Chapter 4: The miRNA expression profile of ES cells depleted of functional *Dgcr8*

4.1 Aim

Having derived two ES cell lines depleted in wild type *Dgcr8* transcript through the insertion of two traps into the *Dgcr8* locus, I wished to demonstrate the abrogation of *Dgcr8* function in these cell lines. In order to do this I conducted a series of experiments to ascertain the extent of the effect of these mutations on the miRNA-processing pathway, by measuring mature miRNA expression in these cell lines and controls.

4.2 Introduction

It is widely accepted and logically expected that the disruption of the miRNA-processing pathway will lead to the broad depletion of mature miRNAs. In the last few years a number of studies have knocked out both Dicer and *Dgcr8* in mouse embryonic stem cells (Calabrese et al., 2007; Kanellopoulou et al., 2005; Murchison et al., 2005; Wang et al., 2007). The miRNA expression profiles of cell lines with either a disrupted Dicer locus or a disrupted *Dgcr8* locus have been determined and both cell types are depleted for the vast majority of miRNAs confirming that the successful targeting of thee two loci will cause a major reorganization of the miRNA expression profile (Babiarz et al., 2008; Calabrese et al., 2007; Wang et al., 2007).

The preferred methods with which to measure broad miRNA expression have changed in the recent past. miRNA microarrays have been designed with a wide

selection of miRNA complementary probes (Babak et al., 2004; Barad et al., 2004; Miska et al., 2004). It is clear, however, that miRNA microarrays are subject to a variety of flaws relating to the short length of the miRNA sequence against which to design specific probes. As a consequence there is little choice when it comes to designing probes for a microarray platform with a uniform melting temperature (T_m) across the complete array, thus probe to probe T_m varies and miRNA microarrays have suffered from problems regarding miRNA target specificity (Miska et al., 2004). To some extent these problems have been reconciled by the advent of locked nucleic acid (LNA) technology. LNAs are nucleic acid analogues, which can be used to increase the thermostabiltity of duplexes via their inclusion within oligonucleotides. It is thus possible to generate arrays upon which probes possess a uniform T_m. Ultimately these arrays can be hybridised with confidence at a high temperature without compromising the sensitivity or specificity of any of the miRNA probes (Castoldi et al., 2006).

An alternative approach to improving miRNA expression profiling was demonstrated by Lu *et al.* (Lu et al., 2005). The authors used a bead-based miRNA profiling system, in which miRNA probes are attached to multicoloured beads, to allow hybridization to be conducted in an environment that "might more closely approximate hybridization in solution". The authors supposed that this might improve the specificity of hybridization of the miRNAs to the probes. Following hybridization of fluorescently labeled, amplified miRNA libraries to the beads, they were flow sorted and the bead colour (associated with a particular miRNA specific probe) and the bead fluorescence (dependent upon the quantity of hybridized miRNA) was used as an indication of the expression level of the miRNA. The authors demonstrated this

increase in accuracy by comparing the specificity of synthetic miRNA detection in both a bead and array format.

The advent of high-throughput sequencing methods now allows the small RNA content of cells to be profiled with extremely high sensitivity and without the requirement of pre-designed, complementary probes. This allows the identification of not only known miRNAs and small RNAs but also the sequences of previously unknown small RNAs. As a consequence this technology is being rapidly applied to the study of miRNAs. The small-RNA complement of both Dicer and *Dgcr8* knockout mouse ES cells have been sequenced using this method and compared to the profiles of wild type and heterozygous cell lines (Babiarz et al., 2008; Calabrese et al., 2007). As a consequence of these efforts, accurate and deeply sequenced mouse ES cell miRNA profiles have been generated. In addition a number of DGCR8-independent miRNAs have been identified and a series of Dicer-dependent small RNAs that were previously not recorded in ES cells, including endogenous siRNAs.

In this chapter, I describe the use of both bead-based miRNA profiling and high-throughput Illumina/Solexa sequencing to determine the miRNA expression profiles of $Dgcr8^{+/+}$, $Dgcr8^{tm1,gt1/+}$, $Dgcr8^{tm1,gt2/+}$, $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells. I demonstrate the broad depletion of mature miRNAs in cells with a trap inserted into each allele of *Dgcr8*, while the cells with a single trapped allele and the wild type cells have comparable miRNA expression profiles. This demonstrates the functional significance of the traps at the *Dgcr8* locus. As a consequence of the sequencing experiment, I also confirm the identity of a number of DGCR8 independent miRNAs

previously described by Babiarz *et al.* in addition to proposing a number of other candidates.

4.3 Results

4.3.1 Use of the Luminex platform to profile mature miRNA expression

Initially the miRNA expression profile of each cell line (*Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$) was determined using a Luminex bead-based miRNA profiling system (Blenkiron et al., 2007; Lu et al., 2005). Briefly, each RNA sample is spiked with a predefined set of control probes (pre-controls). This allows the efficiency of the miRNA cloning reactions to be determined. Subsequently, the miRNAs and controls are purified and cloned in a series of ligation reactions to adapter oligos. The samples are then reverse transcribed, and PCR amplified using biotinylated primers. Finally, each cloned sample is spiked with a further set of control oligos. These post-controls allow an assessment of whether the subsequent hybridization is successful. Each sample is hybridised to beads bearing miRNA specific oligo probes. The bead sets are exposed to streptavidin-phycoerythrin and flow sorted. The bead sets contain beads with a myriad of colours. Each colour corresponds to a miRNA specific probe. Therefore by measuring the colour and fluorescence of each bead it is possible to determine the relative expression of each miRNA between samples. In order to maximize the number of miRNAs that can be profiled in a single experiment, given the limited coloured bead selection available, the samples are hybridized to four bead sets. Each set contains a different set of probes corresponding to different miRNAs associated with each of the beads. Spiked n controls are profiled by all 4 bead sets. The flow sorter returns a set of raw median

fluorescence intensity (MFI) values, each corresponding to a specific miRNA or control.

I conducted the labeling and hybridization reactions on two separate occasions. Initially, I grew up a single flask of each cell line to be tested, purified its RNA and labeled and hybridized these samples as two technical replicates simultaneously (referred to as replicates 1A and 1B below). Subsequently, I grew up two new sets of cells for each cell line, simultaneously (replicates 2 and 3). I purified the RNA from these fresh samples and labeled and hybridized these samples at the same time. Upon examining the results obtained from these two separate experiments it is clear that there are some fundamental differences seen between the two sets of samples (Fig.4.1).

The most striking difference is seen in the levels of the pre-control probe MFIs both between the two sample sets and within each set. It is clear that for at least 5 of the samples prepared for replicates 2 and 3 there is a loss of the spiked-in pre-control signal, this is despite an approximately constant signal from each of the post-controls. This would suggest that a significant proportion of the miRNA fraction from these samples was lost in the sample preparation process. In contrast, samples from replicates 1A and 1B all retain approximately equal levels of pre-control signal, implying a roughly equal retention of the miRNA population between samples. There is also a discrepancy in the signal gleaned from each of the individual pre-controls and post-controls in the two independent experiments. When a comparison is made between each of the controls relative to the other controls it seems that there may be a difference in the quantity of each individual spiked in pre-control and post-control on

the two separate occasions. As the pre-controls were to be used to normalise the miRNA signal between the samples, this discrepancy makes it difficult to combine the two experiments into a single analysis.

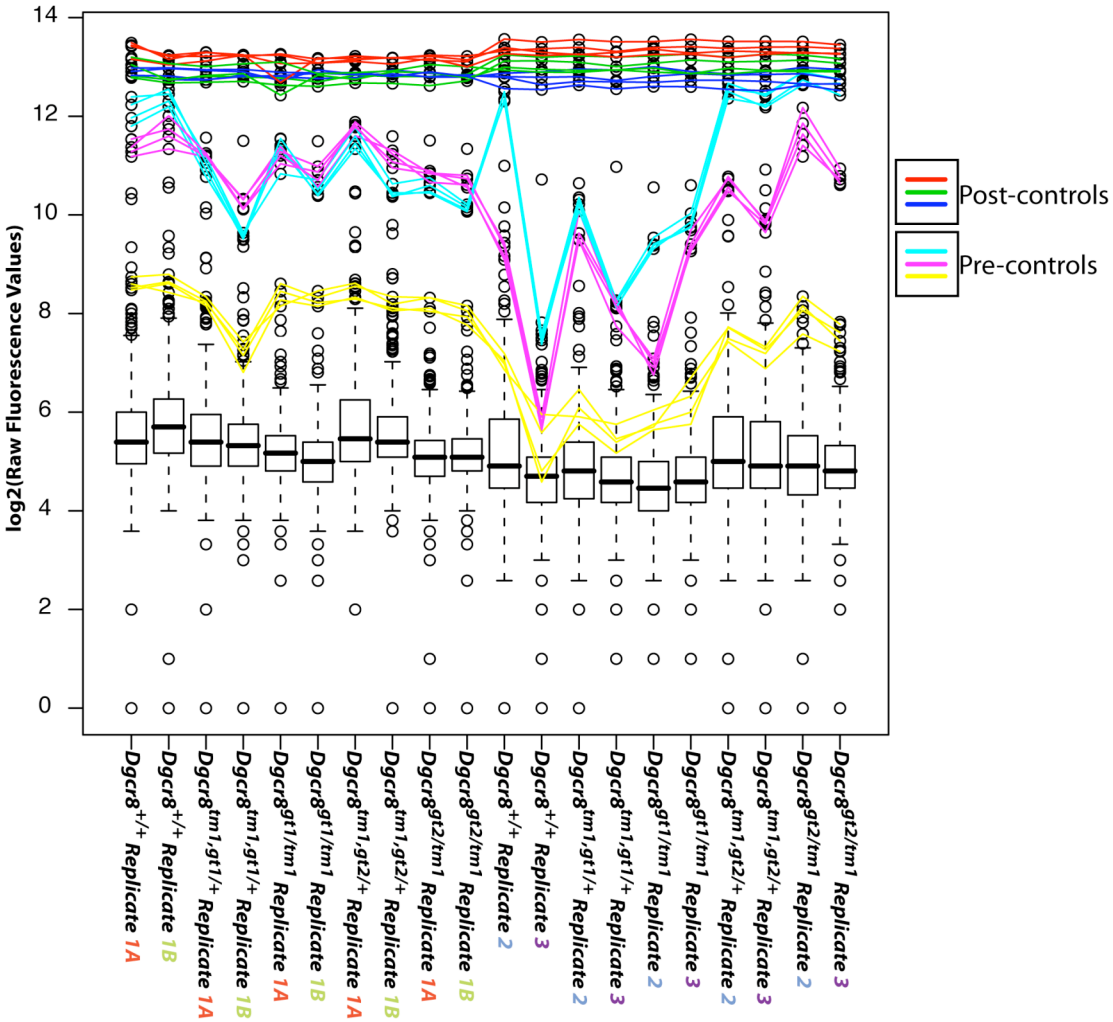
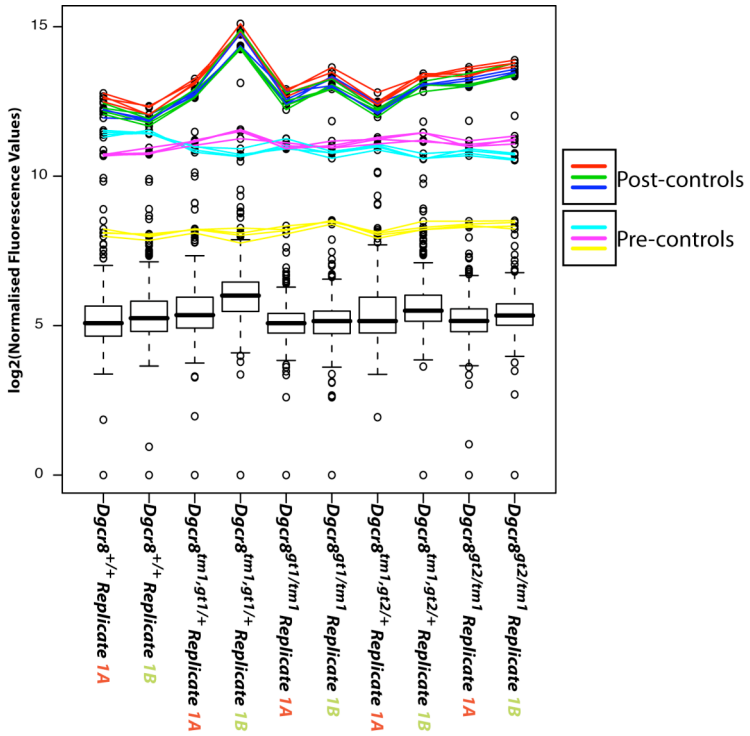


