# The role of microRNA in regulation of lineage-specific gene expression through the cell cycle

by

Bryony Jane Graham

A thesis submitted to Imperial College London

for the degree of Doctor of Philosophy

MRC Clinical Sciences Centre

Imperial College School of Medicine

May 2012

I, Bryony Jane Graham, declare that the work presented in this thesis is my own, and that any work carried out by others has been acknowledged and appropriately referenced in the text.

# Abstract

The establishment and maintenance of highly specialised cell lineages is fundamental to the development of multicellular organisms. Cell identity is determined by specific transcriptional profiles, which are mediated by sequence-specific, chromatin-based and post-transcriptional mechanisms of gene regulation. Changes in cell morphology and chromatin structure which occur during the cell cycle present a challenge to the maintenance of lineage-specific gene expression profiles. This study investigates the role of post-transcriptional regulation by microRNAs in stabilising cell-specific gene expression through the process of cell growth and division. Data presented here show that microRNAs are inherited through mitosis in mammalian cells, and are capable of regulating target gene expression in recipient daughter cells. Genome-wide expression analysis indicates that key developmentally regulated genes marked by a bivalent chromatin signature are globally upregulated in microRNA-deficient ES cells. Binding sites for ES cell-specific microRNAs are significantly enriched in the 3'UTR of these transcripts compared to the rest of the transcriptome, strongly suggesting that microRNAs contribute to maintenance of ES cell identity by co-ordinately regulating multiple lineage inappropriate genes. Finally, analysis of the expression of validated microRNA targets and bivalent genes throughout the cell cycle shows that transcripts from these genes accumulate in G2/M to a greater extent in microRNA-deficient ES cells than in wildtype cells. Taken together, data presented in this study support a role for microRNAs in regulation of lineage-specific gene expression through the cell cycle.

# Acknowledgements

First and foremost, I would like to thank Matthias and Mandy for the opportunity to work in such a stimulating and rewarding environment. I am particularly grateful to Matthias for all his guidance and support over the past three and half years; for challenging me and giving me the confidence and the drive to pursue this career further. This work would not have been possible without financial support from the MRC, for which I am extremely thankful.

I will be forever indebted to Antoine for his interminable patience; for always making time to answer even the most inane of questions; and for his unwavering optimism. Every single member of the Lymphocyte Development Group has made my time in the lab enjoyable. Thanks to Hegias for advice and enthusiasm; to Francesco for continually reminding me of the British stereotype (and how I fit it); for Thais for always having a smile to share; to Isabel for words of comfort and support; to Ludovica for therapeutic runs and bike rides; to Luke for saying exactly the right thing at exactly the right time; to Ben for making me laugh and teaching me everything I know with a pipette; and to Vlad for thoughtful discussions on many topics. I am extremely grateful to Jorge and David for critical reading of this manuscript, and to absolutely everyone – Rory, Ziwei, Hakan, Irene, Andy, Preksha, Cynthia, Feng, Tomomi, Sue, Maria, Helle and Amelie – for a fantastic time in the lab; you have all made the hard times bearable, and the good times great. I would also like to thank Kelly for always making time for coffee; and never complaining when I don't let her get a word in edgeways. I would also like to acknowledge everyone who has had the misfortune of living with me over the past four years – for putting up with the early starts and the late finishes, and the tumultuous experiment-dependent mood swings. In particular, thanks to Abbie - for picking up the pieces on more occasions than I would care to recall.

There are no words to describe the support that I have had from Tom. Thank you for never doubting me; never questioning why; and quietly but firmly keeping me going – I could not have done this without you.

Finally I would like to thank my wonderful family – Mum, Dad, Caroline and Archie – whose relentless optimism, love and just the right level of ambivalence have kept everything in perspective right from the very beginning. You are all my inspiration, and I dedicate this thesis to you.

# Table of Contents


Abstract 3
Acknowledgements 4
Table of Contents 5
Figures and Tables 9
Abbreviations 12

**1.0 Introduction 15**
1.1 The role of microRNA in gene regulation 15
1.1.1 MicroRNA biogenesis 16
1.1.2 Mechanism of microRNA function 18
1.1.3 Regulation of microRNA activity 21
1.1.4 Biological role of post-transcriptional gene regulation by microRNAs 23
1.2 Epigenetic regulation of gene expression 25
1.2.1 Nuclear architecture and chromatin structure 25
1.2.2 DNA methylation 27
1.2.3 Histone modifications 28
1.2.4 RNA inheritance 32
1.3 Regulation of gene expression through the cell cycle 33
1.3.1 Overview of the mammalian cell cycle 34
1.3.1.1 MicroRNA-mediated cell cycle regulation in ES cells 35
1.3.2 Propagation of gene expression through the cell cycle 36
1.4 Aims of study 39

**2.0 Materials and Methods 40**
2.1 Materials 40
2.1.1 Oligonucleotides 40
2.1.2 Antibodies 40
2.1.3 Tools for bioinformatics analysis 41
2.2 Methods 42
2.2.1 Cell culture 42
2.2.1.1 Lymphocytes 42
2.2.1.2 ES cells 43
2.2.1.3 3T3 fibroblasts 43
2.2.2 Fluorescence activated cell sorting (FACS) analysis 44
2.2.2.1 CFSE staining 44

2.2.2.2 Cell surface marker analysis 44
2.2.2.3 Cell cycle analysis 44
2.2.3 Transfection methods 45
2.2.3.1 Amaxa Nucleofection™ 45
2.2.3.2 Lipofectamine™ 45
2.2.3.3 DharmaFECT® 45
2.2.4 Genomic PCR 46
2.2.5 RNA analysis 46
2.2.5.1 Quantification of mRNA 46
2.2.5.2 Quantification of miRNA 48
2.2.6 Sample preparation for microarray analysis of gene expression 48
2.2.7 Mitotic arrest and release assay 50
2.2.8 Elutriation 51
2.2.9 Western blot 52
2.2.10 Chromatin Immunoprecipitation 52
2.2.11 Immunofluorescence 54

**3.0 Mitotic inheritance of microRNA 56**
3.1 Introduction 56
3.2 Optimisation of mitotic inheritance assay in primary mouse $CD4^+$ T cells 56
3.3 Mitotic inheritance of fluorescently labelled microRNA mimic 58
3.4 Mitotic inheritance of unmodified non-T cell expressed microRNA mimic 59
3.5 Mitotically heritable microRNAs downregulate target gene expression in recipient cells 61
3.6 Discussion 66
3.6.1 Identification of potential novel targets of miR-142-3p in $CD4^+$ T cells 66
3.6.2 Mitotic inheritance of miRNA in mammalian cells: evaluation of system 67

**4.0 Analysis of gene expression through the cell cycle in microRNA-deficient mouse embryonic stem cells 69**
4.1 Introduction 69
4.2 Analysis of gene expression in Dicer WT/KO ES cells in early G1 69
4.2.1 Optimisation of mitotic arrest and release assay 69
4.2.2 Genome-wide analysis of gene expression in asynchronous populations of Dicer WT/KO ES cells and in early G1 74
4.2.2.1 Microarray Analysis (1): G1-specific deregulation of histone genes in Dicer KO ES cells 78
4.2.2.2 Microarray Analysis (2): Deregulation of bivalent genes in Dicer KO ES cells 80

4.2.2.3 Microarray Analysis (3): Identification of a subset of predicted miRNA targets deregulated to a greater extent in Dicer KO cells in early G1 compared to an asynchronous population 85

4.2.3 Discussion (1) 87

4.2.3.1 Expression of histone gene expression is co-ordinately and specifically deregulated in Dicer KO ES cells in early G1 87

4.2.3.2 Identification of a potential role for miRNA in regulation of protein phosphorylation in early G1 89

4.2.3.3 Re-evaluation of mechanisms governing regulation of lineage-specific gene expression in pluripotent stem cells 89

4.3 Analysis of gene expression in Dicer WT/KO ES cells throughout the cell cycle 92

4.3.1 Optimisation of elutriation system 92

4.3.2 Quantification of gene expression through the cell cycle 94

4.3.3 Kinetics of histone gene expression in elutriated populations 96

4.3.4 Expression of validated microRNA targets through the cell cycle in Dicer KO ES cells 96

4.3.5 Expression of bivalent genes through the cell cycle in Dicer KO ES cells 98

4.3.6 Discussion 99

4.3.6.1 Use of elutriated populations to detect trends in gene expression through the cell cycle 99

4.3.6.2 Accumulation of microRNA targets through the cell cycle in Dicer KO ES cells 99

4.4 Summary and concluding remarks 101

**5.0 MicroRNA-dependent transcriptional and post-transcriptional regulation of bivalent genes 103**

5.1 Introduction 103

5.2 Evidence for direct post-transcriptional regulation of bivalent genes by microRNAs 103

5.3 Expression of PRC2 components is not altered in Dicer KO ES cells 104

5.4 Levels of Ezh2 binding and H3K27me3 are reduced at bivalent genes in Dicer KO ES cells 105

5.5 Production of full length nascent RNA from bivalent genes is increased in Dicer KO ES cells 107

5.6 Kinetics of expression of mature and nascent RNA from bivalent genes in conditional Dicer KO ES cells suggests that miRNAs function at both the transcriptional and post-transcriptional level in regulation of PRC2 bound genes 108

5.7 Discussion 112

5.7.1 Loss of PcG binding at bivalent genes in Dicer KO ES cells results in transcriptional deregulation of genes marked by both H3K4me3 and H3K27me3 112

5.7.1.1 A direct role for miRNA in transcriptional gene silencing? 116

5.7.1.2 Does dephosphorylation of Ezh2 abolish ncRNA-mediated PcG recruitment to bivalent genes? 116

5.7.1.3 Is Polycomb redistributed to genomic regions that become hypomethylated in cells with reduced DNA methyltransferase expression? 118

5.7.1.4 A role for RNAi in long-range interactions between PcG-repressed genes? 120

**6.0 General Discussion and Future Perspectives 122**

6.1 Mammalian microRNAs are inherited through cell division and are functional in daughter cells 122

6.2 MicroRNAs regulate expression of lineage-appropriate genes through the cell cycle in ES cells 122

6.3 Incorporation of combinatorial miRNA-mediated regulation of bivalent genes into regulatory networks governing the pluripotent state 124

**7.0 References 126**

**8.0 Appendices 143**

I. Derivation of stable and conditional Dicer KO 3T3 fibroblasts 143

a. Derivation of conditional Dicer KO 3T3 fibroblasts 143

b. Derivation of stable Dicer KO 3T3 fibroblasts 145

c. Optimisation of mitotic arrest and release assay in Dicer WT and KO 3T3 fibroblasts 146

II. MicroRNA-mediated regulation of the actin cytoskeleton 148

III. Primer sequences 154

# Figures and Tables

**List of Figures**


Figure 1.1: The microRNA biogenesis pathway in mammalian cells. 18

Figure 1.2: Proposed models for mechanistic basis of miRNA-mediated gene silencing. 20

Figure 1.3: MicroRNAs confer robustness to lineage-specific gene expression. 25

Figure 1.4: Epigenetic mechanisms of gene regulation. 31

Figure 1.5: Overview of the mammalian cell cycle. 34

Figure 3.1: Experimental system to investigate mitotic inheritance of small RNAs. 57

Figure 3.2: Stabilised fluorescently labelled oligonucleotides are inherited through cell division in primary $CD4^+$ T cells. 57

Figure 3.3: Fluorescently labelled miR-181a is transmitted from parent to daughter cells through mitosis in primary mouse $CD4^+$ T cells. 59

Figure 3.4: Neuronal-specific miRNA mimics expressed at endogenous levels in lymphocytes are transmitted through cell division in primary mouse $CD4^+$ T cells. 60

Figure 3.5: Isolation of microRNA-deficient $CD4^+$ T cells from CD4Cre $Dicer^{lox/lox}$ mice. 61

Figure 3.6: Validated miR-142-3p target *Ac9* is not upregulated in activated Dicer KO $CD4^+$ T cells. 62

Figure 3.7: Identification of potential targets of miR-142-3p in $CD4^+$ T cells. 63

Figure 3.8: T-cell specific miRNAs are inherited through mitosis in primary microRNA-deficient $CD4^+$ T cells and downregulate target gene expression in recipient cells. 65

Figure 4.1: Differential cell cycle profiles of T cells and ES cells. 70

Figure 4.2: Derivation of feeder-independent wildtype and microRNA-deficient mouse embryonic stem cells. 71

Figure 4.3: Culture of Dicer WT and KO ES cells in feeder-independent conditions in combined media does not compromise the pluripotent state. 73

Figure 4.4: Mitotic arrest and release yields populations of Dicer WT and KO ES cells in early G1. 74

Figure 4.5: Prinicipal Component Analysis (PCA) of Dicer WT/KO Async/G1 microarray samples. 75

Figure 4.6: Gene expression changes in Dicer KO cells are typically less than two-fold and similar in early G1 and asynchronous populations. 77

Figure 4.7: Motifs for ES-specific microRNA families are significantly associated with upregulated genes in Dicer KO ES cells. 77

Figure 4.8: Histone genes are significantly downregulated in Dicer KO cells specifically in early G1. 79

Figure 4.9: GSEA analysis of microarray data identifies global upregulation of bivalent genes in Dicer KO ES cells. 82

Figure 4.10: Validation of microarray data by qPCR confirms upregulation of bivalent genes in Dicer KO ES cells. 82

Figure 4.11: Simultaneous inhibition of all members of miR-290 family in WT ES cells recapitulates cell cycle defect in Dicer KO ES cells and results in upregulation of validated miR-290 target genes. 83

Figure 4.12: Inhibition of miR-290 family in WT ES cells identifies bivalent genes as potential novel miRNA targets. 84
Figure 4.13: Identification of predicted targets of the miR-290 family upregulated to a greater extent in Dicer KO ES cells in early G1. 86
Figure 4.14: Identification of predicted miR-290 targets which are deregulated to a greater extent in Dicer KO ES cells in early G1 than in an asynchronous population. 86
Figure 4.15: Inhibition of the miR-290 family in WT ES cells identifies a subset of G1-deregulated genes as potential targets of ES-specific miRNAs. 87
Figure 4.16: Model for role of miRNA in G1-specific regulation of histone gene expression. 88
Figure 4.17: Isolation of cells at distinct stages of the cell cycle using elutriation. 93
Figure 4.18: Comparison of endogenous and exogenous RNAs for normalisation of gene expression through the cell cycle. 95
Figure 4.19: Cell cycle-dependent changes in histone gene expression can be detected in elutriated populations of WT ES cells. 96
Figure 4.20: Transcripts from validated miR-290 targets *Cdkn1a*, *Rbl2* and *Lats2* accumulate through the cell cycle to a greater extent in Dicer KO ES cells compared to WT. 97
Figure 4.21: Transcripts from bivalent genes predicted to be targets of the miR-290 family accumulate through the cell cycle to a greater extent in Dicer KO ES cells compared to WT. 98
Figure 5.1: Reintroduction of the miR-290 family into Dicer KO ES cells indicates that transcripts from bivalent genes are directly regulated by ES-specific miRNAs. 104
Figure 5.2: Expression of core PRC2 components and global levels of histone modifications H3K4me3 and H3K27me3 do not change in Dicer KO ES cells. 105
Figure 5.3: Loss of Ezh2 binding at deregulated bivalent genes in Dicer KO ES cells correlates with reduced levels of H3K27me3 but not H3K4me3 at target genes. 106
Figure 5.4: Loss of PcG-binding and H3K27me3 at bivalent genes in Dicer KO ES cells correlates with increased full-length nascent RNA from target genes in a miR-290 dependent manner. 108
Figure 5.5: Characterisation of feeder-independent conditional Dicer KO ES cell system. 109
Figure 5.6: Kinetics of expression of mature and nascent RNA from bivalent genes after acute miRNA depletion. 111
Figure 5.7: Incorporation of post-transcriptional regulation by miRNA into regulatory networks governing expression of bivalent genes in pluripotent cells. 115
Figure 5.8: Model (1): Inhibition of CDK1/CDK2 activity in Dicer KO ES cells results in loss of Ezh2 phosphorylation and abolishes ncRNA-mediated recruitment of PRC2 to target genes. 117
Figure 5.9: Model (2): Downregulation of Dnmts in Dicer KO ES cells creates novel hypomethylated sites which recruit PcG away from canonical target genes. 119
Figure 8.1: Immortalisation of $Dicer^{wt/wt}$ and $Dicer^{lox/lox}$ 3T3 fibroblasts. 144
Figure 8.2: Optimisation of tamoxifen-mediated acute depletion of miRNAs in ERT2Cre $Dicer^{lox/lox}$ 3T3 fibroblasts. 145
Figure 8.3: Derivation of stable Dicer KO 3T3 fibroblasts. 146

Figure 8.4: The mitotic shakeoff and release assay is not suitable for obtaining populations of Dicer KO 3T3 fibroblasts immediately post-cell division. 147
Figure 8.5: Morphological differences in Dicer KO 3T3 fibroblasts compared to WT cells. 148
Figure 8.6: Systematic analysis of actin distribution in Dicer WT and KO 3T3 fibroblasts. 150
Figure 8.7: Polarised actin distribution is impaired in Dicer KO 3T3 fibroblasts. 151
Figure 8.8: Dicer KO 3T3 cells are generally greater in size than Dicer WT 3T3, independent of cell morphology. 151
Figure 8.9: Increased cell size in Dicer KO 3T3 fibroblasts correlates with increase in total actin content. 152
Figure 8.10: mRNA levels of key members of the actin regulatory proteins are deregulated in Dicer KO 3T3 fibroblasts. 153
Figure 8.11: Location and sequences of primers used for analysis of genomic deletion of RNase III domain in cells derived from ERT2Cre x Dicer$^{lox/lox}$ mice. 154

**List of Tables**

Table 1: Key post-translational modifications of core histone tails. 28
Table 2: Summary of Dicer KO vs WT microarray data in asynchronous cells and in early G1. 76
Table 3: Applied Biosystems (AB) Assay Ids for TaqMan® miRNA primer/probe pairs used for qPCR quantification of miRNAs. 154
Table 4: Primer sequences for amplification of mRNA transcripts by qPCR. 154
Table 5: Primer sequences for amplification of specific genomic regions in chromatin immunoprecipitation analyses. 156
Table 6: Primer sequences for amplification of the 5' end of nascent RNA transcripts by qPCR. 157
Table 7: Primer sequences for amplification of the 3' end of nascent RNA transcripts by qPCR. 157

# Abbreviations

| | |
|---|---|
| µ | Micro |
| 2i | Small molecule inhibitors of GSKβ and MEK signalling |
| 3C | Chromosome conformation capture |
| 4-OHT | 4-hydroxytamoxifen |
| AGO | Argonaute protein family member |
| APE 1 | AP-endonuclease 1 |
| bp | Base pair |
| CDK | Cyclin dependent kinase |
| cDNA | Complementary DNA |
| CDS | Coding sequence |
| CFSE | Carboxyfluorescein diacetate succinimidyl ester |
| ChIP | Chromatin immunoprecipitation |
| CpG | Cytosine-guanine dinucleotide |
| cRNA | Complementary RNA |
| C(t) | Threshold cycle |
| d | Day |
| DAPI | 4',6-diamidino-2-phenylindole |
| DMEM | Dulbecco's Modified Eagle's Medium |
| DNA | Deoxyribonucleic acid |
| DNMT | DNA methyltransferase |
| dNTP | Deoxyribonucleotide triphosphate |
| DTT | Dithiothreitol |
| dUTP | 2´-Deoxyuridine, 5´-Triphosphate |
| EDTA | Ethylene diamine tetraacetic acid |
| EGTA | Ethylene glycol tetraacetic acid |
| ERT2 | Oestrogen receptor |
| ES | Embryonic stem |
| EtOH | Ethanol |
| FACS | Fluorescence activated cell sorting |
| FC | Fold change |
| FCS | Foetal calf serum |
| FDR | False discovery rate |
| FISH | Fluorescent *in situ* hybridisation |
| g | Gram |
| GO Term | Gene Ontology Term |
| GSEA | Gene Set Enrichment Analysis |
| h | Hour |
| H | Histone |

| | |
|---|---|
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| IF | Immunofluorescence |
| IL | Interleukin |
| IMDM | Iscove's Modified Dulbecco's Medium |
| IP | Immunoprecipitation |
| IVT | *In vitro* transcription |
| K | Lysine |
| KO | Knockout |
| L | Litre |
| LIF | Leukaemia inhibitory factor |
| LINES | Long interspersed nuclear elements |
| LNA | Locked nucleic acid |
| lncRNA | Long noncoding RNA |
| m | Milli |
| M | Molar |
| me3 | Trimethyl group |
| miRNA | MicroRNA |
| min | Minute |
| mRNA | Messenger RNA |
| ncRNA | Noncoding RNA |
| NP-40 | Nonylphenyl Polyethylene Glycol |
| NTP | Nucleoside triphosphate |
| ORF | Open reading frame |
| PBS | Phosphate buffered saline |
| PcG | Polycomb group proteins |
| PI | Propidium iodide |
| PIPES | Piperazine-N,N′-bis(2-ethanesulfonic acid) |
| PRC1/2 | Polycomb repressive complex 1/2 |
| Pre-miRNA | Precursor microRNA |
| Pri-miRNA | Primary microRNA |
| PCR | Polymerase chain reaction |
| PVDF | Polyvinylidene difluoride |
| qPCR | Quantitative polymerase chain reaction |
| RBP | RNA-binding protein |
| rcf | Relative centrifugal force |
| RIPA | Radioimmunoprecipitation assay buffer |
| RISC | RNA-induced silencing complex |
| RNA | Ribonucleic acid |
| RNAi | RNA interference |
| RNAP II | RNA polymerase II |

| | |
|---|---|
| RNA-seq | RNA extraction followed by high-throughput sequencing |
| rpm | Rotations per minute |
| *S. cerevisiae* | *Saccharomyces cerevisiae* (budding yeast) |
| *S. pombe* | *Saccharomyces pombe* (fission yeast) |
| SDS | Sodium dodecyl sulphate |
| SINES | Short interspersed nuclear elements |
| siRNA | Short interfering RNA |
| snoRNA | Small nucleolar RNA |
| s.t.d. | Standard deviation |
| TCR | T cell receptor |
| TdT | Terminal deoxynucleotidyl transferase |
| TGS | Transcriptional gene silencing |
| Tris | tris(hydroxymethyl)aminomethane |
| TrxG | Trithorax group proteins |
| TSS | Transcription start site |
| U | Units |
| UDG | Uracil-DNA glycosylase |
| UTR | Untranslated region |
| WT | Wildtype |

# 1.0 Introduction

The emergence of multicellular organisms marks a significant event in the course of evolution. The development of progressively complex organisms from primitive multicellular life is dependent on the establishment and maintenance of distinct, specialised cell types, which function co-ordinately to facilitate the increasingly diverse phenotypes observed through evolutionary history. Whilst current knowledge of the distinct mechanisms which govern lineage specification in different organisms is extensive, much remains to be established in terms of how this information is maintained through the cell cycle, and how cell identity is faithfully transmitted from parent cell to progeny. In lower organisms, mounting evidence supports a role for small noncoding RNAs in this process (Voinnet and Baulcombe, 1997; Molnar et al., 2010; Fire et al., 1998). The ubiquitous expression and evolutionary conservation of many noncoding RNAs suggests that these molecules may play a role in many cellular processes in mammals that are yet to be elucidated.

