Chapter 3: Disrupting the *Dgcr8* locus

3.1 Aim

In order to establish a cell based system with which to identify miRNA target genes in the absence of endogenous miRNAs expression, I wished to generate cell lines that were impaired in their ability to process miRNAs. Therefore the aim of this chapter is to describe the creation of mouse ES cell lines bearing a gene trap construct in each allele of the *Dgcr8* gene truncating *Dgcr8* transcripts and abrogating DGCR8 function. These lines were characterised at the molecular and phenotypic level and compared to similar lines described by previous studies.

3.2 Introduction

To create a cellular system for the identification of miRNA targets in a fashion resembling the work of Giraldez *et al*. in the zebrafish (Giraldez et al., 2006), the first step is to disrupt the maturation of endogenous miRNAs. If successful, this will allow the re-introduction of miRNAs individually into a background that will not be conducive to combinatorial regulation or functional redundancy.

miRNAs are embedded within longer RNA molecules, pri-miRNAs, that are processed by two rounds of RNase III digestion. The first enzyme, Drosha, operating in concert with a dsRNA binding protein, DGCR8, releases the miRNA precursor molecule as a hairpin. This hairpin is exported to the cytoplasm where it is further processed by the second enzyme, Dicer, removing the loop of the hairpin and liberating the mature miRNA as one strand of the

hairpin stem. At the time of project inception, *Dgcr8* seemed to be the most likely candidate gene for the generation of a cell line with a specific interruption to the miRNA-processing pathway, because its function appeared to be restricted to miRNA processing. In contrast, in addition to its role in miRNA processing Dicer is known to be involved in the cleavage pathway required for the production of siRNAs (Bernstein et al., 2001; Hutvagner et al., 2001), while Drosha was initially proposed as an enzyme involved in rRNA processing and was only later ascribed a role in miRNA biogenesis (Wu et al., 2000). The identification of DGCR8 as a probable miRNA specific processing enzyme has since been supported by others working in the field (Wang et al., 2007).

Dgcr8 is required for the canonical processing of miRNAs (Gregory et al., 2004). DGCR8 contains 2 dsRNA binding domains and a WW domain (Fig.3.1). By deleting subregions, a structural analysis of the DGCR8 protein has been conducted (Yeom et al., 2006). This analysis revealed the C-terminus (Residues 739-750) is required for its association with Drosha. Point mutations and further deletion experiments found that both dsRNA binding domains are required to effectively bind the pri-miRNA. Finally the N-terminus of the protein was found to be responsible for nuclear localisation.

Gene trap mutagenesis is reviewed by Stanford *et al.* (Stanford et al., 2001). In brief, the principle is to insert an exon cassette (the "gene trap") containing a splice acceptor, a selectable marker and polyadenylation site randomly into the genome. Insertions into an intron of an expressed gene, in the correct orientation, will activate the selectable marker through splicing into the host transcript. The polyadenylation site of the gene trap will cause the truncation of the host transcript and lead to the expression of a truncated host protein

fused to the selection marker. In most cases, this will abrogate the host gene's function. The host gene into which the gene trap has inserted can be identified by the sequencing of Rapid amplification of 5' cDNA ends (5'RACE) products, generated with primers specific to the gene trap exon.

There has been an ongoing international effort to disrupt a sizeable fraction of the genes in the mouse genome by gene trap mutagenesis (International Gene Trap Consortium (IGTC)). Currently, this resource covers ~36% of the genes annotated in v52 of Ensembl. Cell lines bearing these mutations in a single allele are available as a public resource, identifiable through both the Ensembl genome browser and publicly accessible websites (Nord et al., 2006; Stryke et al., 2003). In order to disrupt the *Dgcr8* gene, I opted to select mouse ES cell lines with a gene trap in the *Dgcr8* locus from BayGenomics, a contributor to the IGTC (Nord et al., 2006; Stryke et al., 2003), and to disrupt the second *Dgcr8* allele with a targeted trap.

The targeted trapping approach that I intended to use to disrupt the second allele has been previously described by Friedel *et al.* (Friedel et al., 2005). Gene trap cassettes are targeted to a region of interest through homologous recombination, no longer relying on the random insertion of the trap into gene structures. This approach is highly efficient for genes expressed in ES cells and allowed me to mutate both *Dgcr8* alleles.

As I will be adding miRNAs back into this system following the disruption of the *Dgcr8* locus in order to identify ES cell miRNA targets, it is important that the mutant ES cells are not only depleted of mature miRNAs, but also do not differentiate as a result of the loss of

mature miRNA expression and hence retain some ES cell properties. Using an undifferentiated ES cell line would limit the secondary effects of miRNA depletion and increase the likelihood of ES cell miRNA targets retaining their expression in the developed system. Broad differentiation would bring with it a broad change in the cellular mRNA expression profile and a reduced likelihood that the ES cell miRNA targets would still be expressed. As previous studies have successfully knocked out Dicer and depleted miRNAs without triggering differentiation, I was confident that this combination of properties was achievable (Kanellopoulou et al., 2005; Murchison et al., 2005). In this chapter, in addition to generating a cell line depleted in DGCR8, I also attempt to address whether ES cell properties are retained.

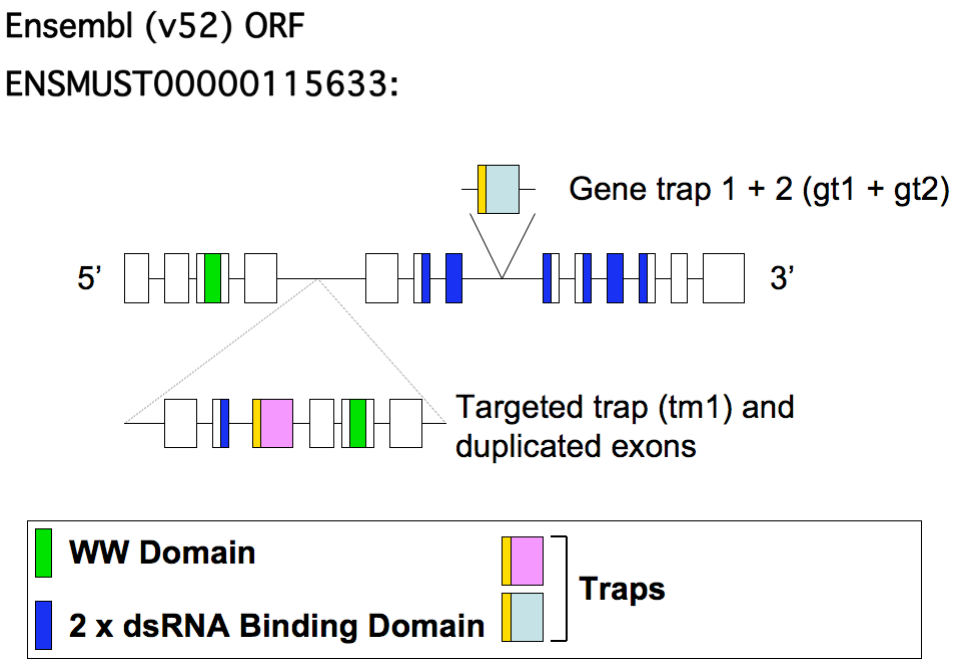
3.3 Disrupting the *Dgcr8* locus

3.3.1 Experimental design

The first step to disrupting *Dgcr8* function in mouse ES cells was to select ES cell lines from the IGTC gene trap resource that contain a random gene trap insertion in one allele of the *Dgcr8* gene. The position of the gene trap in each cell line was confirmed by RT-PCR. The cell lines were subsequently sub-cloned to ensure their homogeneity. Next, I constructed a targeting vector that would insert a gene trap cassette specifically into the second *Dgcr8* allele in these cell lines. The process involves the cloning of a PCR amplified genomic fragment of the *Dgcr8* locus into a vector containing Gateway sites. This fragment was subsequently transferred to a Gateway modified insertion type targeting vector using L/R clonase. The gene trap was designed to integrate into the homologous target sequence by a gap repair mechanism. A gap was introduced into the homologous region and the linearised plasmid construct was then electroporated into each cell line. Additionally this insertion

should lead to the duplication of the target sequence. Cell lines were screened to identify clones that contain the randomly inserted gene trap in the first allele and a targeted trap in the second allele, which were then selected. If successful, this strategy should lead to the truncation of both *Dgcr8* transcripts and the production of a *Dgcr8* hypomorphic or null cell line.

A



B

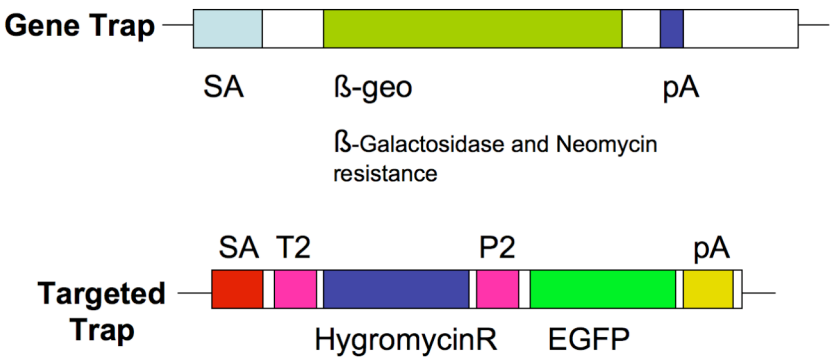


Fig.3.1: A) The structure of the ORF of the *Dgcr8* gene and the positions of inserted traps relative to protein domains. The boxes represent exons and the lines introns. Exons coding for the various protein domains of the DGCR8 protein are shaded. The positions within this gene structure that the BayGenomics gene traps are inserted (gt1 + gt2, light blue) and the intended target site of the targeted trap (pink) are indicated. In addition the exons duplicated by the targeted trap insertion are depicted alongside this trap. Following the insertion of the gene traps the DGCR8 peptide will be truncated and the 3' domains will not be expressed.
B) The structure of the gene trap cassettes used to disrupt the *Dgcr8* locus. Both traps contain a splice acceptor, to ensure splicing into the *Dgcr8* transcript, and a poly-adenylation site to truncate the transcript once inserted. The BayGenomics gene trap (top) contains a region coding for a βgeo peptide. When translated this will form a fusion protein with the remaining upstream DGCR8 domains, thus conferring G418 resistance to the

cell line, and staining the cells blue with the addition of Xgal reagents. The targeted trap (bottom) contains T2 and P2 regions. These are of viral origin and cause the elongating ribosomes to skip during translation of the fusion transcript, releasing the upstream peptide and beginning a new peptide from this point. Subsequently independent hygromycin resistance peptides and enhanced green fluorescence protein (EGFP) peptides are translated from this trap. If the two traps were to be inserted into the same allele of *Dgcr8* it would be expected that the expression of selectable markers would be limited to those coded by the upstream trap as the transcript would no longer splice into the downstream cassette. This allows this targeted insertion event to be distinguished from one in which traps are inserted within separate alleles using Xgal staining, which requireβ-galactosidase expression.

	Position in ENSMUST00000115633	Position in ENSMUST00000009321	Primer Orientation
Features			
Bay Genomics Gene Trap	Between Exons 9 and 10	Between Exons 8 and 9	NA
Duplicated Region	Exon 4 to Exon 8	Exon 3 to Exon 7	NA
Targeted Trap Cassette	Between Exons 8 and 9	Between Exons 7 and 8	NA
Primers			
1	Exon 9	Exon 8	5' -> 3'
3	Exon 8	Exon 7	5' -> 3'
21	Exon 3	Exon 2	5' -> 3'
22	Exon 3	Exon 2	5' -> 3'
2	BayGenomics Gene Trap	BayGenomics Gene Trap	3' -> 5'
4	BayGenomics Gene Trap	BayGenomics Gene Trap	3' -> 5'
23	BayGenomics Gene Trap AND Tageted Trap Cassette	BayGenomics Gene Trap AND Tageted Trap Cassette	3' -> 5'
24	BayGenomics Gene Trap AND Tageted Trap Cassette	BayGenomics Gene Trap AND Tageted Trap Cassette	3' -> 5'
5	Exon 11	Exon 10	3' -> 5'
6	Exon 11/12	Exon 10/11	3' -> 5'

Table 3.1: A description of the relative positions of important features within the structure of the *Dgcr8* transcript. Noted are the positions of the gene trap cassettes and the region duplicated as a consequence of the insertion of the second targeted trap vector. In addition, the relative positions and orientations of the primers used to check the gene structure by RT-PCR are also included.

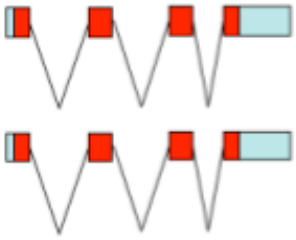
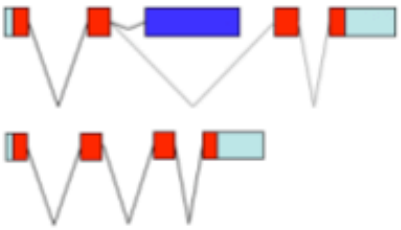
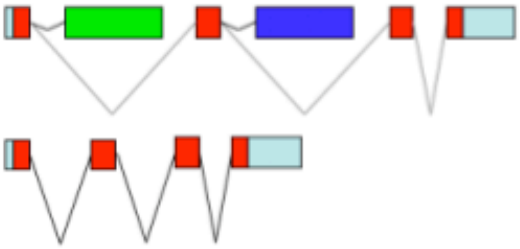
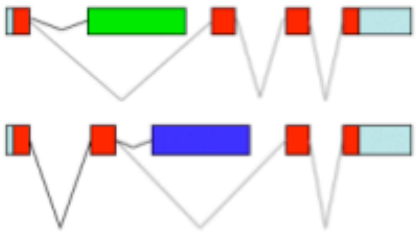
Cell line genotype	Nomenclature when considering cells with a similar genotype	Schematic of *Dgcr8* transcript structures
$Dgcr8^{+/+}$	NA	
$Dgcr8^{gt1/+}$	$Dgcr8^{gt/+}$	
$Dgcr8^{gt2/+}$		
$Dgcr8^{tm1,gt1/+}$	$Dgcr8^{tm1,gt/+}$	
$Dgcr8^{tm1,gt2/+}$		
$Dgcr8^{gt1/tm1}$	$Dgcr8^{gt/tm1}$	
$Dgcr8^{gt2/tm1}$		

Table 3.2 A summary of the genotype nomenclature used throughout this thesis. The first column lists the genotype of each cell line used in this study. The cell lines that are derived from the same original BayGenomics cell line contain the same initial gene trap ("gt1" OR "gt2"). "tm1" refers to the targeted trap insertion. The second column lists the nomenclature used to refer to sets of cell lines with similar genotypes when analyses are performed that consider two independently derived cell lines as biological replicates and in which their data is combined. In these cases the number is removed from the gene trap names. The third column provides a schematic representation of the gene structures at the *Dgcr8* locus. Red boxes represent exons, light blue boxes are UTRs, lines represent introns, dark blue boxes refer to gene traps and green boxes to targeted traps.