Fig.4.1: A box plot of the raw fluorescence values for each miRNA specific bead derived from the Luminex miRNA profiling system across multiple samples. The mature miRNA expression profiles of the *Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines were measured using the Luminex bead-based system. Replicates 1A and 1B were measured on the same day from the same RNA samples. Replicates 2 and 3 were grown in parallel and RNA was analysed on a second day. Coloured lines represent the fluorescence values for each of the spiked control sequences as indicated (see section 2.10.2.1).

As the initial experiment contained a more complete set of efficiently cloned samples, it was the samples from this experiment that were selected for further analysis. I attempted to normalise replicates 2 and 3 according to the levels of the pre-controls in order to make use of these samples. However, it seems that in the cases where the pre-control levels appear to be relatively depleted, normalising the bulk of the data according to these spiked in oligos leads to the distortion of the sample expression signals (Fig.4.2B). This is seen most clearly in the case of the *Dgcr8*$^{+/+}$ replicate 3. Following normalization, the bulk of the fluorescence values for this sample are much higher than those of the other, more successfully cloned samples. As a consequence of this distortion of signal I chose to exclude these replicates from further analysis.

A



B

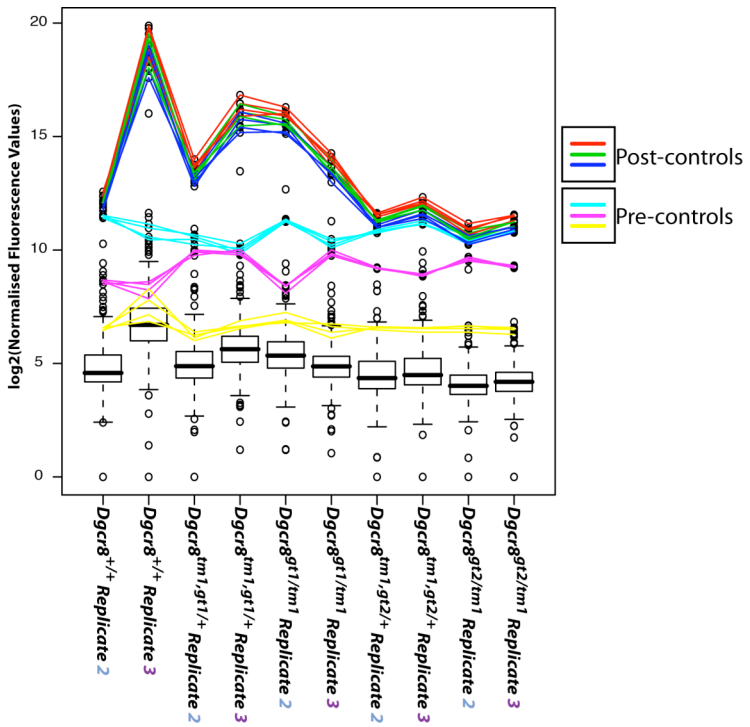


Fig.4.2 Box plots of bead/miRNA-associated MFIs following normalization. Replicate sets 1A and 1B (Fig.4.2A) and sets 2 and 3 (Fig.4.2B) were normalised as separate groups. Samples were

normalised according to the pre-control fluorescence levels in each sample. Coloured lines represent the fluorescence values of spiked in pre and post-controls.

A comparison of *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ miRNA expression and the expression values of the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells (Fig 4.3) demonstrates a broad depletion of a large subset of miRNAs. For this subset, the MFI associated with each miRNA specific probe demonstrates a negative log fold change in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells compared to the control cell lines. Among this subset are a large number of miRNAs known to be expressed in wild type muse ES cells (Houbaviy et al., 2003). These are miRNAs for which a clear reduction of bead fluorescence would be expected in this comparison given the reduced *Dgcr8* expression.

However, it seems that the fluorescence detected by other miRNA specific beads remains relatively constant between the cells with these differing phenotypes. It is very difficult, using this system, to determine which of the probes may be detecting noise rather than a clear miRNA signal, and there is no obvious fluorescence level at which to place a detection threshold cut off to remove noisy expression data. It seems obvious that the signal from a large bulk of the probes with an average $\log_2$ expression level of approximately 5.5 will be below the accurate detection threshold of the method and may not be expressed in these cell lines at all. It is also clear that cross-hybridisation between probes and non-specific miRNAs or other interfering nucleic acids can be misinterpreted as a specific miRNA induced signal as would be expected by any oligo based detection method. The miR-191 probe is a case in point, which is known to generate a high background signal "in the absence of any prepared target" (Lu et al., 2005).

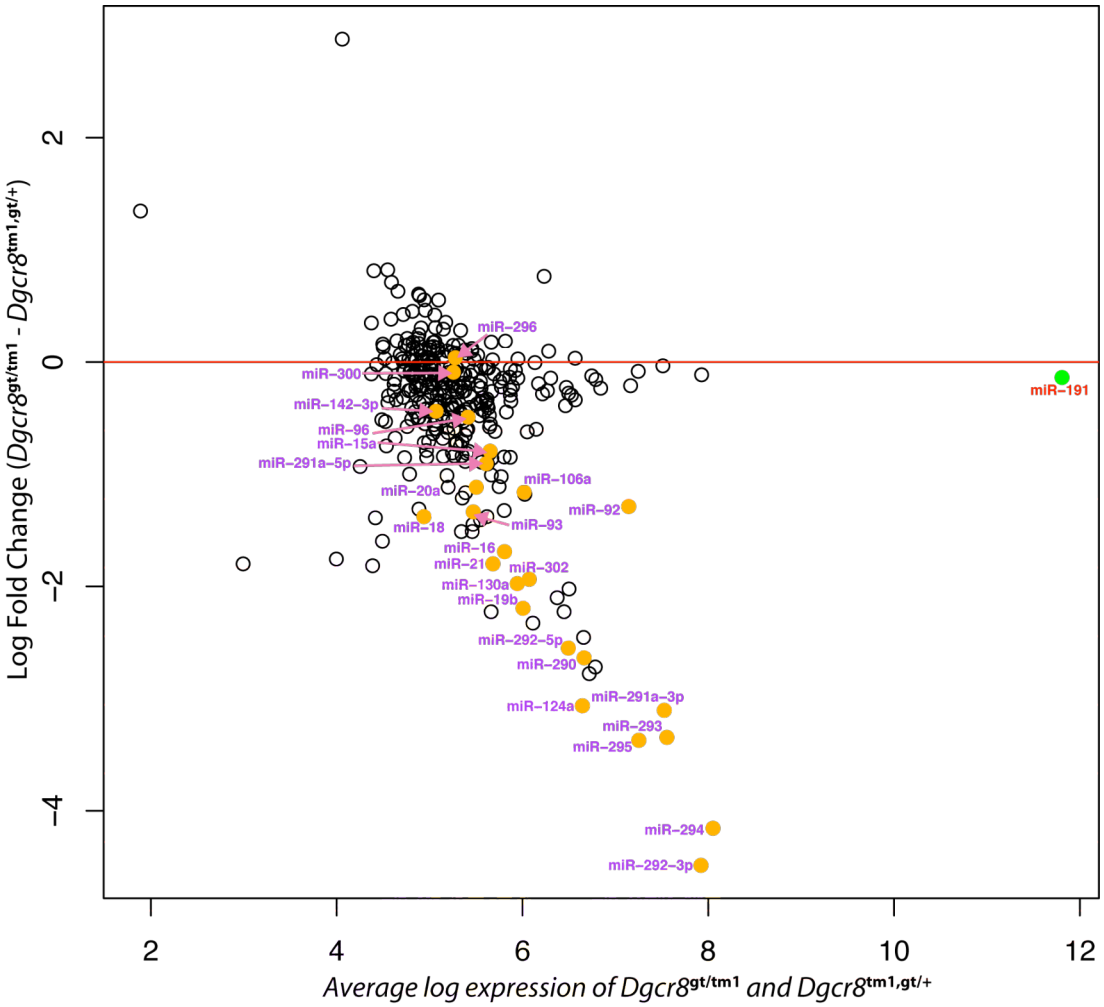


Fig.4.3: A comparison of the expression of miRNAs in *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$cells compared to *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells. The averaged log fluorescence of beads from of *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$cells and *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells from replicates 1A and 1B were compared to each other in two sets according to broad genotype (*Dgcr8*$^{gt/tm1}$ vs. *Dgcr8*$^{tm1,gt/+}$). The expression of each probe was compared between these sets with each dot representing a miRNA. Orange dots are those miRNAs shown to be expressed in mouse ES cells cultured in feeder free conditions by Houbaviy *et al.* (Houbaviy et al., 2003), with sequences that correlate between the aforementioned study and the probes of the Luminex set. The miR-191 probe (green) is known to register high background readings (Lu et al., 2005).

