disruption of multiple targets of the ES cell core transcriptional network which may again go some way to explaining both the differences described between *Dgcr8*$^{gt/tm1}$ and *Dgcr8*$^{tm1,gt/+}$ expression profiles and the phenotypes described in Chapter 3.

In addition to tumour suppressors among the predicted targets there are also oncogenes. *Whsc1L1* is one such gene that appears in both target lists with a large number of seed sequences within its 3'UTR, as well as being a miR-291a-3p predicted target by both TargetScan and miRBase Targets. It is within amplified chromosomal regions in a number of cancer types, while siRNAs directed against it lead to a 50% fewer soft-agar colonies in anchorage-independent growth assays (Tonon et al., 2005).

Finally, within the target lists there are a large number of genes involved in signal-transduction in some way, as judged by associated GO terms and KEGG pathway

terms. These include *Adam23*, *Plekhm1*, *Rab8b*, *Plekhg3*, *Gnaq*, *Gli2*, *Rgma* and *Prkar1a* within the miR-25 list and *Arhgef18*, *Gnb5*, *Il7*, *Rab11a* and *Cdkn1a* in the miR-291a-3p list. *Rgma* is potentially one of the most interesting genes in this list. BMP signaling is important in the maintenance of mouse embryonic stem cells in culture. *Rgma* is involved in the regulation of this signaling pathway by increasing BMP signaling, allowing BMP2 and BMP4 to bind the ACTR2A receptors more effectively in addition to the BMPR2 receptor (Xia et al., 2007). Clearly alterations within this signaling pathway may have a knock on effect on downstream gene regulation and phenotype.

Undoubtedly there are a number of other associations and conclusions that can be garnered from these gene lists of potential miRNA targets. It is worth noting that all conclusions are speculation and would require further experiments in order to support and confirm any subsequent hypotheses. However, these are miRNA targets supported by experimental evidence and are no longer solely computational predictions. Therefore as a resource for constructing these hypotheses it is superior to purely predictive methods.

5.4 Discussion

The depletion of miRNAs in ES cells due to the disruption of the miRNA-processing pathway leads to significant disruption of the mRNA expression profile of these cells. GO analysis identified an enrichment of genes with GO terms associated with development and morphogenesis amongst the genes up regulated in *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells leading me to conclude that miRNAs probably play a fundamental role in the stabilization of the undifferentiated stem cell state. Indeed, by comparing

the differential mRNA expression changes identified upon the insertion of a trap into both alleles of *Dgcr8*, to the transcriptional targets of transcription factors integral to the determination of the ES cell fate, it appears that there are significant perturbations of these TF target gene sets induced by the depletion of miRNAs. In the context of these alterations in gene expression it is perhaps less surprising that the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells exhibit some of the phenotypes presented in Chapter 3. The exact nature of the miRNA mediated control of these core transcriptional networks warrants considerable further study as it may prove fundamental to understanding ES cell pluripotency. It is worth bearing in mind, however, that there is considerable scope with which to improve the ChIP-Chip and ChIP-Seq data. As noted in Liu *et al.* there are major differences seen between available ChIP-Chip and ChIP-Seq datasets and further repeats and replications of these results may be required to refine transcriptional target sets (Liu et al., 2008b).

Subsequently I reintroduced two miRNAs into $Dgcr8^{gt1/tm1}$ cells. These miRNAs are highly expressed in wild type E14 cells, have seed sequences enriched amongst those genes up-regulated upon DGCR8 depletion and consequential disruption of miRNA processing and do not share similar seed regions. By introducing these miRNAs into cells that have a reduced complement of endogenous miRNAs, I was able to generate a list of predicted miRNA target genes for each miRNA that I may not have been able to generate in the context of knockdown or over expression experiment in wild type cells. This is because within wild type cells such experiments are susceptible to the saturation of endogenous targets by expressed miRNAs or functional redundancy by miRNAs with similar seed sequences, which may mask the effect of the loss of the expression of a single, or limited set, of miRNAs.

These lists enriched for true miRNA targets do not necessarily concur with computational target predictions. This is not necessarily surprising as the computational algorithms accept that the lists they present will likely be neither comprehensive nor noise free. The discrepancies seen between the predictions of different algorithms are an illustration of the problems inherent with computational extrapolation of experimentally derived targeting rules (Sethupathy et al., 2006). Indeed I believe it is widely accepted that the rules used to predict miRNA targets are not necessarily comprehensive as demonstrated by the continuous improvements that are made to these predictive methods as the data available improves and expands. Despite this there are a number of previously described miRNA targets that do not appear in my target lists such as *Bim* for miR-25 and *Lats2* and *Rbl2* for miR-291a-3p (Wang et al., 2008) as well as a set of predicted targets for the miR-290 cluster presented Sinkkonen *et al.* identified by a method similar to that presented here (Sinkkonen et al., 2008). Although there could be a number of reasons why various targets are not identified within my target lists, including an absence of mapped probes on the array and expression and functional differences between cell lines, it is also likely that the target lists presented here are not comprehensive. It is probable that the number of genes identified as targets would increase if further miRNA transfection replicates were conducted, which would improve the statistical power of the significant probe identification. Although it is possible that, as a consequence of this limitation, these lists are enriched for the most significant miRNA targets, if longer and less noisy lists were desired, I would recommend conducting these replicates at a later date.

However, despite these differences the method presented in this chapter also successfully confirmed a previously identified miRNA target gene as a probable miR-291a-3p target (*Cdkn1a*) as well as independently identifying 9 genes from the Sinkkonen *et al.* predicted miR-290 cluster targets as miR-291a-3p potential targets (*Hp1bp3*, *Dctn4*, *Tbcel*, *D030056L22Rik*, *Il7*, *Pfn2*, *Tnfaip1*, *Zbtb41* and *Cdkn1a*) (Sinkkonen et al., 2008). In addition the gene lists presented here predict a large number of additional, functionally relevant, miRNA target relationships and allow the description of a large number of hypotheses that warrant further investigation or confirmation.

Ultimately, the system presented here is simple and inherently scaleable to generate a large number of gene lists enriched for a specific miRNA's target genes. This would provide a useful resource for generating target lists based on experimental evidence from which to draw patterns and hypotheses for further study.