3.3.2 Selecting *Dgcr8* trapped cell lines

Initially, 2 mouse ES cell lines containing a gene trap within the *Dgcr8* locus were selected from the IGTC website (http://www.genetrap.org). These would subsequently form the basis of the second allele targeting experiments. All available IGTC gene trapped cell lines are annotated in Ensembl as a DAS track, with the trap in each cell line mapped to the genome through the sequencing of 5'RACE products, to identify exons upstream of the intron into which the gene trap is inserted. To maximise the possibility that these gene traps create a null allele, cell lines were selected that contain gene traps positioned as near as possible to the 5' end of the *Dgcr8* transcript. Consequently, truncated fusion protein products are unlikely to retain wild type function. The two cell lines selected (XG058 and XH157) are independent gene trapped cell lines derived from separate trapping experiments. However, both of these cell lines contain a gene trap within the same intron of *Dgcr8* (Fig.3.1A).

To ensure that the position of the traps had been annotated correctly, the 5'RACE sequences provided by BayGenomics were analysed by BLAST against mouse cDNA sequences in Ensembl (v31, m33) using default settings. As expected, the RACE sequences from each gene trap aligned to exons 6-8 of the *Dgcr8* Ensembl transcript (ENSMUST00000009321), confirming their correct annotation by the BayGenomics pipeline.

Next a manual comparison of the gene trap insertion site with the Pfam protein domain structure of the *Dgcr8* gene (Ensembl v31, m33) revealed that both gene traps are inserted between exons coding for the first of DGCR8's two double stranded RNA binding domains. As a result, it is expected that truncating the endogenous protein at this point and creating a

fusion protein should have a catastrophic effect on the function of DGCR8 protein coded by mutant transcripts from these trapped alleles (Yeom et al., 2006).

3.3.3 Confirmation of *Dgcr8* gene trapped cell lines.

To confirm that the selected gene trap lines were generating the expected fusion transcript from the disrupted *Dgcr8* allele and a wild type transcript from the unaffected allele, a RT-PCR approach was adopted. Nested primers were designed for RT-PCR to amplify cDNAs expressed from the gene trap and wild type alleles (Fig.3.2). A set of 5' primers that anneal to an exon upstream of the gene trap was partnered to a set of primers within the gene trap cassette and to a further set of primers that anneal to an exon downstream of the gene trap insertion site. As a positive control, a further set of primers were designed to bind to exons either side of a 187bp intron in the *Arsa* housekeeping gene. These primers would allow me to perform a control amplification to judge genomic contamination of the cDNA used in the RT-PCR reaction, as contamination would result in a PCR product of a larger than expected size.

RNA was purified from the two independently trapped BayGenomics cell lines (henceforth known as $Dgcr8^{gt1/+}$ and $Dgcr^{gt2/+}$ (Table 3.2)). The RNA was quantified and subsequently reverse transcribed into cDNA. PCR reactions using the primers described above were used to confirm the identity of each cell line and the position of the gene trap (Fig.3.2). All products were of the expected size, implying that the gene traps were indeed within intron 8 of the Ensembl transcript ENSMUST00000009321, although there was evidence of a significant second amplification product of a larger than expected size in the wild type transcript lane (see below).

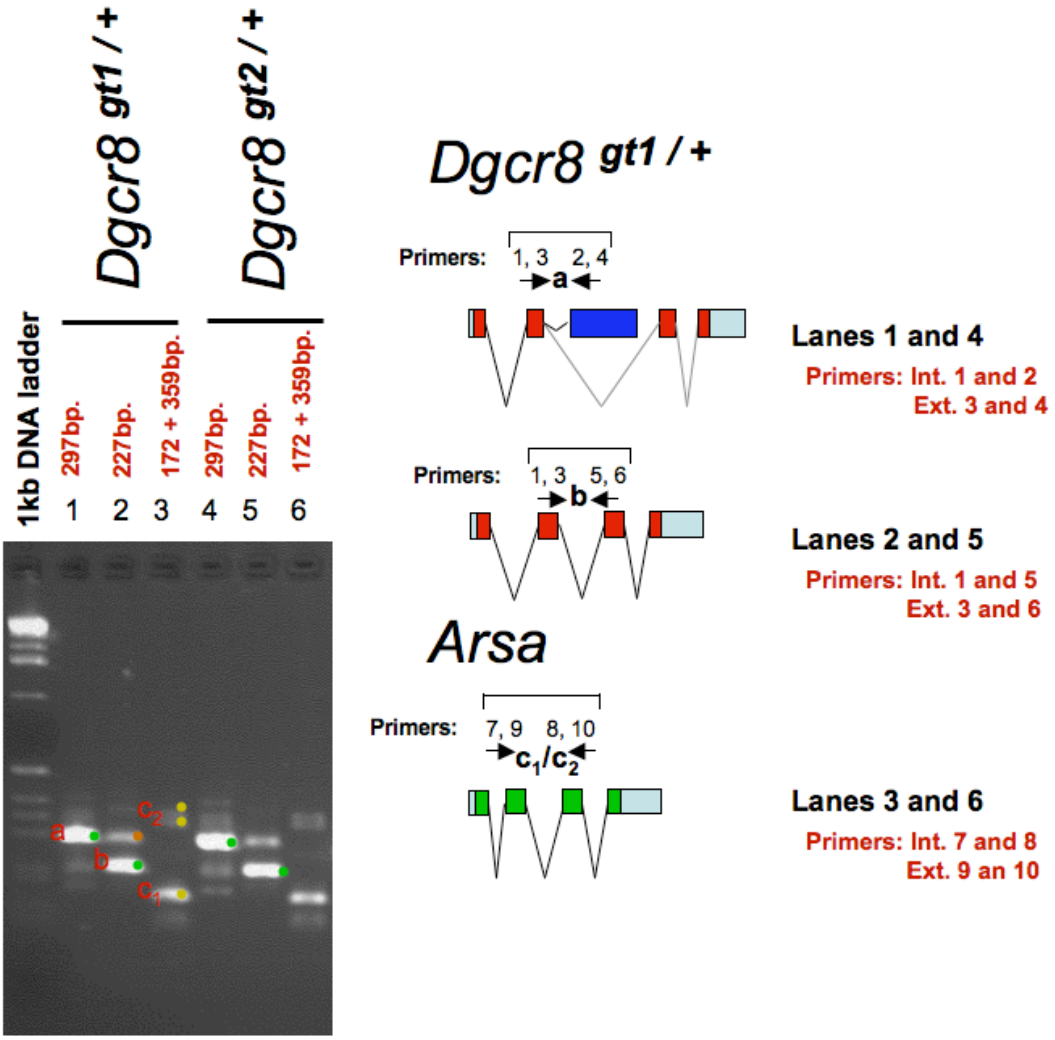


Fig 3.2: Nested RT-PCR to confirm the gene trap position within two BayGenomics *Dgcr8*$^{gt/+}$ cell lines: Nested primers were designed to amplify fragments across exon-exon boundaries. Set 'a' (Primers 1, 2, 3 and 4) anneal to an exon upstream of the gene trap and to the gene trap itself (Table 3.1), amplifying an expected fragment of 297bp. Set 'b' contains primers 1 and 3 which anneal to the upstream exon, but also two primers that anneal to an exon downstream of the gene trap insertion site (Primers 5 and 6) (Table 3.1). These are expected to amplify a PCR product corresponding to wild type splice events, 227bp in length. Primer set 'c' is a control set of primers. These anneal either side of a short intron 187bp in length within *Arsa* cDNA. Fragments amplified from reverse transcribed, spliced RNA will be 172bp in length (c_1). Fragments amplified from genomic DNA contamination of the cDNA sample will be 359bp in length (c_2). The PCR reactions run in lanes 1, 2 and 3 are amplified from the *Dgcr8*$^{gt1/+}$ cDNA template. Lanes 4, 5 and 6 are amplified from the *Dgcr8*$^{gt2/+}$ template cDNA. The bands in Lanes 1 and 4 are amplified with the gene-trap to exon primers. The bands in Lanes 2 and 5 are amplified with the exon to exon primers. Lanes 3 and 6 contain the *Arsa* control amplification fragments. Green spots represent transcripts of the expected length subsequently re-amplified and sequenced. The orange spot represents an unknown band, also re-amplified and sequenced. Yellow spots in lane 3 are bands later re-amplified and discussed in section 3.3.6.

Products amplified from trapped and wild type transcripts (Fig.3.2, marked in green) were re-amplified, and sequenced. Where possible, non-specific bands were also purified and sequenced in order to characterise them (Fig.3.2, marked in orange). Each RT-PCR product was sequenced from both ends using the internal PCR primers with which it was amplified. The sequence results were compared to the expected sequence in Gap4.

The PCR products equivalent to the green bands in lanes 1 and 4 (Fig.3.2) were as expected. The sequences from both ends of the products overlapped, providing representative sequence for the entire length of the fragment. The sequence alignment also allowed me to confirm position of the splice junction between the *Dgcr8* transcript and gene trap. Likewise, the sequences from the PCR products of the expected sizes amplified from the WT transcripts (Lanes 2 and 5, marked in green (Fig.3.2)) matched their templates without any significant discrepancies and once again the primer sequences were evident at each end of the sequenced PCR products.

The ~300bp, non-specific band marked in orange (Fig.3.2) was also sequenced. Gap4 analysis revealed a 70bp repeat sequence that was probably derived from an RT-PCR artefact, as, when the consensus sequence was folded with RNAfold (Default settings) (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi), it contained a substantial secondary structure.

The control fragments amplified from *Arsa* cDNA produced three bands. Two of these were of the expected sizes. The intermediate band implies that there may be some genomic contamination of my reverse transcribed cDNA. The largest band was of unknown origin. A

more in depth assessment of the identities of these bands was conducted at a later date (see section 3.3.6). Any genomic contamination of the cDNA did not affect the identification of the correctly amplified spliced gene trap products.

3.3.4 Subcloning *Dgcr8* cell lines

Dgcr8$^{gt1/+}$ and *Dgcr8*$^{gt2/+}$ were sub-cloned to ensure that they were homogenous cell populations and did not contain wild type cell contamination. Following subcloning the RT-PCR procedure explained above was repeated in order to confirm the identity of each sub-clone. An E14 wild type control was included in these reactions and as expected the PCR performed on the cDNA template generated from these cell lines only generated a product between the primers pairs which annealed to the 5' and 3' exons and not between the pairs designed to amplify between the upstream exon and the gene trap.

3.3.5 Generating a targeting vector

The *Dgcr8*$^{gt1/+}$ and *Dgcr8*$^{gt2/+}$ cell lines originate from the E14 ES cell line, which is derived from the 129P2 mouse strain. Previous studies have demonstrated that the isogenicity of the homologous fragment within a targeting vector can have a profound effect on the efficiency of targeted insertion (te Riele et al., 1992). Therefore it is necessary to ensure that the targeting fragment is amplified from the same strain template DNA to reduce the occurrence of single nucleotide polymorphisms (SNPs) that may interfere with the efficiency of homologous recombination. To this end, a BAC of 129 origin (129S7/AB2.2 BAC clone (bMQ-62C21) (Adams et al., 2005)) that spans the *Dgcr8* gene was identified from Ensembl. The end sequences, by which bMQ-62C21 was mapped, were checked to ensure the clone

had been annotated correctly. Subsequently bMQ-62C21 was used as the template for the amplification of the homologous region of the targeting vector.

The intended placement of the targeted trap within the *Dgcr8* locus placed several constraints upon the design of the primers used to amplify the homologous fragment (Primers 11 and 12, Fig.3.4). Therefore following the repeat masking of the template region, primers were designed such that:

- Primers were restricted to a region downstream of ENSMUST00000009321 exon 2 and upstream of exon 8. As a consequence there would be sufficient sequence 5' of the homologous region within which to design primers to confirm the insertion of a second allele by RT-PCR. In addition the second trap would also be targeted upstream of the gene trap currently inserted in the *Dgcr8* locus, which would cause a loss of marker expression from the first gene trap if the second trap inserted into the same allele rather than into the wild type allele. Therefore it would be simple to distinguish these two events (Fig.3.1).

- Primers were designed within intronic sequences. This ensures that the trapped allele continues to splice as expected and that the splice acceptor of the trap resides in an intron and is available to the splicing mechanisms. The amplified region's ends were at least 120bp from any splice junctions. The primers were positioned between exons 2 and 3 and 7 and 8 of ENSMUST00000009321.

- The amplified fragment was greater than 5000bp in length (6083bp) to allow efficient targeting of the trap.

- Following primer design, an *AscI* restriction site was added to the 5' end of each primer followed by a short "AATT" sequence. This restriction site would later be used in order to clone the fragment into a Gateway vector.

The chloramphenicol resistance gene and ccdB cassette were removed from between the attR3 and attR4 Gateway sites of the pR3R4AsiSI Gateway vector (Fig.3.3) by *AscI* restriction digest and the plasmid backbone was gel purified and de-phosphorylated. The fragment for cloning was amplified by Long PCR from *NotI* digested bMQ-62C21 DNA using primers 11 and 12 (Fig.3.4) and gel purified. The *AscI* digested targeting fragment was then ligated into the pR3R4AsiSI vector and transfomed into MACH1 *E. coli*.