As a consequence it is unclear whether any miRNAs are still expressed in the *Dgcr8* depleted cell lines and to what extent. It is clear, however, that those mature miRNAs known to be expressed in ES cells are strongly down regulated in these knockout

cells. These results concur with the results of my Northern blots in Chapter 3 and extend them to a much broader set of ES cells expressed miRNAs.

4.3.2 miRNA expression profiling of cell lines with Illumina/Solexa high throughput sequencing

In order to avoid the problems inherent to array analysis which is limited by probe design, probe specificity and sensitivity, I prepared small RNA libraries for each of the *Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines for Illumina/Solexa massively parallel sequencing (the sequencing itself was conducted by the Sanger Institute Sequencing Core). Following the sequencing procedure the resultant reads were quality controlled and mapped by Dr. Cei Abreu-Goodger.

Initially all of the reads for each library were mapped to reference sequences comprising of a range of non-coding RNAs (including miRNAs, tRNAs, snoRNAs, snRNAs, rRNAs, mitochondrial tRNAs, mitochondrial rRNAs and miscellaneous other ncRNAs) all derived from the Ensembl database (for non-miRNA ncRNAs) (Flicek et al., 2008) or miRBase (miRNA hairpins) (Griffiths-Jones et al., 2008) (Table 4.1). Subsequently, for each base of each of these reference sequences the number of Illumina/Solexa reads representing that base was calculated. The maximum number of reads representing a single base within a ncRNA sequence was then used to represent the expression level of that RNA. The use of this value means that the length of the RNA sequence does not affect its representative value on any subsequent plot. As the sequences associated with the ncRNAs did not appear to be randomly distributed across the length of the sequences, this would seem to produce the fairest reflection of expression.

Sample	Filtered read depth of library	Depth of sequence that maps to ncRNAs
Dgcr8$^{+/+}$	1865369	1409029
Dgcr8$^{tm1,gt1/+}$	2581471	1842631
Dgcr8$^{gt1/tm1}$	1569525	728577
Dgcr8$^{tm1,gt2/+}$	2389852	1756044
Dgcr8$^{gt2/tm1}$	1073911	483343

Table 4.1: The filtered read depth of each of the small RNA libraries and the number of reads that subsequently map to the Ensembl and miRBase derived non-coding RNAs. For each sample the first column represents the number of reads remaining following the removal of the 3' adapter sequence from each read and the removal of reads that were comprised of a single nucleotide for >75% of their length or that were less than 16 bp long following the removal of the adapter sequence. The second column represents the number of the remaining reads that subsequently mapped to the ncRNA sequences derived from Ensembl and miRBase (See section 2.10.3.2 for details). These initial stages of library processing were performed by Dr. Cei Abreu Goodger.

Of the reads which were mapped against this ncRNA set, 71-76% of the reads from libraries derived from the *Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ cell lines matched a library sequence and passed all of the associated thresholds imposed (see section 2.10.3.2). By contrast only 45-46% of the reads from libraries derived from the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines matched to the libraries. As can be seen in Fig.4.4, of the reads that do match to the library sequences a dramatically reduced proportion of the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell line libraries can be considered as miRNA derived sequences. It seems that the reduced proportion of ncRNA matching reads in these libraries may be accounted for by a reduction in the number of miRNA matching reads in the library (Fig.4.4).

In parallel, an Illumina/Solexa miRNA sequence library generated for wild-type mouse ES cell lines in a previous study (Babiarz et al., 2008) was remapped to the same ncRNA set, following the same criteria as presented here. In order to

demonstrate consistency between the experimental approaches used in these two studies, I calculated the Spearman correlation (0.731) between the maximum sequence depths for all of the miRNAs with a sequence depth greater than 8 in both the wild type ES cell set of Babiarz *et al.* and that which is presented in this thesis for the *Dgcr8*$^{+/+}$ cell line. This comparison confirms that the expression profile for miRNAs generated for mouse ES cells in this study is consistent with that seen in previous studies, bearing in mind variation introduced by experimental procedure. A closer comparison of the expression of individual miRNAs within each library confirmed that there were also no startling expression differences between miRNAs at this level (Data not shown).

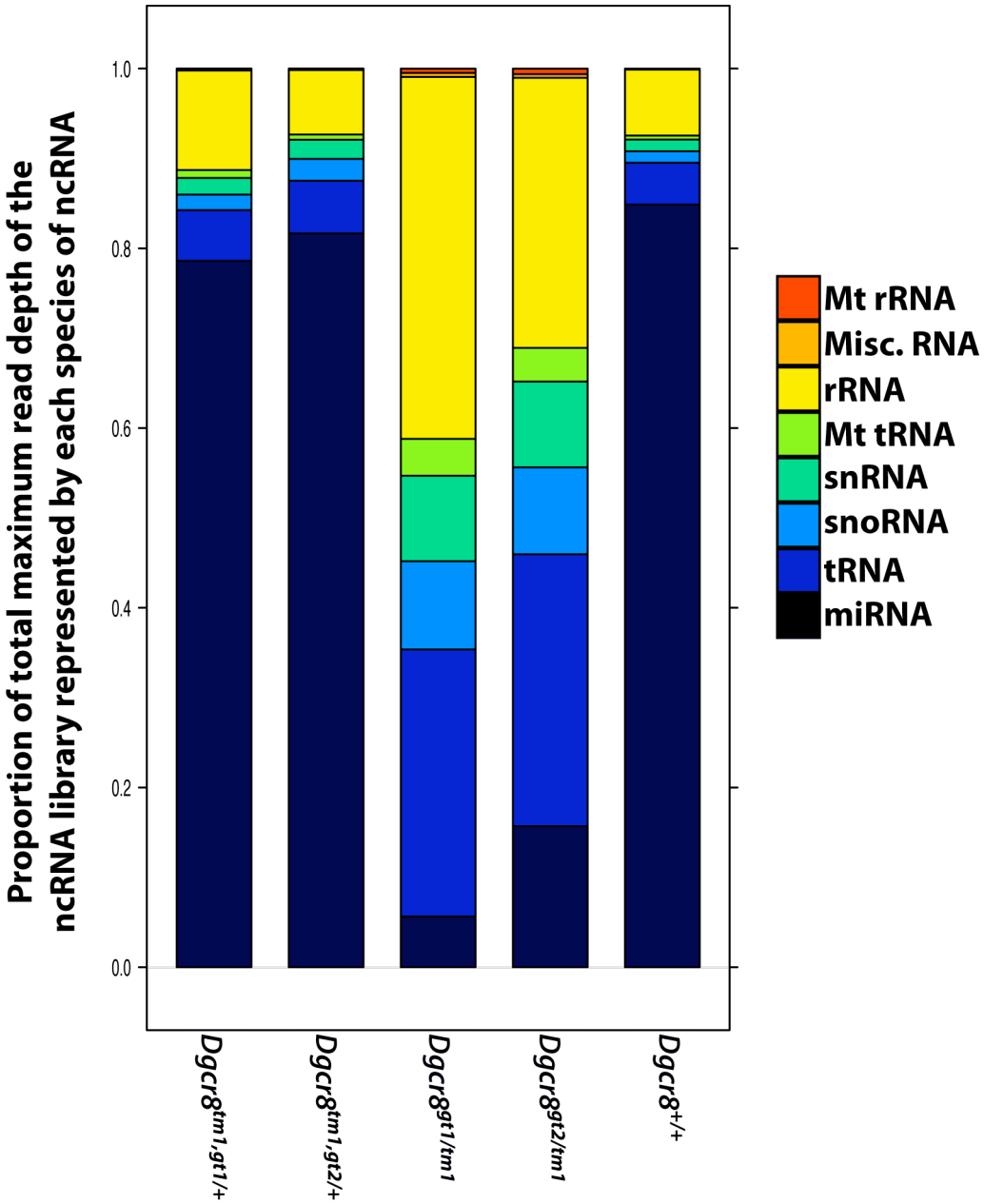


Fig.4.4: The proportion of ncRNA matching reads in each library that are accounted for by each ncRNA type.

In order to confirm that the two *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ cell lines behave consistently, in a manner which not dependent on their cell line of origin, the maximum sequence depths for each of the ncRNAs was compared between the two heterozygotes and the homozygous mutant cell lines (Fig.4.5A). As can be seen in these plots the cell lines with corresponding genotypes behave very similarly. However, it should be noted there is a degree of variation in both cases. This is

particularly obvious in the case of the *Dgcr8*$^{gt/tm1}$ cell line comparison, where *Dgcr8*$^{gt1/tm1}$ is consistently attributed with slightly lower counts for miRNAs relative to the other non-coding RNA species. It is unclear whether this difference is attributable to a genuine difference in the effectiveness of the disruption of the *Dgcr8* locus or is a consequence of a difference in the library preparation in each case. Further replicates of each library would allow an effective estimate of the variance in results due to the experimental procedure and would help to clarify the source of these differences.

When the ncRNA populations for the *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ libraries were compared to the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell line libraries pairwise it was clear that there is a linear relationship between the max sequence depths for the non-miRNA sequences in each case (Fig.4.5B), in addition to a substantial reduction in miRNA associated reads in the case of the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell line libraries (Discussed below). It therefore seems likely that none of these non-miRNA ncRNAs species are DGCR8 dependent. It is of particular note that although Drosha has been proposed to play a role in rRNA processing (Wu et al., 2000), there was no clear dependence on wild type *Dgcr8* for the expression rRNAs in these cells. Therefore all of these DGCR8 independent RNAs were used to normalise the read depths for each sample. This normalization will allow fair cross sample comparisons to judge the effect of the inserted traps on *Dgcr8* function.