## 1.1 The role of microRNA in gene regulation

The chemical properties of ribonucleic acid (RNA) were first found to be distinct from those of DNA in the early 1900's, but technical limitations meant that the biological relevance of this molecule was not fully appreciated until over 50 years later. The fundamental importance of RNA in molecular biology was encapsulated by Francis Crick in 1970 with his proposition of the 'Central Dogma of Molecular Biology', which illustrates that RNA is the critical intermediate between genetic information and protein production (Crick, 1970). The identification of protein-coding messenger RNA (mRNA) (Jacob and Monod, 1961; Brenner et al., 1961) was swiftly followed by the discovery of transfer RNA (tRNA) (Holley et al., 1965) as the entity responsible for physically linking individual amino acids in a sequence-dependent manner.

The identification of tRNA introduced the concept of non-coding RNA (ncRNA) species which do not, unlike mRNA, generate protein products. Since then, numerous classes of ncRNA have been identified: ribosomal RNAs (rRNA); small nuclear RNAs (snRNA); and most recently, the advent of deep sequencing technology has facilitated the detection of small regulatory RNAs which affect the stability, translation, and possibly even transcription of mRNA molecules. Genome-wide sequencing has also led to the discovery of long noncoding RNAs (lncRNA), which are thought to participate in

recruitment of chromatin remodelling proteins to specific genomic sites. However, the function of the vast majority of these transcripts is unknown, and their biological significance is subject to debate (Kowalczyk et al., 2012). Further to the increasingly diverse species of ncRNA now known to exist in mammalian cells, additional layers of complexity are added to elucidating the biological roles of these molecules by their ability to form extensive secondary structures, be subjected to editing processes (Nishikura, 2010), modified post-transcriptionally (Katoh et al., 2009; Heo et al., 2008), and interact with RNA-binding proteins (Burns et al., 2011).

This study will focus on one of the many subclasses of noncoding RNA: microRNAs. The fundamental role of microRNA in regulation of lineage specification has been apparent since the discovery of the first microRNA, *lin-4*, in 1993 (Lee et al., 1993). *Lin-4* and its target, *lin-14,* were named as such when it was discovered that mutations in these genes lead to defects in the development of several cell lineages in *C. elegans* (Sulston and Horvitz, 1981) . The finding that small RNAs processed from the *lin-4* transcript were able to negatively regulate expression of *lin-14* via sequence complementarity to the 3'UTR was the first indication that small RNA molecules (20-22 nucleotides in length) are involved in gene regulation (Lee et al., 1993). At first, the significance of this discovery was underestimated as it was assumed to be an idiosyncratic mechanism unique to *C. elegans*. It was not until the discovery of an additional regulatory RNA in *C. elegans*, *let-7*, seven years later (Reinhart et al., 2000) combined with the critical finding that this RNA is highly conserved, even in vertebrates (Lagos-Quintana et al., 2001; Pasquinelli et al., 2000), that the widespread relevance of gene regulation by microRNAs was acknowledged.

### 1.1.1 MicroRNA biogenesis

Until recently, the production of mature miRNAs in mammals was thought to be dependent on a series of processing steps, mediated by the enzymes Drosha and Dicer. Whilst this canonical pathway is still currently believed to be the most prevalent mechanism of generating mature miRNAs (Yang and Lai, 2011), it is now understood that there are also several alternative pathways of miRNA biogenesis, which involve distinct combinations of enzymes and cofactors. The key pathways currently thought to be involved in miRNA biogenesis are discussed below.

In the canonical pathway of miRNA biogenesis (Fig. 1.1A), miRNA genes are transcribed in the nucleus by RNA polymerase II (Lee et al., 2004) to generate primary miRNA transcripts (pri-

miRNAs). Multiple mature miRNAs can be produced from a single pri-miRNA, owing to the ability of these transcripts to form distinct stem-loop/hairpin secondary structures known as precursor miRNAs (pre-miRNAs), which are the substrate for the Drosha/DGCR8 enzymatic complex (Lee et al., 2003). Drosha cleaves pri-miRNA transcripts at the base of the stem-loop to release the hairpin, which is transported to the cytoplasm via the Exportin 5 complex (Lund et al., 2004).

Once in the cytoplasm, the loop is cleaved from the stem of the hairpin by Dicer (Hutvagner et al., 2001; Bernstein et al., 2001; Knight and Bass, 2001), generating the mature double-stranded miRNA. One strand of the duplex, the guide strand, is then incorporated into the RNA-induced silencing complex (RISC) and the other is released (known as the passenger, or miRNA*, strand). The mechanisms which govern strand selection in mammals are not yet fully understood (Czech and Hannon, 2011).

An alternative pathway of miRNA biogenesis, which is still dependent on Dicer but independent of Drosha/DGCR8 activity, generates a subclass of miRNAs known as mirtrons (Fig. 1.1B). These miRNAs are produced from pre-miRNAs which are generated from the introns of protein-coding genes during splicing of mRNA transcripts (Okamura et al., 2007; Ruby et al., 2007). Whilst the mirtron pathway is far less prevalent than the canonical pathway of miRNA biogenesis, examples of these intron-derived precursors can be detected in multiple lineages including mammals (Berezikov et al., 2007), suggesting that this pathway is not inconsequential.

Additionally, recent studies have also identified a Dicer-independent, Ago-dependent mechanism of pre-miRNA processing to mature miRNAs (Fig. 1.1C); but currently this has only been shown to be applicable to a single miRNA, miR-451 (Cheloufi et al., 2010; Cifuentes et al., 2010). In spite of the infrequency with which this phenomenon has been observed, potential mechanisms of miRNA biogenesis that avoid the canonical pathway have significant implications on interpreting the phenotypes of systems which infer deletion of Drosha or Dicer as rendering a cell ‘microRNA-deficient’, and thus should not be overlooked.

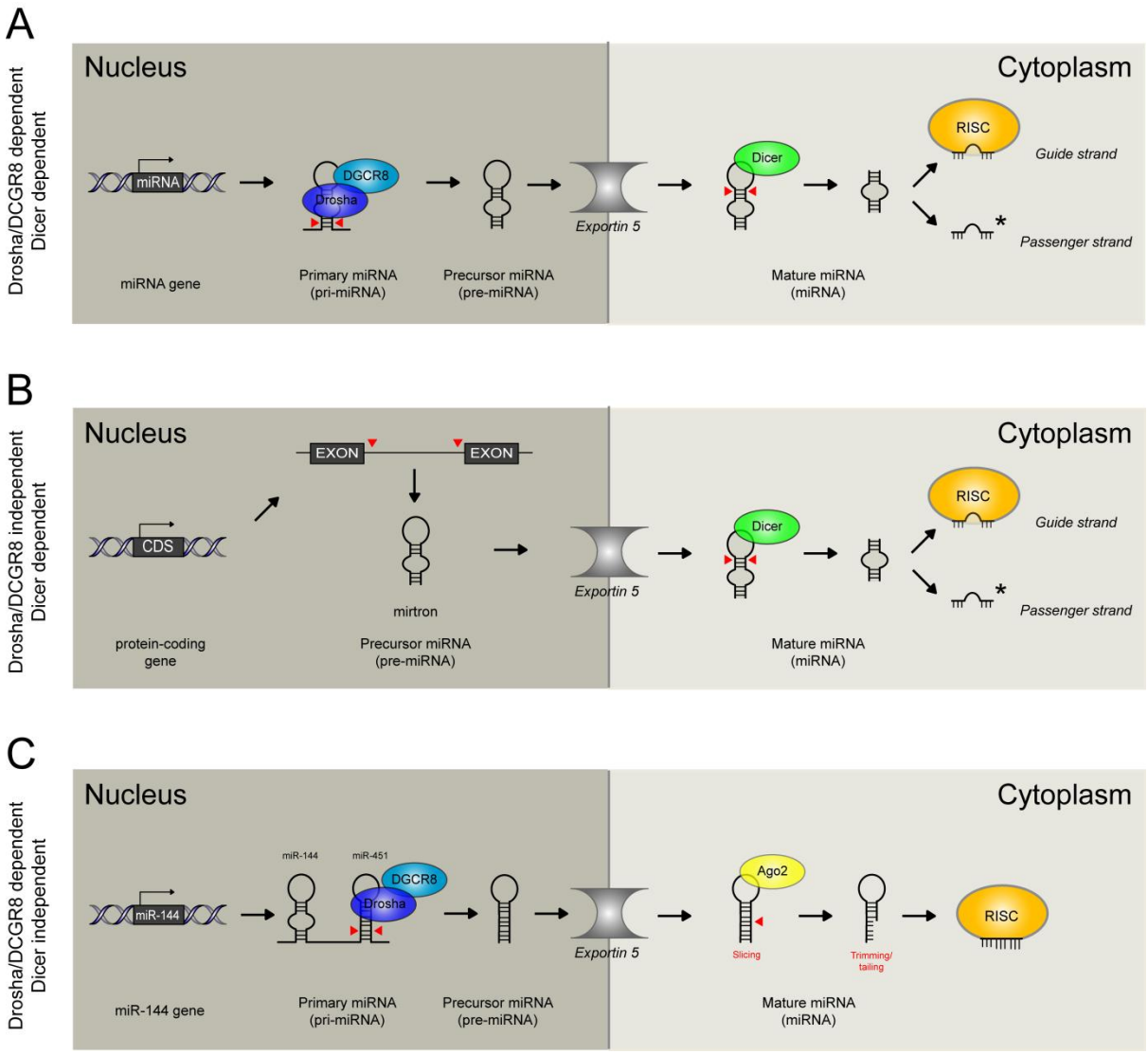



**Figure 1.1: The microRNA biogenesis pathway in mammalian cells.**
(A) The canonical miRNA biogenesis pathway. MiRNA genes are transcribed by RNAP II to generate primary miRNA transcripts (pri-miRNA) which form hairpin secondary structures. The Drosha/DCGR8 complex cleaves the hairpin at the base of the stem loop, generating a precursor miRNA (pre-miRNA) which is exported to the cytoplasm via the Exportin 5 complex. The loop is cleaved from the hairpin by Dicer in the cytoplasm. One strand of the mature miRNA duplex is then incorporated into the RISC complex, whilst the other is degraded. (B) The mirtron biogenesis pathway. Precursor miRNAs (mirtrons) are generated during splicing of introns from protein-coding genes, which form hairpin structures and are subsequently recognised and processed by Dicer as in the canonical pathway. (C) Ago2-dependent miRNA biogenesis. Both miR-144 and miR-451 are processed from the miR-144 pri-miRNA. The miR-451 precursor is still cleaved from the pri-miRNA by Drosha/DGCR8, but after export to the cytoplasm, the hairpin is bound by Ago2 which cleaves one strand of the duplex. The loop is then presumed to be removed by exonuclease activity.

### 1.1.2 Mechanism of microRNA function

Once incorporated into RISC, miRNAs target the complex to specific mRNA transcripts via complementarity to sequences usually found in the 3'UTR of the target, although examples of miRNAs binding to the coding sequence have also been identified (Tay et al., 2008). Until recently, functional miRNA binding was thought to be dependent on perfect complementarity between nucleotides 2-7 of the miRNA (known as the seed sequence) and the binding site in the target mRNA,

whilst only partial complementarity to the target mRNA was seen to be sufficient for interaction of the rest of the miRNA (Bartel, 2009). However, target prediction algorithms based on sequence complementarity to the miRNA seed sequence often result in false positive rates above 50% (Baek et al., 2008) suggesting that this rule, whilst clearly predictive in some cases, is not universally applicable to all miRNA:mRNA target interactions.

In fact, the HITS-CLIP method (high throughput sequencing of RNAs isolated by crosslinking immunoprecipitation), in which miRNAs and mRNAs simultaneously associated with Argonaute are identified, has shown that 27% of miRNA:mRNA interactions do not follow the seed sequence rule in mammalian cells (Chi et al., 2009). In light of this, the group which developed the technique reanalysed the HITS-CLIP dataset, and examined all Ago-bound sites in the crosslinked mRNAs which did not follow the canonical seed rule (Chi et al., 2012). The authors found that around 15% of mRNAs bound by miR-124 in the brain have a 'pivot' nucleotide at position 5-6, generating a bulged site which still results in a functional miRNA:mRNA interaction. Further studies are likely to uncover even more mechanisms by which miRNAs target specific mRNA transcripts.

Whilst miRNAs are required for targeting specific mRNA transcripts, it is the RISC complex of proteins which functionally inhibits the expression of these RNAs. The mechanism by which this is achieved is not fully understood, but potential models which have been proposed to explain how miRNAs downregulate target gene expression are discussed below.

It has been observed in plant cells that miRNAs can mediate mRNA endonucleolytic cleavage and degradation of target transcripts (Fig. 1.2A); but this interaction is dependent on perfect complementarity between the miRNA and its target (Llave et al., 2002). In mammalian cells, this phenomenon is rare; and inhibition of mRNA function is usually mediated by either repression of translation or mRNA deadenylation and subsequent transcript degradation.

Regarding the latter mechanism, multiple studies have shown that the GW182 family of proteins, which interact with AGO proteins, recruit the CAF1/CCR4/NOT1 deadenylase complex to the RISC-bound mRNA (Braun et al., 2011; Chekulaeva et al., 2011; Fabian et al., 2011). Deadenylation of the transcript is followed by decapping of the 5' end (Eulalio et al., 2007), which leaves RNA susceptible to 5'-3' degradation by the exonucleases (Fig. 1.2B).

In terms of establishing how miRNA activity inhibits translation of target miRNAs, much controversy remains. Some groups find that miRNAs act to inhibit translation initiation (Pillai et al., 2005) (Fig. 1.2C) and others post-initiation (Petersen et al., 2006) (Fig. 1.2D). Whilst these discrepancies could potentially be attributed to the different experimental systems used, these results could also plausibly indicate that miRNAs repress translation at multiple levels. Further work is required to understand this mechanism in greater detail.

A fundamental question in fully establishing the mechanistic function of miRNA in post-transcriptional gene regulation is whether translational inhibition and mRNA degradation occur independently or sequentially. In 2010, Guo *et al* used ribosome profiling to show that changes in mRNA level on introduction of specific miRNAs corresponded with over 80% of changes in protein levels, in support of a model for transcript deadenylation/degradation before inhibition of translation. Conversely, two very recent papers indicate that inhibition of translation initiation is the first step in miRNA-mediated gene repression, followed by deadenylation and target degradation (Djuranovic et al., 2012; Bazzini et al., 2012). However, as these studies were carried out in *Drosophila* and zebrafish, this mechanism remains to be confirmed in higher eukaryotes. Further studies are required to fully resolve the kinetics of miRNA-mediated gene regulation.

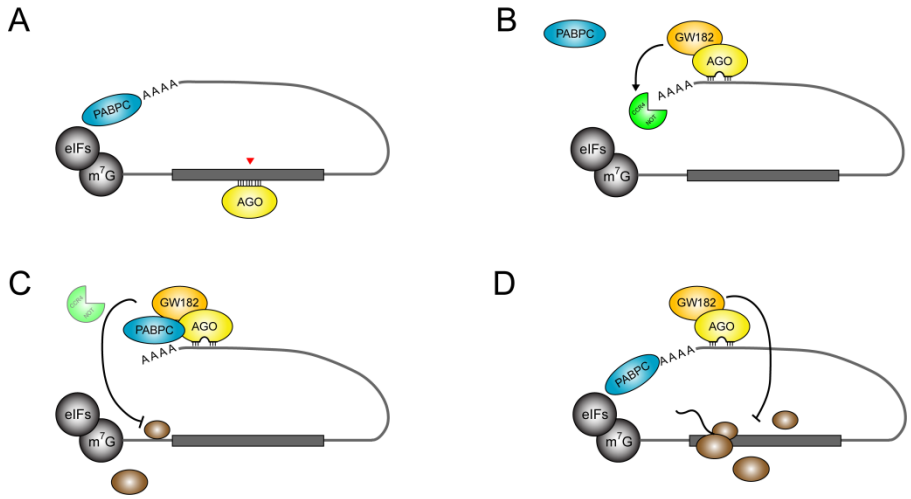



**Figure 1.2: Proposed models for mechanistic basis of miRNA-mediated gene silencing.**
(A) Perfect complementarity of miRNA to target mRNA results in direct endonucleolytic cleavage of the transcript catalysed by the Argonaute proteins (AGO). (B) MiRNA-mediated deadenylation and degradation of target mRNAs. GW182 binds to Argonaute proteins and recruits the CCR4-NOT complex, which deadenylates the mRNA. This results in loss of PABPC binding (cytoplasmic poly(A) binding protein), decircularisation of the mRNA and therefore inhibition of translation. (C) MiRNA-mediated inhibition of translation initiation. Binding of GW182 has also been shown to directly inhibit translation initiation independently of polyadenylation by the CCR4-NOT complex, potentially by directly interacting with PABPC itself, therefore inhibiting circularisation and initiation of translation. (D) MiRNA-mediated inhibition of translation elongation. (D) MiRNA-mediated inhibition of translation elongation. It has been suggested that targeting of AGO proteins to mRNA transcripts causes ribosome drop-off post-translation initiation, but no mechanism has been described for this phenomenon. Adapted from (Tritschler et al., 2010; Pasquinelli, 2012).

It should also be mentioned that a small number of studies have recently reported that conversely, miRNAs can also upregulate translation of target mRNAs (Orom et al., 2008; Roberts et al., 2011). More generalised studies are required in multiple systems to elucidate whether this mechanism is a truly global function of miRNA or restricted to a few highly specific circumstances.

### 1.1.3 Regulation of microRNA activity

MicroRNAs are expressed in a highly cell-specific and developmental stage-specific manner. Additionally, the ability of miRNAs to target multiple transcripts simultaneously makes them prime candidates for regulating transitions between developmental stages and mediating co-ordinated cellular responses to changes in extracellular conditions. As a result, miRNA activity itself must be extremely tightly regulated, and discrepancies between pri-miRNA, pre-miRNA and mature miRNA levels indicate that this regulation can occur at multiple steps in the miRNA biogenesis pathway.

At the level of transcription, miRNA gene expression is regulated in the same way as protein coding genes (in the case of the canonical biogenesis pathway). Cell-specific transcription factors bind and activate transcription of pri-miRNA transcripts by RNA polymerase II – for example, in mouse ES cells, the pluripotency factors Oct4 (*Pou5f1*), Nanog and Sox2 have been shown to bind the promoter of the ES-specific miR-290 family cluster (Marson et al., 2008). In addition, there are numerous examples of both positive and negative feedback loops involving miRNA-mediated regulation of the proteins which activate their transcription (Chang et al., 2009; Raver-Shapira et al., 2007; He et al., 2007). These mechanisms ensure precise temporal and cell-specific expression of both transcription factor and miRNA, buffering levels of both components against fluctuations in the internal cellular state and therefore stabilising cell function.