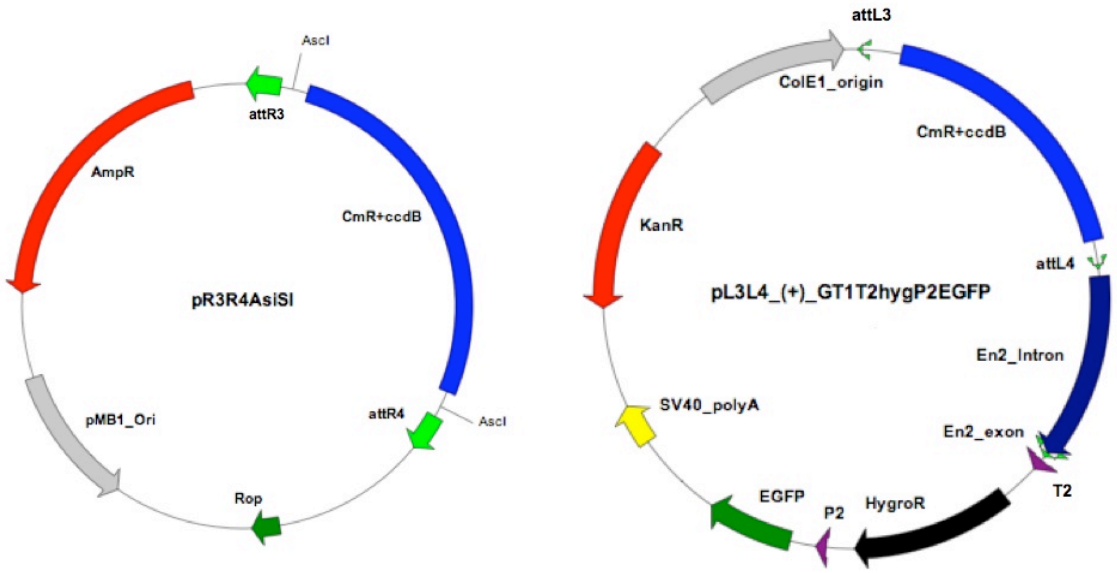
MACH1 cell colonies containing plasmids with the correctly ligated fragments were identified by colony PCR with a set of primers that amplify a region from within the ligated fragment (Primers 13, 14, 15 and 16 (See Fig.3.4)). It is essential for trap function that its splice acceptor be situated upstream of the trap within the *Dgcr8* transcript. As the orientation of the genomic fragment within the pR3R4AsiSI plasmid would determine the orientation within the trap targeting vector following Gateway transfer, a further colony PCR was conducted with primers from both within the fragment and within the flanking regions (Primers 17 and 18, See Fig.3.4). A single colony was identified containing the correctly inserted fragment and the ends of the fragment were sequenced primed by primers 14, 15, 17 and 18. BLASTN was used to compare the consensus of the resulting sequence contigs

against the mouse genomic sequence. As expected the sequences matched *Dgcr8* with a very high percentage alignment (98.96-99.79%) and were positioned as expected.

The targeting fragment was transferred from the cloning vector into the trap vector (pL3/L4_(+)_GT1T2hygroP2EGFP (Fig.3.3)) through an *in vitro* L/R clonase reaction. Once inserted into the genome, it is essential that the phase of the reading frame of the trap is the same as that of the upstream exon of the targeted transcript, so that once transcribed and spliced the trap gene is translated such that the selectable markers are expressed in frame. The phase of the splice junction of the upstream exon (Exon 7, ENSMUST00000009321) was "1", hence the corresponding phase trap vector was used.

After transformation of the L/R reaction, kanamycin resistant colonies were tested by colony PCR for the presence of the amplified *Dgcr8* fragment (Primers 14, 17, 15, 18). Subsequently, a second colony PCR confirmed the orientation of the inserted fragment within the targeting vector in a subset of the positive colonies (Primers 14, 19, 15, 20) (Fig.3.4).

One of the correct colonies was selected and prepared for electroporation. Prior to electroporation, a unique *HindIII* restriction site approximately in the middle of the homologous fragment, was used to linearise the plasmid. This insertion-type vector will recombine into the ES cell genome and duplicate the target sequence.



Plasmid and Gene Trap Characteristics:	
Plasmid	**Features**
pGT1lxf Gene Trap	En2-intron / En2-exon (Splice acceptor) β–geo (Fusion with endogenous peptide - confers neomycin resistance and stains blue with X-gal staining reagents) SV40 pA (Polyadenylation site)
pR3/R4AsiSI	R3-*Asc*I-CmR-ccdB-*Asc*I-R4 (Gateway R3-4 Chloramphenicol resistance and ccdB cassette) pMB1 origin rop (Reduces copy No.) AmpR (Ampicillin resistance)
pL3/L4(+)GT1T2hygP2EGFP	L3-CmR-ccdB-L4 (Gateway L3-4 Chloramphenicol resistance and ccdB cassette) En2-intron / En2-exon (Splice acceptor) T2A (Causes ribosome to skip) HygroR (Hygromycin resistance) P2A (Causes ribosome to skip) EGFP (Enhanced green fluorescent protein) SV40 pA (Polyadenylation site) KanR (Kanamycin resistance) ColE1 origin

Fig.3.3: Gene traps and vectors used to disrupt the *Dgcr8* locus. Plasmids generously provided by the Skarnes laboratory. Vector maps were drawn with Savvy v0.1 (http://www.bioinformatics.org/savvy/). Red text within the plasmid description refers to elements inserted into the*Dgcr8* transcript.

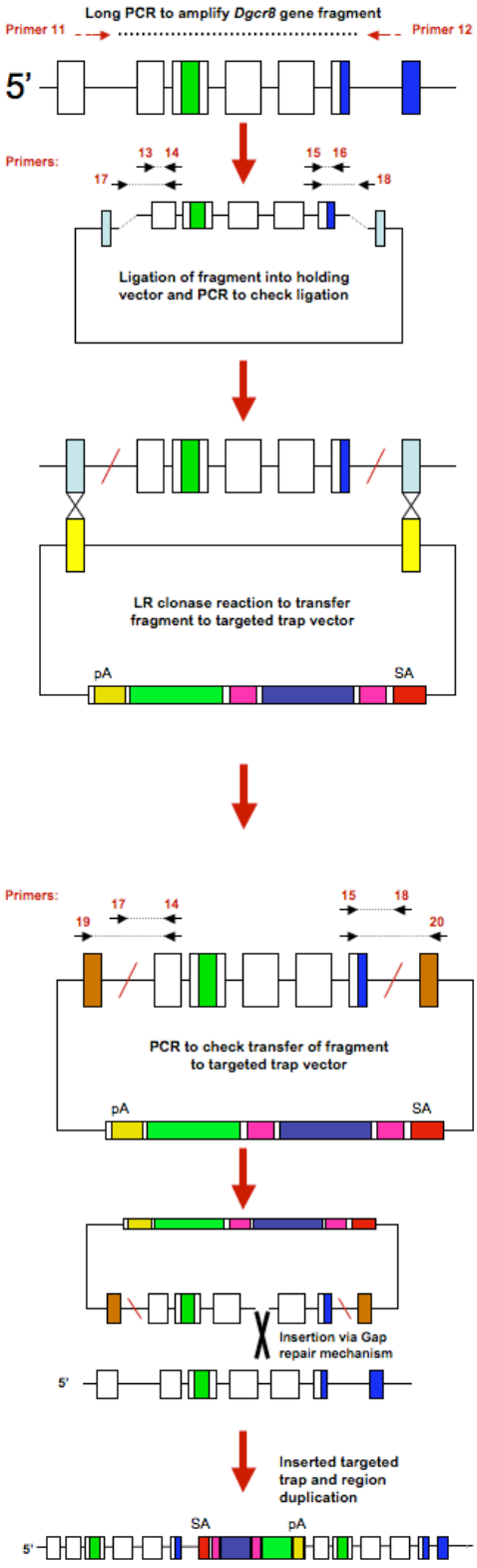


Fig.3.4: Principles of targeting vector construction: The fragment required to target the vector to the *Dgcr8* locus is amplified by long PCR from a BAC clone derived from the *Dgcr8* region (Primers 11 and 12). These

primers contain *AscI* restriction sites, which are used to clone the fragment into the pR3/R4AsiSI plasmid in place of the CmR-ccdB cassette. The fragment is transferred to the pL3/L4(+)GT1T2hygP2EGFP targeting vector, which contains the trap cassette, by an L/R clonase reaction. The target vector is linearised through the introduction of a gap into the centre of the homologous region by restriction digest. This will allow the vector to insert into the *Dgcr8* locus through a gap repair mechanism. The targeting vector is introduced to mouse ES cells by electroporation. The vector then inserts into the region homologous to the amplified fragment, causing a duplication of the region itself. At various stages primers were designed to amplify fragments with which to confirm fragment insertion, transfer and orientation within the plasmids. These are illustrated as red numbers above the schematic vectors.

3.3.6 Identification of cells with a successfully targeted *Dgcr8* locus

Confirmed sub-clones of $Dgcr8^{gt1/+}$ and $Dgcr8^{gt2/+}$ were electroporated with the linearised plasmid pL3/L4_(+)_GT1T2hygroP2EGFP, containing the cloned fragment of *Dgcr8*. Hygromycin resistant colonies from each cell line were picked, expanded, archived and lysed for the preparation of RNA.

A

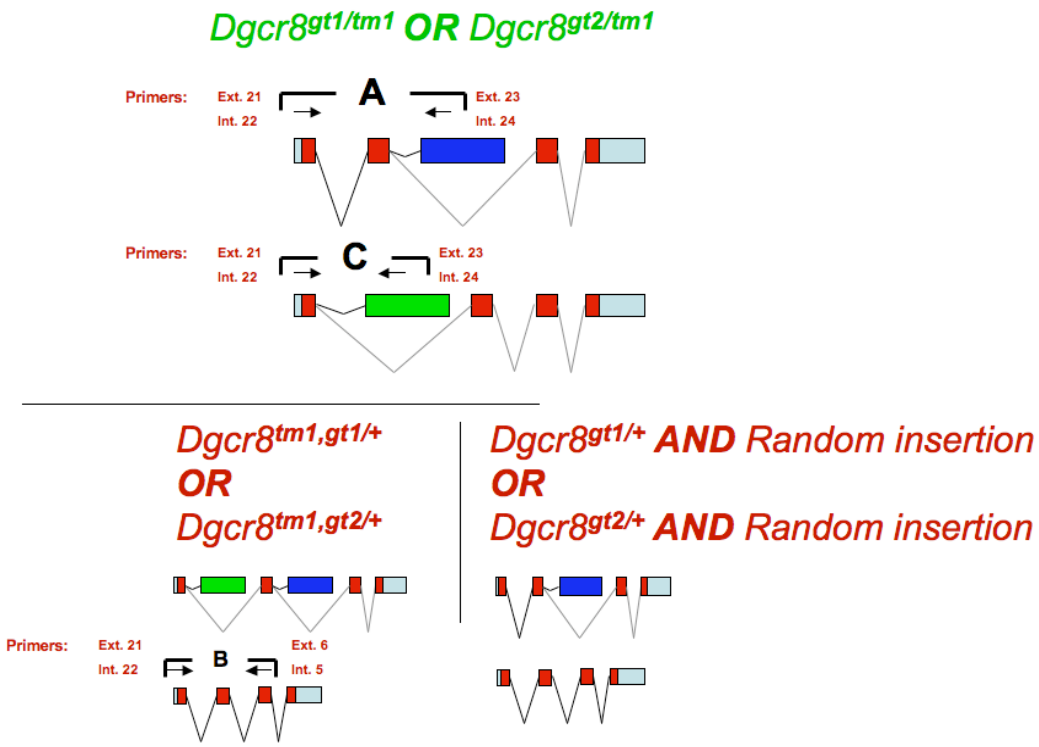


Fig.3.5: A) A schematic of the expected outcomes of the targeted vector insertion and the primers designed to distinguish these events. The targeted trap can insert as intended into the second allele, resulting in the disruption of both alleles (*Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ (Table 3.2)) (top), it can insert into the same allele as the initial gene trap (*Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ (Table 3.2)) (bottom left), which will disrupt the expression of selectable markers from the initial gene trap, or it can insert at a random point in the genome (*Dgcr8*$^{gt1/+}$ and *Dgcr8*$^{gt2/+}$ AND Random insertion) (bottom right). Primers were designed to distinguish these three genotypes as marked. A single primer set (Primers 21, 22, 23 and 24) will amplify fragments from cDNAs containing either gene trap. Primer 21 ad 22 are designed to anneal to exons upstream of the region duplicated by the targeted trap insertion while primers 23 and 24 are an overlapping nested set designed to anneal to sequences common to both traps' splice acceptors. These fragments will be of a different size depending upon the traps expressed in the cell line (larger if amplifying from cDNA containing the downstream gene trap and smaller if amplifying from the targeted trap transcripts). Primers were also selected to amplify fragments from cDNAs demonstrating wild type splicing events, with forward primers annealing upstream of the traps and downstream primers annealing 3' of both traps (Primers 21, 22, 5 and 6).