A

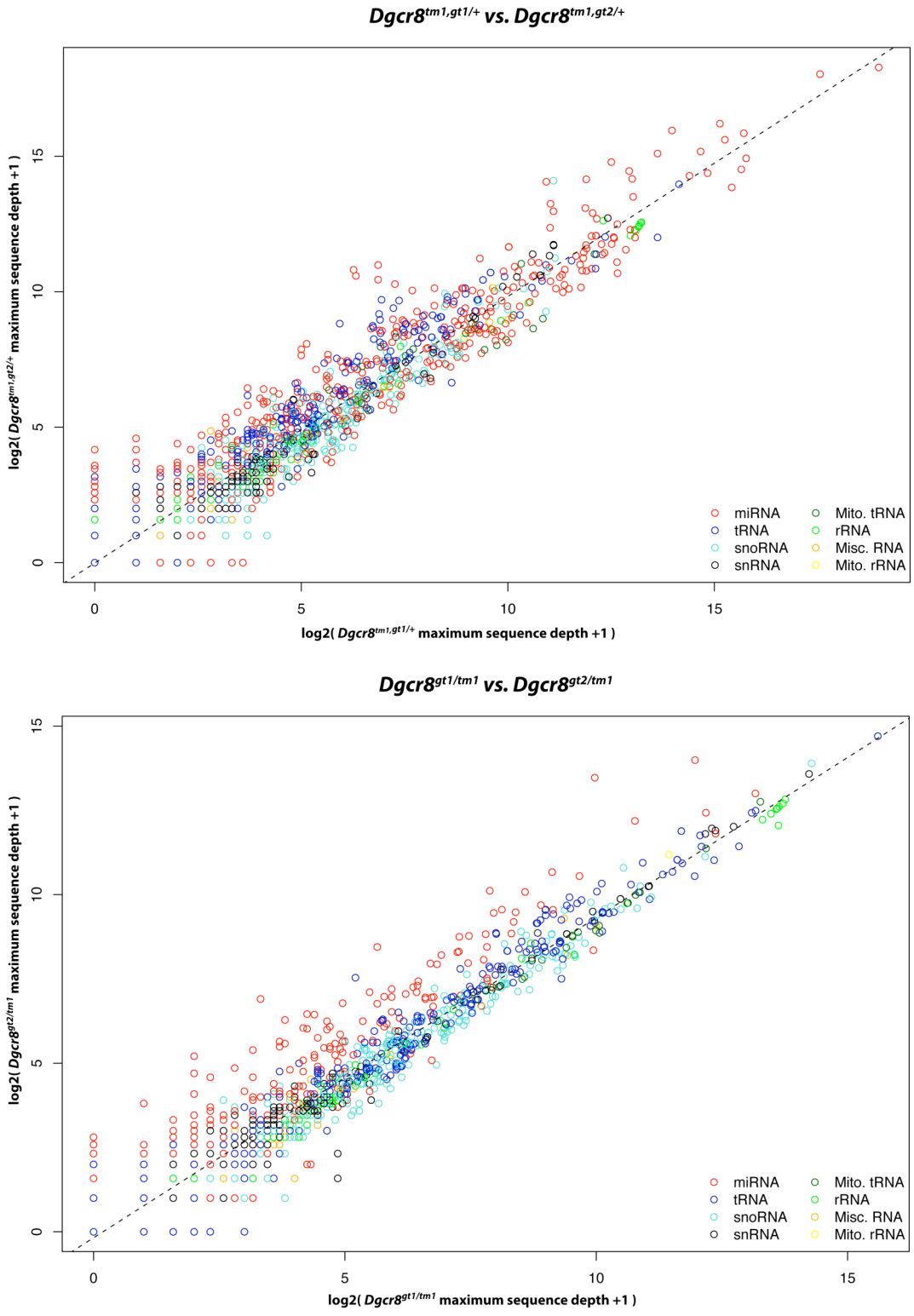


Fig.4.5: A) A Comparison of the small RNA libraries derived from the replicate cell lines. For each cell line maximum sequence depth for individual miRNAs, tRNAs, snoRNAs, rRNAs, snRNAs, mitochondrial tRNAs and rRNAs and the Ensembl class of miscellaneous RNAs are plotted. The dotted line represents a linear model fitted to the non-miRNA sample maximum read depths, based on ncRNAs with a maximum read depth greater than 8 in both samples.

B

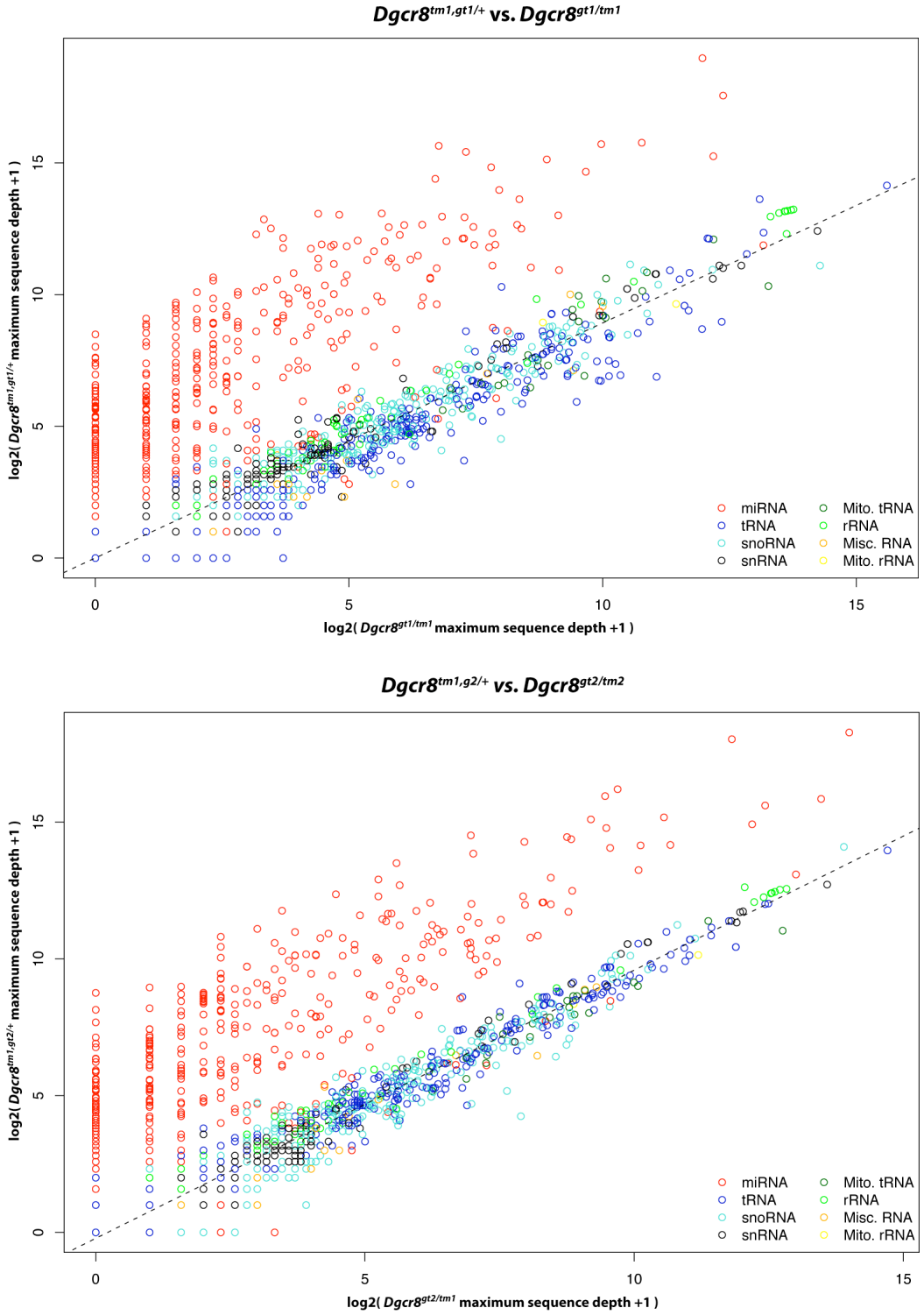


Fig.4.5: B) A wide selection of ncRNAs are not affected by the depletion of functional *Dgcr8* and are thus used to normalise the samples. In a comparison between the maximum read depth (pre-normalisation) of a set of ncRNAs between the *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$cells and the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines, tRNAs, snoRNAs, rRNAs, snRNAs, mitochondrial tRNAs and rRNAs and the Ensembl class of miscellaneous RNAs all demonstrate a clustered linear relationship between the samples. The dotted line represents a linear model fitted to the non-miRNA sample maximum read

depths, based on ncRNAs with a maximum read depth greater than 8 in both samples. It is widely accepted that DGCR8 does not play a role in the processing of the majority of these RNA classes. This analysis demonstrates that DGCR8 independent processing extends to a much larger set of ncRNAs, demonstrating that all of these non-miRNA ncRNA species behave in a similar fashion upon the depletion of DGCR8.

Having normalised the maximum sequence depth values for all of the ncRNAs from each cell line it is clear that there is a fundamental depletion of miRNAs in the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cell lines when compared to the control lines (Fig.4.6) (Appendix A (CD)). In addition, following normalization, comparisons were made between the $Dgcr8^{tm1,gt1/+}$ and $Dgcr8^{tm1,gt2/+}$ libraries and the $Dgcr8^{+/+}$ library (Fig.4.7A) in addition to the comparisons between the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells and their respective control cell line (Fig.4.7B). From these comparisons it is evident that there is a broad reduction in the level of expression of miRNAs in both the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells as would be expected when compared to the expression levels in control cell lines. It also seems apparent that *Dgcr8* is not functionally limiting in the processing of miRNAs as there seems to be a limited change in miRNA expression between the $Dgcr8^{tm1,gt1/+}$ and $Dgcr8^{tm1,gt2/+}$ cells and the $Dgcr8^{+/+}$ control.