During miRNA processing, numerous factors have been implicated in regulation of both Drosha and Dicer, thus attenuating levels of pre-miRNA and mature miRNA respectively. Both enzymes are regulated by levels of their interacting partners, RNA binding proteins DGCR8 (in the case of Drosha) (Han et al., 2009) and TRBP (which associates with Dicer) (Melo et al., 2009). These autoregulatory interactions ensure that Drosha/DGCR8 and Dicer/TRBP activity is maintained within a level appropriate for normal cellular function. A significant number of other proteins have also been shown to influence the function of Drosha and Dicer – including p53 (He et al., 2007; Raver-Shapira et al., 2007), SMAD proteins (Davis et al., 2008), and estrogen receptor alpha (ERα) (Yamagata et al.,

2009) - which serve to either activate or repress miRNA biogenesis. These interactions illustrate that the rate of miRNA production is influenced by multiple factors, and highlights the tight regulation of this process.

There are also several examples of modulation of miRNA activity by RNA-binding proteins (RBPs) which target the same mRNA transcript. These interactions have been shown to have either a negative (DND1 (Kedde et al., 2007)) or positive (PUF-9 (Nolde et al., 2007; Galgano et al., 2008)) effect on miRNA function; or in the case of some RBPs (HuR), their impact on miRNA activity is dependent on the mRNA target (Bhattacharyya et al., 2006; Kim et al., 2009), and is probably influenced by the context of binding (e.g. relative proximity of miRISC/RBP binding sites within 3'UTR; secondary structure of the target mRNA transcript).

The effect of miRNAs on their targets can also be influenced by active alteration of the length of the 3'UTR of target mRNA transcripts. Shortening of the 3'UTR, via use of alternative polyadenylation sites, results in loss of the miRNA binding sites and therefore alleviation of miRNA-mediated repression of these genes. Currently, examples of this mechanism have been identified in activated lymphocytes (Sandberg et al., 2008) and cancer cells (Mayr and Bartel, 2009). It remains to be seen whether the phenomenon is representative of a ubiquitous mechanism of regulating miRNA activity, or is restricted to specialised circumstances.

Finally, the mature miRNAs themselves can be subject to post-maturation editing processes, which can modify their seed sequences and therefore target specificity (Kawahara et al., 2007b). This process involved the conversion of adenosine residues to inosine by adenosine deaminases that act on RNA (ADARs) (Krol et al., 2010b). MicroRNA editing via this mechanism has also been shown to inhibit processing of pri-miRNAs (Kawahara et al., 2007a). In mice, deep sequencing of RNAs present in the mouse brain identified eight miRNAs with edited bases in the seed sequence (Chiang et al., 2010). However, recent controversy surrounding the extent of RNA editing in mammalian genomes (Schrider et al., 2011; Li et al., 2011) suggests that these data should perhaps be interpreted with caution pending further analyses.

Alternatively, modifications of miRNAs which do not alter their sequence have been shown to affect stability of the RNA. Addition of a single adenosine residue to the 3' end of miR-122 by the GLD-2 poly(A) polymerase in mouse liver cells protects the miRNA from degradation (Katoh et al.,

2009); whereas the addition of a methyl group to the 3' end of miRNAs by the Hen1 methyltransferase in plants (Yu et al., 2005) has been shown to have a similar effect (Li et al., 2005) . MicroRNA stability appears to be tightly regulated in a cell-, developmental-, and miRNA-specific manner; illustrated by the identification of miRNAs in different systems which are either highly stable, or conversely, subject to rapid turnover (Krol et al., 2010a). Whilst much remains to be established regarding the factors which participate in this process, the small RNA degrading nuclease (SDN) family of 3' – 5' exonucleases have been shown to target single-stranded miRNAs for degradation in plants (Ramachandran and Chen, 2008); and in animals, the 5'-3' exonuclease XRN-2 plays a similar role (Chatterjee and Grosshans, 2009).

The ability of microRNAs to target multiple transcripts simultaneously means that changes in miRNA levels can have a significant impact on the cell function. This brief overview of the multiple mechanisms which are known to regulate microRNA function serves to illustrate that stringent regulation of these noncoding RNAs is a fundamental component of maintenance of cell homeostasis. In support of this, loss or gain of miRNA activity has been associated with a plethora of disease states (Lu et al., 2008).

### 1.1.4 Biological role of post-transcriptional gene regulation by microRNAs

MicroRNA-mediated gene regulation has been shown to be essentially ubiquitous, with evidence for miRNA-mediated gene regulation in a multitude of cell types, tissues, signalling pathways and regulatory networks – a notion supported by the estimate that around 50% of protein-coding genes in the mammalian genome are regulated by microRNAs (Krol et al., 2010b). This universal and fundamental role of microRNAs in cellular biology is exemplified by embryonic lethality of Dicer knockout (KO) mice (Bernstein et al., 2003).

Despite high conservation of the sequence of individual microRNAs throughout animal evolution, however, miRNA:mRNA target interactions are surprisingly poorly conserved (Chen and Rajewsky, 2006). As a result, inferring biological functions to miRNAs across different experimental systems is challenging. In spite of this, several themes in gene regulation by microRNAs are beginning to emerge.

As previously discussed, initial pivotal studies in *C. elegans* which identified the first microRNAs showed that mutations in *lin-4* (initially assumed to be regulatory protein, and later identified as an RNA) resulted in dramatic defects in lineage specification (Arasu et al., 1991; Chalfie et al., 1981). Examples of miRNA-mediated gene regulation of this kind, in which miRNAs act as 'genetic switches' (Flynt and Lai, 2008) are rare; in fact, a study which systematically mutated 85 miRNAs in *C. elegans* showed that very few mutants exhibited gross phenotypic abnormalities or affected viability of the organism (Miska et al., 2007).

Combined with the knowledge that a single microRNA can target multiple transcripts, these studies led to the generally accepted view that microRNAs are often functionally redundant and act co-ordinately within gene regulatory networks. Supporting this idea, two elegant genome-wide studies published in 2008 illustrated that the impact of knocking down or introducing individual microRNAs on global levels of mRNA and protein is typically mild, providing evidence that microRNAs function as biological rheostats to fine-tune gene expression (Selbach et al., 2008; Baek et al., 2008). These findings support a role for microRNAs in conferring robustness to cell-specific gene expression profiles, acting as a buffer for inherent fluctuations in transcriptional noise, and in response to changes in extracellular states.

MicroRNAs execute this function as components of gene regulatory networks, which operate in two major ways. Firstly, co-expression of microRNAs with target mRNAs fine-tunes cell-specific gene expression by buffering levels of expressed genes against internal variation in cellular conditions, ensuring expression of these transcripts remains within a threshold tolerable to the cell (Fig. 1.3A). Secondly, mutually exclusive expression of miRNAs and their targets confers robustness to cell-specific gene expression by ensuring that if lineage-inappropriate genes are aberrantly transcribed, miRNAs are present within the cell to downregulate these transcripts before they are able to produce protein and therefore potentially compromise cell identity (Stark et al., 2005) (Fig. 1.3B).

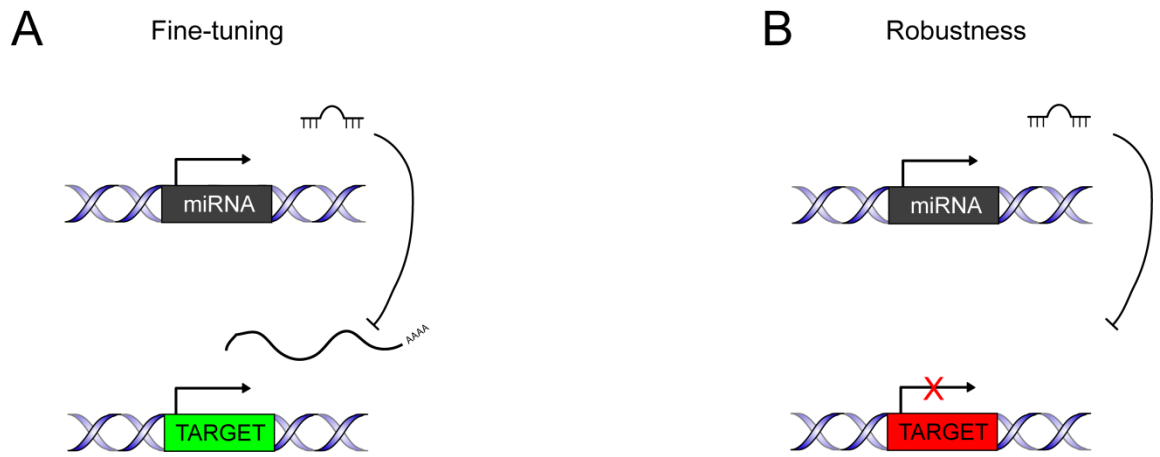



**Figure 1.3: MicroRNAs confer robustness to lineage-specific gene expression.**
(A) Co-expression of miRNA and target mRNA fine-tunes levels of the target transcript. (B) Mutually exclusive expression of miRNA and target mRNA ensures that translation of the target is inhibited even in the event that it is aberrantly transcribed.

The role of microRNAs in stabilising lineage-specific gene expression at the post-transcriptional level via both networks described above is critical given that even as a result of recurrent and essential biological processes, mechanisms regulating cell-specific gene expression are potentially compromised. The mechanisms which establish lineage-specific gene expression, and circumstances which may challenge their fidelity, are discussed in the following sections.

## 1.2 Epigenetic regulation of gene expression

Establishing lineage-specific gene expression patterns is fundamental to defining cell identity and function. The influence of genetic elements in this process, predominantly in terms of binding of transcription factors to specific sites in the genome, is relatively well characterised and will not be discussed in depth here. In recent years, however, epigenetic mechanisms of gene regulation have been shown to arguably be as instrumental in determining lineage specification as the DNA sequence itself. These mechanisms are classed as epigenetic if they have a phenotypic impact on cell biology but do not alter the underlying genomic sequence itself. Examples of epigenetic mechanisms of gene regulation which contribute to establishment of lineage-specific gene expression are discussed below.

### 1.2.1 Nuclear architecture and chromatin structure

The identification of regions of heterochromatin and euchromatin using early microscopy techniques was the first indication that the DNA molecule is able to form distinct structures within the nucleus (Cooper, 1959), although it was not until much later that a correlation between gene

expression and these domains was discovered. It is now known that chromosomes occupy discrete territories within the nucleus (Stack et al., 1977) and they are able to interact at the periphery of these regions (Branco and Pombo, 2006). Evidence that physical positioning of chromosomes within the nucleus can impact gene expression was provided by a study showing that localisation of immunoglobulin genes to the nuclear periphery correlated with their repression (Kosak et al., 2002). This concept has now been extended genome-wide, with the identification of lamin associated domains (LADs) which are transcriptionally inactive regions of the genome associated with the nuclear periphery (Guelen et al., 2008). Interestingly, it has recently been shown that activated genes within these domains can loop out and dissociate from the periphery (Peric-Hupkes et al., 2010), illustrating the dynamic nature of nuclear organisation and the impact this has on gene regulation.

Many studies of nuclear architecture employ the fluorescence in situ hybridisation (FISH) technique, which allows high resolution single-cell analysis of gene positioning and activity. However, this method is not easily applied genome-wide, and therefore studies using FISH are often candidate-based analyses. The development of the chromosome conformation capture (3C) technology has provided a powerful alternative to FISH (Dekker et al., 2002), and the derivation of numerous techniques based on the original 3C protocol now facilitates the interrogation of chromatin structure at high resolution within a defined genomic region (for example, the human α-globin (Baù et al., 2011) and β-globin (Dostie et al., 2006) loci) and recently, even genome-wide. Analysis of the 3D structure of the entire genome has currently only been possible in yeast; application of the HiC technique (3C combined with high-throughput next generation DNA sequencing) to both the *S. cerevisiae* and *S. pombe* genomes confirmed localisation of chromosomes in distinct territories, but the two studies dispute whether functionally related genes are found in clusters within the nucleus (Duan et al., 2010; Tanizawa et al., 2010). Further analyses will be required to resolve this discrepancy.

One key limitation of 3C-derived technologies is the population-based nature of the analysis. Given the large amount of evidence documenting the dynamic nature of inter- and intra- chromosomal interactions, and the kinetics of associated genomic regions with nuclear domains, it is likely that interactions detected by 3C are not representative of what is actually occurring in a given cell at a defined point in time. Developing tools to perform single-cell 3C-based analyses would be hugely

informative in comprehensively elucidating the role of nuclear architecture in gene regulation, and would undoubtedly reveal even further layers of complexity to chromatin structure within the nucleus.

### 1.2.2 DNA methylation

Arguably the most well characterised of all known epigenetic modifications, DNA methylation refers to the addition of a methyl group to cytosine residues in the DNA molecule, and is predominantly associated with gene silencing. DNA methylation is found in almost all eukaryotes, and is estimated to be found at 70-80% of symmetric GC sites in the mammalian genome (Ehrlich et al., 1982). Conversely, the remaining GC sites are hypomethylated, and are often associated with gene promoters. Clusters of hypomethylated symmetric GC sites are known as CpG islands (Bird, 1986). In mammals, DNA methylation is established and maintained by members of the DNA methyltransferase family of proteins (Dnmts) (Chen and Li, 2004). The critical role of DNA methylation in embryonic development is illustrated by the finding that whilst Dnmt-deficient ES cell lines are viable, with no obvious phenotypic defects, Dnmt KO embryos derived from these lines die mid-gestation (Li et al., 1992). It is now understood that this phenotype is attributable to the dynamic changes in patterns of DNA methylation first during gametogenesis, and secondly during embryogenesis (He et al., 2011), although the mechanisms which govern DNA demethylation during this process are not yet understood (Chen and Riggs, 2011).

In addition to playing a critical role in embryonic development, DNA methylation is also implicated in lineage specification. As mentioned above, despite having no effect on viability, loss of Dnmt activity in ES cells results in severe defects in differentiation (Lei et al., 1996). In somatic cells, however, loss of Dnmt function causes chromosomal instability, proliferation defects and cell death (Dodge et al., 2005) indicating that DNA methylation is essential for normal cellular functions. In support of this, genome-wide analysis of DNA methylation in ES cells and foetal fibroblasts illustrated significant differences in methylation patterns (Dodge et al., 2005), and in an independent study, it was shown that methylation of *Oct4* and *Nanog* is required for silencing of pluripotency genes during differentiation (Farthing et al., 2008). These results illustrate that DNA methylation is essential for the specification of distinct lineages during development, and also plays a critical role in maintenance of lineage specific gene expression in terminally differentiated cells.

### 1.2.3 Histone modifications

Early *in vitro* assays investigating chromatin structure first described the nucleosome:DNA complex as 'beads on a string' (Olins and Olins, 1974); an observation which was initially assumed to represent a purely structural mechanism to facilitate packaging of DNA within the nucleus. It is now understood that nucleosomes, and in particular their constituent histone proteins, in fact play an essential and extensive role in gene regulation (Campos and Reinberg, 2009).

The histone proteins H2A, H2B, H3 and H4 are highly conserved, and form an octameric complex around which approximately 147bp of DNA is wound (Zhou et al., 2011). Each histone has an N-terminal 'tail' which extends from the nucleosome and is subject to a wide range of post-translational modifications including acetylation, phosphorylation, methylation, ubiquitination, SUMOylation and ADP-ribosylation (Strahl and Allis, 2000), which are summarised in Table 1.

| Histone | Acetylation | Phosphorylation | Methylation | Ubiquitination |
|---|---|---|---|---|
| H2A | K5, K9, K13, K15, K36 | S1, S120 | K99, K119 | K119 |
| H2B | K5, K12, K15, K20, K85, K108, K116, K120 | S14 | K5 | K20, K120 |
| H3 | K4, K9, K14, K18, K23, K27, K36, K56 | T3,T10, T11, S28, S41, T45 | R2, K4, R8, R17, K23, R26, K27, K36, K37 | |
| H4 | K5, K8, K12, K16, K20, K77, K19, K91 | S1, S47 | R3, K12, K20, K59, R92 | |

**Table 1: Key post-translational modifications of core histone tails.**
Summary of key post-transcriptional modifications to specific residues in the N-terminal tail of each of the four histone proteins which form the nucleosome. Modifications to histone H1 are not shown. Where formally demonstrated, modifications which have a direct effect on gene activation are shown in green, and modifications which contribute to gene repression are shown in red. K – lysine; S – serine; T – threonine; R – arginine. Adapted from (Strahl and Allis, 2000; Portela and Esteller, 2010).

Broadly speaking, actively transcribed euchromatin is marked by high levels of acetylation, combined with trimethylated H3K4, H3K36 and H3K79. Conversely, silenced regions of heterochromatin are characterised by low levels of acetylation, and high levels of methylated H3K9, H3K27 and H4K20 (Portela and Esteller, 2010). The influence of specific histone modifications on expression of individual genes within euchromatic regions, however, is far more difficult to dissect. The effect of post-translational histone modifications on transcription is influenced by multiple factors, including the context of the modification within the histone tail itself. The functional dependency of

histone marks on the status of adjacent residues is known as 'the histone code' (Jenuwein and Allis, 2001) and adds a considerable degree of complexity to understanding mechanisms of gene regulation in mammalian cells.

One relatively well characterised combination of histone modifications is the 'bivalent' signature of concurrent H3K4me3 and H3K27me3 modifications at the same gene promoter in ES cells (Bernstein et al., 2006; Azuara et al., 2006). Bivalent genes are key developmental regulators, and some of the first genes to be expressed when an ES cell receives a signal to differentiate. A role for histone modifications in lineage specification is clearly exemplified by the resolution of bivalency on differentiation; where loss of H3K4me3 from lineage-inappropriate genes results in high levels of H3K27me3, thus ensuring these loci are stably repressed (Bernstein et al., 2006; Mazzarella et al., 2011), and conversely, lineage-specific genes lose H3K27me3 and gain high levels of H3K4me3. Additionally, another recent study shows that in multipotent primitive endoderm progenitor cells, genes required for distinct terminally differentiated cell types within the endoderm lineage are marked by a unique bivalent pattern of chromatin modification: H3Ac and H3K27me3 (Xu et al., 2011). The authors speculate that these chromatin marks 'prepattern' relevant lineage-specific genes to ensure correct specification of defined cell types from an early stage in development. This study thus provides further evidence for a key function of epigenetic histone modifications in lineage commitment.

Much remains to be elucidated in terms of the proteins and mechanisms involved in establishment, maintenance and removal of histone modifications, and due to the wide range of chemically distinct modifications which are possible (Table 1), the scope of current knowledge of these will not be covered in depth here. However, of particular relevance to this study are the Polycomb (PcG) and Trithorax (TrxG) groups of proteins, which are involved in the establishment and maintenance of bivalent domains in ES cells.

PcG proteins were first identified in *Drosophila,* where mutations in these genes results in defects in body patterning and segmentation (Lewis, 1978). Mice deficient in key PcG proteins are embryonic lethal (Carroll et al., 2001; Pasini et al., 2004; Faust et al., 1998) illustrating the critical role of these proteins in mammalian development. The molecular basis for this phenotype was later identified when the targets of PcG in mouse ES cells were shown to be key developmental regulators

(Boyer et al., 2006). PcG proteins have been found in several protein complexes, but in mammals, the best characterised of these are the Polycomb-repressive complexes 1 and 2 (PRC1/PRC2). The core components of PRC1 are ubiquitin ligases Ring1A and Ring1B, whilst PRC2 activity is dependent on Eed, Suz12 and catalytic subunit Ezh2, which mediates di- and tri- methylation of H3K27 (Klymenko et al., 2006). The mechanisms by which PRC2 is targeted to specific genomic loci are not yet fully understood, but recruitment of PRC2 via interactions with long ncRNA is supported by studies of the X chromosome (Zhao et al., 2008) and Hox gene loci (Rinn et al., 2007), and recent genome-wide RIP-seq (RNA-immunoprecipitation followed by high throughput sequencing) analyses (Zhao et al., 2010). Until recently, it was thought that PRC1 binding is downstream of PRC2, as it is recruited to the H3K27me3 mark (Margueron and Reinberg, 2011). However, genome-wide studies have identified multiple PRC2 binding sites that lack the H2A ubiquitination mark catalysed by PRC1 (Ku et al., 2008), and this year, two separate groups identified PRC1 complexes containing RYBP (Ring1/YY1-binding protein) which are capable of targeting genomic loci independently of PRC2 (Tavares et al., 2012; Gao et al., 2012). Another study published simultaneously provides evidence that PRC1 binding could be mediated by interaction with transcription factors (Yu et al., 2012). Taken together, these studies serve to illustrate that much remains to be established in terms of the mechanisms which regulate PcG activity in ES cells, and that further investigation in this field is critical to elucidate the role of these proteins in regulation of bivalency and lineage specification in mammalian development.

Whilst PcG proteins establish epigenetic chromatin marks which repress gene activity, TrxG proteins catalyse trimethylation of H3K4; a mark associated with active transcription. TrxG proteins were first identified in *Drosophila* as positive regulators of the same genes repressed by PcG (Ingham, 1983), but remain less well understood than their previously identified counterparts. These proteins usually function as part of large, multisubunit complexes, within which the Set1/2 and MLL families of proteins play important roles in generating the H3K4me3 mark and targeting the complex to specific genomic sites (Schuettengruber et al., 2011). As with PcG proteins, the mechanisms governing recruitment of TrxG to target loci are incompletely understood; but evidence for interactions with transcription factors (Muntean et al., 2010), long ncRNA (Wang et al., 2011), and histone modifications (Milne et al., 2010) is provided by studies using candidate-based approaches.