B

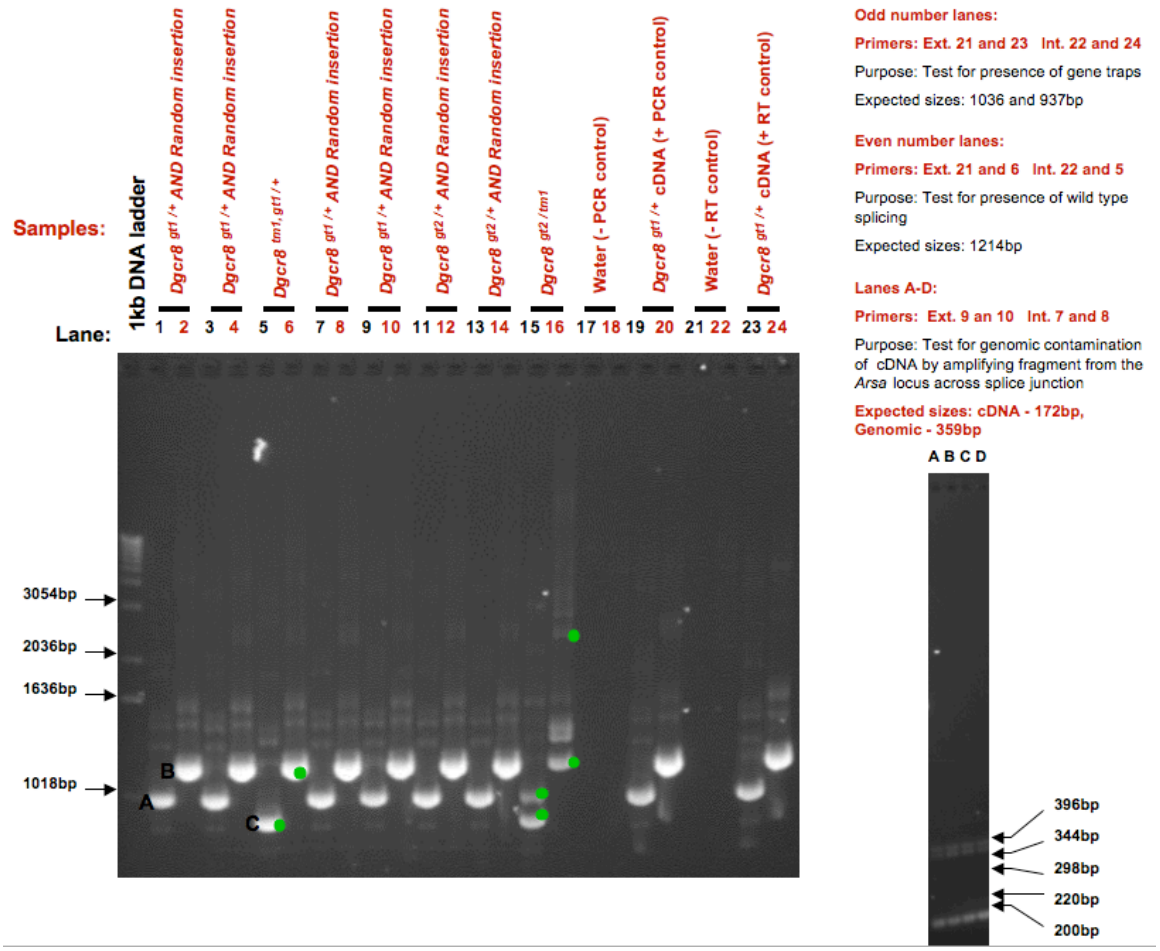


Fig.3.5: B) RT-PCR was used to distinguish the genotypes of ES cell lines resulting from the electroporation of the gene trap targeting vector. Nested RT-PCR was conducted with the primer combinations described in Fig.3.5A with RNA derived from each of the hygromycin resistant colonies collected from the electroporation of the targeting construct. Odd number lanes are expected to contain either a single fragment (Fig.3.5A (top) fragments A or C) or both of these fragments. If both templates are present in the lysate, and hence both fragments are amplified in the PCR reaction, the cell line is expected to contain a trap in each allele (*Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$). The presence of only a single fragment in this lane indicates that either the targeted trap has inserted upstream of the gene trap at this locus (smaller fragment) or the targeted trap has randomly inserted elsewhere in the genome (larger fragment). Therefore only a single gene trap is expressed from the *Dgcr8* locus. As explained previously (see section 3.3.3), primers 7, 8, 9 and 10 amplify a fragment of the *Arsa* cDNA in order to judge the presence of genomic contamination of the cDNA. As previously noted, this last set of primers amplifies an unknown band in addition to the two expected bands.

Correctly targeted clones were identified by RT-PCR performed on cDNA produced from a set of 25 *Dgcr8*$^{gt1/+}$ and 23 *Dgcr8*$^{gt2/+}$ derived hygromycin resistant colonies (Table 3.3) (Fig.3.5B). Nested primer sets were designed to distinguish between cell lines with a correctly inserted targeted trap from those in which the traps have inserted randomly or into the same allele as the BayGenomics gene trap (Fig.3.5A). This process identified 4-5 colonies associated with RT-PCR results consistent with the event of a targeted trap inserting into non-gene-trapped allele of *Dgcr8* (Table 3.3), of which 2-3 derived from each parent cell line. The bands amplified in one of the 5 cases were weak, so the PCR was deemed inconclusive.

	Random Insertion	**Gene trapped allele insertion**	**2nd Allele Insertion**	**Failed PCR**
Dgcr8$^{gt1/+}$	17	4	2 (+1)	1
Dgcr8$^{gt2/+}$	16	5	2	NA
Total	33	9	4 (+1)	1

Table.3.3: A summary of the number of each kind of insertion event resulting from the targeted gene trap electroporation experiment. The genotype of each cell line at the *Dgcr8* locus was determined by RT-PCR.

It might be expected that in the case of potential null cell lines, the primers designed to amplify fragments from wild type cDNA (Primers 21, 22 and 5, 6) would not amplify products; as transcripts from both alleles should be truncated by traps. However, these products were seen (Fig.3.5B). It should be noted that this aberrant splicing through the traps does not necessarily imply that the experiment has been unsuccessful as such transcripts may not be full length or translated. RT-PCR is also notoriously non-quantitative with these products perhaps being derived from an insignificant quantity of residual wild type transcript. However, this does raise the possibility that the cell lines may be hypomorphic for *Dgcr8*

rather than null. This issue was addressed by Northern blot as described later (see section 3.4.1).

Once again the control RT-PCR reaction specific for the *Arsa* gene transcript showed products derived from both cDNA and genomic DNA templates for all samples. The presence of genomic DNA in the reactions is not a problem as all PCR primers used were designed to amplify products across multiple exon-exon splice junctions and any products derived from this contamination would be of a larger than expected size.

Interestingly, there are three bands present in the lane corresponding to the *Arsa* primer PCR reaction, the third being slightly larger than the expected genomic DNA size. This reaction was repeated on cDNA derived from DNase treated RNA from wild type cells. Again there were three bands evident in the lane. A PCR reaction performed on simultaneously prepared cDNA, which omitted the reverse transcriptase failed to produce any products. These results imply that none of the bands are amplified from genomic DNA, even those that correspond to the genomic band size. It is therefore likely that the template for this product is generated from unspliced RNA molecules. All three bands were sequenced. As expected, sequences derived from the smaller of the two bands corresponded to the expected splice form of the cDNA. The band that corresponded to the unspliced cDNA also aligned to the genomic template as expected. Finally, sequences from the largest of the three bands did not align to each other easily and were not of a high enough quality to ascertain their origin. However, since these products are not amplified from non-reverse transcribed template, it would suggest that they originate from an RNA source and hence the exact origin of the sequences

is not essential in order to fulfil the purpose of the control PCR. In future a non-reverse transcribed template should prove a better control for genomic DNA contamination.

Subsequently, the RT-PCR was repeated for a representative sample of the cell lines originally tested in this PCR. The cDNA for these tests was prepared from Trizol purified, DNase treated RNA. Amplified PCR products were purified for both primers 21, 23 and 22, 24 and primers 21, 6 and 22, 5 equivalent to those marked green in Fig.3.5B. On re-amplification, the PCR product for the wild type transcript primers amplified a larger than expected product from a template based on cDNA from one of the cell lines believed to contain traps in both alleles. This larger band corresponded to the size product that might be expected if the cDNA template skipped the gene trap but contained all of the duplicated exons. This band was also purified.

All purified products were sequenced and the sequences aligned to the ENSMUST00000009321 sequence as expected. As noted above, in the case of cell lines containing a trap in each allele of *Dgcr8*, two fragments are expected to be amplified in a single PCR reaction. It is important therefore to ensure that the larger of the two fragments sequenced from these lanes is indeed amplified from cDNA originating from transcripts containing the first gene trap and is not a fragment amplified by the external nested primers from targeted trap transcripts. Importantly, the sequence contig for the larger of the two fragments extended into exon 8 of ENSMUST00000009321, and thus confirmed the identity of the band as this exon is downstream of the intron within which the targeted trap has inserted.

Sequence derived from the large band amplified by primers 21, 6 and 22, 5, that was suspected to contain the duplicated exons caused by the targeted insertion, aligned to the relevant exons of *Dgcr8*. However, the sequences obtained were not long enough to determine if the product was indeed generated from duplicated exons, although the fact that the sequences derive from *Dgcr8,* in addition to the length of the PCR product, supports this hypothesis.

3.3.7 Determining the functional significance of trap insertions through the use of miRNA Northern blots

The purpose of disrupting the *Dgcr8* locus is to cause an interruption in miRNA processing and a reduction of mature miRNAs in mouse embryonic stem cells in order to provide a resource for miRNA target identification in the absence of endogenous miRNA expression. To determine whether the cell lines that appeared to contain a trap in each *Dgcr8* allele should be pursued further, the functional significance of the RT-PCR results was examined through a series of Northern blots for miRNAs known to be expressed in mouse ES cells (Houbaviy et al., 2003).

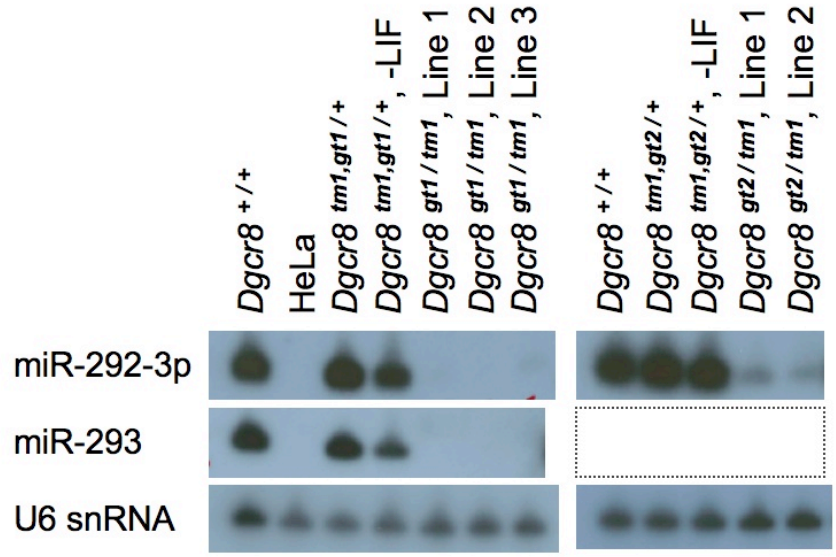


Fig.3.6: Northern blot to judge miRNA expression within a range of cell lines. At this early stage the cells were cultured and RNA was prepared in a fashion that differed slightly from that described in the Methods section. Cells were cultured under hygromycin selection, where suitable. *Dgcr8*$^{+/+}$ cells were cultured in non-selective media. HeLa S3 cells were cultured under specific conditions (see section 2.7.4). The hygromycin selection was removed as cells were plated for RNA preparation. Cells were lysed 2 days after plating. Cell lines plated without LIF were seeded at half the cell number and cultured for 3-4 days prior to lysis. RNA was prepared with Trizol reagent. At this stage the confluence of the cells on RNA lysis was also not tightly controlled. Generally 20ug of total RNA was run per lane. However, where this was not possible, as much RNA was used as feasible (*Dgcr8*$^{+/+}$ - 17.1ug, *Dgcr8*$^{tm1gt1/+}$ - 15.1ug, *Dgcr8*$^{tm1gt1/+}$ - 16.3ug). These Northern blots were probed with oligos complementary to miR293, miR-292-3p and U6 loading control. The missing panel was not probed.

The mutant cells were maintained under hygromycin selection throughout their culture to prevent homologous recombination between the duplicated regions that flank the targeted trap, which could potentially restore DGCR8 function (an event which may produce a selective advantage in untreated cells). For each parent cell line (*Dgcr8*$^{gt1/+}$ and *Dgcr8*$^{gt2/+}$), hygromycin resistant colonies with two traps in the same allele were chosen as the controls, as these cells would also control for the effect of hygromycin treatment and for other potential effects caused by the point of insertion (*Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ (Table

3.2)). These cells were cultured under the same selective conditions as the homozygous mutant cells with a trap in each allele (*Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ (Table 3.2)). An apparent growth defect was observed in the 5 homozygous mutant cell lines (*Dgcr8*$^{gt/tm1}$) as they appeared to replicate more slowly than the control cell lines. At this early stage these cell lines were generally split at a much lower ratio than the heterozygous cell lines and were initially not fed every day (although these conditions were changed as I became more accustomed to their growth patterns). The 2 heterozygous cell lines (*Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$) were also grown up in the absence of LIF to trigger differentiation. This control was included in case the hygromycin selection was having a more profound effect on the differentiation status of the homozygous mutant cells. Inducing significant differentiation in the heterozygous cells through the removal of LIF from the media and then comparing the ES cell specific miRNA expression to that of the homozygous cells should help to determine whether low level differentiation is the root cause of any expression changes seen between the heterozygous and homozygous cells. As a further positive control for the ES cell miRNA expression, RNA derived from wild type E14 cells was included on the Northern blots. In addition RNA derived from HeLaS3 cells was also included as a negative control for the miRNAs profiled.

For Northern blotting, RNA was separated by 15% TBE-Urea gel electrophoresis suitable for discriminating small RNA molecules and transferred to a nylon membrane. The membrane was hybridised with radiolabelled DNA oligos complementary to the mature form of the miRNA of interest. The results clearly demonstrate a significant reduction of the ES cell specific miRNAs tested in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines (Fig.3.6), to the point at which the miRNAs cannot be clearly discerned on the blot without developing the blot for an

extended period. Subsequently, the miRNA signal from the *Dgcr8*$^{tm1,gt/+}$ and *Dgcr8*$^{+/+}$ samples was so strong that it is difficult to know if the faint signal in the *Dgcr8*$^{gt2/tm1}$ lane was produced by residual mature miRNA expressed in these cells or by slight contamination of these wells during the gel loading process. The extent of the mature miRNA depletion is addressed in section 3.4.2 and further in Chapter 4. The HeLaS3 sample is devoid of miRNA signal. This is expected, as the miRNAs selected are thought to be specifically ES cell expressed, thereby increasing the confidence in the specificity of the probes. There also appears to be a reduction in the level of these ES cell miRNAs upon differentiation (in the absence of LIF) although the reduction in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ is far greater than that seen in the cells cultured without LIF. This implies that the reduction seen in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells is beyond that expected from a relative increase in the proportion of differentiated cells in the culture. In all these results suggested that the introduction of the traps into the *Dgcr8* locus was having a significant effect on the processing of miRNAs in the stem cells.