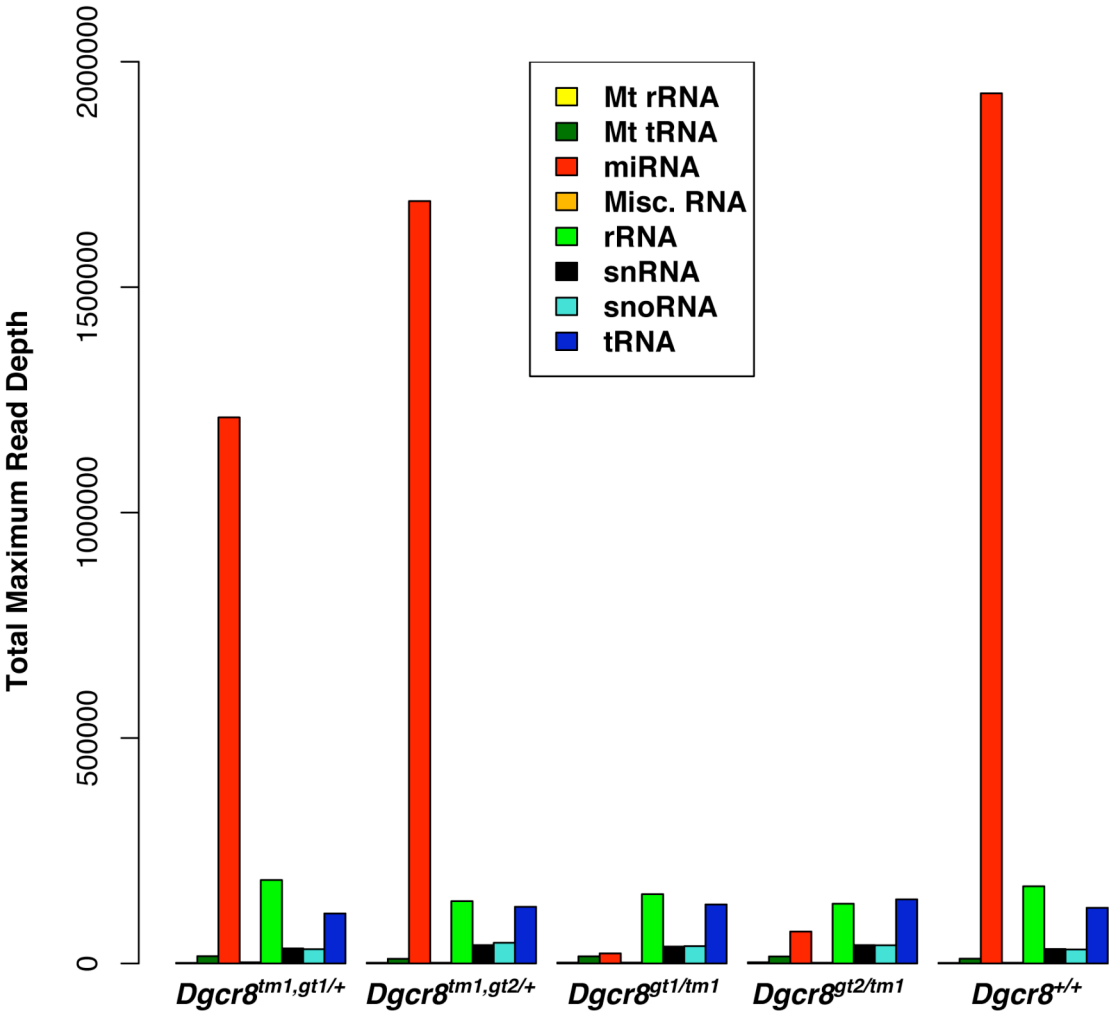
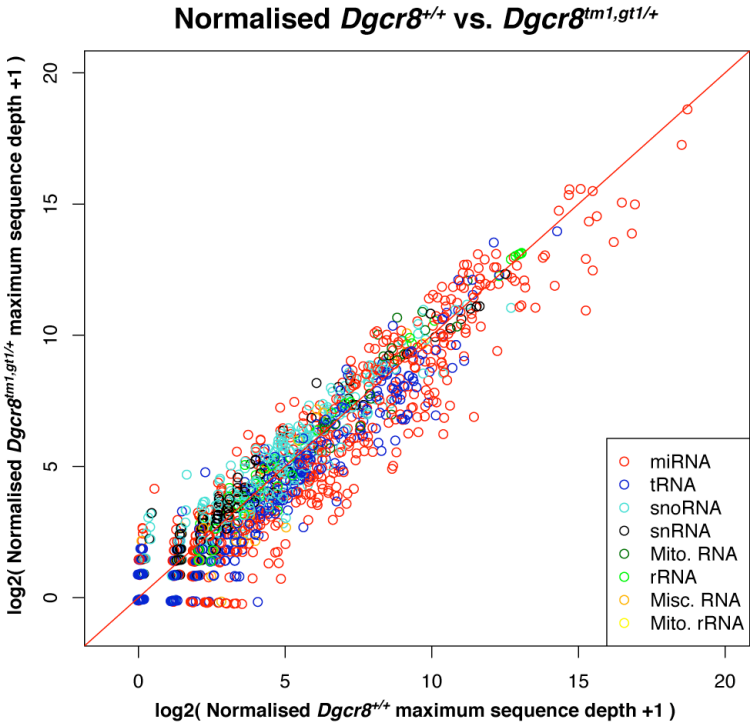
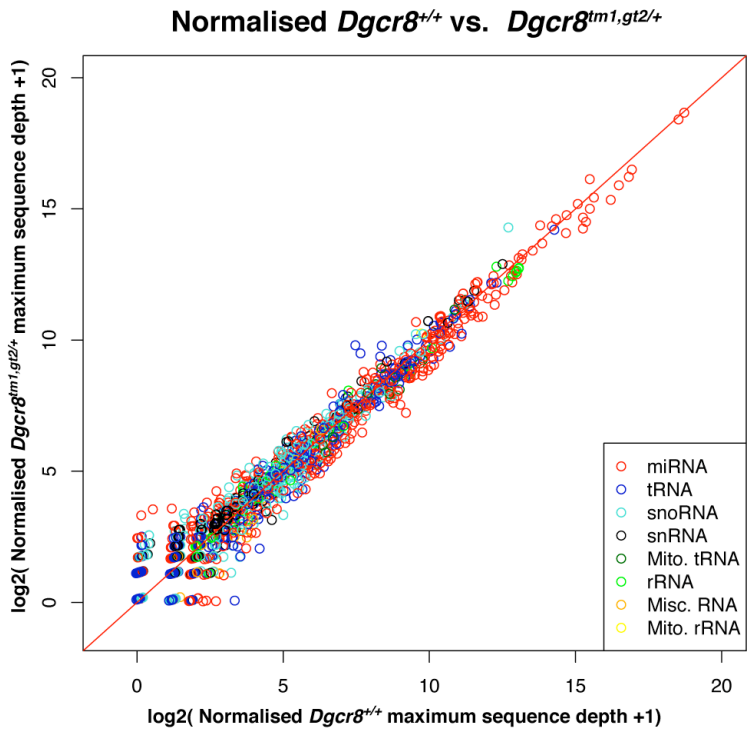


Fig.4.6: Normalised total maximum sequence depths for each ncRNA species in each cell line. Each column represents the total of the maximum sequence depths for a ncRNA species following the loess normalization of the data based on non-miRNA ncRNA species.

A


Normalised Dgcr8+/+ vs. Dgcr8tm1,gt1/+
log2(Normalised Dgcr8tm1,gt1/+ maximum sequence depth +1)
log2(Normalised Dgcr8+/+ maximum sequence depth +1)
0
5
10
15
20
miRNA
tRNA
snoRNA
snRNA
Mito. RNA
rRNA
Misc. RNA
Mito. rRNA


Normalised Dgcr8+/+ vs. Dgcr8tm1,gt2/+
log2(Normalised Dgcr8tm1,gt2/+ maximum sequence depth +1)
log2(Normalised Dgcr8+/+ maximum sequence depth +1)
0
5
10
15
20
miRNA
tRNA
snoRNA
snRNA
Mito. tRNA
rRNA
Misc. RNA
Mito. rRNA