Systematic, genome-wide analyses are required to further discern the mechanisms underlying TrxG recruitment.

The opposing roles of the PcG and TrxG groups of proteins demonstrate that epigenetic regulation of gene expression via histone modifications is complex; dependent not only on the crosstalk between modifications on the same histone tail, but also by the interplay between the protein factors which implement, modify and remove these marks. Extending this, as the amount of genome-wide data on epigenetic marks increases, functional and mechanistic links between the mechanisms described in this section begin to emerge. For example, the LADs identified by Guelen *et al* were also enriched for repressive chromatin marks H3K27me3 and H3K9me2 (Guelen et al., 2008). Combining increasingly high resolution genome-wide datasets for epigenetic mechanisms of gene regulation will, in the future, provide exciting insights into the functional interdependency between all the regulatory systems discussed in this section.

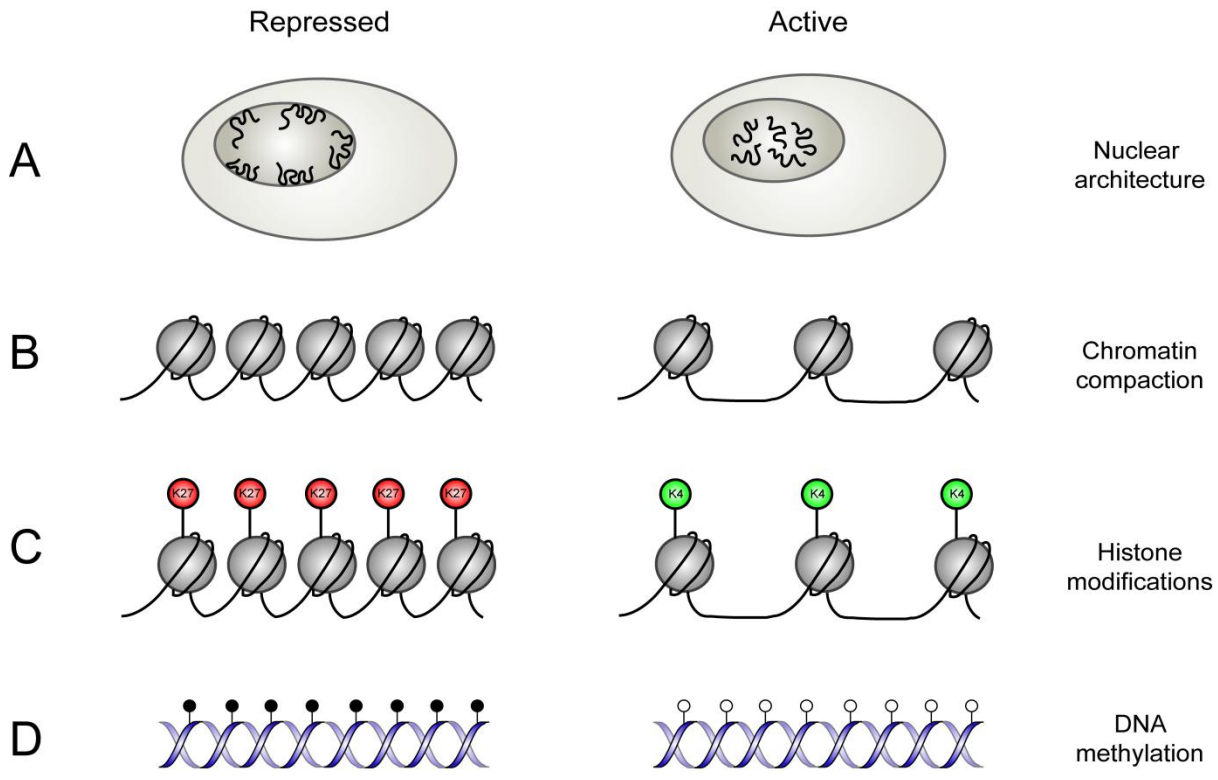



**Figure 1.4: Epigenetic mechanisms of gene regulation.**
Summary of chromatin-based epigenetic mechanisms of gene regulation. (A) Association of chromatin with the nuclear lamina generally leads to gene repression, whereas dissocation facilitates gene activation. (B) Highly compacted chromatin is associated with inactive regions of the genome, whereas decondensed chromatin generally corresponds with active loci. Grey circles: nucleosomes. (C) Post-transcriptional histone modifications contribute to gene repression (e.g. H3K27me3 - red) or activation (e.g. H3K4me3 - green). Grey circles: nucleosomes. (D) Methylated DNA (black spots) is associated with gene repression, whereas hypomethylated DNA (white spots) is associated with gene expression.

### 1.2.4 RNA inheritance

In addition to the chromatin-based epigenetic mechanisms described in the sections above, recent studies have identified inter-cellular and inter-generational transmission of RNA as a novel mechanism of epigenetic gene regulation. This phenomenon was first identified in maize, in response to observation of non-Mendelian inheritance of phenotypic states. In wildtype plants, transcription of a repetitive element upstream of the *b1* locus is required for the paramutation phenotype of lighter pigmentation (Stam et al., 2002). The discovery that the RNA-dependent RNA polymerase (RDRP) *Mop1* is required for this phenomenon (Alleman et al., 2006) first gave rise to the idea that RNA, as well as DNA, could functionally mediate transmission of phenotypic information between generations.

Following on from this, the correlation between loss of silencing of the *b1* locus and mutations in the RDRP *Mop1* led the authors to speculate that RNAi-induced epigenetic silencing could mechanistically be involved in this phenotype. The role of the RNAi pathway in establishment and maintenance of centromeric heterochromatin is well characterised in *S. pombe (*fission yeast), where transcripts from centromeric repeats are processed by Dcr1 to yield small RNAs, which subsequently target chromatin remodelling proteins to the region and induce epigenetic silencing (Grewal, 2010). Whilst evidence for a similar mechanism in plants has been identified (Zilberman et al., 2003; Neumann et al., 2007) this process is far more challenging to dissect than in yeast due to the significant increase in the number of core components involved (for example: *S. pombe* has one Dicer homolog, *Dcr1*, whereas plants have four different Dicer-like proteins (Djupedal and Ekwall, 2009)). Similarly, studies in mammals which have attempted to identify a function for the RNAi pathway in heterochromatin formation have yielded conflicting results (Wang et al., 2006; Maison et al., 2002; Kanellopoulou et al., 2005). Despite the technical difficulties in investigating the role of these proteins in higher organisms, the conservation of key members of this pathway suggests that there may be many additional functions of this pathway in mammals that are yet to be discovered.

The phenomenon of paramutation has also been observed in mammals. Mice mutant for the *Kit* gene (a tyrosine kinase receptor) have a characteristic phenotype of white spots, feet and tail tips. It was observed that crosses between parents heterozygous for this mutation produced wildtype offspring at a greatly reduced frequency than expected (Rassoulzadegan et al., 2006). This effect was shown to be mediated by RNAs from the parental generation, via injection of RNA extracted from

either heterozygote somatic or germ cells into wildtype one-cell embryos. These embryos carried to term and approximately 50% of offspring exhibited the characteristic *Kit* mutant phenotype despite being genetically wildtype. Whilst the study of paramutation at the *Kit* locus in mice remains an isolated example of this phenomenon in mammals, and the mechanism underlying RNA-mediated silencing of the region in genetically wildtype progeny remains unknown, these data indicate that the RNA-mediated paramutation phenomenon is not restricted to plants.

In addition to the above evidence for transgenerational RNA inheritance, studies have also suggested that RNAs are also transmitted between cells within the same organism; identifying a novel mechanism of cell-cell communication. Systemic movement of RNA of this kind has been identified in plants (Voinnet and Baulcombe, 1997) and in *C. elegans* (Fire et al., 1998). One key difference between these organisms is the mechanism by which the RNA travels; whilst in plants, the molecules move via physical connections between cells (Voinnet et al., 1998), *C. elegans* are able to take up RNA by feeding or existing in a solution containing the molecule via the membrane transporter protein SID-1 (Winston et al., 2002). One recent study in plants illustrates that systemic movement of small RNAs can instigate epigenetic modifications in recipient cells (Molnar et al., 2010), attributing functionality to this process. Evidence for cell-cell communication via RNA in mammals is limited, but it has been shown that exogenously introduced RNAs can silence targets in distal tissues in mice (Soutschek et al., 2004) and circulating miRNAs complexed with AGO2 have been identified in vesicles purified from human blood plasma (Arroyo et al., 2011).

These findings suggest that systemic and intergenerational movement of RNA may also be prevalent in higher organisms, but much remains to be established in the novel field of RNA-based epigenetic inheritance. Whether RNAi plays a role in heterochromatin formation or establishment of epigenetic modifications in mammalian systems remains an elusive area of investigation, despite extensive evidence for this phenomenon in plants and lower organisms.

### 1.3 Regulation of gene expression through the cell cycle

Epigenetic mechanisms of gene regulation must, by definition, be heritable through generations. As a result, they must also be stably propagated through cell division, to ensure maintenance and fidelity of cell-specific states.

### 1.3.1 Overview of the mammalian cell cycle

The mammalian cell cycle can be divided into four distinct stages (Fig. 1.5A). Transition between these phases is regulated by highly conserved cyclin dependent kinases (CDKs), which are functionally dependent on interactions with cyclin subunits for their enzymatic activity. Over 20 CDK-related proteins have been identified in mammals (Satyanarayana and Kaldis, 2009), illustrating the wide range of potential CDK-cyclin complexes that control cell cycle progression. It is thought that in somatic cells, CDK levels remain relatively constant throughout the cell cycle, whilst oscillation of cyclin subunits mediates transitions from one stage to the next (Hochegger et al., 2008) (Fig. 1.5C).

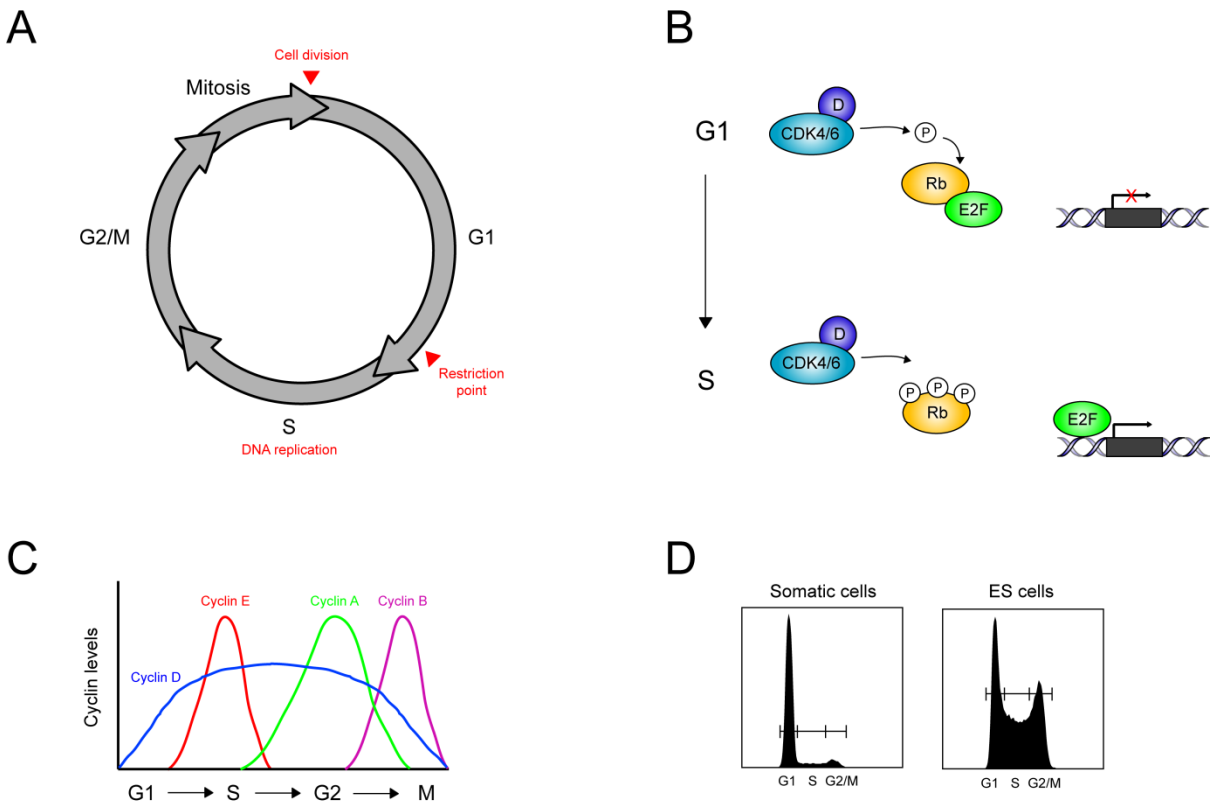



**Figure 1.5: Overview of the mammalian cell cycle.**
(A) Overview of the key stages of the cell cycle. (B) Progression from G1 to S phase is mediated by phosphorylation of Rb by the CDK4/6-cyclin D complex in late G1. In a hypophosphorylated state, the E2F family of transcription factors are bound to Rb. Phosphorylation of Rb results in release of the E2F proteins, which are then able to translocate to the nucleus and activate E2F target genes (including S phase genes). (C) Cell-cycle dependent oscillations in cyclin levels (somatic cells only). (D) Representative cell cycle profiles of somatic and pluripotent (ES) cells. Somatic cells cycle slowly and the majority of the population is therefore found in G1; ES cells proliferate rapidly and subsequently have higher proportions of cells in S phase and G2/M.

Of significance to this study is the unique cell cycle structure of pluripotent stem cells. One key property of ES cells is their capacity to self-renew; an attribute which is facilitated by their ability to remain in a rapidly proliferative state. These properties are distinctive features of pluripotent cells, and in contrast to somatic cells, ES cells are not subject to contact inhibition and do not undergo

senescence (White and Dalton, 2005). The ES cell cycle profile is characterised by a reduced proportion of cells in G1, and a greater proportion in S phase when compared to somatic cells (Fig. 1.5D)). The molecular basis of this difference is thought to be attributed to the fact that in ES cells, cyclin levels remain relatively constant throughout the cell cycle, as opposed to the oscillating levels observed in somatic cells (Stead et al., 2002). In addition, expression of CDK inhibitors such as the INK family of proteins is extremely low in ES cells (Stead et al., 2002) resulting in high levels of CDK activity, driving rapid proliferation of these cells.

There are several key checkpoints during the cell cycle which ensure that the cell can only divide if external conditions are favourable and if the genomic integrity is adequate (Kastan and Bartek, 2004). In the latter case, checkpoints facilitate assessment and repair of DNA damage before DNA replication (in the case of G1 and S checkpoints) and additionally ensure proper and accurate segregation of chromosomes in G2/M. An example of such a checkpoint is the G1-S transition, known as the restriction (R) point. Before the restriction point, cells are able to choose to remain in an unproliferative state (quiescence) referred to as G0, or to undergo cell division. Progression beyond the restriction point commits the cell to undergoing an entire cell cycle (Pardee, 1989) which can continue in the absence of external signals (Polager and Ginsberg, 2008). Progression through the restriction point is dependent on a signalling cascade involving activation of the retinoblastoma (Rb) family of proteins by the CDK4/6-cyclin D complex (Fig. 1.5B). Phosphorylation of Rb results in dissociation of these proteins from the E2F transcription factors, facilitating activation of E2F target genes. Notably, the restriction point is absent in ES cells (Savatier et al., 1994); another factor contributing to the rapid progression of pluripotent cells through the cell cycle. Recently, it has been shown that repression of this transition in ES cells is mediated, at least partially, by microRNAs.

#### 1.3.1.1 MicroRNA-mediated cell cycle regulation in ES cells

MicroRNA-deficient ES cells are characterised by a distinctive cell cycle defect, where cells proliferate slower and accumulate in G1 (Wang et al., 2007). To attempt to elucidate the specific miRNAs and their targets responsible for this defect, Blelloch and colleagues employed a screening strategy to re-introduce individual microRNA mimics into miRNA-deficient ES cells and assess which, if any, rescued the proliferation defect (Wang et al., 2008). The authors found that the miR-290 family of miRNAs, highly expressed in mouse ES cells, directly regulate and inhibit the expression of key cell

cycle regulators p21, Rbl2 and Lats2. Upregulation of these genes in KO cells results in inhibition of the G1-S transition and therefore accumulation of miRNA-deficient ES cells in G1.

Subsequently, the same group have shown that the let-7 family of miRNAs is upregulated on differentiation, and downregulates target genes involved in ES cell proliferation (Melton et al., 2010). Thus, miRNAs play a critical role in regulation of self-renewal in ES cells. To date, little is known regarding the role of microRNAs in regulation of the cell cycle in somatic cells; but based on the ubiquitous expression of key cell cycle regulators, and cell-type specificity of miRNAs, it seems likely that a similar role for other miRNAs in non-pluripotent cells will be identified in the future.

### 1.3.2 Propagation of gene expression through the cell cycle

As discussed previously, lineage-specific expression of selected genes is established by both sequence-dependent and epigenetic mechanisms. At various stages of the cell cycle, however, chromatin-based systems of gene regulation could potentially be compromised, due to the conformational changes which the DNA molecule must physically undergo in order to facilitate cell growth, DNA replication, and cell division. For example: during mitosis, chromosome condensation results in dissociation of many transcription factors from DNA (Zaidi et al., 2011; Martínez-Balbás et al., 1995); in early G1, these factors must reassociate with specific genomic sites in order to accurately re-establish lineage-specific gene expression; and during S phase, epigenetic modifications must be accurately restored on histone proteins incorporated into the newly synthesised DNA molecule. The precise mechanisms which facilitate these processes are incompletely understood, however current models which have been proposed to explain how epigenetic information is faithfully propagated through the cell cycle are discussed below.

In comparison to current understanding of how histone modifications are transmitted through cell cycle stages, maintenance of DNA methylation at specific genomic loci through DNA synthesis is relatively well characterised. Dnmt1 recognises hemimethylated CpG sites on DNA molecules composed of parental and newly synthesised strands and adds a methyl group to the unmarked nucleotide (Stein et al., 1982; Hermann et al., 2004). In light of this knowledge, it has been proposed that maintenance of histone modifications through S phase exploits a similar mechanism. Given that each nucleosome consists of two H3-H4 dimers (forming a tetrameric complex) and two H2A-H2B dimers, a semi-conservative model of inheritance of histone marks has been suggested; whereby a

parental H3-H4 dimer (complete with modifications) is incorporated into each copy of the replicated DNA molecule, and is used as a template to catalyse the same modifications on the newly constructed H3-H4 dimer (Probst et al., 2009). Alternatively, as early biochemical assays suggest that histone inheritance is in fact conservative (Leffak et al., 1977) it has also been proposed that parental histones are retained on one parental strand only, therefore leaving the other copy of the DNA molecule composed of entirely newly synthesised histones. Re-establishment of histone modifications on the newly synthesised histones in this scenario would require inter-strand crosstalk, which has yet to be experimentally demonstrated. Finally, some also suggest that distribution of parental and newly synthesised histones between DNA strands is a stochastic process, and maintenance of the epigenetic mark is achieved by use of adjacent nucleosomes as templates (Probst et al., 2009). Currently, all of these models remain plausible – and are not necessarily mutually exclusive. Further detailed biochemical analyses will be required to elucidate this process fully.

Another stage during the cell cycle where chromatin conformation is significantly altered is the G2-M to G1 transition. As previously mentioned, chromosome condensation during the later stages of the cell cycle results in loss of nuclear architecture and higher order chromatin structure. It has been suggested that compaction of the DNA molecule results in dissociation of DNA-binding proteins from chromatin (Martínez-Balbás et al., 1995) but evidence for retention of sequence-specific transcription factors to mitotic chromosomes has also been shown (Young et al., 2007). Several examples of both exclusion and retention of DNA-binding factors from mitotic chromatin have now been identified (summarised in (Egli et al., 2008)), but more recent analyses have shown that the localisation of proteins in question may also be dependent on cell-cycle dependent post-transcriptional modifications of the transcription factors themselves, and G2/M-specific chromatin modifications. One the most well characterised examples of this is dissociation of HP1α from mitotic chromatin due to phosphorylation of the histone 3 serine 10 residue by the kinase Aurora B, which is only active in G2/M (Fischle et al., 2005). These studies undoubtedly demonstrate that the significant physical and structural remodelling that occurs during the G2/M to G1 transition has a significant impact on chromatin-dependent gene regulation, although the precise mechanisms underlying these changes remains to be established, and may well prove to be protein-specific.

In light of the challenge that mitosis represents to maintenance of lineage-specific gene expression, it is interesting to note examples of histone modifications being stably inherited through this transition which have been proposed to act as 'mitotic bookmarks', facilitating the transmission of epigenetic information (Zaidi et al., 2011). For example, retention of H3K9 methylation through mitosis ensures maintenance of facultative heterochromatin on the inactive X chromosome in female cells (Peters et al., 2002). Alternatively, incorporation of the variant histone protein histone H3.3 into the promoter region of the *MyoD* gene has been shown to be functionally involved in mediating 'memory' of the active transcriptional status of the gene through cell divisions (Ng and Gurdon, 2008).