At this point the choice of miRNA deficient cell lines for further study was refined to a single representative line for each of the original clones (*Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$). Each of these mutants was paired with a heterozygous control cell line with a second trap inserted upstream of the first gene trap within the same allele, (*Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$).

3.3.8 Confirmation of the position of inserted targeted traps by gene-trap specific staining

Due to the position of the traps within each of the cell lines, it should be possible to distinguish individual insertion events by Xgal staining. As shown in Fig.3.3, the initial gene trap within each line (gt1 and gt2) will code for a ß-geo fusion protein that will cause cells that express it to stain blue upon staining with X-gal. The insertion of the second trap upstream of this trap in the same allele, as is the case for $Dgcr8^{tm1,gt1/+}$ and $Dgcr8^{tm1,gt2/+}$, should cause the loss of positive Xgal staining, while $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ should retain their Xgal staining phenotype as the trap should be inserted into the previously wild type allele. As a control for this experiment, a cell line was selected that was derived from $Dgcr8^{gt1/+}$ in the electroporation experiment but with a second, expressed random insertion of the targeting vector elsewhere in the genome. Once again this line would be expected to stain blue. Wild type E14 cells ($Dgcr8^{+/+}$) should stain white following the Xgal staining procedure.

All of the cell lines stained in the expected manner (Fig.3.7). This reinforces the veracity of the original genotyping performed by RT-PCR. Notably, the $Dgcr8^{gt1/+}$ cell line stained slightly less strongly than $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$. This pattern replicates data gathered in other X-gal staining experiments (data not shown). It may be that some form of *Dgcr8* auto-regulation is disrupted when *Dgcr8* is knocked down within the system, resulting in an up-regulation of the *Dgcr8* promoter. Indeed, a slightly different form of *Dgcr8*-Drosha co-regulation has recently been described that may account for these changes in *Dgcr8* expression level (Han et al., 2009). *Dgcr8* mRNA contains embedded RNA hairpins within the 5'UTR and the 5' end of the ORF. The mechanism proposed by Han *et al.* would predict

that a decrease in DGCR8 protein would cause a destabilisation of Drosha. The loss of microprocessor activity would lead to a reduction in the cleavage of these embedded hairpins and an up-regulation of *Dgcr8* mRNA, and hence fusion proteins, when compared to control cells.

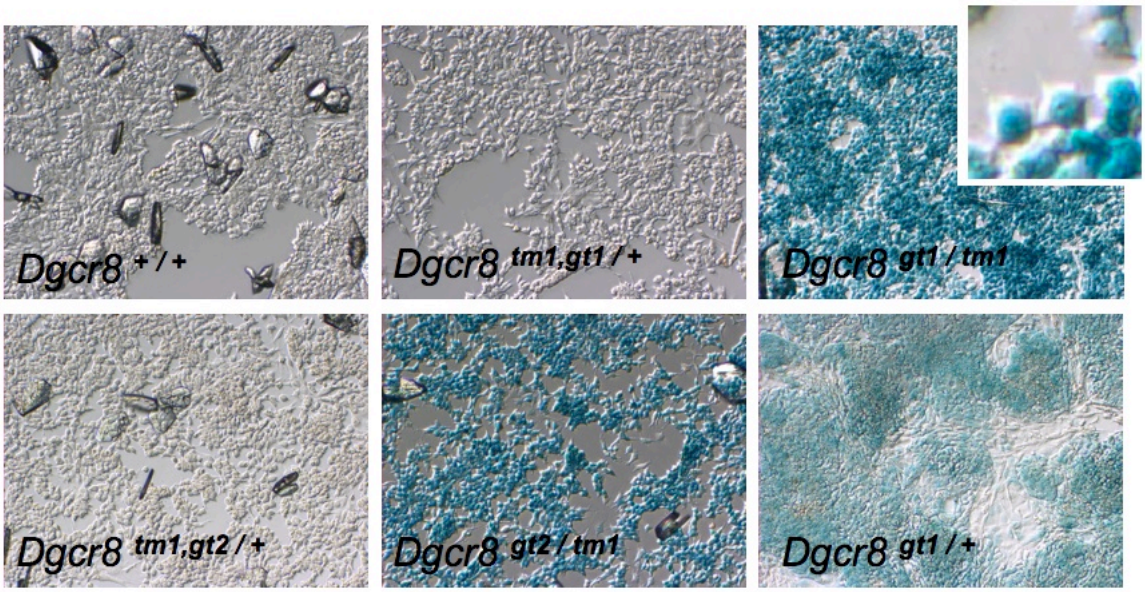


Fig.3.7: Xgal staining of cell lines to determine β-geo (β-galactosidase) activity. *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{gt2/tm1}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$ and *Dgcr8*$^{gt1/+}$ (+ Random-insert), were all maintained under hygromycin selection until stained. *Dgcr8*$^{+/+}$ cells were cultured in non-selective media. The insert shows nuclear localisation of β-galactosidase activity of the β-geo fusion protein. Note crystals were formed in some wells during staining. However, these were seen in both control and *Dgcr8*$^{gt1/tm1}$/ *Dgcr8*$^{gt2/tm1}$ wells.

3.4 Characterisation of *Dgcr8* expression and a broader spectrum of miRNAs

3.4.1 Detection of wild type and truncated *Dgcr8* transcripts

Given that the results of the RT-PCR suggest the presence of wild type splicing events in *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$, it is important to use a more quantitative approach to determine whether there is significant reduction in *Dgcr8* wild type transcripts in these cell lines. I designed probes to bind upstream of the duplicated region and downstream of the two trap insertions (411bp and 423bp respectively) within the *Dgcr8* transcript (ENSMUST00000115633 /ENSMUST00000009321) and cloned them into pGEM-T-Easy, confirming them by sequencing. It could be expected that the probes designed to bind upstream should reveal changes in the *Dgcr8* transcript sizes between the cell lines, caused by the insertion of traps into the gene; the resulting splice events truncate the wild type coding sequence and add gene trap sequence in its place. The downstream probes should demonstrate a depletion of the wild type transcripts in *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines.

7 cell lines were analysed by Northern blot: *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{gt2/tm1}$, *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$ and *Dgcr8*$^{+/+}$ cells were selected. In addition, samples of RNA from the parental gene trap cell lines (*Dgcr8*$^{gt1/+}$ and *Dgcr8*$^{gt2/+}$) were also used. These parental cell lines had been grown under slightly different growth conditions but this issue will be addressed in Chapter 5 (see section 5.3.1).

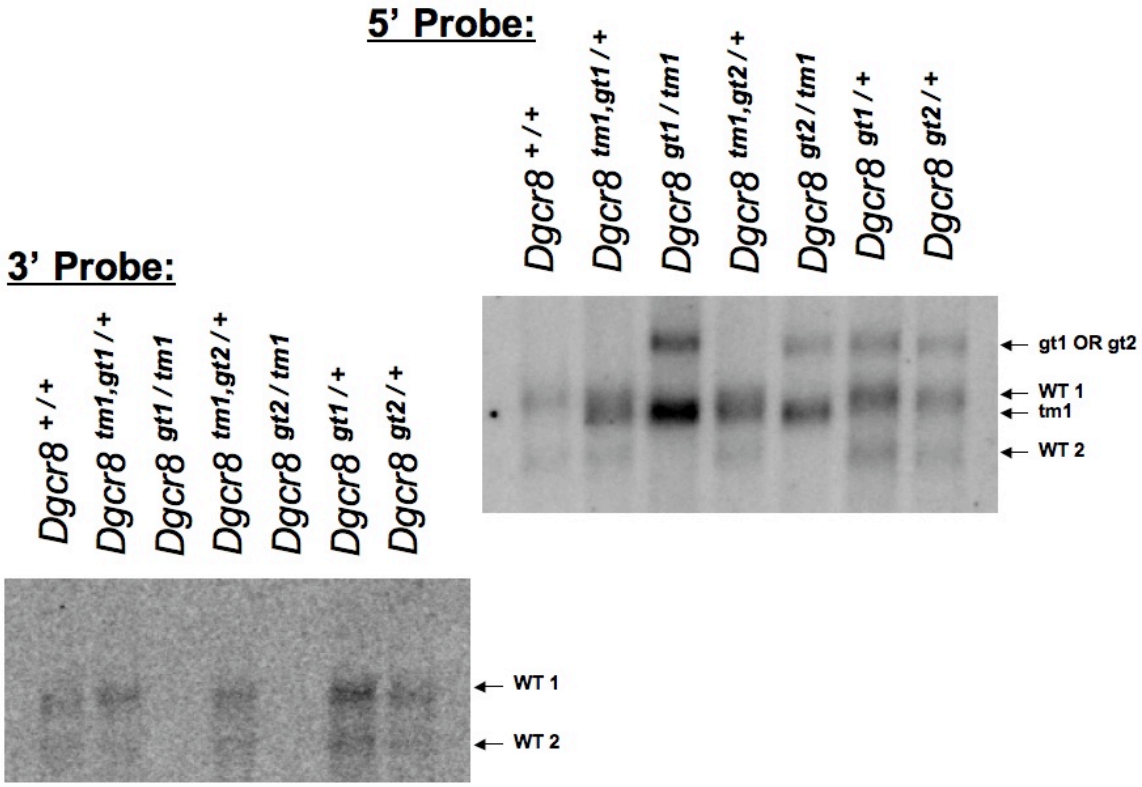


Fig.3.8: Northern blot to demonstrate the loss of wild type *Dgcr8* transcript in *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines and the effect of the traps on transcript length. RNA from each cell line was size separated, transferred to a membrane and hybridised with either a radiolabeled probe specific to the 5' end of the *Dgcr8* transcript (top) or the 3' end (bottom). The expected transcript lengths are: gt1 and gt2 - ~6.4kb, tm1 - ~4.2kb, WT1 - ~4.3kb (ENSMUST00000009321) or 4.5kb (Shiohama et al., 2003), WT2 - ~3.5kb (Shiohama et al., 2003). Each of these transcripts is marked with an arrow on the blots. Note the truncated transcripts are only labeled when the 5' probe is used.

The Northern blot was probed with both the 5' and 3' probe by Dr Claudia Kutter in Dr Duncan Odom's laboratory in the Cambridge Research Institute (CRI). Transcript sizes were consistent with those predicted. Most significant is the lack of wild type transcript in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ lanes when probed with the 3' probe, in contrast to the strong fusion transcripts seen when the 5' probe is used on the same blot (Fig.3.8). All other samples expressed detectable levels of wild type transcript identifiable with both the 5' and 3' probes.

The 5' probe also clearly identifies the transcripts for the gene trap and the targeted trap events in the *Dgcr8*$^{gt1/+}$, *Dgcr8*$^{gt2/+}$ and *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$ samples respectively. I was unable to identify a clone representing the 3.5kb wild type transcript reported in a previous study (Shiohama et al., 2003). However, Gregory *et al.* identified a carboxyl truncated version of the DGCR8 protein in HEK-293 cells (Gregory et al., 2004). This may be coded for by the smaller transcript, which would imply a 3' alteration to the transcript structure. Such an alteration may explain why the wild type transcripts appear to be of two alternative sizes, while trapped transcripts are of a single length, as both wild type transcripts would share the 5' exons.

It is clear from this blot that there is a significant reduction in wild type *Dgcr8* transcripts in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells when compared to the control cell lines. It follows that there is minimal splicing across the inserted traps to derive these functional transcripts, and demonstrates that the RT-PCR result identifying potential wild type transcripts in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells was not quantitative. Ultimately, this result supports the conclusion that the *Dgcr8* locus has been disrupted by the insertion of traps to a point at which DGCR8 function is impaired, as confirmed by the absence of processed miRNAs.

3.4.2 The expression of ES cell miRNAs

In parallel to establishing the allelic structure of *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{gt2/tm1}$, *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$, further miRNA Northerns were performed for these cells to investigate the functional significance of the allelic alterations in more detail.

As I grew more accustomed to the growth properties of *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells the growth conditions altered slightly from those used for the previous miRNA Northern blot analyses. Hygromycin selection was removed from the cells 2 passages (4 days) prior to Trizol lysis, to allow the cells to recover from the effects of the drug. At the stage at which the cells were plated for RNA they were plated in quantities that allowed sub-confluent cells to be harvested 2 days later from all of the samples. Due to the slower growth rate apparent amongst *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells the quantity of cells plated was higher than in the case of *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$. At this stage a protocol was established at which *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ cells were plated at a higher density than *Dgcr8*$^{+/+}$ cells. Although no significant differences were seen between the properties of the *Dgcr8*$^{tm1,gt/+}$ cells and wild type cells throughout my research, synchronising cell lines that are maintained with or without selection, that grow at different rates and which have different phenotypes makes it unlikely that perfectly matched cell densities between the cell lines used in these experiments is achievable. However, by following these plating ratios the cells reached similar degrees of confluence by the time of lysis so these plating densities were maintained throughout for consistency.

The set of miRNAs to be tested were previously identified as ES cell expressed (miR-292as, miR-293, miR-16, miR-130, miR-21and miR-92a) (Houbaviy et al., 2003). In addition, let-7a was selected as the let-7 family (Landgraf et al., 2007) is specifically repressed in ES cells but is widely expressed throughout differentiated tissues. The U6 small nuclear RNA (snRNA) probe was used as a loading control for all the Northerns. (Fig.3.9) The results clearly demonstrate the loss of miRNA expression in *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells.