B

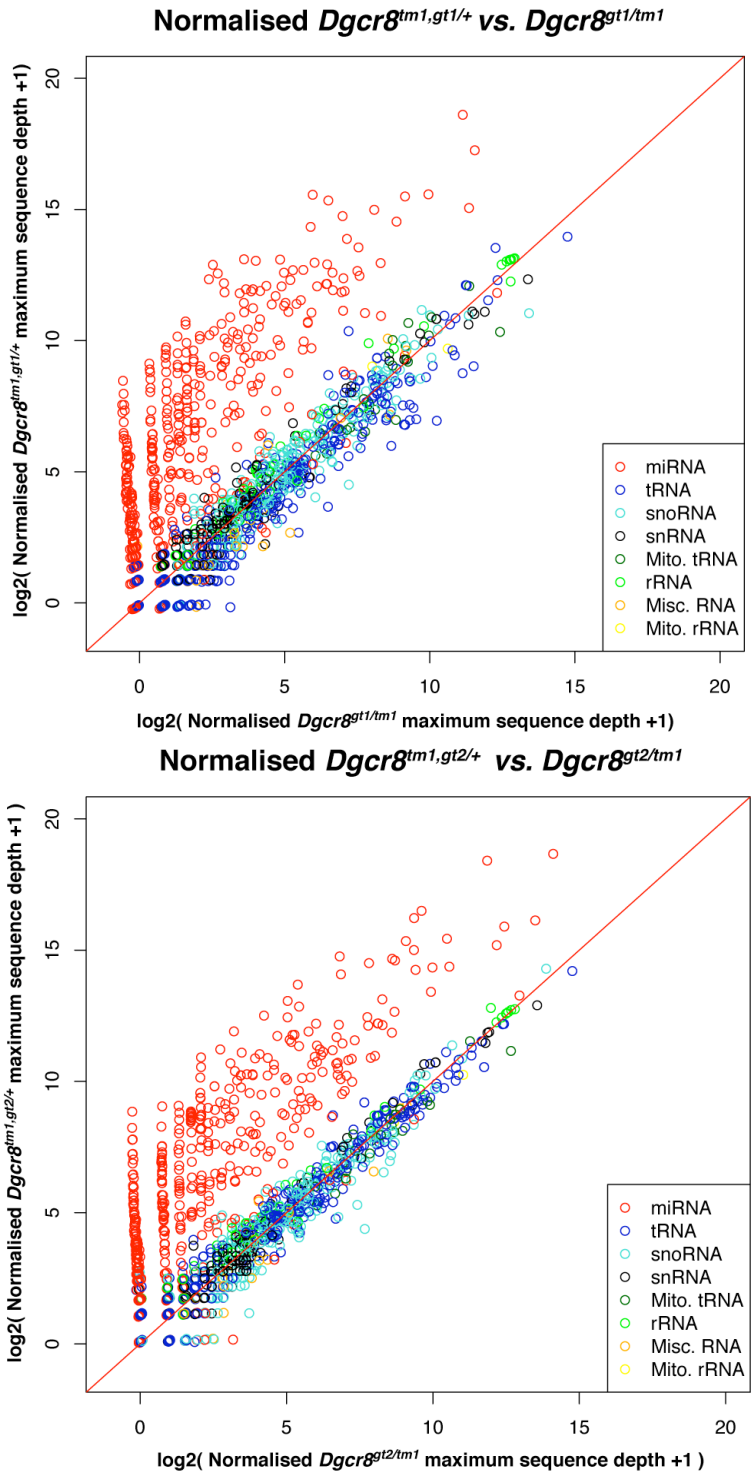


Fig.4.7: The normalised maximum sequence depth for the library of ncRNAs was compared between samples. Following normalisation the maximum sequence depths for each of the ncRNAs in the assembled sequence library were compared in a pairwise fashion. A) The ncRNA read counts were compared between the *Dgcr8*$^{+/+}$ cells and the heterozygous *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ control cells.

B) A comparison between the expression of ncRNAs in $Dgcr8^{tm1,gt1/+}$ and $Dgcr8^{tm1,gt2/+}$ control cells and $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells depleted in functional DGCR8. The red line represents a 1:1 relationship between the samples.

In order to develop a clearer picture as to the significance of the miRNA reduction induced by the insertion of a trap into each *Dgcr8* allele, I considered the $Dgcr8^{tm1,gt1/+}$ and $Dgcr8^{tm1,gt2/+}$ cells and the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells as biological replicates of each other ($Dgcr8^{tm1,gt/+}$ and $Dgcr8^{gt/tm1}$) and conducted a statistical analysis of the expression changes between these two sets (Fig.4.8). The vast bulk of the mature miRNAs are clearly down regulated upon the depletion of DGCR8 (Shifted left, Fig.4.8). This analysis revealed that 414/580 miRBase annotated mature miRNAs are significantly down regulated (adjusted *P*-value < 0.05, LFC = $\log_2(2)$) (Appendix A (CD)), while no miRNAs seem to be significantly up regulated upon DGCR8 depletion. The average maximum read depth included a wide range of expression levels varying from a maximum read depth of greater than 10,000 reads per miRNA (Fig.4.8, orange) down to ~10 reads (Fig.4.8 black). Across this entire expression spectrum the same pattern of down regulation in the $Dgcr8^{gt/tm1}$ is evident.

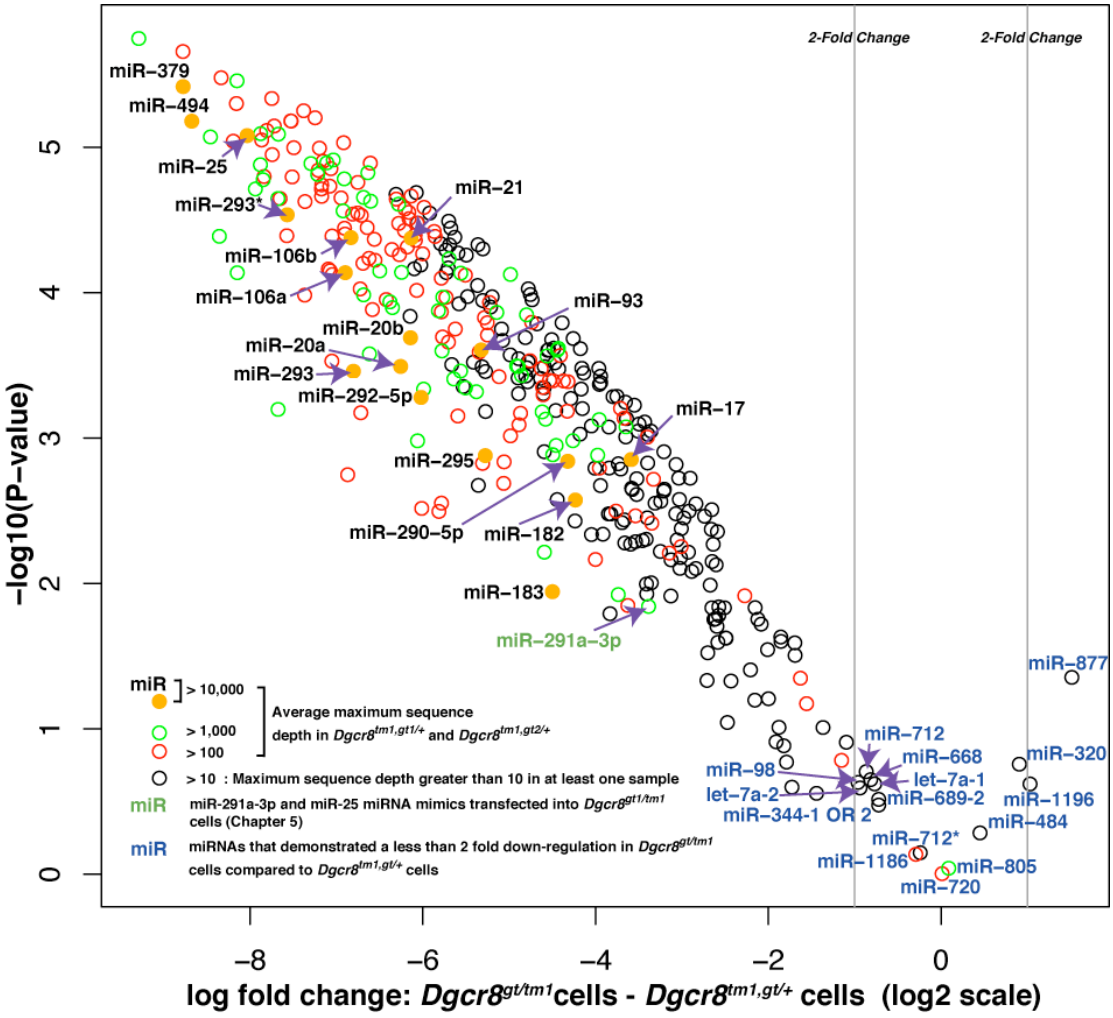


Fig.4.8: A demonstration of the significance of the change in expression of miRBase annotated miRNAs between *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ genotyped cells. A volcano plot comparing the log fold change of miRNA expression, as assessed by maximum read depth, between the *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ cell lines and the statistical significance of these changes. To be plotted miRNAs must have a maximum depth of at least 10 in one cell line library.

It is interesting to consider the prospect that some miRNAs may be processed in a DGCR8 independent way. Indeed there is a clear precedent for this with the discovery of mirtrons (Okamura et al., 2007; Ruby et al., 2007). These miRNAs reside in short introns and are excised from their host transcript through splicing rather than RNase III excision. In addition, further examples of DGCR8 independent miRNAs have also been proposed by Babiarz *et al.* (Babiarz et al., 2008); a study also conducted with *Dgcr8* knockout ES cell lines. Here, considering only miRNAs with a post-

normalisation maximum read depth greater than 10 in any of the samples to remove noise and with the same loose criteria as Babiarz *et al.*, considering miRNAs whose expression changes less than 2-fold following *Dgcr8* disruption as DGCR8 independent, 16 miRNAs maintain a relatively constant expression level in cell lines of both genotypes. Of these miR-320, miR-344, miR-668, miR-877 and miR-484 were all identified as DGCR8 independent by Babiarz *et al.* (Babiarz et al., 2008) (Table 4.2).

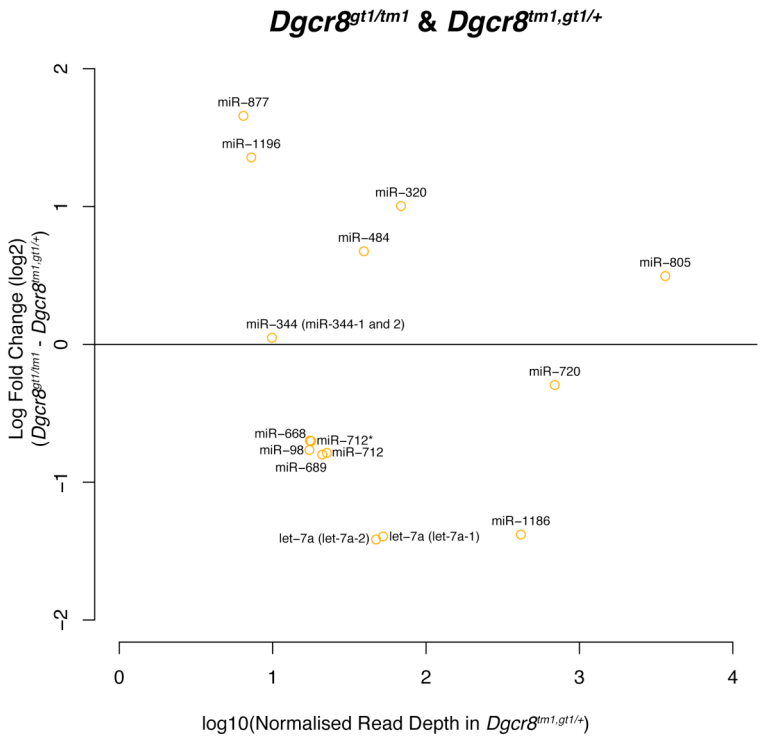
miRNA hairpin	Intronic/Intergenic	Associated Gene	Suggested mirtrons	Suggested DGCR8 independent
miR-320	Intergenic			Babiarz *et al.*
miR-720	Intergenic			
miR-877	Intronic (sense)	Abcf1	Berezikov *et al.*	Babiarz *et al.*
miR-344-1	Intergenic			Babiarz *et al.* ?
miR-344-2	Intergenic			Babiarz *et al.* ?
let-7a-1	Intergenic			
miR-689-2	Intronic (antisense)	Zc3h7a		
miR-1196	Intergenic			
miR-1186	Intergenic			
let-7a-2	Intergenic			
miR-712 (miR-712*)	Not Mapped			
miR-484	Intergenic			Babiarz *et al.*
miR-98	Intronic (sense)	Huwe1		
miR-668	Intergenic			Babiarz *et al.*
miR-805	Intergenic			

Table 4.2: miRNAs that maintain relatively constant expression between the *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ genotyped cell lines. miRNAs with a maximum read depth greater than 10 in any sample and with a negative fold change less than 2 between the *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ cell lines are considered as processed in a potentially DGCR8 independent manner and their details are given here. In addition, intronic status and corroborative information from previous studies is also noted.