Understanding mechanisms of maintaining epigenetically-regulated gene expression through the cell cycle is not only relevant to ensuring transmission of defined transcriptional states within a given lineage – but also to circumstances in which this process may be artificially reversed. Somatic cell reprogramming refers to the process by which pluripotent cells can be derived from terminally differentiated lineages, and holds immense promise for cell-based therapeutics. Whilst a range of different methods have currently been attempted to achieve this goal, the efficiency of the process remains extremely low (Plath and Lowry, 2011), and cells often retain an 'epigenetic memory' of the lineage from which they were derived (Kim et al., 2010). Understanding the mechanisms which underlie this phenomenon is therefore critical to optimising these methods further.

Following on from this, reprogramming experiments using somatic cell nuclear transfer have identified that cells arrested in mitosis are more amenable to being reprogrammed (Egli et al., 2007; Ganier et al., 2011). The mechanistic basis of this phenomenon has not yet been established, but given the above discussion regarding the challenge that the G2/M to G1 transition presents to maintenance of lineage-specific gene expression, this stage of the cell cycle could plausibly represent a time in which the genome is 'vulnerable' to external influences: cell identity is not as stringently regulated at this point in the cell cycle as during interphase. Taken together, these studies illustrate that the G2-M to G1 transition presents a significant challenge to maintenance of lineage-specificity; or, taken from the view of reprogramming experiments, a window of opportunity to induce a distinct transcriptional programme and cellular identity.

## 1.4 Aims of study

In this study, I will investigate the idea that post-transcriptional regulation by microRNAs could act as a buffer to maintenance of lineage-specific gene expression through the cell cycle; ensuring that challenges to chromatin-based mechanisms of gene regulation do not result in translation of any lineage-inappropriate transcripts produced during these transitions. The high in vivo stability of microRNAs combined with evidence in lower organisms that small RNAs can be transmitted between cells provides an attractive model for inheritance of microRNAs throughout the cell cycle, including the G2/M to G1 transition, serving to ensure the stable transmission of lineage-specific gene expression from a parent cell to its progeny. To investigate this hypothesis, I will initially aim to establish whether microRNAs are inherited through mitosis in mammalian cells. Using primary mouse lymphocytes as a model system, where parent and daughter cells can be distinguished by flow cytometry, I will ask whether transmission of noncoding RNAs through cell division impacts expression of target genes in daughter cells.

A role for microRNAs in regulation of lineage-specific gene expression is supported by several mechanistic aspects of their function; firstly by regulating noise in expression levels of expressed genes, secondly by conferring robustness to cell-specific gene expression via mutually exclusive expression of miRNA and target mRNAs, and finally by the ability of a single miRNA to target multiple mRNAs simultaneously. In addition, the processes which facilitate propagation of epigenetic regulatory signatures through transitions between cell cycle stages are incompletely understood. In this study I will investigate the hypothesis that microRNAs function to maintain lineage-specific gene expression through points during the cell cycle where transcriptional regulation is compromised. To explore this idea, gene expression analysis in wildtype and microRNA-deficient ES cells will be undertaken in populations of cells at different stages in the cell cycle, using drug-mediated and density based approaches for cell cycle fractionation.

The role of non-coding RNAs in mammalian cell biology is a rapidly expanding field of interest where much remains to be elucidated. In lower organisms, where experimental manipulation of these RNAs is technically less challenging, an emerging role for small RNAs in epigenetic inheritance of phenotypic information is gathering increasingly convincing evidence. Establishing whether a similar function can be attributed to these RNAs in mammals remains an elusive and compelling topic of research.

# 2.0 Materials and Methods

## 2.1 Materials

### 2.1.1 Oligonucleotides

| Oligo | Manufacturer | Sequence | Modifications | Catalogue No. |
|---|---|---|---|---|
| siGlo Red Transfection Ctrl | Dharmacon | *Not available* | DY547 | D-001630-02 |
| miR-181a | Dharmacon | AAC AUU CAA CGC UGU CGG UGA GU (double stranded custom RNA oligo) | 5' DY547 (on sense strand) | n/a |
| miR-124a | Dharmacon | UAA GGC ACG CGG UGA AUG CCU U (double stranded custom RNA oligo) | None | n/a |
| miR-142-3p | Dharmacon | UGU AGU GUU UCC UAC UUU AUG GA (double stranded) | None | C-310420-07 |
| *Tgfbr1* siRNA SMARTpool | Dharmacon | 4 pooled siRNAs | None | L-040617-00 |
| *Cfl2* siRNA SMARTpool | Dharmacon | 4 pooled siRNAs | None | L-041155-01 |
| Mimic Negative Control #1 | Dharmacon | UCA CAA CCU CCU AGA AAG AGU AGA (double stranded) | None | CN-001000-01 |
| LNA miR-290 inhibitor | Exiqon | AGCACTTT (custom LNA oligo) | Phosphorothioate backbone | n/a |
| miR-291a-3p | Dharmacon | AAA GUG CUU CCA CUU UGU GUG C (double stranded) | None | C-310470-05 |

### 2.1.2 Antibodies

**H57-957** Hamster monoclonal anti-mouse TCR β-chain (553166; BD Biosciences, Franklin Lakes, NJ, USA).

**CD28** Hamster monoclonal anti-mouse CD28 (553294; BD Biosciences, Franklin Lakes, NJ, USA).

**CD4-PercP** PercP-conjugated anti-mouse CD4 (100431; BioLegend, San Diego, CA, USA). Used for FACS analysis of primary lymphocytes at 1:300 dilution.

**CD8-APC** APC-conjugated anti-mouse CD8 (100711; BioLegend, San Diego, CA, USA). Used for FACS analysis of primary lymphocytes at 1:300 dilution.

**Dicer (H-212)** Rabbit monoclonal anti-mouse Dicer (sc-30226; Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA). Used for western blot at 1:500 dilution.

**Ezh2** Rabbit polyclonal anti-mouse Ezh2 (pAb-039-050; Diagenode, Denville, NJ, USA). Used for western blot at 1:1000.

**Suz12** Rabbit polyclonal anti-mouse Ezh2 (pAb-039-050; Diagenode, Denville, NJ, USA). Used for western blot at 1:1000.

**Jarid2** Rabbit polyclonal anti-mouse Jarid2 (ab48137: Abcam, Cambridge, UK). Used for western blot at 1:2000.

**H3K4me3** Rabbit polyclonal anti-mouse H3K4me3 (07-473; Millipore, Billerica, MA, USA). Used for western blot at 1:1000.

**H3K27me3** Rabbit polyclonal anti-mouse H3K27me3 (07-449; Millipore, Billerica, MA, USA). Used for western blot at 1:1000.

**Lamin B (c-20)** Goat polyclonal anti-mouse Lamin B (sc-6216; Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA). Used for western blot at 1:20,000 dilution.

**H3** Rabbit polyclonal anti-mouse histone H3 (ab1791; Abcam, Cambridge, UK).

### 2.1.3 Tools for bioinformatics analysis

TargetScan: http://www.targetscan.org/mmu_50/

PicTar: http://pictar.mdc-berlin.de/

GSEA: http://www.broadinstitute.org/gsea/index.jsp

DAVID: http://david.abcc.ncifcrf.gov/

miReduce: http://www.mdc-berlin.de/en/research/research_teams/systems_biology_of_gene_regulatory_elements/projects/mireduce/index.html

## 2.2 Methods

### 2.2.1 Cell culture

All cell lines were maintained at 37°C and 5% (v/v) $CO_2$.

#### 2.2.1.1 Lymphocytes

*Isolation of primary lymphocytes*

Inguinal, brachial, axillary, mesenteric and salivary lymph nodes were extracted from 8-10 week old mice, passed through a 70µm cell strainer and washed in 10ml phosphate buffered saline without calcium or magnesium (PBS -/-) (Life Technologies, NY, USA) or media (Iscove's Modified Dulbecco's Medium (IMDM) (Life Technologies, NY, USA)), supplemented with 10% v/v foetal calf serum (FCS) (Biosera, East Sussex, UK), 2mM L-glutamine (Life Technologies, NY, USA), antibiotics (100U/ml penicillin/ 100µg/ml streptomycin), and 50µM β-mercaptoethanol (Life Technologies, NY, USA) by centrifugation at 1200rpm for 5 minutes. Cells were then resuspended in PBS -/- or media according to requirements.

*Activation of primary lymphocytes*

Cell culture vessels were coated with 500ng/ml anti-T cell receptor (TCR) antibody (4°C overnight). Culture vessels were washed three times with PBS -/- before use. Primary lymphocytes were plated at 2 x $10^6$ cells/ml in media (see above) supplemented with 5ng/ml recombinant mouse IL-2 (R&D Systems, Inc., MN, USA) and 2µg/ml anti-mouse CD28 antibody.

*Magnetic cell sorting: purification of $CD4^+$ T cells*

Single positive $CD4^+$ T cells were purified from populations of total lymphocytes using the magnetic cell sorting (MACS) protocol (Miltenyi Biotec, CA, USA). Primary lymphocytes were centrifuged at 300 x g for 10 minutes. Cells were resuspended in 90µl buffer (PBS -/- supplemented with 2% v/v FCS and 2mM EDTA (Life Technologies, NY, USA)) and 10µl CD4 MicroBeads (#130-

049-201; Milenyi Biotec, CA, USA) per 1 x $10^7$ cells, and incubated at 4°C for 15 minutes. Cells were washed once by the addition of 2ml per 1 x $10^7$ cells buffer followed by centrifugation at 300 x g for 10 minutes. Up to 1 x $10^8$ cells were resuspended in 500µl buffer.

MACS LS columns (#130-042-401; Milenyi Biotec, CA, USA) were placed in a MACS Separator (#130-042-302; Milenyi Biotec, CA, USA) and rinsed with 3ml of buffer before cell suspension was added. The column was then washed three times by the addition of 3ml of buffer. Effluent (containing buffer and unlabelled cells) was discarded. The column was then removed from the magnet (MACS Separator) and placed in a 15ml collection tube. Labelled cells were flushed from the column by adding 5ml of buffer and using the syringe supplied with the column.

#### 2.2.1.2 ES cells

Feeder dependent ES cells were maintained in Knockout Dulbecco's Modified Eagle Medium (KO DMEM) (Life Technologies, NY, USA) supplemented with 10% v/v FCS (Life Technologies, NY, USA), 2mM L-glutamine, antibiotics (100U/ml penicillin/ 100µg/ml streptomycin), 50µM β-mercaptoethanol, 1mM non-essential amino acids (NEAA) (Life Technologies, NY, USA) and 1000U/ml leukaemia inhibitory factor (LIF).

Feeder-independent ES cells were either maintained in KO DMEM media (see above) N2B27 + 2i, or a combination of both (see 4.2.1). The composition of N2B27 + 2i is as follows: Neurobasal media supplemented with 50% v/v DMEM:F12 media, 0.5% v/v N2 supplement. 1% v/v B27 supplement, 2mM L-glutamine, antibiotics (100U/ml penicillin/ 100µg/ml streptomycin) and 100µM β-mercaptoethanol (all from Life Technologies, NY, USA). Chemical inhibitors (Ying et al., 2008) were then added at a final concentration of 1µM (PD032590) and 3µM (CHIR99021) (Stemgent, MA, USA). Feeder-independent cells were cultured on vessels coated with 0.1% gelatin (Sigma-Alrich, MA, USA).

#### 2.2.1.3 3T3 fibroblasts

3T3 fibroblast cell lines (see Appendix I) were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% v/v foetal calf serum (FCS) (Biosera, East Sussex, UK), 2mM L-glutamine (Life Technologies, NY, USA), and antibiotics (100U/ml penicillin/ 100µg/ml

streptomycin). During derivation of the immortalised 3T3 lines used in this study, the above media was supplemented with 50μM β-mercaptoethanol.

### 2.2.2 Fluorescence activated cell sorting (FACS) analysis

#### 2.2.2.1 CFSE staining

CFSE was diluted to a final concentration of 62.5μg/ml in PBS -/- and was kept in the dark at all times during the staining process. Up to 10 x $10^6$ cells were resuspended in 1ml of CFSE + PBS -/- and incubated at 37°C for 8 minutes. 7ml of wash buffer (PBS -/- supplemented with 5% v/v FCS) was added and cells were centrifuged for 5 minutes at 12000 rpm. This was repeated an additional three times to remove all CFSE in solution. Cells were then resuspended in PBS -/- for use in further analyses. Cells were analysed using a FACSCalibur™ flow cytometer or sorted using a FACSAria™ (both from BD Biosciences). Data were analysed using FloJo software.

#### 2.2.2.2 Cell surface marker analysis

Each antibody (CD4-Percp and CD8-APC) was resuspended in wash buffer at a 1:300 dilution. Up to 1 x $10^6$ cells were resuspended in 50μl of the wash buffer + antibody, and incubated for 20 minutes at room temperature in the dark. Cells were then washed once in PBS -/- by centrifuging the cells for 5 minutes at 12000 rpm, removing the antibody solution, resuspending in wash buffer, and centrifuging again. Cells were then resuspended in wash buffer for further analysis. Cells were analysed using a FACSCalibur™ flow cytometer or sorted using a FACSAria™. Data were analysed using FloJo software.

#### 2.2.2.3 Cell cycle analysis

Cells were collected, counted, and up to 1 x $10^6$ cells fixed by resuspension in 1ml of 70% EtOH. For propodium iodide (PI) staining, fixed cells were resuspended in PI buffer (PBS -/- supplemented with 50μg/ml propidium iodide (Sigma-Alrich, MA, USA), 1mg/ml RNase A (Life Technologies, NY, USA), and 0.05% v/v NP40 (Calbiochem, Merck Millipore) and incubated on ice for 20 minutes. Cells were analysed using a FACSCalibur™ flow cytometer. Data were analysed using FloJo software.

### 2.2.3 Transfection methods

#### 2.2.3.1 Amaxa Nucleofection™

The Amaxa Nucleofection™ protocol was used for all lymphocyte transfections described in Chapter 3. All reagents are from Lonza Group, Switzerland except oligos (see 2.1.1). Up to 10 x $10^6$ cells were centrifuged at 90 x g for 10 minutes and resuspended in 100µl Nucleofector™ solution. The required quantity of oligo was then added to the cell suspension, and the mixture transferred to a cuvette. Cells were electroporated using the X-01 program (see Nucleofector™ manual for details). After electroporation, 500µl of Nucleofector™ medium was added to the cuvette, then the entire contents were transferred to a pre-warmed culture vessel and incubated at 37°C for the required period of time (see text). Nucleofector™ medium was removed after 4-6 hours and replaced with standard T cell culture media (2.2.1.1).

#### 2.2.3.2 Lipofectamine™

Lipofectamine™ 2000 was used for transfection of LNA miRNA inhibitors into ES cells. All reagents are from Life Technologies except oligos (see 2.1.1). 5-6 hours before transfection, 0.5 x $10^6$ cells per well were plated in antibiotic-free media in 6-well plates coated with 0.1% gelatin. Inhibitor and control oligos were diluted in 250µl of Opti-MEM to a final concentration of 500nM, and 5µl of Lipofectamine™ 2000 reagent was diluted separately also in 250µl of Opti-MEM. The separate solutions were incubated at room temperature for 5 minutes before being combined. The combined solution was then incubated at room temperature for a further 20 minutes, before being added to each well of cells. Cells were incubated at 37°C for 5 hours before transfection media was removed and replaced with ES cell media (2.2.1.2).

#### 2.2.3.3 DharmaFECT®

DharmaFECT® Reagent #1 (Thermo Scientific, Dharmacon) was used for transfection of miRNA mimics into ES cells. All reagents are from Life Technologies except oligos (see 2.1.1). 5-6 hours before transfection, 0.5 x $10^6$ cells per well were plated in antibiotic-free media in 6-well plates coated with 0.1% gelatin. Mimic and control oligos were diluted to a final concentration of 100nM in serum-free media, and 12µl of DharmaFECT® Reagent #1 was added to 438µl of serum-free media

in a separate tube. Both tubes were incubated at room temperature for 5 minutes before being combined. The combined solution was then incubated at room temperature for a further 20 minutes, before being added to each well of cells. Cells were incubated at 37°C for 5 hours before transfection media was removed and replaced with ES cell media (2.2.1.2).

### 2.2.4 Genomic PCR

*Genomic DNA extraction*

One million cells were lysed by resuspension in 100µl of lysis buffer (25mM NaOH, 0.2mM EDTA in $dH_2O$) followed by incubation at 95°C for 30 minutes. DNA was then precipitated by adding 100µl 40mM Tris-HCl (pH 5.0).

*Genomic PCR*

Genomic DNA was amplified using 1.25 U of HotStarTaq™ DNA Polymerase (Qiagen) in a final reaction volume of 25µl, with 1 X HotStarTaq™ buffer (Qiagen), 0.2mM dNTPs (Life Technologies, NY, USA), 0.2mM primers, and 1µl DNA as prepared above. PCR amplification was performed using the following program: 15 minutes at 95°C, followed by 40 cycles of denaturation at 95°C for 30 seconds, primer annealing at 55°C for 30 seconds, and elongation at 72°C for 30 seconds. PCR products were separated by electrophoresis on 1% agarose gels and visualised by ethidium bromide staining. Primers for analysis of genomic Dicer deletion are given in Appendix III.

### 2.2.5 RNA analysis

#### 2.2.5.1 Quantification of mRNA

RNA was extracted using the RNA-Bee™ reagent (Tel-Test, Inc., Friendswood, TX, USA) according to manufacturer's instructions. Two million cells were lysed in 1ml RNA-Bee and incubated on ice for 30 minutes. 200µl chloroform was added, and samples were mixed by shaking tubes for 30 seconds. Samples were then centrifuged at 12,000g for 15 minutes at 4°C. 400µl of the upper phase was transferred to a new tube and 1µl of GlycoBlue™ (Ambion) was added, followed by 400µl isopropanol. Samples were mixed, incubated at room temperature for 10 minutes, and then centrifuged at 12,000g for 10 minutes at 4°C. The supernatant was removed, leaving blue RNA pellet

at the bottom of the tube, which was then washed by adding 1ml 70% EtOH and centrifuging again at 12,000g for 10 minutes at 4°C. The supernatant was removed, and the pellet allowed to air-dry for 10 minutes before being dissolved in 30µl nuclease-free water. Residual DNA was eliminated using Turbo™ DNase (Ambion) according to manufacturer's instructions before being reversed transcribed into cDNA.

RNA was then reversed transcribed using the SuperScript® III reverse transcriptase (Life Technologies, NY, USA). 1µg of RNA was diluted in 12µl $dH_2O$ for quantification of mature transcripts, and 3µg used for analysis of nascent RNA. RNA dilutions were supplemented with 1µl dNTPs (10mM) and 1µl random primers (0.25µg/µl) (both from Life Technologies, NY, USA). Samples were incubated at 65°C for 5 minutes, then 4°C for 5 minutes. Then, 2µl of 0.1M DTT, 1µl of ribonuclease inhibitor RNase OUT™ (both from Life Technologies, NY, USA), 1µl of SuperScript® III and 2µl of 10X SuperScript® III buffer were added to give a final volume of 20µl. Samples were then heated at 25°C for 10 minutes, 50°C for 50 minutes, and 70°C for 15 minutes. cDNA was diluted 1:10 before use in qPCR reactions.

qPCR quantification of genes of interest was performed using a Chromo4 Opticon Monitor qPCR machine, and Opticon 3 software (MJ Geneworks Inc., Waltham, MA, USA). A reaction mix composed of 2X BioRad iQ™ SYBR® Green Supermix, 0.5mM primers (see Appendix III), and 2µl of cDNA template in a final volume of 10µl was used. Real-time PCR was performed using the following program: reactions were heated to 95°C for 3 minutes, then a cycle of incubations at 94°C for 15 seconds, 60°C for 30 seconds, and 72°C for 30 seconds was repeated 40 times. Fluorescence was read at 72°C, 74°C, 76°C and 80°C. The melting curve was determined from 70-95°C at 0.5°C intervals. Quantification of the relative abundance of detected sequences was calculated using the ΔΔC(t) method (Pfaffl, 2001) where C(t) is the number of cycles above which fluorescence of PCR products can be detected above background. If the amplication efficiency of each cycle is close to 2, the relative abundance of the gene of interest (C(t)1) compared to a housekeeping control (C(t)2) can be calculated as $2^{-\Delta C(t)1}/2^{-\Delta C(t)2}$. Housekeeping genes used for normalisation are indicated where relevant in the text.

All primers used for gene expression analysis by qPCR are listed in Appendix III. Primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3/) and checked against the genome

browser UCSC using *in silico* PCR tools (http://genome.ucsc.edu/) to ensure only one amplification product was produced using the given primer pair. The amplification efficiency of each primer pair was determined using sequential 2-fold dilutions of either cDNA (for testing of primers for mature RNA) or genomic DNA (for testing of qPCR primers nascent RNA). Primers yielding poor linear fits of the C(t) versus logarithm of the primer concentration ($R^2 < 0.99$) or amplification efficiencies outside of the range of 1.8 – 2.2 were discarded.