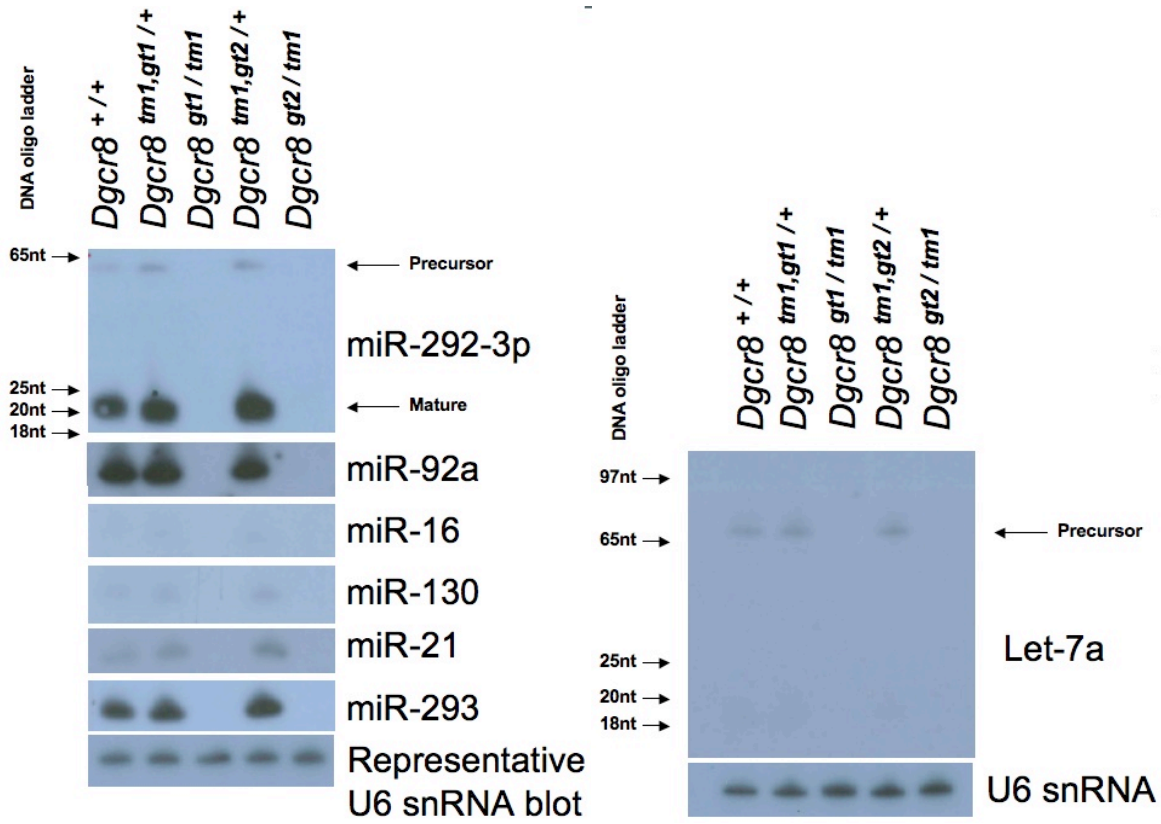


Fig.3.9: miRNA Northern blots demonstrating the loss of ES cell miRNA expression in *Dgcr8*$^{gt1/tm}$ and *Dgcr8*$^{gt2/tm1}$ cells. 12.7µg to 20µg of total RNA from each cell line was size separated per gel and transferred to a membrane before being hybridised with radio-labelled DNA oligos complementary to the miRNA sequences. All samples per gel consisted of the same quantity of RNA. All filters were hybridised with U6 snRNA as a loading control to ensure that all samples were loaded in comparable quantities. The upper blot on each panel is labelled with a ladder to the left. To the right of each panel are the names of each miRNA/snRNA tested and the labelled mature and precursor sequences where appropriate.

An intriguing feature demonstrated by these Northems is that the precursor let-7a band is apparent in the wild type, *Dgcr8*$^{tm1,\ gt1/+}$ and *Dgcr8*$^{tm1,\ gt2/+}$ samples. This is to be expected, as the let-7 family of miRNAs is expressed in ES cells but remains unprocessed at either the primary transcript or precursor stages (Heo et al., 2008; Newman et al., 2008), due to a let-7 specific block instigated by the Lin-28 protein. These results would seem to be consistent with a block at the stage of precursor processing. Let-7a expression is not evident in the homozygous mutant *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines. As let-7 is a widely expressed

miRNA family, this argues against the possibility that differentiation of the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines may account for the changes seen in miRNA expression when compared to the control cell lines.

3.5 Investigation of ES cell properties in *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines

A possible explanation for the lack of ES cell miRNA expression in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines could be that the disruption of the canonical miRNA-processing pathway and the associated protein expression changes could trigger partial differentiation of the ES cells resulting in a change in the population of miRNAs expressed. As a result the expression profile of miRNAs in these cells could be expected to be altered not because of the lack of a functioning DGCR8, but due to the triggering of a new set of cell specific promoters. Taking into account the intention to use these cells for ES cell miRNA target identification, it is preferable that the cells retain ES cell qualities post-electroporation and *Dgcr8* disruption.

3.5.1 Expression of core ES cell transcription factors

3.5.1.1 Oct4 expression measured by Western blot

Stem cells, by definition, are capable of continuous self-renewal and are pluripotent (capable of differentiating into many, specialised cell types). At the centre of the transcriptional network responsible for maintaining this cell state is a set of core TF, including Oct4, *Sox2* and *Nanog* (Avilion et al., 2003; Kopp et al., 2008; Mitsui et al., 2003; Niwa et al., 2000). To investigate whether the expression of a core TF in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines

is altered, protein was purified from *Dgcr8*$^{gt1/tm1}$, *Dgcr8*$^{gt2/tm1}$, *Dgcr8*$^{tm1,gt1/+}$ *Dgcr8*$^{tm1,gt2/+}$ and wild type cells and a Western blot was performed for the transcription factor Oct4 (Fig.3.10). The results are consistent with the expectation that *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cell lines maintain their ES cell identity despite the significant reduction of *Dgcr8*, as the expression of Oct4 appears unchanged between cell lines. α-tubulin was used as a loading control.

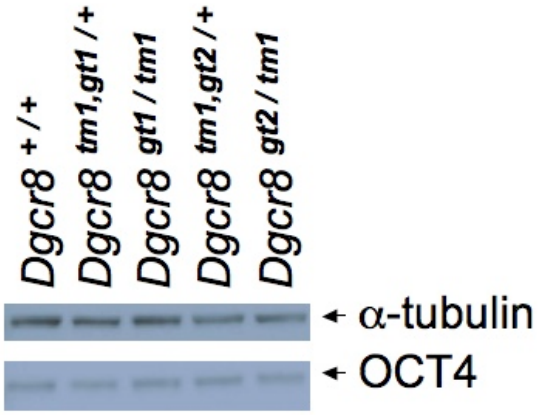
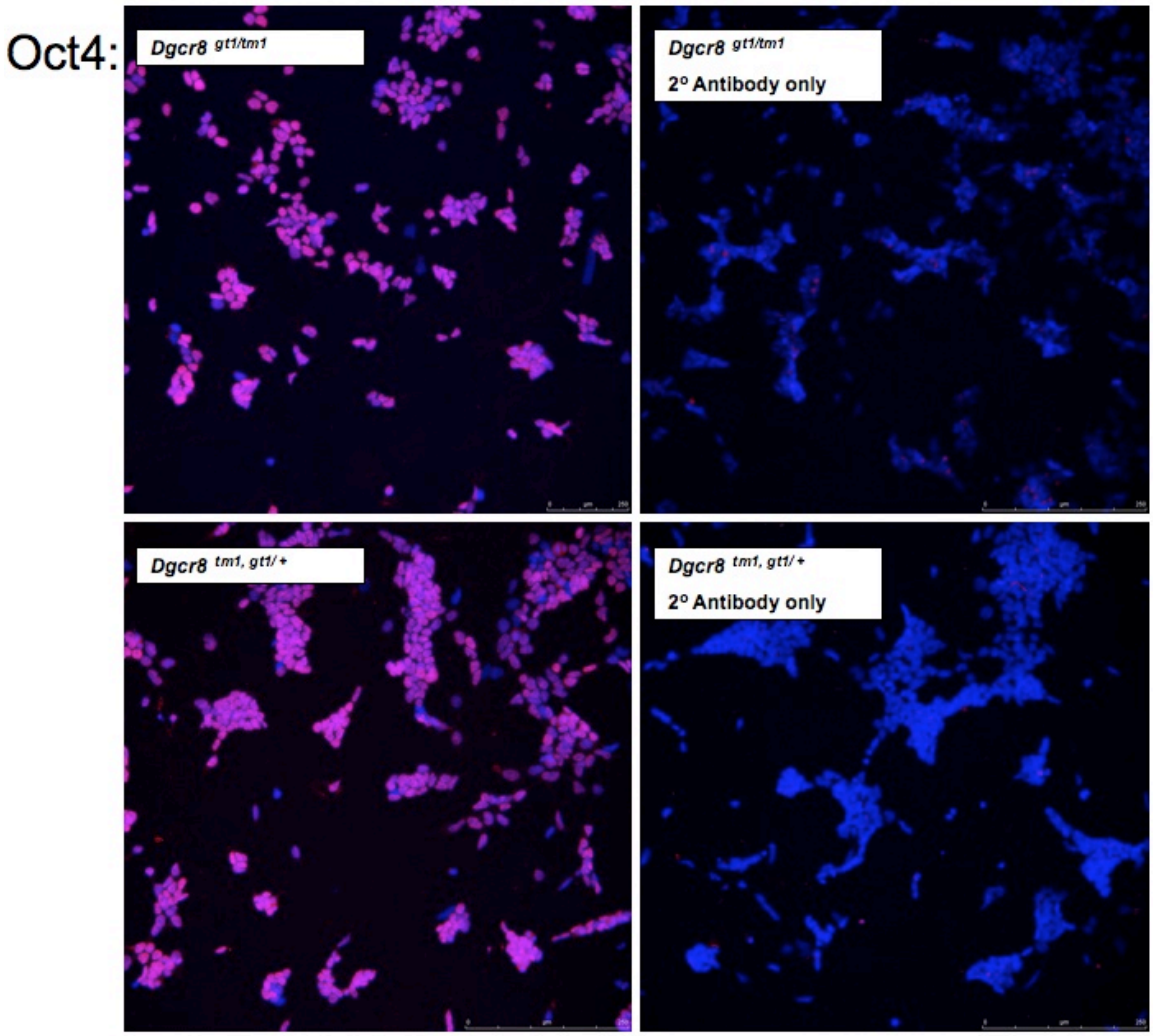


Fig.3.10: Western blot to demonstrate comparable Oct4 expression between cell lines. The same protein blot was probed with both an Oct-4 antibody and an α-tubulin antibody as a loading control.

3.5.1.2 Comparing Oct4 and SOX2 expression by the immuno-staining of cell cultures

The above analysis of the core transcriptional network was extended by culturing *Dgcr8*$^{gt1/tm1}$ and the *Dgcr8*$^{tm1,gt1/+}$ cells on slides and immuno-staining with antibodies to Oct4 and SOX2. One important consideration is that the excitation spectrum of enhanced green fluorescent protein (EGFP), potentially expressed by both cell lines from the targeted trap, overlaps with that of Alexa-488 used to detect the SOX2 antibody. However, earlier experiments with fluorescent microscopy suggested that EGFP is expressed at very low levels in these lines (data not shown). In order to determine if background fluorescence interfered with the antibody specific fluorescence signals, fluorescence sections were taken through control

wells containing cells stained solely with secondary antibodies at the same experimental settings. None of these wells produced significant signals. This would suggest that although EGFP may be expressed from this trap, the level is insufficient to interfere with the results of the cell staining.



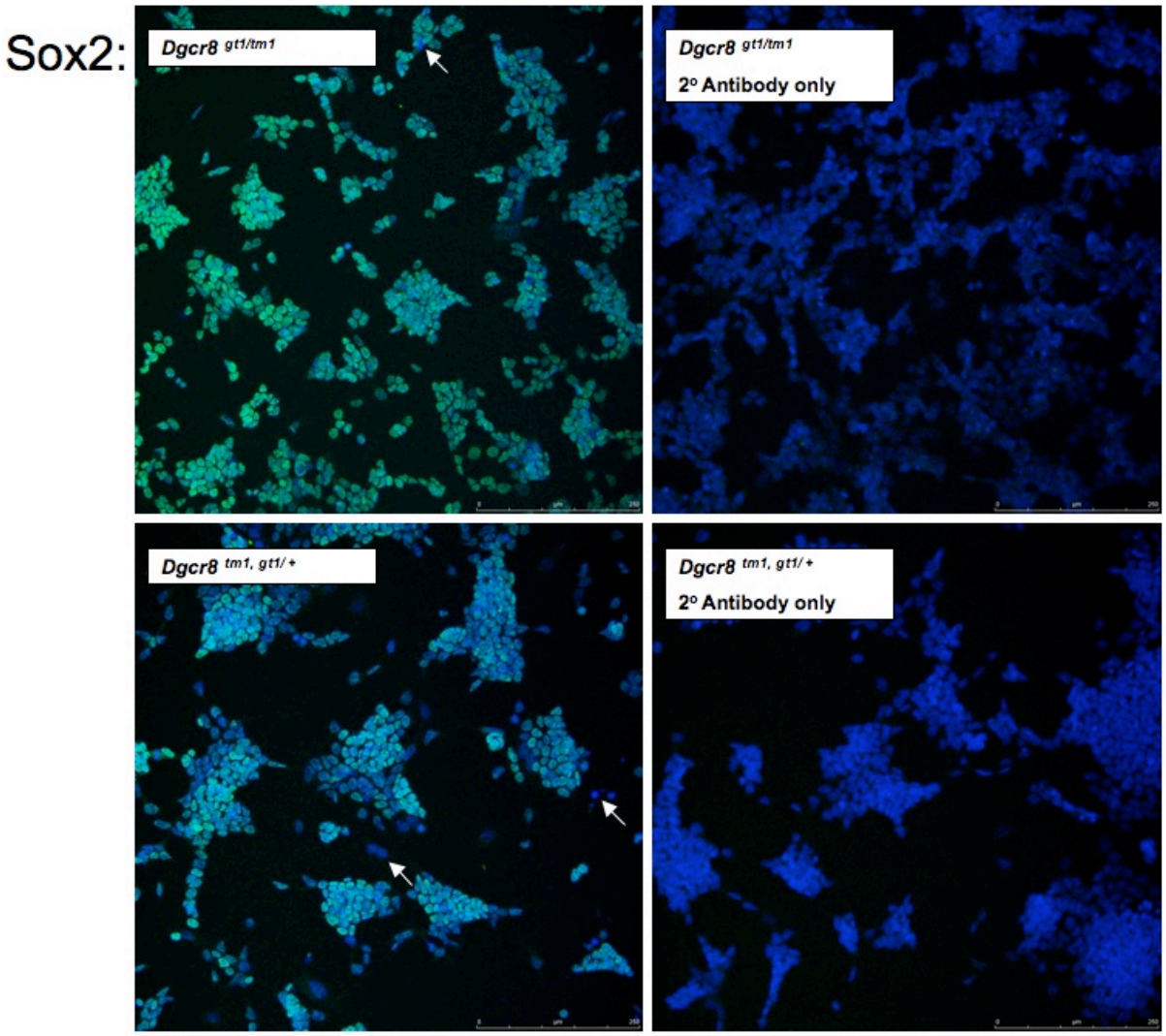


Fig.3.11: Immuno-staining of *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{gt1/tm1}$ cells to demonstrate consistent expression of both Oct4 and *Sox2* within the cultures. Slides of *Dgcr8*$^{tm1,\ gt1/\ +}$ and *Dgcr8*$^{gt1/tm1}$ were prepared. Slide wells were stained with primary antibodies specific to Oct4 or SOX2 and Alexa 594 (Oct4, red) or Alexa 488 (SOX2, green) conjugated secondary antibodies. A z-series of ten images was captured for test samples and the maximum intensity at each x,y coordinate was projected to produce a flat image. This image was overlaid with DAPI staining (blue). Control images are a maximum projection of a series of 5 slices taken through any visible fluorescence, again overlaid with DAPI staining. White arrows indicate cells that appear not to express SOX2.