In order to determine whether the miRNAs identified as potentially processed in a DGCR8 independent manner in this combined analyisis are consistently identified as

such by both *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{gt/tm1}$ cell line pairs and to gain a clearer indication of the sequence depth of these miRNAs, I conducted a comparison of the expression changes associated with these 16 miRNAs in each case (Fig.4.9). As expected, the majority of these miRNAs appear to comply with the criteria I have used to define this set in both cases. However, some fall outside the cutoff in one of the cases. These include miR-344 (miR-344-1 and miR-344-2), let-7a (hairpin let-7a-1 and let-7a-2) and miR-1186. More stringent criteria could be used to remove these from the set of interest, however miR-344 has also been identified by Babriaz *et al.* as processed in a DGCR8 independent fashion. Interestingly the read depth associated with these miRNAs spans several orders of magnitude (Fig.4.9 and Appendix A (CD)).

A



B

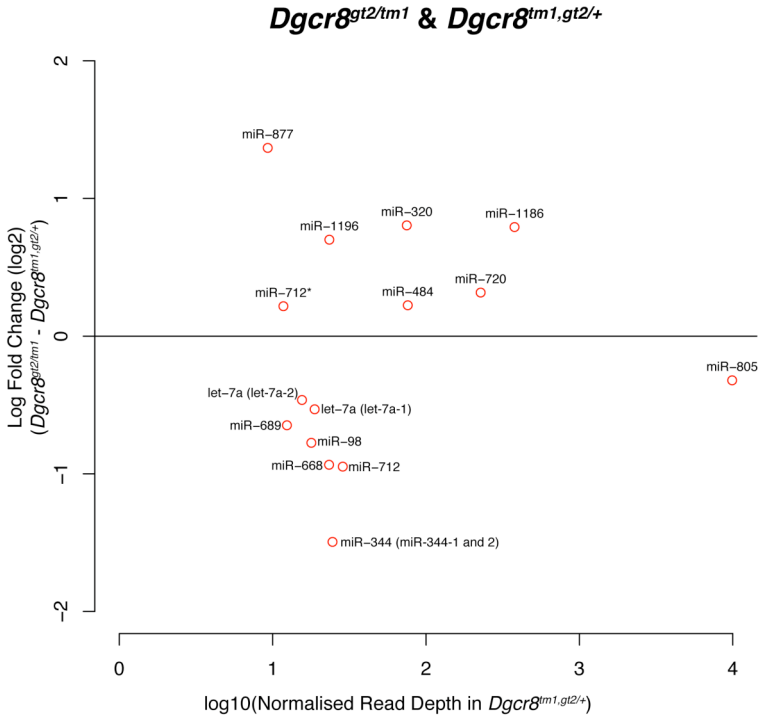


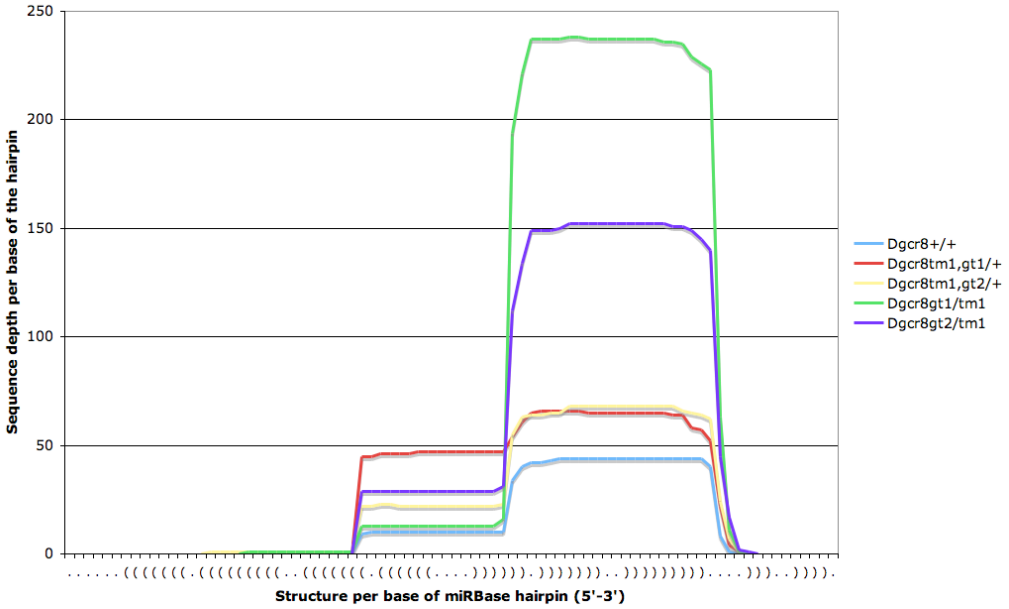
Fig.4.9: The maximum read depth and log fold changes of miRNAs that are potentially processed in a DGCR8 independent manner in pairwise comparisons between each *Dgcr8*$^{gt/tm1}$ cell line and their respective heterozygous control. Each spot represents the expression of an individual miRNA. Where a miRNA can be derived from multiple hairpins the hairpin number is included in brackets. A) A comparison between the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{tm1,gt1/+}$ cell lines. B) A comparison between the *Dgcr8*$^{gt2/tm1}$ and *Dgcr8*$^{tm1,gt2/+}$ cell lines.

Previously, *miR-877* has been proposed as a mirtron (Berezikov et al., 2007) and resides in the intron of *Abcf1*. Together with the work of Babiarz *et al.* here I confirm that this miRNA does appear to be expressed via a mechanism that does not require *Dgcr8*, confirming its identity as a mirtron. Although not significantly up regulated, this miRNA had the greatest positive log fold change following the depletion of DGCR8 activity. This increase in expression when DGCR8 is depleted may be due to a lack of competition for the remainder of the processing machinery from canonically processed miRNAs precursors leading to more efficient release of *miR-877* mature miRNAs. It could also be caused by a lack of competition from other miRNAs for the miRNP complexes that may protect mature miRNA sequences from nucleolytic degradation.

Babiarz *et al.* noticed that the vast majority of reads mapping to the *miR-320* hairpin were associated with the 3' arm (Babiarz et al., 2008). The sequence reads which map to this hairpin from my libraries do not refute this observation, with the greatest sequence depth clearly associated with the 3' hairpin arm (Fig.4.10A). However, the read depth generated from my experiments is not as great as that published by Babiarz *et al.* so it is difficult to be as confident of the strand bias. The strand bias is much clearer in the Babiarz *et al.* libraries (Fig.4.10B). Babiarz *et al.* hypothesise that the concentration of miR-320 hairpin matching sequence reads to the 3' end side of the miR-320 hairpin may be the result of a lack of the 5' phosphate group at the 5' end of the hairpin, which is left by RNase III like cleavage of the primary sequence and which is required for successful cloning. This supports the notion of an alternative mechanism for processing the pri-miRNA. A Northern Blot confirmed the expression

of miR-320 in *Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines (Fig.4.11). Despite the faint signal it seems that there is greater expression of this miRNA in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines. This further confirms the use of Illumina high through put sequencing as a quantitative indication of expression.

A



B

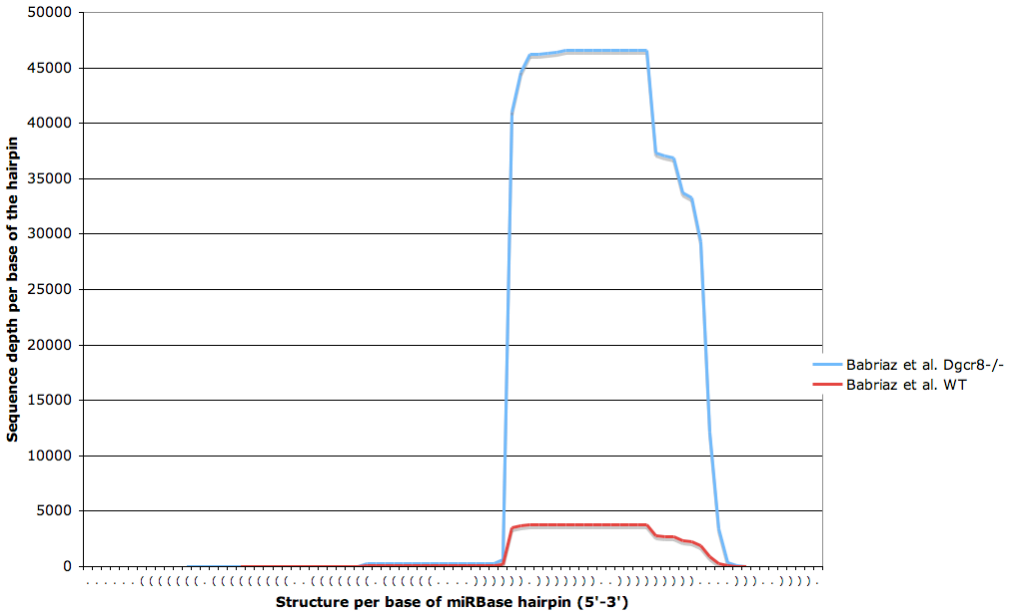


Fig.4.10: The maximum depth of reads which match the miR-320 hairpin sequence. A) The depth of reads across the miR-320 miRBase annotated hairpin from libraries generated as part of this study. B) The depth of sequence reads across the miR-320 miRBase annotated hairpin from libraries generated by Babiarz *et al.* (Babiarz et al., 2008).