#### 2.2.5.2 Quantification of miRNA

Small RNAs (less than 200 nucleotides in length) were extracted using the *mir*Vana™ miRNA Isolation Kit (Ambion) according to manufacturer's instructions. For reverse transcription, 22ng of RNA was diluted in 11µl $dH_2O$. To 5µl of this RNA dilution, the following was added: 0.15µl 100mM dNTPs, 1µl MultiScribe™ reverse transcriptase, 1.5µl 10X MultiScribe™ buffer and 0.19µl RNase inhibitor (all part of the TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems)). 3µl of primers specific for the miRNA of interest were added (Applied Biosystems), followed by 4.16µl of $dH_2O$ to give a final volume of 15µl. The reaction was then incubated at 16°C for 30 minutes, followed by 30 minutes at 42°C, and finally 5 minutes at 85°C.

For qPCR analysis of miRNAs of interest, the Chromo4 Opticon Monitor qPCR machine, and Opticon 3 software (MJ Geneworks Inc., Waltham, MA, USA) was used. The PCR reaction mix was as follows: 10µl TaqMan® 2X Universal PCR Master Mix, 1µl TaqMan® probe specific to the miRNA of interest (both from Applied Biosystems), 2µl template (produced in reverse transcription reaction described above) and 8µl $dH_2O$ to a final reaction volume of 20µl. The following program was used for amplification: samples were heated to 95°C for 10 minutes, followed by 40 cycles of incubation at 95°C for 15 seconds, and 60°C for 1 minute. Expression of all miRNAs analysed was normalised to levels of snoRNA-135 using the same method of relative quantification described in 2.2.4.1.

### 2.2.6 Sample preparation for microarray analysis of gene expression

mRNA was extracted as described in 2.2.4.1. Quality of RNA was assessed before sample preparation using the Agilent 2100 Bioanalyser (Agilent Technologies). Samples with a RNA integrity number (RIN score) of greater than 9 were used for microarray analysis. Samples were prepared for hybridisation to Affymetrix Mouse GeneChip® 1.0 ST Arrays according to manufacturer's instructions.

First strand cDNA synthesis, second strand cDNA synthesis, cRNA synthesis, and hydrolysis of cRNA were carried out using reagents from the GeneChip WT cDNA Synthesis and Amplification Kit (Affymetrix). Clean up of cRNA and cDNA was carried out using the GeneChip Sample Clean-up Module. Fragmentation and labelling of single stranded cDNA was carried out using the GeneChip WT Terminal Labelling Kit.

For the first cycle of first strand cDNA synthesis, 0.2µl of T7-$(N)_6$ primers (2.5µg/µl), 0.1µl of poly A controls and 1.7µl of RNase free $dH_2O$ were added to 200ng of RNA to a final volume of 5µl. The sample was then heated to 70°C for 5 minutes, cooled to 4°C for 2 minutes, and then kept on ice for no longer than 10 minutes. Then, the following was added to each sample: 2µl 5X first strand buffer, 1µl 0.1M DTT, 0.5µl 10mM dNTP mix, 0.5µl RNase inhibitor, and 1µl SuperScript II (added last). Samples were then subjected to the following incubations: 10 minutes at 25°C, 1 hour at 42°C, 10 minutes at 70°C, and 2 minutes at 4°C. Samples were then put on ice.

For second strand cDNA synthesis, a mastermix was prepared containing the following (per sample): 4.8µl of RNase free $dH_2O$, 4µl of 17.5mM $MgCl_2$ (dilution made fresh), 0.4µl 10mM dNTP mix, 0.6µl DNA Polymerase I, and 0.2µl RNase H (enzymes were added last). 10µl of the reaction mix was added to each sample, which were then incubated in a thermocycler at 16°C for 2 hours without a heated lid, and then incubated at 80°C for 10 minutes in a thermocycler with a heated lid (100°C). Samples were then cooled to 4°C.

For cRNA synthesis, the following was added to each sample of cDNA: 5µl IVT buffer, 20µl NTP mix, and 5µl enzyme mix (added last). Samples were then incubated at 37°C for 16 hours in a thermocycler with a heated lid (100°C). cRNA was then column purified and eluted in 15µl of RNase free $dH_2O$. The 260nm and 280 nm absorbance of each sample was measured. The yield of the first cRNA cycle should be greater than 16µg.

10µg of cRNA from the first cycle of cRNA synthesis was used for the second cycle of first strand cDNA synthesis (using the same kit as above). 1.5µl of 3µg/µl random primers were added to 10µg cRNA from the previous step, and the final volume of the reaction brought to 8µl with RNase-free $dH_2O$. Samples were then heated to 70°C for 5 minutes, cooled to 25°C for 5 minutes, and brought to 4°C. A mastermix was prepared containing the following (per sample): 4µl 5X first strand buffer, 2µl 0.1M DTT, 1.25µl 10mM dNTP + dUTP, and 4.75µl SuperScript II (added last). 12µl of the

mastermix was added to each sample, which were then heated to 25°C for 10 minutes, 42°C for 90 minutes, 70°C for 10 minutes, and then cooled to 4°C.

Hydrolysis of cRNA was carried out by adding 1µl of RNase H to each sample containing single stranded cDNA, which were then heated to 37°C for 45 minutes, 95°C for 5 minutes, and then cooled to 4°C. cDNA was then column purified and eluted in 28µl. 260nm and 280nm absorbances were measured to ensure cDNA yield was greater than 5.5µg. An aliquot of 1.5µl cDNA was for size analysis of amplified DNA using the Agilent Bioanalyser 2100.

Fragmentation of single stranded cDNA was carried out by adding 10µl RNase free $dH_2O$, 4.8µl 10X cDNA fragmentation buffer, 1µl 10U/µl UDG and 1µl 1,000U/µl APE 1 to 5.5µg of cDNA from the previous step. Samples were then heated to 37°C for 60 minutes, 93°C for 2 minutes, and then cooled to 4°C. A 1.5µl aliquot was taken for size analysis, and run alongside the unfragmented sample (taken in the previous step) on the Agilent Bioanalyser 2100. Fragmented cDNA size should be 40-70bp.

Labelling of fragmented cDNA was carried out by adding 12µl 5X TdT buffer, 2µl TdT and 1µl 5mM DNA labelling agent to 45µl of fragmented cDNA. Samples were then incubated at 37°C for 60 minutes, 70°C for 10 minutes, and then cooled to 4°C. Samples were frozen at -80°C for storage and hybridised to Affymetrix GeneChip® Gene 1.0 ST Arrays by the MRC CSC Genomics Laboratory. Initial processing and analysis of the differential gene expression was carried out by Dr. Tom Carroll (MRC CSC, currently at CRUK, Cambridge Research Unit).

### 2.2.7 Mitotic arrest and release assay

ERT2Cre x $Dicer^{lox/lox}$ and ERT2Cre x $Dicer^{\Delta/\Delta}$ ES cells were cultured in N2B27 + 2i and DMEM + FCS (see 2.2.1.2) at a 1:1 ratio, in T75 flasks (Nunc, Thermo Scientific) coated with 0.2% gelatin. One T75 flask of each cell type was divided into five new flasks 24 hours before the experiment.

A day after cells were plated, 20ng/ml demecolcine (D1925, Sigma) was added to four of the five flasks, and cells were incubated at 37°C for 4 hours. The remaining flask was collected as an asynchronous control. Then, media was carefully removed from each flask (taking care not to detach adherent cells) and replaced with 15ml cold PBS -/-. Flasks were then agitated gently to release

mitotic cells from the culture vessel. The PBS -/- containing mitotic cells was pooled from the four flasks, and centrifuged at 4°C for 5 minutes at 12000rpm. Samples were resuspended in 1ml demecoline-free culture media and counted. 100µl of cells were fixed in 70% EtOH for cell cycle analysis, and the remaining 900µl was divided between two gelatinised wells of a 6-well plate in 3ml demecolcine-free culture media per well.

After two hours incubation at 37°C, non-adherent cells in the 6-well plates were removed and adherent cells (i.e. cells which had progressed from G2/M to G1) were collected for further analysis.

### 2.2.8 Elutriation

The elutriation chamber and centrifuge was set up as described in (Banfalvi, 2008). Before cells were loaded, the entire system was washed with 200ml 70% EtOH, followed by 500ml $dH_2O$, and then 200ml PBS -/-. The system was then equilibriated by running elutriation buffer (PBS -/- supplemented with 1% v/v FCS and 0.01% w/v EDTA) through the chamber until cells were ready to be loaded. The centrifuge chamber was pre-cooled to 4°C and all buffers kept on ice as far as possible.

Eight T75 flasks of cells were collected by trypsinisation and centrifuged at 12000rpm for 5 minutes. Cells were resuspended in 15ml of cold elutriation buffer and counted. A minimum of 100 x $10^6$ ES cells are required for each elutriation. Cells were passed through an 18G syringe pre-loading to ensure a single cell suspension was obtained. The cell suspension was loaded into the chamber at a flow rate of 6ml/min and cells were allowed to settle in the chamber for 30 minutes, in order to form a gradient of cells ordered by size before the first cells were eluted.

The flow rate was then gradually increased at increments of 1ml/min. 200ml of eluted cells was collected at each flow rate from 7ml/min to 17ml/min, and kept at 4°C. After all fractions were collected, samples were centrifuged at 4°C for 5 minutes at 12000rpm. Cells were resuspended in 1ml cold PBS -/- and counted. Then, 100µl of cell suspension was fixed in 1ml 70% EtOH for cell cycle analysis, whilst the remaining 900µl was aliquoted into a 1ml centrifuge tube, spun briefly at max speed to pellet cells, and then snap frozen in liquid nitrogen.

### 2.2.9 Western blot

Whole cell extracts were prepared by resuspending one million cells in 50µl of PBS -/-, then adding 50µl of 2X sample buffer ($dH_2O$ supplemented with 100mM Tris (pH 6.8), 2% sodium dodecyl sulphate (SDS), 20% glycerol, 0.002% bromophenol blue, 1.43M β-mercaptoethanol) and incubating samples at 95°C for 5 minutes. 10µl of lysate (equivalent to 100,000 cells) was loaded per lane onto an acrylamide (BioRad) stacking gel (4% w/v acrylamide, 0.125M Tris (pH 6.8), 0.1% w/v SDS, 0.1% w/v ammonium persulfate, and 0.1% v/v N, N, N', N'-tetramethylethylenediamine (TEMED)). Proteins were separated in an 8% acrylamide resolving gel (8% w/v acrylamide, 0.4M Tris (pH 8.8), 0.1% w/v SDS, 0.1% w/v ammonium persulfate, and 0.1% TEMED) using Tris-glycine electrophoresis buffer (1.5% w/v Tris, 7.2% glycine, 0.5% w/v SDS). Gels were run at 25mA / 140V per gel (50mA / 140V for two gels). Proteins were transferred to a polyvinylidene fluoride (PVDF) membrane (Millipore) using the Trans-Blot® SD Semi-Dry Transfer Cell (BioRad) in transfer buffer (48mM Trisma base, 39mM glycine, 0.037% w/v/ SDS and 20% v/v methanol) . PVDF membranes were prepared immediately prior to use by incubating for 30 seconds in methanol, 5 minutes in $dH_2O$, and 5 minutes in transfer buffer. Proteins were transferred for 1 hour and 45 minutes at 140mA / 25V per gel (280mA / 25V for two gels).

Membranes were then incubated for one hour in blocking buffer (10% w/v fat free milk powder (Marvel), 1.2g/L Tris pH 7.4, 8.75 g/L NaCl) followed by incubation with the primary antibody (diluted in blocking buffer) for 2 hours at room temperature with agitation. Membranes were then washed three times for ten minutes in wash buffer (1.2g/L Tris pH 7.4, 8.75 g/L NaCl) before incubation with the fluorophore-conjugated secondary antibody (Life Technologies) (also diluted in blocking buffer) in the dark for one hour at room temperature with agitation. Fluorescence was then quantified using the LI-COR Odyssey® Imager.

### 2.2.10 Chromatin Immunoprecipitation

50 x $10^6$ cells were centrifuged at 4°C for 5 minutes at 12000rpm, and resuspended in 18ml of complete medium (2.2.1.2). 2ml of fixation solution was then added (50mM Hepes (pH 8.0), 1mM EDTA (pH 8.0), 0.5mM EGTA (pH 8.0), 100mM NaCl and 10% formaldehyde). Cells were incubated for 10 minutes at room temperature on a rotating platform, then 2.5ml of 1.25M glycine was added. Fixed cells were then centrifuged at 4°C for 5 minutes at 550 rcf, before cells were resuspended in

10ml cold PBS -/- and transferred to a 15ml tube. Cells were then washed twice by centrifugation at 4°C for 5 minutes at 550 rcf and resuspended in 10ml cold PBS -/- each time. After the final wash, all supernatant was removed and cell pellet was snap frozen in liquid nitrogen.

All buffers used for cell lysis, sonication and immunoprecipitation contain protease inhibitors (1:25) (Roche) unless otherwise stated. Pellets of fixed cells were then thawed on ice, resuspended in 1ml of cell lysis buffer (5mM PIPES (pH 8.0), 85mM KCl, 0.5% NP40), and incubated on ice for 20 minutes. The suspension was then transferred to a 1.5ml centrifuge tube and centrifuged at 4°C for 10 minutes. The supernatant was removed, and the pellet resuspended in 400µl nuclear lysis buffer (50mM Tris-Cl (pH 8.1), 10mM EDTA (pH 8.0), and 1% SDS). Samples were incubated on ice for 10 minutes.

To sonicate, 600µl RIPA buffer (10mM Tris-HCl (pH 7.5), 1mM EDTA (pH 8.0), 0.5mM EGTA (pH 8.0), 1% Triton X-100, 0.1% SDS, 0.1% Na Deoxycholate and 140mM NaCl) was added to each sample. Samples were sonicated for 2 x 15 minutes at 4°C with 30 seconds on, and 30 seconds off using a Bioruptor® (Diagenode). Chromatin fragment size was assessed on 1% agarose gels and was accepted if between 500-1000bp. Sonicated lysates were then centrifuged at 10°C for 20 minutes to remove insoluble material, and the supernatant transferred to a new tube. DNA content was quantified by spectrophotometric analysis, and sonicated chromatin aliquoted and stored at -80°C.

For immunoprecipitation (IP): for Ezh2, 300µg chromatin was used with 50µl of Dynabeads® Protein A (Life Technologies) and 10µg of antibody (Diagenode) in a final volume made up to 1ml with RIPA buffer without SDS. For histone proteins and modifications (histone H3 (Abcam), H3K4me3 and H3K27me3 (both from Millipore)) 20µg of chromatin was used with 25µl of Dynabeads® Protein A and 5µg of antibody in a final volume made up to 250µl with RIPA buffer without SDS.

In both cases, prior to immunoprecipitation the required volume of beads was first washed by resuspending beads in 1ml ice cold RIPA buffer, then placing tubes in a magnetic field, removing the supernatant and repeating the process twice more. Washed beads were then incubated with the required quantity of antibody in RIPA buffer on a rotating platform at 4°C for 6 hours. Sonicated chromatin was then thawed and 10% of the total chromatin required for the IP was removed and frozen as the input sample. Next, beads coupled with antibody were added to the remaining quantity

of sonicated chromatin, the total volume of the IP reaction was made up to the required volume as described above, and samples were incubated on a rotating platform at 4°C overnight.

The following day, beads were washed by placing the tubes in a magnetic field, removing the supernatant, and then resuspending the beads in the required buffer. Tubes were then incubated in the buffer on a rotating platform at 4°C for 5 minutes before the process was repeated. The following sequence of buffers was used to wash beads: two washes in RIPA buffer; two washes in high salt RIPA (10mM Tris-HCl (pH 7.5), 1mM EDTA (pH 8.0), 0.5mM EGTA (pH 8.0), 1% Triton X-100, 0.1% SDS, 0.1% Na Deoxycholate and 500mM NaCl), one wash in RIPA with lithium chloride (10mM Tris-HCl (pH 7.5), 1mM EDTA (pH 8.0), 0.5mM EGTA (pH 8.0), 1% Triton X-100, 0.1% SDS, 0.1% Na Deoxycholate and 250mM LiCl), and two washes in Tris-EDTA buffer (10mM Tris-Hcl, 1mM EDTA).

Chromatin was eluted by resuspending washed beads in 150µl of elution buffer (20mM Tris-HCl (pH 7.5), 5mM EDTA (pH 8.0), 50mM NaCl, 1% SDS, 50µg/ml proteinase K, and 100µg/ml RNase A) and incubating at 37°C for two hours, followed by incubation at 68°C with strong agitation (1000 rpm: 30 seconds on, 30 seconds off) overnight. 150µl elution buffer was also added to thawed input samples, which were subject to the same incubations. The following morning, beads were placed in a magnetic field and the supernatant transferred to a new tube. Beads were resuspended in an additional 150µl elution buffer (and 150µl elution buffer was also added to input samples) and incubated at 68°C with strong agitation for a further 5 minutes. A phenol-chloroform protocol was used to extract DNA, which was then precipitated by adding two volumes of 100% EtOH, 50µl sodium acetate (pH 5.2) to each sample followed by incubation for 30 minutes at -80°C. Samples were then centrifuged at 4°C for 30 minutes at maximum speed to pellet DNA, which was then washed by removing the supernatant, adding 1ml of 70% EtOH, and centrifuging samples again at 4°C for 10 minutes at maximum speed. EtOH was then removed, and the DNA pellet was allowed to air-dry for 10 minutes at room temperature before being resuspended in 100µl of TE buffer. 2µl of ChIP DNA was used per qPCR reaction.

### 2.2.11 Immunofluorescence

The following protocol is relevant to data presented in Appendix II.

Sterile glass coverslips were placed in 6-well plates and 500µl of poly-L-lysine (Sigma) was added to each for 1 minute. 3ml of PBS -/- was then added to each well for 10 minutes to wash. 3T3 fibroblasts were then trypsinised and 0.1 x $10^6$ cells plated onto each poly-L-lysine coated coverslip. Cells were cultured overnight in media (2.2.1.3), washed once in PBS -/-, and fixed by adding 2.5ml 2% paraformaldehyde per well and incubating for 20 minutes at room temperature. Coverslips were then washed three times (3 minutes each) with PBS -/-, and cells were permeabilised by incubation in 0.4% Triton-X-100 (Sigma) for 6 minutes.

Cells were then washed twice (3 minutes each) in PBS -/-, and once in wash buffer (PBS with calcium and magnesium (+/+) (Life Technologies) supplemented with 0.2% w/v bovine serum albumin (BSA) (Sigma), 0.05% v/v Tween-20 (Sigma)). Coverslips were then incubated in blocking buffer (PBS +/+ supplemented with 2.5% w/v BSA, 0.05% v/v Tween-20, and 10% v/v normal goat serum (Vector)) for 30 minutes at room temperature in a humid chamber. Coverslips were washed once in wash buffer and then incubated in the primary antibody, Alexa Fluor® 568 Phalloidin (Life Technologies), diluted 1:50 in blocking buffer for 2 hours at room temperature in a humid chamber. Cells were then washed twice in wash buffer, once in PBS -/-, and mounted in Vecatshield (Vector) with DAPI (0.1µg/ml).

Samples were visualised using a TCS SP5 Leica laser scanning confocal microscope. Images were processed using Leica Confocal software and Adobe Photoshop CS2.

For immunofluorescence analysis using the CYTOO chips™, the same protocol was used as described above except CYTOO micropatterned coverslips were used instead of poly-L-lysine coated coverslips. In addition, instead of being cultured overnight on coverslips, 3T3 fibroblasts were allowed only four hours to adhere to CYTOO chips before fixation. Automated image collection and collective data analysis was done by Dr. Dirk Dormann, Microscope Facility, MRC Clinical Sciences Centre, London.

# 3.0 Mitotic inheritance of microRNA

## 3.1 Introduction

As previously discussed, cell to cell transmission of microRNAs in mammalian organisms has not been comprehensively investigated. In order to elucidate whether a similar capacity can be assigned to miRNAs in higher eukaryotes, we developed an assay to ask whether miRNAs are transmitted from parent to daughter cells in mammalian systems.

## 3.2 Optimisation of mitotic inheritance assay in primary mouse $CD4^+$ T cells

First, it was necessary to identify a method to dissociate parental cells from progeny within a total population of dividing cells. The fluorescent dye carboxyfluorescein diacetate succinimidyl ester (CFSE) has been used for monitoring lymphocyte proliferation both *in vitro* and *in vivo* (Lyons and Parish, 1994; Lyons, 2000). CFSE is a membrane permeable dye which binds non-specifically to free amine groups on cytoplasmic proteins. Using CFSE, a parental population of undividing cells are stained with the dye and then induced to proliferate. Upon each cell division, the dye is equally partitioned to daughter cells. Each generation of progeny will therefore emit on average half of the fluorescence of their parent population, allowing dissociation of undivided and divided cells. The specific amenability of lymphocytes to this protocol is attributable to the fact that at the population level, these cells are extremely homogenous in terms of cell size. This means that when stimulated to divide, all daughter cells of a CFSE-labelled parental population are also very similar in size, and therefore take up a similar amount of the dye. Parental and daughter cell populations therefore emit distinctive amounts of fluorescence, facilitating visualization of sequential generations of the labeled parental population.

Having identified a protocol by which parent and daughter cells could be identified within a proliferating population, an assay was designed to assess whether this system could be used to ask whether small oligonucleotides are transmitted through cell division. To do this, I isolated total lymphocytes from wildtype mice. These cells were stained with CFSE, activated for 24 hours, and then transfected with a fluorescently labelled synthetic molecule (siGlo) before stimulated cells had undergone the first cell division. SiGlo is designed as a transfection control and therefore modified in order to maximise transfection efficiency and to resist degradation *in vitro*. Analysis of the quantity of

siGlo in parental and daughter populations was assessed by fluorescence activated cell sorting (FACS).