The results suggest that there is not a substantial difference in the expression of Oct4 and *Sox2* between the *Dgcr8*$^{gt1/tm1}$ and the *Dgcr8*$^{tm1,gt1/+}$ cell lines, implying that these cells do broadly maintain their ES cell identity as judged by the expression of pluripotency critical transcription factors (Fig.3.11). It should be noted that although both cell lines do appear to

express these proteins, it is very difficult to quantitate the absolute level of expression by this method and hence subtle differences in protein levels cannot be excluded (as may be the case with *Sox2*). The apparent low-level variation could be the result of differential staining in different regions of the well, or residual fluorescence from non-specific antibody that may vary to a small extent between wells. However, one benefit of the use of immuno-staining in this way, rather than by Western blots is that it is possible to see that the expression of these proteins remains consistent throughout the cell populations and that there are not significant populations of differentiated cells within these cultures.

3.5.2 Morphological phenotype of *Dgcr8*$^{gt/tm1}$ cells

During the course of this study, it became apparent that a subset of cells in the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cultures had a subtly different morphology to those of the control cell lines (Fig.3.12). This phenotype has either not been seen or not described by previous studies with impaired miRNA processing pathways (Kanellopoulou et al., 2005; Murchison et al., 2005; Sinkkonen et al., 2008; Wang et al., 2007). However, in some of these studies ES cell cultures were maintained on a supportive layer of mouse embryonic fibroblasts (MEFs) (Wang et al., 2007). Indeed one study noted that these supportive cells improved the phenotypes they saw upon the disruption of the miRNA processing pathway (Murchison et al., 2005). Nevertheless, I continued to culture the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells in the absence of MEFs as neither the *Dgcr8*$^{gt1/tm1}$ or *Dgcr8*$^{gt2/tm1}$ cells showed the loss of ES cell marker expression after an extended period in culture. In addition, it was my intention to use these cells for the generation of ES cell specific miRNA target lists. For these experiments, a MEF free culture would be desirable to obtain the most specific signal possible. In the future, it would be interesting to culture these cells with MEFs in order to see if they would

ameliorate this morphological phenotype. ES cells lines cultured for extended periods and mice of different strain backgrounds can also harbour copy number variations or other mutational changes that may cause phenotypic differences. These effects could account for differing phenotypes and changes in expression between different knockout cell lines (Liang et al., 2008; Sibilia and Wagner, 1995).

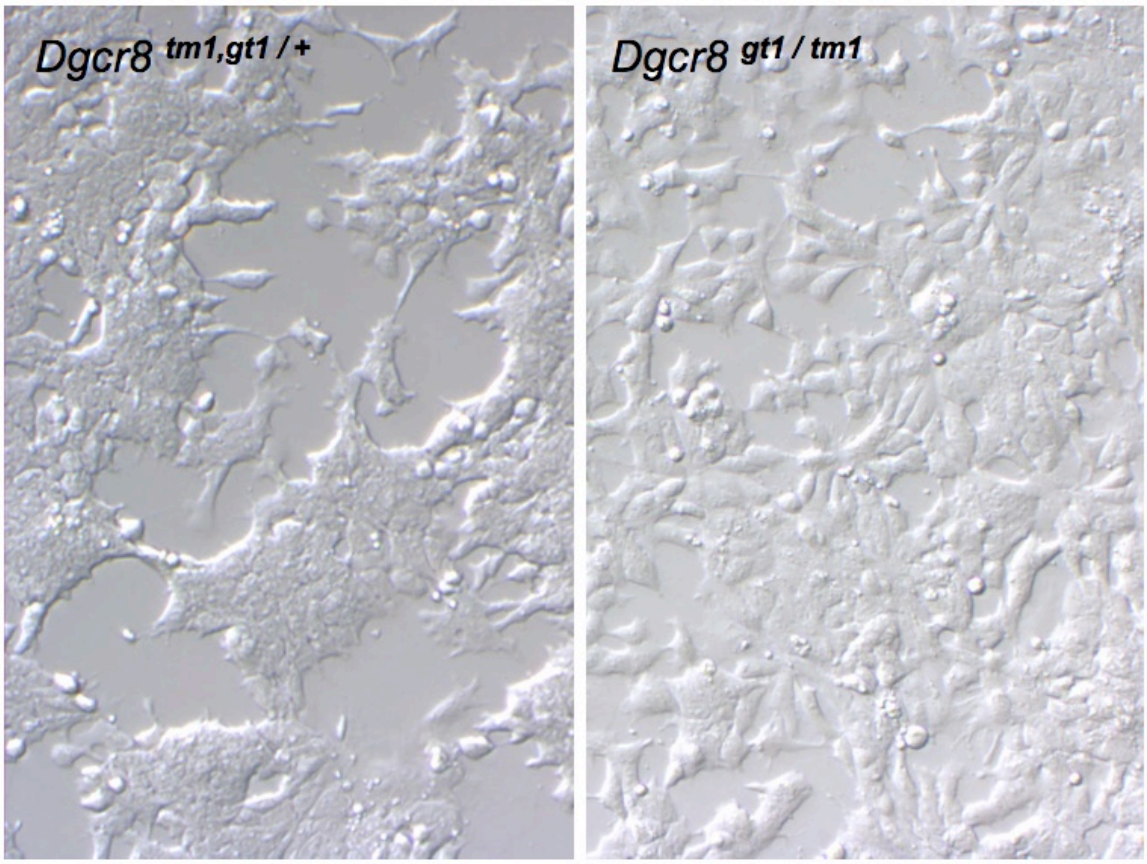


Fig.3.12: An example of the morphological phenotype seen among *Dgcr8*$^{gt1/tm1}$ cells when compared to *Dgcr8*$^{tm1,gt1/+}$ controls. Cells were cultured for 4 days with no selection prior to images being captured (10x relief contrast). A proportion of the *Dgcr8*$^{gt1/tm1}$ cells have a broader and slightly flatter morphology than the *Dgcr8*$^{tm1,gt1/+}$ cells, while the *Dgcr8*$^{tm1,gt1/+}$ cell morphology resembles that of wild type cells.

3.5.3 Investigation of *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{tm1,\ gt1/\ +}$ cell pluripotency

Differentiation via embryoid bodies (EBs) is an *in vitro* technique used to simulate the early embryonic differentiation of embryonic stem cells (Desbaillets et al., 2000). Within the balls

of cells that develop, ES cells differentiate into cells derived from all three embryonic germ layers.

Upon plating into a non-tissue culture treated, round bottom plate, in the absence of LIF, wild type cells aggregate into balls of cells. When transferred into low attachment plates these EBs will continue to grow and after 7-8 days they will begin to beat and contract due to the development of cardiomyocytes within these aggregates. When transferred to gelatinised plates after 8 days, the embryoid bodies attach and begin to spread across the plate, revealing a myriad of morphologically distinct cell types. Patches of these cells continue to beat, again betraying their lineage. The addition of retinoic acid (RA) can induce a greater proportion of cells within the aggregates to follow the neuronal lineage. EBs cultured with RA at the correct stage of differentiation and in the correct concentration will no longer beat and a large number of neuritic outgrowths are evident when the aggregates spread on gelatinised plates (Guan et al., 2001).

I attempted to optimise the technique for *Dgcr8*$^{gt1/tm1}$ differentiation. First I plated a cell number equal to and double that used to differentiate wild type cells in order to compensate for their slower rate of growth. It rapidly became apparent that the *Dgcr8*$^{gt1/tm1}$ cells formed EBs less successfully than wild type cells. At the time of transfer to 24 wells, the *Dgcr8*$^{gt1/tm1}$ EBs had formed as multiple foci in each well as opposed to forming a single large EB. Each EB was tiny and was not spherical. The phenotype was worse in EBs formed from $1x10^3$ cells and these EBs were discarded. By day 8 the EBs formed from $2x10^3$ cells had begun to disintegrate. When transferred to standard wells the mutant EBs did not stick to the plates. Next, I increased the number of *Dgcr8*$^{gt1/tm1}$ cells added to each 96 well to $3x10^3$ and $4x10^3$

cells, reasoning that the formation of tiny EBs may be due to a growth deficit in the *Dgcr8*$^{gt1/tm1}$ cells. These EBs still seemed to form as multiple foci in each well. At the stage of transfer the EBs formed from 3x10^3 cells were still small with a few rare examples of reasonably sized EBs. The 4x10^3 wells seemed to follow the same trend of improved phenotype with increased cell number, but still formed as multiple foci of various sizes. The *Dgcr8*$^{gt1/tm1}$ EBs were still not spherical and were "hairy". By day 8 the RA treated EBs appeared to be less coherent than those without RA. By day 12 a very small proportion of the EBs formed from 4x10^4 cells had began to stick to the well and spread. The EBs without RA had spread to a greater extent as is seen with wild type EBs. No contraction of *Dgcr8*$^{gt1/tm1}$ cells without RA was seen. The majority of the remaining EBs were loosely attached, although some had attached more firmly than others. Clearly this was a limited set of experiments and the specific phenotypes relating to –LIF/+RA and –LIF EBs would have to be followed up by a far more comprehensive study in order to define them more confidently.

Following the optimisation of the *Dgcr8*$^{gt1/tm1}$ differentiation and with a slight alteration in the protocol by which the cells were spun out of trypsin and resuspended in media without LIF prior to plating to 96 wells, in order to remove any traces of LIF in the cultured cell media, *Dgcr8*$^{+/+}$, *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{gt1/tm1}$ cells were all differentiated via EBs, both with and without RA. Both the control cell lines to which I compared to the *Dgcr8*$^{gt1/tm1}$ cells (*Dgcr8*$^{+/+}$ and *Dgcr8*$^{tm1,gt1/+}$) were plated at densities of 1x10^3 cells per 96 well. These control cell lines developed as expected. Following the removal of LIF, a large proportion of beating EBs were observed, indicating the presence of differentiated cardiac muscle. This differentiation pathway was broadly negated by the addition of RA and once plated these EBs spread to reveal a greater presence of tight bundles of neuronal cells (Table 3.4). The single beating

colony amongst those of RA treated *Dgcr8*$^{tm1,gt1/+}$ EBs suggests that future experiments should be repeated with a higher concentration of RA to achieve more significant commitment to neuronal development to the detriment of mesodermal development. The *Dgcr8*$^{gt1/tm1}$ cells were plated at a density of $4x10^3$ cells per 96 well as this was the concentration that resulted in the largest proportion of EBs spreading on the standard gelatinised tissue culture plates once the EBs were plated. Once again the *Dgcr8*$^{gt1/tm1}$ cells formed as multiple foci of many varying sizes within the wells of the 96 well plates. A rough average of the number of foci per well, taken across ten 96 wells, was 26, as opposed to the single foci seen in the case of the wild type and *Dgcr8*$^{tm1,\ gt1/+}$ cells.

Dgcr8$^{+/+}$		
Conditions	Beating EBs	Total EBs
-LIF	18	30
-LIF +RA	0	36

Dgcr8$^{tm1,gt1/+}$		
Conditions	Beating EBs	Total EBs
-LIF	26	29
-LIF +RA	1	38

Dgcr8$^{gt1/tm1}$		
Conditions	Loose EBs	Attached EBs
-LIF	~361	~12.75
-LIF +RA	~300	~3.25
	Average of 2 wells	Average of 4 wells

Table 3.4: Summary of the phenotypes seen amongst EB colonies after 12 days of culture in the absence of LIF. The vastly increased number of smaller EBs in the case of the *Dgcr8*$^{gt1/tm1}$ cells caused me to only count the number of EBs in a sample of wells, whereas all the control EBs plated were considered. All of the *Dgcr8*$^{+/+}$ and *Dgcr8*$^{tm1,gt1/+}$ EBs attached to the plates successfully, while none of the *Dgcr8*$^{gt1/tm1}$ EBs began to beat as cells differentiated into cardiomyocytes.

Once plated to low attachment plates, there was evidence of EBs fusing both amongst the control EBs and the *Dgcr8*$^{gt1/tm1}$ EBs as expected. By day 8, the control cells had formed large round EBs (Fig.3.13). There was clear evidence of twitching cardiac muscle cells amongst

the *Dgcr8*$^{+/+}$ and *Dgcr8*$^{tm1,gt1/+}$ EBs without LIF or RA. The *Dgcr8*$^{gt1/tm1}$ EBs were much smaller and poorly defined (Fig.3.13). In some cases the balls of cells were so small that it was difficult to discern the EBs from cellular debris that had accumulated in the wells.