Babiarz *et al.* also saw a similar strand bias associated with the *miR-484*. It is difficult to compare my results for this miRNA currently as Babiarz *et al.* used a miRNA precursor hairpin that differs from the miRBase annotation; basing their new version

on local secondary structure. In order to address these differences I would need to map all the sequence reads against the genome or a different hairpin structure. This is beyond the scope of the current study.

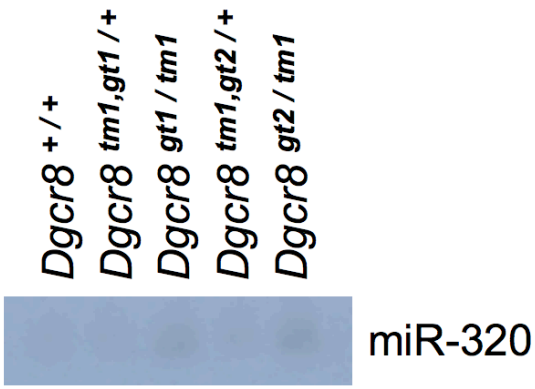


Fig.4.11: A Northern blot for miR-320 expression in *Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$, *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines.

A further set of miRNAs, previously unidentified as DGCR8 independent were also selected by these loose criteria as potentially unaffected by the loss of *Dgcr8* function, in my experiments. These are let-7a-1, let-7a-2, miR-98, miR-720, miR-689-2, miR-712, miR-1196, miR-712*, miR-1186, miR-805 and miR-344-1 or 2 (Babiarz *et al.* did not define which loci of miR-344 they suspected to be DGCR8 independent (Babiarz et al., 2008)) (Table 4.2). Due to the nature of Illumina/Solexa sequencing and miRNA families it is not always possible to determine the locus from which a miRNA sequence will have been derived as multiple miRNA loci can comprise a set of highly related precursors and as a result sequence reads can match each equally (e.g. *let-7a-1* and *-2*). Where this is the case we halved the number of identical sequences that map to each loci and add this number to the sequence depth at each. While this is not an ideal solution to the problem it seemed like the simplest and fairest method by which to divide the identical reads between highly similar loci. As a consequence, however, it is not entirely clear as to which locus the let-7a miRNA

reads will have necessarily derived from and which primary sequences, if either, will likely be processed in a DGCR8 independent manner.

Of this remaining set of miRNAs that fall within my criteria for DGCR8 independence, the majority are intergenic (7/11) and hence are not expected to be mirtronic in their origin. The two miRNAs miR-689-2 and miR-98 are both intronic, but the miR-689-2 miRNA is antisense with relation to the *Zc3h7a* gene and the miR-98 locus is within a large intron of the *Huwe-1* so, again, neither are expected to be processed as mirtrons. Finally the locus associated with the miR-712 hairpin is not annotated in miRBase. Further investigation of these loci and continued experimentation would be required to confirm the enzymatic dependencies of these miRNAs and to identify alternative routes of miRNA processing.

4.4 Discussion

Through the use of the Luminex bead-based miRNA expression-profiling system I have demonstrated the broad depletion of ES cell associated mature miRNAs in the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cell lines. I have then confirmed this knockdown through the generation and sequencing of small RNA cloned libraries with Illumina/Solexa high through put sequencing technology. The latter method offers several improvements over the former probe based technology including a greatly increased sensitivity and fewer specificity-associated issues. The sequencing results therefore clarified the differential expression of a far larger set of miRNAs, in a fashion that was not possible with the bead experiment alone due to the inherent noise associated with the probe based Luminex method.

The sequence results demonstrate a broad and inclusive reduction of mature miRNA expression in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines. The vast majority of miRNAs are significantly depleted in these cell lines confirming that the insertion of a trap into each *Dgcr8* allele has depleted DGCR8 function in these cells. These results concur with the results of the miRNA Northern blots presented in Chapter 3 which demonstrate a reduction in the expression of a more select group of ES cell expressed miRNAs (Fig.4.12 reference to Fig.3.6 and Fig.3.9). However, sequence reads from canonically processed miRNAs are not totally absent despite the depletion of DGCR8 and there is a low base level of miRNA expression in the *Dgcr8*$^{gt/tm1}$ cells. It is possible that there is a small amount of residual wild type *Dgcr8* expression in these cells, undetected by mRNA Northern blots, from transcripts that splice over the traps from upstream exons to downstream exons. This may explain why there is not a complete absence of mature canonical miRNA expression in these cells, as judged by library sequencing, as residual DGCR8 may process some canonical miRNAs. It is also plausible that there may be alternative processing pathways or dsRNA binding proteins that are able to compensate for the loss of DGCR8 to some extent. Finally due to the sensitivity of the method it is not inconceivable that low level cross-library contamination could lead to residual miRNA associated sequences in the *Dgcr8*$^{gt/tm1}$ samples.

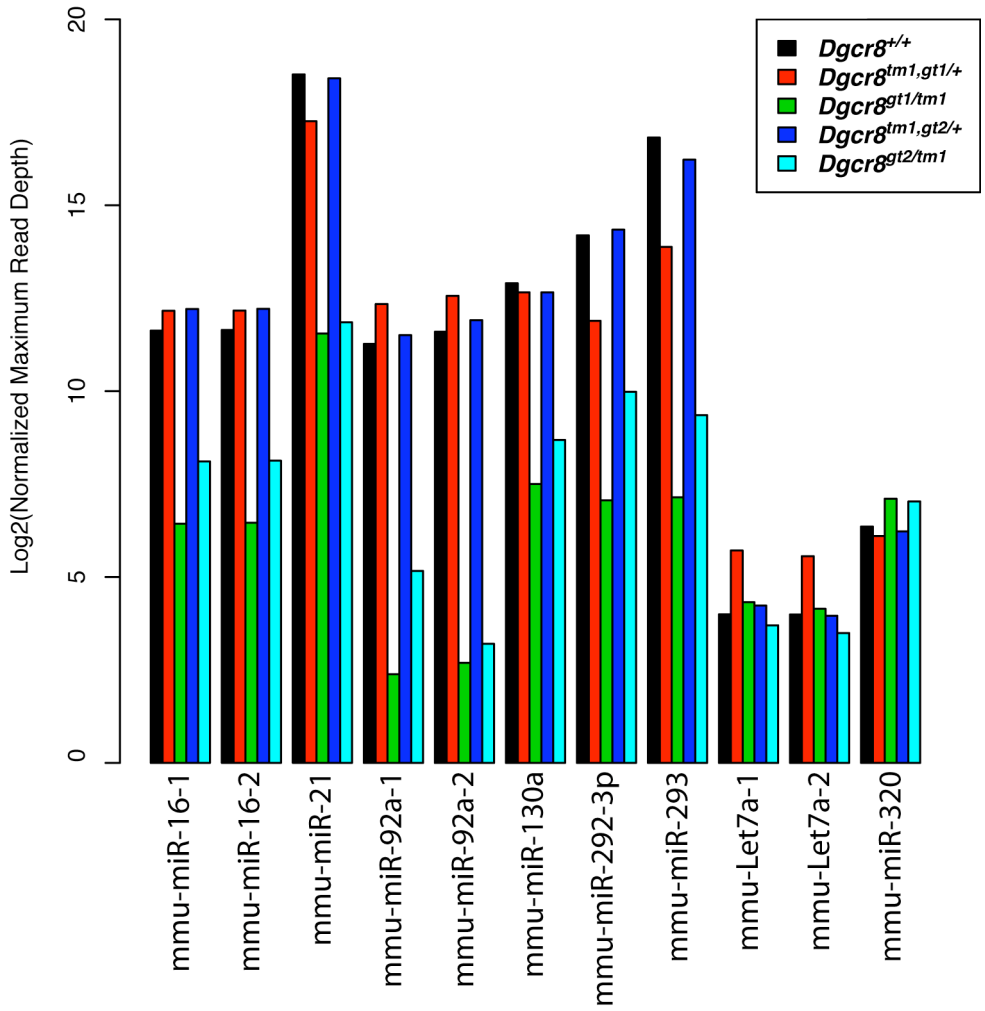


Fig..4.12: Maximum read depth for all miRNAs for which the expression pattern has been demonstrated by Northern blot in this thesis. Normalised maximum read depth for each of the miRNAs tested by Northern Blot (Fig.4.11, Fig.3.6 and Fig.3.9) in *Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{tm1,gt2/+}$ and *Dgcr8*$^{gt2/tm1}$ cells.

In addition to demonstrating the substantial depletion of the majority of miRNAs, the sequencing experiment seems to suggest that both the vast majority of other ncRNAs amongst the set in the sequence libraries used to map Solexa data, and a small number of miRNAs are processed in a DGCR8 independent manner. With reference to the broader set of ncRNAs, this result is potentially significant as a dsRNA binding protein may be expected to be involved in further RNA processing pathways, especially since Drosha (its RNase III partner protein) has been implicated in rRNA processing (Wu et al., 2000).

The DGCR8 independent miRNAs include miRNAs that have been identified in a previous study as processed in the absence of this protein in mouse ES cells. These include miR-320, miR-344, miR-668, miR-877 and miR-484. Of these miR-877 is likely to be a mirtron, released from its primary transcript by the mRNA splicing mechanism (Babiarz et al., 2008; Berezikov et al., 2007). It is also intriguing to note that recently miR-320 has been proposed as capable of directing transcriptional repression (Kim et al., 2008a). Consequently it would seem that this miRNA does not conform to a number of canonical rules for miRNA processing and function and as such is deserving of further research.

In addition to these miRNAs, let-7a, miR-98, miR-720, miR-689, miR-712, miR-1196, miR-712*, miR-1186 and miR-805 are also potential DGCR8 independent candidates. These miRNAs were identified with a loose set of criteria (Greater than a maximum sequence depth of 10 in at least one cell line and less than a 2 fold down-regulation of mature miRNA levels in the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cell lines). More stringent criteria would refine this list, but it may be interesting to initially consider this complete set of miRNAs for further study.

In this chapter I have confirmed the functional significance of the traps inserted into the *Dgcr8* locus in Chapter 3. Subsequently it is possible to use these cells for miRNA re-addition experiments in an environment depleted of endogenously expressed miRNAs. On conducting such experiments I can be confident that, to a very large extent, interference in post re-introduction expression profiles from miRNA-target saturation and complicated expression networks caused by miRNAs which target the same transcripts will have been dramatically reduced.