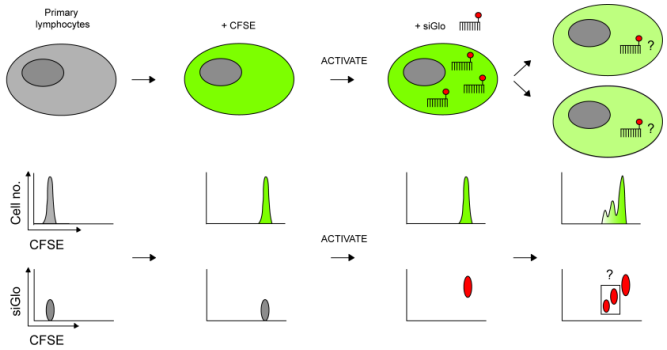



**Figure 3.1: Experimental system to investigate mitotic inheritance of small RNAs.**
Schematic of assay designed to investigate mitotic inheritance of small RNAs. Total primary lymphocytes were extracted from 8-10 week old wildtype mice. Cells were stained with 62.5µg/ml CFSE and activated for 24 hours with α-TCR (500ng/ml) and α-CD28 (2µg/ml). Cells were then collected and transfected with 10µM siGlo using the Amaxa Nucleofector protocol. Cell proliferation and siGlo fluorescence was assessed by FACS 48 hours later. Hypothetical graphs show a representation of anticipated FACS output. Undivided, parental cells appear as a single, CFSE positive peak. Successfully transfected cells within this peak will appear siGlo positive. If siGlo is transmitted from parent to daughter cells, fluorescence will also be detected in sequential populations of dividing cells identified by CFSE.

Fig. 3.2A illustrates that using this assay, siGlo can be detected in undivided and divided populations of both $CD4^+$ and $CD8^+$ T cells; indicating that the oligo is transmitted through cell division. The quantity of siGlo in each sequential population of daughter cells appears to approximately half in both classes of T cells (Fig 3.2B); suggesting that the molecule does not undergo degradation *in vitro* during the course of the experiment (as would be expected due to its synthetic nature).

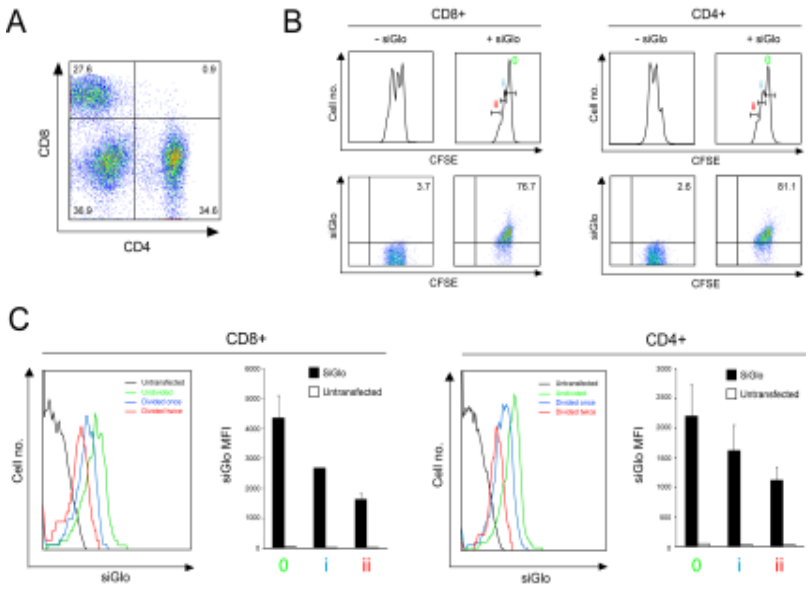



**Figure 3.2: Stabilised fluorescently labelled oligonucleotides are inherited through cell division in primary $CD4^+$ T cells.**
(A) Identification of CD4+ and CD8+ T cells in population of total lymphocytes stained with CSFE and transfected with siGlo (72h post-activation; 48h post-transfection). (B) Top panel: analysis of cell proliferation by CFSE dilution (0: undivided cells; i: one cell division; ii: two cell divisions). Bottom panel: distribution of siGlo fluorescence in transfected cells (plus control). (C) Distribution (histograms) and quantification (bar graphs) of siGlo fluorescence in populations identified in (B). Bar graphs show mean fluorescence of the indicated populations. Error bars: s.t.d. of three biological replicates. MFI: mean fluorescence intensity (arbitrary units).

It was observed that cells transfected with siGlo exhibit proliferation defects compared to untransfected controls (- siGlo) which were not subject to the nucleofection procedure. This data indicates that the transfection protocol impedes proliferation of primary T cells and was a factor subsequently taken into account when adapting the assay for further experiments.

### 3.3 Mitotic inheritance of fluorescently labelled microRNA mimic

Next, having optimised an experimental system capable of detecting transmission of small oligonucleotides from parent cells to progeny, I then addressed the question of whether more biologically relevant molecules are inherited in a similar manner. To do this, instead of transfecting siGlo (which bears no sequence homology to any known siRNA or miRNA); a fluorescently labelled mimic of the T-cell expressed miR-181a was used. Whilst the oligo is still synthetic (as in the case of siGlo, and conjugated with the same fluorophore DY547 for consistency) the mimic was not modified to increase transfection efficiency or stability *in vitro.* I therefore anticipated that the miR-181a mimic would function more as an endogenous miRNA; in terms of incorporation into RISC, binding to target mRNAs, and being subject to cellular degradation processes.

In this experiment, analysis was restricted to $CD4^+$ cells only as these cells constitute a greater proportion of total lymphocytes (Fig. 3.2A), facilitating higher accuracy and greater reliability of analysis due to the increased number of cells in each subpopulation defined by CFSE staining. Figure 3.3A shows that like siGlo, miR-181a-DY547 can be detected in parent and daughter populations of transfected $CD4^+$ T cells. As with siGlo, the amount of the fluorophore detected in each sequential cell division approximately halves (Fig. 3.3B), supporting the notion that this oligo is indeed transmitted through cell division.

The transfection efficiency and *in vitro* stability of the miR-181-DY457 oligo was shown, as anticipated, to be greatly reduced compared to siGlo. This is indicated by the two-fold reduction in miR-181a-DY547 oligo 24h post-transfection in undivided cells, compared to the consistent detectable amount of siGlo at both timepoints (Fig. 3.2C). This data suggests that the miR-181a-DY547 oligo is degraded within the cell, and therefore subject to endogenous biological processing which siGlo is designed to be resistant to. Inheritance of miR-181a-DY547 provides more support for the notion that miRNA inheritance is a biologically relevant phenomenon.

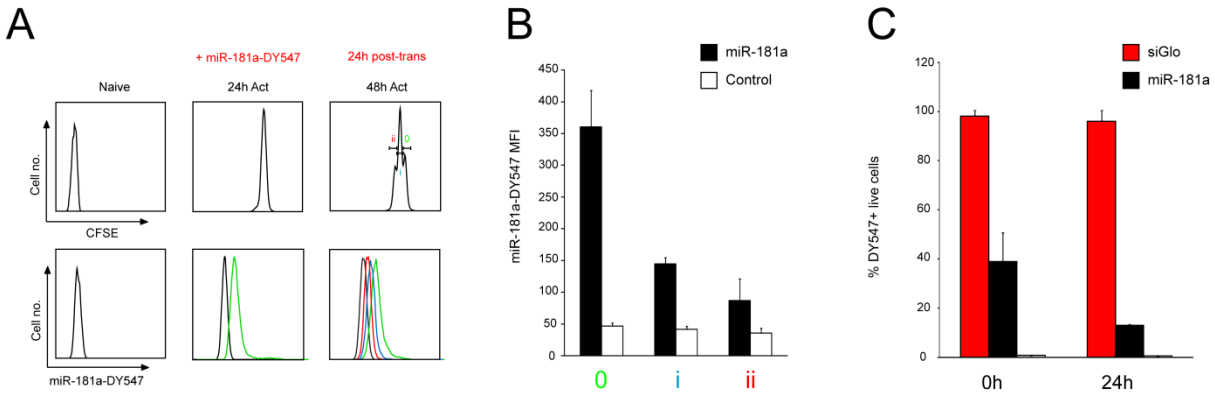



**Figure 3.3: Fluorescently labelled miR-181a is transmitted from parent to daughter cells through mitosis in primary mouse CD4$^+$ T cells.**
(A) Top panel: analysis of cell proliferation by CFSE dilution pre- and post- activation/transfection (0) undivided cells; i) one cell division; ii) two cell divisions). Bottom panel: miR-181a-DY547 fluoresence in populations identified by CFSE (Overlay: black - untransfected cells; green – undivided cells (i); blue – one cell division (ii); red – two cell divisions(iii).) (B) Quantification of mean DY547 fluorescence in populations shown in (A). (C) Quantification of mean DY547 in siGlo and miR-181a transfected CD4$^+$ T cells 0 hours and 24 hours post-transfection (values shown for undivided cell populations). Error bars: s.t.d. of three biological replicates. MFI: mean fluorescence intensity.

## 3.4 Mitotic inheritance of unmodified non-T cell expressed microRNA mimic

The low transfection efficiency of the miR-181a-DY547 oligo (Fig. 3.3C) meant that a significantly greater amount of oligo had to be transfected in comparison to siGlo in order to be able to detect fluorescence by FACS in divided cells (40μM miR-181a-DY547 compared to 10μM siGlo). As a result, it is possible that the transmission of miR-181a-DY547 from parent to daughter cells shown in Fig. 3 is attributable to saturation of the parent population with the fluorophore, as opposed to active and functional mitotic inheritance of the miRNA mimic.

Therefore, to improve the biological relevance of this assay, I modified the protocol in order to allow quantification of the transfected miRNA in parent and daughter populations by qPCR. qPCR is a more sensitive method of analysis, capable of detecting low levels of DNA and RNA, compared to FACS which relies on high levels of expression of the protein of interest in a given experimental system. Use of qPCR to detect the transfected RNA therefore meant that the initial concentration of oligo could be reduced, alleviating potential misleading results due to saturation of the parent population with the exogenously introduced RNA.

As the qPCR based method will also detect and quantify endogenously expressed miRNAs, I selected an miRNA which is not expressed in T cells: neuronal-specific miR-124a (Lagos-Quintana et al., 2002). Fig. 3.4A shows that miR-124a is usually expressed at almost undetectable levels in Dicer

WT CD4$^+$ T cells, and that transfection of miR-124a mimic at a final concentration of 10uM does not result in levels of expression greater than that of endogenous, T-cell expressed miRNAs miR-142-3p, miR-21 and miR-92.

As before, cells were stained with CFSE, activated for 24 hours and transfected with the miRNA mimic, but instead of gating on parent and daughter populations post-acquisition, these populations were physically isolated based on CFSE fluorescence using FACS sorters (Fig. 3.4B). I then extracted RNA from undivided and divided cells, and the levels of mimic quantified by qPCR.

qPCR analysis of levels of miR-124a in sorted populations of transfected cells which have either not divided or have divided once suggests that similarly to siGlo and miR-181a-DY547, unmodified miR-124a mimics transfected at a physiological level are inherited from parent to daughter cells in mouse T cells, as indicated by detection of approximately half the quantity of oligo in divided cells as present in the parent population (Fig. 3.4C). Due to the biologically relevant quantity of oligo initially transfected, it is therefore likely that presence of miR-181a-DY547 in daughter cells is not due to the saturation of the parent cells with the exogenously introduced RNA.

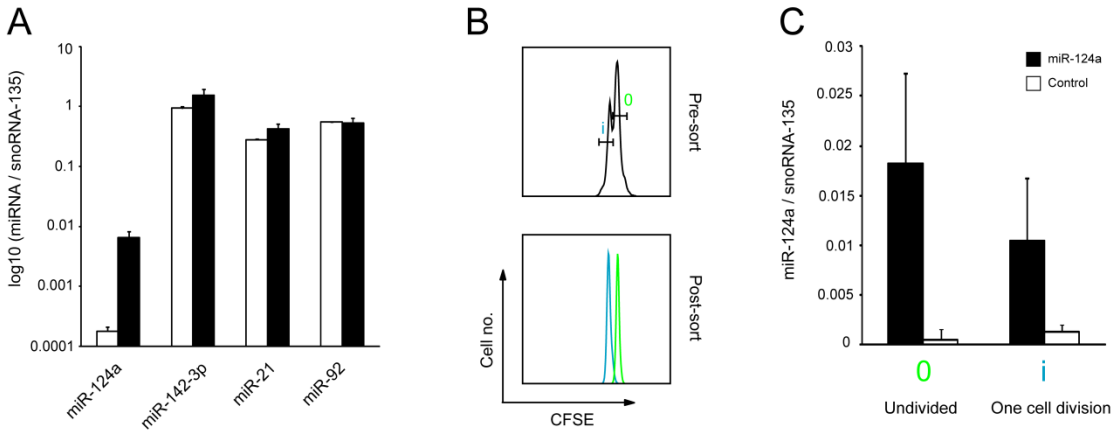



**Figure 3.4: Neuronal-specific miRNA mimics expressed at endogenous levels in lymphocytes are transmitted through cell division in primary mouse CD4$^+$ T cells.**
(A) Quantification of miRNA levels in WT CD4$^+$ T cells 24 hours post-transfection with 10μM miR-124a mimic (black bars) and control oligo (white bars). (B) Top panel: analysis of proliferation 24 hours post-activation of miR-124a transfected cells by CFSE dilution.(0) and (i) show gates used for sorting of parent and daughter cells respectively. Lower panel shows overlay of CFSE profiles of populations (0) (blue) and (i) (green) post-sort. (C) Quantification of miR-124a levels by qPCR in sorted populations of transfected cells. Error bars: s.t.d. of three biological replicates.

### 3.5 Mitotically heritable microRNAs downregulate target gene expression in recipient cells

Having demonstrated that synthetic oligos which mimic non-T cell expressed miRNAs can be transfected at endogenous levels and inherited through cell division in $CD4^+$ T cells, I was next interested to investigate whether inherited miRNAs are functional in recipient cells, in terms of downregulation of target mRNAs.

To do this, I used microRNA-deficient T cells, in which direct targets of T-cell specific miRNAs should be upregulated. If mitotically inherited miRNAs are functional in recipient cells, it would be expected that the transmission of the transfected miRNA would result in downregulation of specific targets genes in the daughter population. The lack of endogenous miRNAs in the recipient cells would indicate that this effect is mediated by exogenously introduced, miotically heritable miRNAs.

To obtain microRNA-deficient lymphocytes, I isolated $CD4^+$ T cells from CD4Cre $Dicer^{lox/lox}$ mice previously developed by members of the lab (Cobb et al., 2005). In this system, the RNase III domain of the Dicer gene is flanked by loxP sites (Fig. 3.5A). The Cre recombinase enzyme which cuts specifically at loxP sites is under the control of the CD4 promoter in these mice, which results in expression of Cre at the double positive stage during lymphocyte development (see Appendix I for a detailed description of the Cre/LoxP system). Dicer is therefore only inactivated in developing lymphocytes expressing the CD4 gene, avoiding deleterious effects of loss of miRNA function in other organs and maintaining viability of the mouse. $CD4^+$ T lymphocytes isolated from these mice show a greater than 80% reduction in levels of Dicer mRNA in both naive and activated cells (Fig. 3.5B), and subsequently are significantly depleted of miRNAs (Fig. 3.5C).

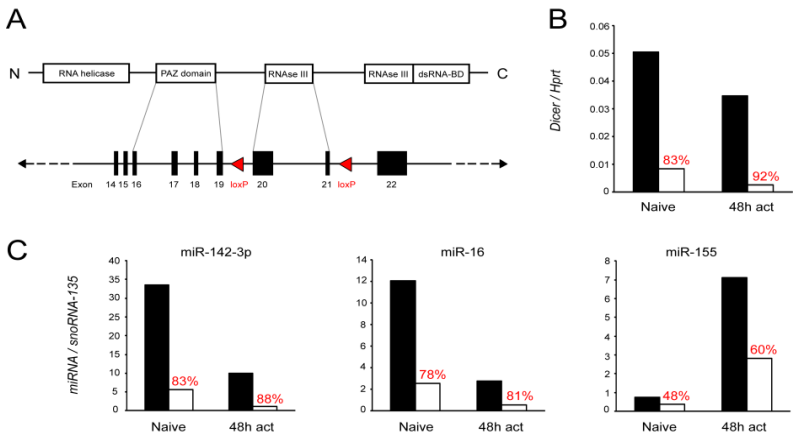



**Figure 3.5: Isolation of microRNA-deficient $CD4^+$ T cells from CD4Cre $Dicer^{lox/lox}$ mice.**
(A) Top panel: representation of key functional domains in the Dicer protein (mouse). Bottom panel: schematic diagram of exons 14-22 of Dicer gene, illustrating location of loxP sites used to delete RNaseIII domain after activation of Cre. Quantification of (B) Dicer mRNA and (C) miRNA expression by qPCR in naive and activated WT (black) and $Dicer^{lox/lox}$ (white) $CD4^+$ T cells.

Next, it was necessary to identify potential targets of T-cell specific miRNAs in order to assess the functionality of mitotically inherited microRNAs in this system. Whilst it has been published that *Ac9* is a direct target of highly expressed miR-142-3p in both $CD4^+$ and $CD4^+CD25^+$ ($T_{reg}$) cells (Huang *et al*., 2009), the study shows that *Ac9* is expressed at a much lower level in $CD4^+$ T cells compared to $T_{regs}$. I confirmed this by qPCR expression analysis of naive purified CD4+ T cells from Dicer WT mice and from $CD4^+CD25^+$ ($T_{reg}$) cells[1] (Fig. 3.6A). However, as the experimental system described in 3.2 relies on quantification of miRNA and target expression in activated $CD4^+$ T cells, and the cells analysed by Huang *et al* were in the naive state, I next quantified *Ac9* expression in Dicer WT and KO $CD4^+$ T cells activated for 48 hours. It was observed that *Ac9* levels are in fact downregulated on activation in both WT and KO cells and additionally, in activated cells, *Ac9* expression is no longer upregulated in Dicer KO cells compared to WT (Fig. 3.6B). Therefore, despite being a validated target of miR-142-3p in naive $CD4^+$ T cells and other T cell subsets, *Ac9* is not appropriate for use as a readout for miRNA function in the experimental system described here.

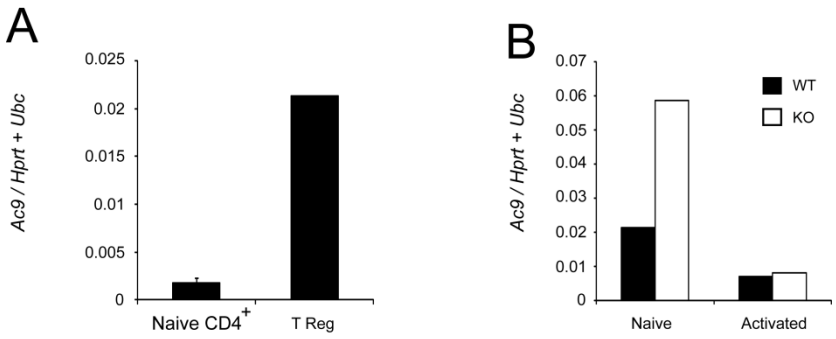



**Figure 3.6: Validated miR-142-3p target *Ac9* is not upregulated in activated Dicer KO $CD4^+$ T cells.**
(A) qPCR analysis of *Ac9* expression in naive WT $CD4^+$ T cells and $T_{reg}$ cells. Error bars: s.t.d. of three biological replicates ($CD4^+$ cells only). (B) qPCR analysis of *Ac9* expression in naive and 48 hour activated WT and Dicer KO $CD4^+$ T cells.

Therefore, to identify novel targets of miR-142-3p, the miReduce algorithm was applied to genome-wide microarray expression data for naive Dicer KO $CD4^+$ T cells compared to WT (A. Marçais, unpublished data). miReduce correlates the logarithmic fold change in expression with motifs found in a specified region of the transcripts of genes (in this case, the 3'UTR), and assigns statistical significance to association of a given motif with a positive or negative fold change in expression (Sood et al., 2006). These motifs were then annotated to identify potential microRNA binding sites. The top five motifs associated with a positive fold change are shown in Fig 3.6A, ranked by significance. As the average change in expression of genes with the miR-142-3p binding site is

[1] cDNA obtained by Dr. L. Bruno.

greatest, I investigated targets of this miRNA further. I identified predicted targets of miR-142-3p using the miRNA prediction algorithms TargetScan and Pictar. The five most upregulated predicted targets of miR-142-3p in Dicer KO CD4$^+$ T cells (based on microarray data) are listed in Fig. 3.7B.