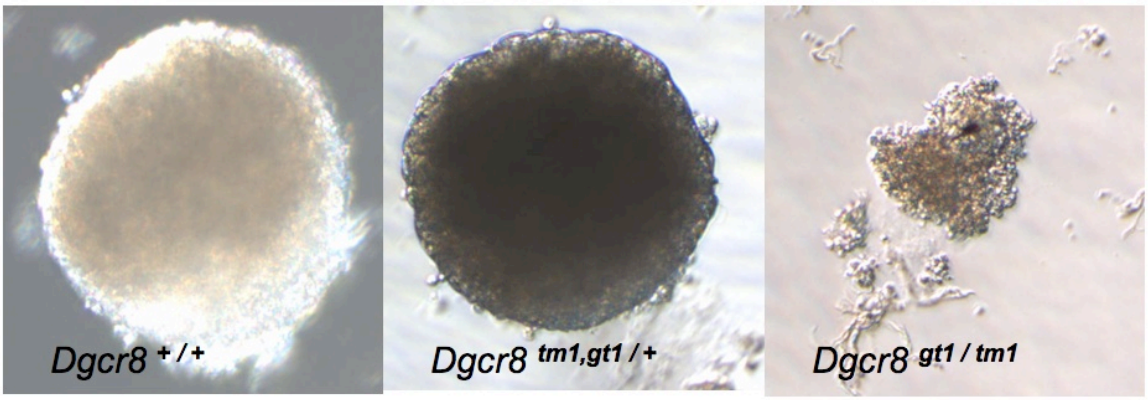


Fig.3.13: Images of EBs cultured on low attachment plates after 8 days of culture in the absence of LIF. All three images are taken at 10x relief contrast.

At day 12, the EBs of each cell line were counted and loosely phenotyped (Table 3.4). The attached and spread EBs of all cell lines were also fixed and representative images taken (Fig.3.14).

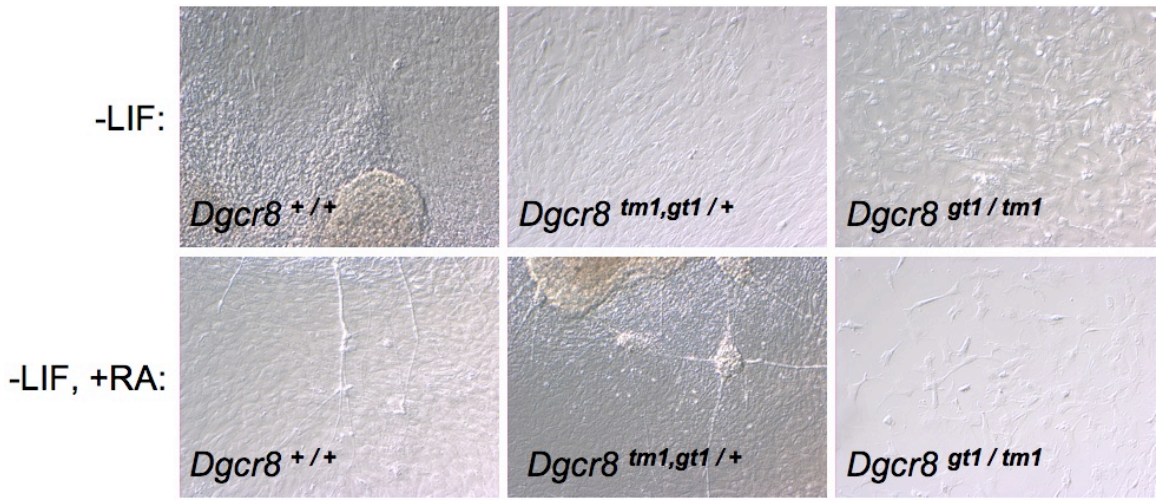


Fig.3.14: Representative images of cell morphologies seen amongst EBs spreading on gelatinised tissue culture plates. Examples are presented of EBs cultured in both the presence and absence of RA. All images are taken at 10x relief contrast. Due to the size difference between the *Dgcr8*$^{gt1/tm1}$ EBs and the control EBs, the spread *Dgcr8*$^{gt1/tm1}$ cells are more sparsely packed on the gelatinised plates.

It is not possible to ascribe a cause to the phenotype seen in the case of the *Dgcr8*$^{gt1/tm1}$ cells. The number of pathways involved in differentiation is vast and any or all of these pathways could be affected in some way by the reduced levels of *Dgcr8* and miRNAs within the cells. Contributing factors could range from deficiencies in cell adhesion to a reduced rate of cellular proliferation. *Dgcr8*$^{gt1/tm1}$ cells are clearly incapable of forming fully functional EBs. They do, however, seem capable of differentiating into a range of morphologically different cells.

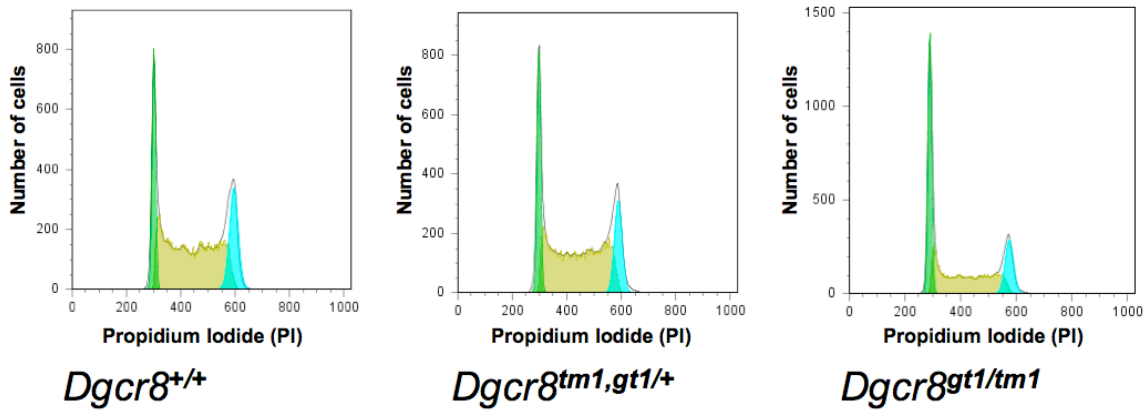
3.5.4 Flow sort for cell cycle

It has been noted that knocking out both Dicer and *Dgcr8* in mouse embryonic stem cells leads to a slower cell cycle and an accumulation of cells in the G1 phase (Murchison et al., 2005; Wang et al., 2007). *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells appeared to grow more slowly than *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ cells. I therefore performed a cell cycle analysis by fluorescence activated cell sorting (FACS) to ascertain whether *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells were progressing through the cell cycle in the normal fashion.

Cells were grown for 2 days without selection and then plated in the same quantities as had been previously used for RNA lysis, into non-selective media. *Dgcr8*$^{gt1/tm1}$ cells were plated at two quantities. Throughout it had been difficult to ensure *Dgcr8*$^{tm1,gt1/+}$ and *Dgcr8*$^{tm1,gt2/+}$ cells and *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells were equally confluent after a further 2 days in culture. As confluence may have an effect on the cell cycle experiments, I decided to plate more than one cell quantity for this analysis in the case of *Dgcr8*$^{gt1/tm1}$. I would then be able to compare the results to see if cell density had a profound effect on cell cycle distribution. For future analysis I would first generate a growth curve for each cell line to be used, to be

sure that the cell cycle profile for each cell line is generated while the cells are in the exponential phase of growth.

Despite these caveats, this experiment does concur with previous cell cycle analyses conducted on Dicer and *Dgcr8* knockout cell lines, with $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells accumulating in G1 phase (34-37%) compared to $Dgcr8^{tm1,gt1/+,}$ $Dgcr8^{tm1,gt2/+}$ and $Dgcr8^{+/+}$ cells (18-22%) (Fig.3.15). It is also clear that the increase in cell density from 304 x 10^4 cells plated to 380 x 10^4 cells had a minimal effect on the cell cycle profile of the $Dgcr8^{gt1/tm1}$ cells. These results along with those of previously published studies concur that miRNAs play an important role in the regulation of the cell cycle in mouse ES cells, and more specifically in the regulation of the G1 to S phase transition.



Cells	G1	S	G2
$Dgcr8^{+/+}$	18.40%	61.25%	20.29%
$Dgcr8^{tm1,gt1/+}$	21.10%	60.59%	16.43%
$Dgcr8^{gt1/tm1}$ (304)	36.30%	45.10%	17.34%
$Dgcr8^{gt1/tm1}$ (380)	36.68%	44.87%	16.63%
$Dgcr8^{tm1,gt2/+}$	19.97%	60.90%	17.71%
$Dgcr8^{gt2/tm1}$	34.49%	49.23%	16.97%

Fig.3.15: ES cells depleted in the expression of *Dgcr8* and mature miRNAs accumulate in the G1 phase of the cell cycle. The upper three panels are a representative sample of the FlowJo cell cycle histograms. The x-axis is a linear measure of the propidium iodide fluorescence associated with each cell, used to quantify cell DNA content. The green, beige and blue segments represent the portion of cells in the G1, S and G2 phases respectively. The lower panel provides a summary of the proportion of cells from each cell line in eachphase of the cell cycle.

3.6 Discussion

Using a targeted trap to target the second allele in heterozygous gene trapped cell lines, I derived multiple cell lines in which both alleles of the *Dgcr8* locus were disrupted. I confirmed the genotypes of these cell lines by RT-PCR to amplify regions from trap specific transcripts, by X-gal staining to reveal the expression of trap specific markers and by Northern blot identifying the presence of trap disrupted transcripts specific to the *Dgcr8* locus. The Northern blot also revealed a severe reduction in wild type *Dgcr8* transcript levels in these cells.

Subsequently, despite subtle changes in morphology and an altered cell cycle profile in the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ lines, I confirmed that they maintained their expression of essential ES cell associated TFs. The $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells express Oct4 at a comparable level not only to $Dgcr8^{tm1,gt1/+}$ and $Dgcr8^{tm1,gt2/+}$ cells but also to wild type cell lines, as assessed by Western blot. In addition, both Oct4 and *Sox2* expression is maintained throughout the culture, as seen with the immuno-stained cultures.

Although it is possible to maintain the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells in culture for extended periods without the loss of ES cell identity, there are fundamental alterations in ES cell phenotype of *Dgcr8* homozygous mutant cells. Most significant is the loss of ES cell miRNA expression, presumably due to reduced microprocessor activity. Through the use of miRNA Northern blots I demonstrated the loss of miR-292-3p, miR-293, miR-130, miR-21 and miR-92a expression in the $Dgcr8^{gt1/tm1}$ and $Dgcr8^{gt2/tm1}$ cells. miRNAs are expected to regulate a large proportion of the mouse transcriptome (Friedman et al., 2009) and the loss of a large number of miRNAs from the system could lead to considerable alterations in the

cellular transcriptional profile. It is not surprising therefore that there are some fundamental changes in both the cell cycle and EB formation, when these cells are compared to *Dgcr8*$^{tm1,gt1/+}$, *Dgcr8*$^{tm1,gt2/+}$ and *Dgcr8*$^{+/+}$ cells. EBs from *Dgcr8*$^{gt1/tm1}$ cells were smaller and less coherent than *Dgcr8*$^{tm1,\ gt1/+}$ and *Dgcr8*$^{+/+}$ aggregates. *Dgcr8*$^{gt1/tm1}$ EBs do not adhere to gelatinised plates consistently although when they can adhere and spread, a variety of differentiated cellular morphologies are identifiable. In order to interpret the inability to form healthy EBs and understand the mechanistic implications, considerable further work would be required to identify candidate genes affected at the core of this process. This has recently been achieved by Sinkkonen *et al.* who demonstrated the role of the miR-290 cluster in the regulation of the methylation of the Oct4 promoter during the course of differentiation (Sinkkonen et al., 2008). The re-addition of this cluster to differentiating Dicer deficient ES cells went some way to rescuing the Oct4 transcriptional silencing defect identified in these cells.

In order to continue this avenue of research, a wide range of experiments could be attempted. Initially lineage marker immuno-staining would provide an insight into the exact lineages to which the *Dgcr8*$^{gt1/tm1}$ cells are capable of contributing. This study should be inclusive of both EBs that adhere to plates and those that remain loose after 12 days of culture. It may also be informative to differentiate *Dgcr8*$^{gt1/tm1}$ cells as a monolayer, removing the more complex aspects of EB differentiation. In parallel, an *in vivo* analysis of *Dgcr8*$^{gt1/tm1}$ cell contribution to lineages in a mouse embryo should be considered. The ß-geo marker expressed by the gene trap in *Dgcr8*$^{gt1/tm1}$ cells would allow these cells to be traced following microinjection into blastocysts. The presence of a β-geo marker is a distinct advantage over previous microinjection studies of similar cell lines (Kanellopoulou et al., 2005). These experiments

would lead to a more complete understanding of the developmental potential of mutant *Dgcr8* cells.

In addition, *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells seem to accumulate in the G1 phase of the cell cycle, 34-37% compared to 18-22% for control cells. Members of the miR-290 cluster have been shown to regulate the cell cycle inhibitor p21 (*Cdkn1a*) and appear to play an active role in increasing the rate of proliferation when reintroduced into *Dgcr8* knock out mouse ES cells (Wang et al., 2008). Other miRNAs have also been found to regulate the cell cycle (Liu et al., 2008a; Petrocca et al., 2008a). It seems that a number of miRNAs may converge in the regulation of the cell cycle in mouse ES cells, thus piecing together the exact nature of the interplay between these factors will require considerable future effort.

The phenotype of the *Dgcr8*$^{gt1/tm1}$ and *Dgcr8*$^{gt2/tm1}$ cells is broadly consistent with previous attempts to disrupt the miRNA-processing pathway in mouse ES cells (Kanellopoulou et al., 2005; Murchison et al., 2005; Wang et al., 2007). I have confirmed the disruption of the *Dgcr8* locus, the loss of ES cell mature miRNAs from the system and the continued expression of stem cell markers. I therefore conclude that I have derived a suitable system into which I can reintroduce miRNAs to determine mouse ES cell specific miRNA targets, in a background within which endogenous miRNA maturation will be depleted.