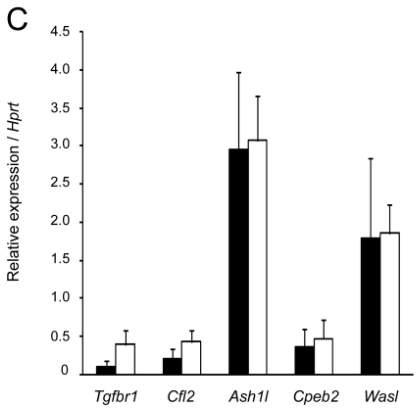
Next, I validated the expression of the top five most deregulated predicted miR-142-3p targets by qPCR in naive Dicer WT and KO CD4$^+$ T cells (Fig. 3.7C). *Tgfbr1* and *Cfl2* were upregulated in naive Dicer KO CD4+ T cells compared to Dicer WT cells. However, as the mitotic inheritance assay described in previous sections is carried out in activated cells, it was then necessary to verify whether the expression of these genes changes in activated cells; to assess whether, as in the case of *Ac9*, these genes are in fact no longer upregulated in activated Dicer KO T cells compared to WT. Analysis of *Tgfbr1* and *Cfl2* expression during a timecourse of activation indicates that whilst the pattern of expression of these genes does change as the cells are activated, upregulation in Dicer KO compared to Dicer WT cells is maintained. These data validate *Tgfrb1* and *Cfl2* as potential novel targets of miR-142-3p in both naive and activated CD4$^+$ T cells, and therefore suitable candidates for further analysis.

A

| 3' UTR motif | p-value | No. of genes | Avg. expr. in Dicer KO | microRNA family |
|---|---|---|---|---|
| **tttgta** | 0 | 6857 | 0.06 | — |
| **tactgt** | 0 | 4883 | 0.07 | miR-101a |
| **gtgggca** | $4.0 \times 10^{-10}$ | 1622 | 0.01 | miR-702 |
| **acactac** | $5.3 \times 10^{-7}$ | 829 | 0.1 | miR-142-3p |
| **acttgaa** | $2.5 \times 10^{-6}$ | 1946 | 0.08 | miR-26a |

B

| Gene | Description | log2 fold change | Adjusted p-value |
|---|---|---|---|
| *Tgfbr1* | Transforming growth factor, β-receptor 1 | 1.17 | 0.04 |
| *Cfl2* | Cofilin 2 | 0.46 | 0.26 |
| *Ash1l* | Ash1-like (*Drosophila*) | 0.31 | 0.56 |
| *Cpeb2* | Cytoplasmic polyA binding element protein 2 | 0.31 | 0.26 |
| *Wasl* | Wiskott-Aldrich syndrome-like (human) | 0.18 | 0.42 |



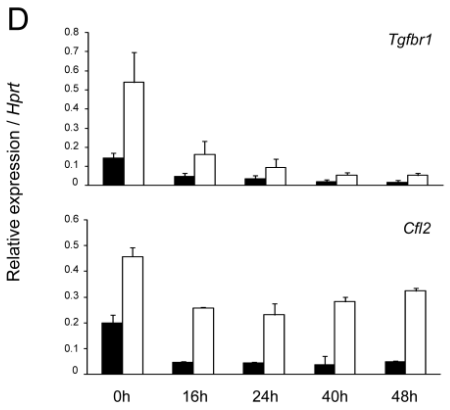




**Figure 3.7: Identification of potential targets of miR-142-3p in CD4$^+$ T cells.**
(A) Five motifs most statistically significantly associated with a positive fold change in expression in naive CD4$^+$ Dicer KO T cells vs. WT microarray data. (B) Top five most deregulated predicted targets of miR-142-3p (PicTar and TargetScan) in Dicer KO CD4$^+$ T cells microarrays. Values shown are calculated from probe values in array data. (C) qPCR validation of expression of predicted miR-142-3p targets identified in (B) in Dicer WT (black) and KO (white) CD4$^+$ T cells. (D) qPCR analysis of *Tgfbr1* and *Cfl2* expression in timecourse of activation in Dicer WT (black) and KO (white) CD4$^+$ T cells. Error bars in (B) and (C): s.t.d. of three biological replicates.

Having identified *Tgfbr1* and *Cfl2* as candidate miR-142-3p targets which could be used to assess whether mitotically inherited miRNAs are functional in recipient cells, I next wanted to improve the assay itself in order to reliably investigate this question. As indicated by the deviation in experiments shown in Fig. 3.4C, whilst the pattern of expression in parent and daughter cells was consistent between replicates, actual quantification of miR-124a by qPCR in these populations was shown to vary considerably between experiments. It was thought that this may be attributable to differential transfection efficiencies between replicates, which could potentially be caused by the fact that in the system described above, all primary lymphocytes are CFSE-stained and transfected, but then only $CD4^+$ cells within this population are analysed. To attempt to improve the consistency of transfection between experiments, I first purified $CD4^+$ cells from total populations of primary lymphocytes (see Materials and Methods) before CFSE staining, activation, and transfection. By only transfecting a smaller, purified population of cells which were relevant to the assay, it was hoped that uniformity of transfection between experiments could be improved.

Next, I needed to identify the quantity of miR-142-3p mimic to transfect in order to reconstitute endogenous WT expression levels of the miRNA, and not result in significant overexpression. As a final concentration of 10μM miR-124a had been found to be reliably detectable by qPCR in undivided and divided cells, this concentration of oligo was initially used. Expression of miR-142-3p was analysed by qPCR in Dicer WT, KO and transfected cells during a timecourse of activation where KO cells were transfected 24 hours post-activation (Fig. 3.8A), to mimic experimental conditions in the mitotic inheritance assay described in previous sections.

Quantification of miR-142-3p in these populations indicates that up to 18 hours post-transfection, levels of miR-142-3p in transfected cells are reminiscent of endogenous WT levels (Fig. 3.8A), but by 24 hours post-transfection, levels of the exogenously introduced RNA are significantly decreased. These results indicate that the optimal point at which to analyse the effect of the miRNA mimic on target gene expression is 13-18 hours post-transfection.

Having optimised experimental conditions to maximise transfection efficiency and ensure that levels of exogenously introduced miRNA in Dicer KO cells reconstitute WT expression levels, the full assay was carried out. I isolated $CD4^+$ cells from 8-10 week old CD4Cre $Dicer^{lox/lox}$ mice, stained the cells with CFSE, and activated them for 24 hours. Activated cells were then transfected with 10μM

final concentration miR-142-3p mimic or relevant control and undivided and divided cells were sorted by CFSE dilution 18 hours later. I extracted long and short RNA from the same sorted populations using the miRVana RNA isolation kit, and levels of miR-142-3p and quantified target gene expression in each population by qPCR.

Analysis of miR-142-3p levels in sorted populations shows that consistent with previous assays, the quantity of miR-142-3p in divided cells is approximately half that detected in the undivided population, indicating that the miRNA is inherited through cell division from parent cells to progeny (Fig. 3.8B). Quantification of *Tgfbr1* and *Cfl2* mRNA levels in sorted populations of transfected cells shows that expression of these genes is significantly downregulated in both undivided and divided cells relative to untransfected controls (Fig. 3.8C). The expression of T cell expressed non-miR-142-3p targets *Cd69* and *Cd4* was not affected (Fig. 3.8D). Taken together, these data show that miRNAs are inherited through cell division in mammalian cells, and are capable of downregulating target gene expression in recipient progeny.

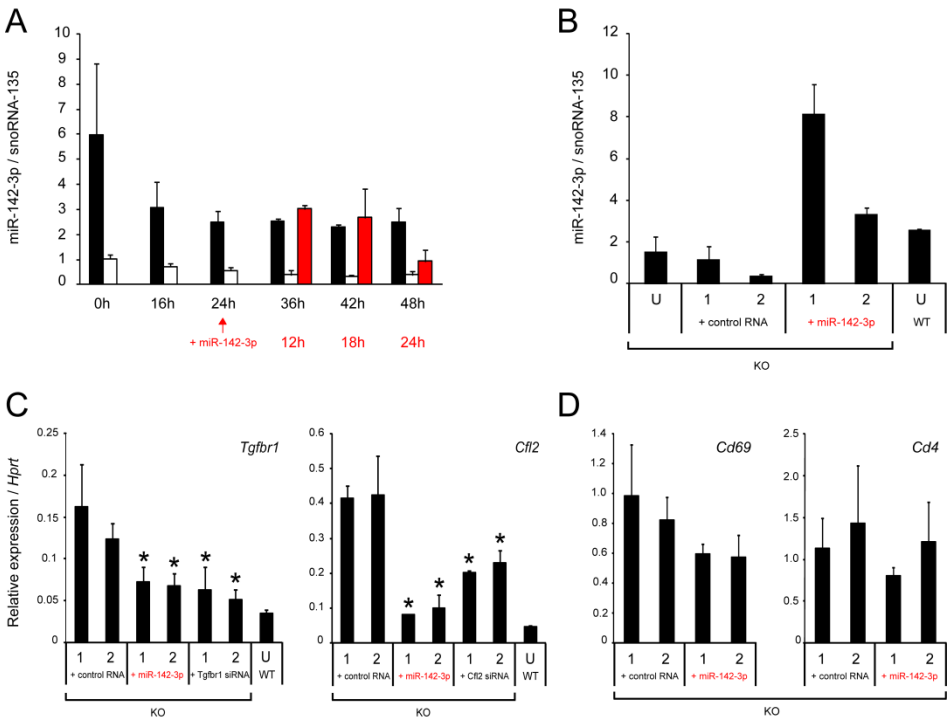



**Figure 3.8: T-cell specific miRNAs are inherited through mitosis in primary microRNA-deficient CD4$^+$ T cells and downregulate target gene expression in recipient cells.**
(A) qPCR analysis of miR-142-3p levels in CD4$^+$ Dicer WT (black), KO (white) and KO transfected with miR-142-3p (red – from 24 hours post-activation). Black numbers indicate hours post-activation; red numbers indicate hours post-transfection. (B) qPCR analysis of miR-142-3p expression in unsorted (U) and sorted (1 – undivided; 2 – divided) populations of CD4$^+$ WT/KO T cells (as indicated). (C) qPCR analysis of predicted miR-142-3p targets *Tgfbr1* and *Cfl2* in populations as described in (B). (D) qPCR analysis of control genes (non miR-142-3p targets) *Cd69* and *Cd4* in populations as described in (B). Error bars: s.t.d. of three biological replicates. Asterisks indicate significance (T-test: * = $p < 0.05$) of downregulation in cells transfected with miR-142-3p mimic or relevant siRNA relative to cells transfected with control RNA.

## 3.6 Discussion

Transgenerational and intercellular transmission of small noncoding RNAs has recently been demonstrated in several different studies in lower organisms. Experiments described in this chapter were designed to address whether a similar function could be attributed to noncoding RNAs in mammalian cells. Using dye-labelled primary mouse lymphocytes and fluorescently labelled synthetic RNAs, data presented here suggest that miRNAs are transmitted between cells via mitotic inheritance, and importantly, are functional in recipient cells.

### 3.6.1 Identification of potential novel targets of miR-142-3p in $CD4^+$ T cells

During optimisation of the assay described in this chapter, it was necessary to identify novel targets of miR-142-3p in $CD4^+$ T cells. Based on evolutionary conservation of miRNA binding sites in the 3'UTR of *Tgfbr1* and *Cfl2*, these genes are likely to be direct miR-142-3p targets; but this has not yet been formally addressed in publicly available literature. Downregulation of mRNA levels of these genes after transfection of a miR-142-3p mimic in Dicer KO cells provides strong evidence that *Tgfbr1* and *Cfl2* are novel direct targets of miR-142-3p (Fig. 3.8C). However, further analysis would be required to conclusively demonstrate this; for example by using the luciferase system, which identifies whether the 3'UTR of a given gene is sufficient to mediate miRNA-dependent repression (Gottwein et al., 2006).

It is interesting to consider why, in terms of biological relevance, *Tgfbr1* and *Cfl2* may be under miRNA mediated control in $CD4^+$ T cells. It is known that regulatory T cells ($T_{regs}$), of which there are several distinct subsets (Akbar et al., 2007), can develop from $CD4^+$ T cells either in the thymus or the periphery (Hsieh et al., 2012). Thymic development of $T_{reg}$ cells has been shown to be dependent on expression of *Tgfbr1* (Liu et al., 2008). Based on this knowledge, it is possible that miRNA-mediated repression of *Tgfbr1* in single positive $CD4^+$ T cells facilitates low-level expression of the gene, which could potentially facilitate rapid upregulation of expression if the cell receives a signal to induce $T_{reg}$ development.

The cofilin family of proteins play a fundamental role in regulation of actin dynamics in higher organisms (Bernstein and Bamburg, 2010b). Cofilin-1 (*Cfl1*) is ubiquitously expressed (Vartiainen et al., 2002), whereas cofilin-2 (*Cfl2*) is predominantly muscle-specific (Ono et al., 1994). Regulation of

*Cfl2* by T-cell specific miRNAs could therefore potentially reflect a mechanism to ensure correct lineage-specific expression of these genes, concurrent with a role for miRNA in conferring robustness to cell-specific transcriptional profiles. In addition, a potential function of miRNAs in actin regulation is discussed further in Appendix I, in response to the observation of distinct morphological defects in Dicer KO 3T3 fibroblasts.

### 3.6.2 Mitotic inheritance of miRNA in mammalian cells: evaluation of system

It should be noted that data presented in this chapter could be interpreted as evidence for partitioning of cytoplasmic contents during cell division, and therefore passive transmission of miRNAs from parent to daughter cell; whereas the concept of mitotic inheritance may imply a more active process. In either case, the ability of the transmitted miRNAs to target endogenous mRNAs in the progeny of transfected cells shows that miRNAs which are present in divided cells are functional, and therefore the process is biologically relevant.

A limitation of the system described in this chapter to investigate the mitotic inheritance of microRNAs is that the transmitted RNAs in question are synthetic, and exogenously introduced. However, in the absence of a technically feasible system to effectively tag endogenous miRNAs, this assay is as biologically relevant as currently possible. Tracking endogenous miRNAs is challenging due to the nature by which the majority of miRNAs are produced within the cell; designing a protein or RNA based tag which does not either inhibit processing by Drosha/Dicer, or is not removed during miRNA biogenesis as a result of the activity of these enzymes, is extremely difficult. However, since these experiments were carried out, technologies capable of tracking endogenous RNAs with RNA-based fluorescent tags have been developed (Paige et al., 2011, 2012). The assay described in this chapter could potentially be developed to utilise this system, by tagging endogenous miRNAs in the parental, undivided population and observing whether these RNAs are present in daughter cells. One potential caveat to this approach is whether binding of the tag affects binding (and therefore function) of the miRNA.

Another approach to provide further evidence for the biological relevance of mitotically inherited exogenously introduced miRNAs is to assess whether these RNAs are incorporated into the RISC complex in daughter cells. The recent development of techniques to simultaneously identify protein-miRNA and protein-mRNA interactions provide a powerful tool to convincingly detect

functional association of miRNAs with target genes (Chi et al., 2009), and would be extremely informative in terms of assigning even greater biological meaning to the assay described here.

However, in accordance with published data (Mantei et al., 2008), it was observed that unless stabilised synthetically, exogenously introduced double-stranded RNAs were rapidly degraded in activated T cells, suggesting that the miRNA mimics used in this study are subject to endogenous processing pathways. Whilst this result was an encouraging indication that the exogenously introduced RNAs were recognised as biologically relevant by transfected cells, it also generated a requirement to transfect high levels of the RNAs into parent cells in order to reliably quantify presence of the RNA in daughter cells; introducing the potential problem of cell-cell transmission as a result of cytoplasmic saturation. This issue was overcome by using qPCR as a highly sensitive method capable of quantifying low levels of unmodified miRNA mimics and target mRNA transcripts from the same population of cells. Taken together, these data indicate that as far as technically feasible, the assay described in this chapter accurately represents biologically relevant inheritance of microRNAs through cell division in mammalian cells.

In the next section of this study, the biological function of mitotically inherited microRNAs will be explored. If miRNAs are transmitted through cell division, is this an indication that noncoding RNAs play a specific role in early G1, throughout the cell cycle, or both? These questions are considered in more detail in the following chapter.

# 4.0 Analysis of gene expression through the cell cycle in microRNA-deficient mouse embryonic stem cells

## 4.1 Introduction

Mechanisms that establish and regulate gene expression are challenged by events that occur during the cell cycle. For example: during S phase, epigenetic modifications must be faithfully re-established on the replicated DNA molecule; and in mitosis, chromosome compaction results in exclusion of sequence-specific DNA binding proteins from chromatin, which must then re-associate with specific genomic sites in early G1. In the following chapter, experimental systems are described which are designed to explore the hypothesis that microRNAs could buffer lineage-specific gene expression through these transitions: firstly, by using cells arrested in mitosis and released into G1 to assess global gene expression in microRNA-deficient cells immediately post cell division; and secondly, using a drug-free cell cycle fractionation assay to analyse gene expression throughout the entire cell cycle. Using these systems, I aim to address the question of whether mitotic inheritance of microRNAs could play a role in stabilising transmission of lineage-specific gene expression from one cell to its progeny.

## 4.2 Analysis of gene expression in Dicer WT/KO ES cells in early G1

### 4.2.1 Optimisation of mitotic arrest and release assay

Initially, the G2/M to G1 transition was identified as an example of a transition between cell cycle stages which could potentially compromise regulation of lineage-specific transcriptional profiles. To interrogate the function of mitotically inherited miRNAs through this transition, an assay was required to isolate populations of cells immediately post-mitosis.

Whilst the T cell system offers the advantage of being able to dissociate parent and daughter cell populations, it is technically challenging to isolate T cells in distinct stages of the cell cycle as even in activated cells, the majority of the population is in G1 (Fig. 4.1). In contrast, due to the high proliferation rate of pluripotent cells, exponentially growing populations of ES cells have approximately equivalent proportions of cells distributed through each stage of the cell cycle (Fig. 4.1), and therefore isolation of cell populations from each phase is relatively more straightforward. As a result, and due to

the availability of previously derived Dicer KO ES cells (Nesterova et al., 2008) the analysis of gene expression through the cell cycle described in this chapter is carried out in ES cells. The 'wildtype' cells used in this study are the parental ERT2Cre Dicer$^{lox/lox}$ cells from which the Dicer$^{\Delta/\Delta}$ (KO) cells were derived (Nesterova et al., 2008).

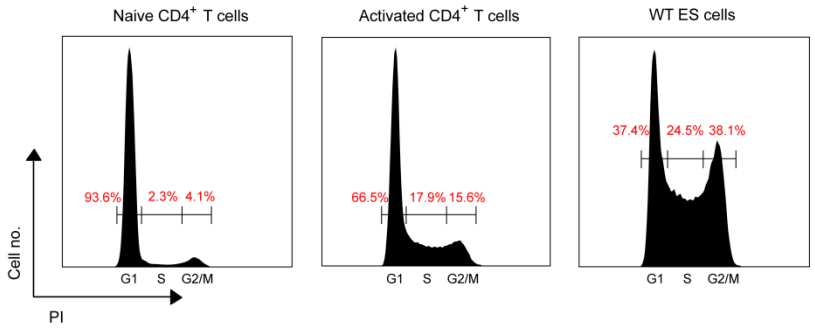



**Figure 4.1: Differential cell cycle profiles of T cells and ES cells.** FACS analysis of cycling populations of asynchronous CD4$^+$ T cells and ES cells. Intensity of propidium iodide is proportional to DNA content, and therefore can be used to generate a distribution indicative of the proportion of the total population in each stage of the cell cycle. Red numbers show the percentage of cells in each stage of the cell cycle as indicated below the graphs.

One method of obtaining cell populations immediately post-cell division is to isolate a population of mitotic cells, and then allow this population to progress into early G1. Asynchronously growing cells can be arrested in mitosis using drugs which inhibit microtubule formation, such as nocodazole and demecolcine (Jackman and O'Conner, 2001). This method has the advantage of generating populations highly enriched for cells in G2-M, but treating cells with drugs often results in off-target effects. Alternatively, a drug-free method of obtaining mitotic populations of adherent cells is known as the mitotic shakeoff (Terasima and Tolmach, 1963; Fox, 2003), which utilises the characteristic rounded shape that adherent cells assume when in G2-M. Due to this morphological change, adherent cells in mitosis are less robustly attached to the culture surface and therefore will detach on gentle agitation of the vessel. Whilst this method avoids potential secondary effects of drug-mediated mitotic arrest, extremely low numbers of mitotic cells are obtained per experiment, which can make reliable analysis of these populations challenging.

Both protocols described above are essentially only applicable to adherent cells, including embryonic stem cells. However, the wildtype and microRNA-deficient ES cell lines derived by Nesterova *et al* are feeder-dependent (Nesterova et al., 2008). The presence of a layer of fibroblasts in the culture vessel prohibits the use of either drug-mediated mitotic arrest or the mitotic shakeoff protocols on these cell lines, as it would be difficult to dissociate the feeder cells from the mitotic population of ES cells.

To address this issue, I derived feeder-independent lines of Dicer WT and KO ES cells from the original lines described by Nesterova *et al* (Fig. 4.2). This was achieved by adapting the cells to grow in N2B27 + 2i media (Ying et al., 2008) where '2i' refers to two chemical inhibitors of the mitogen-activated protein kinase (MAPK) and glycogen-synthase kinase-3 (GSK3) signalling pathways. It has been shown that ES cells can grow in this 'defined' media without feeder cells and maintain the pluripotent state (Ying et al., 2008).

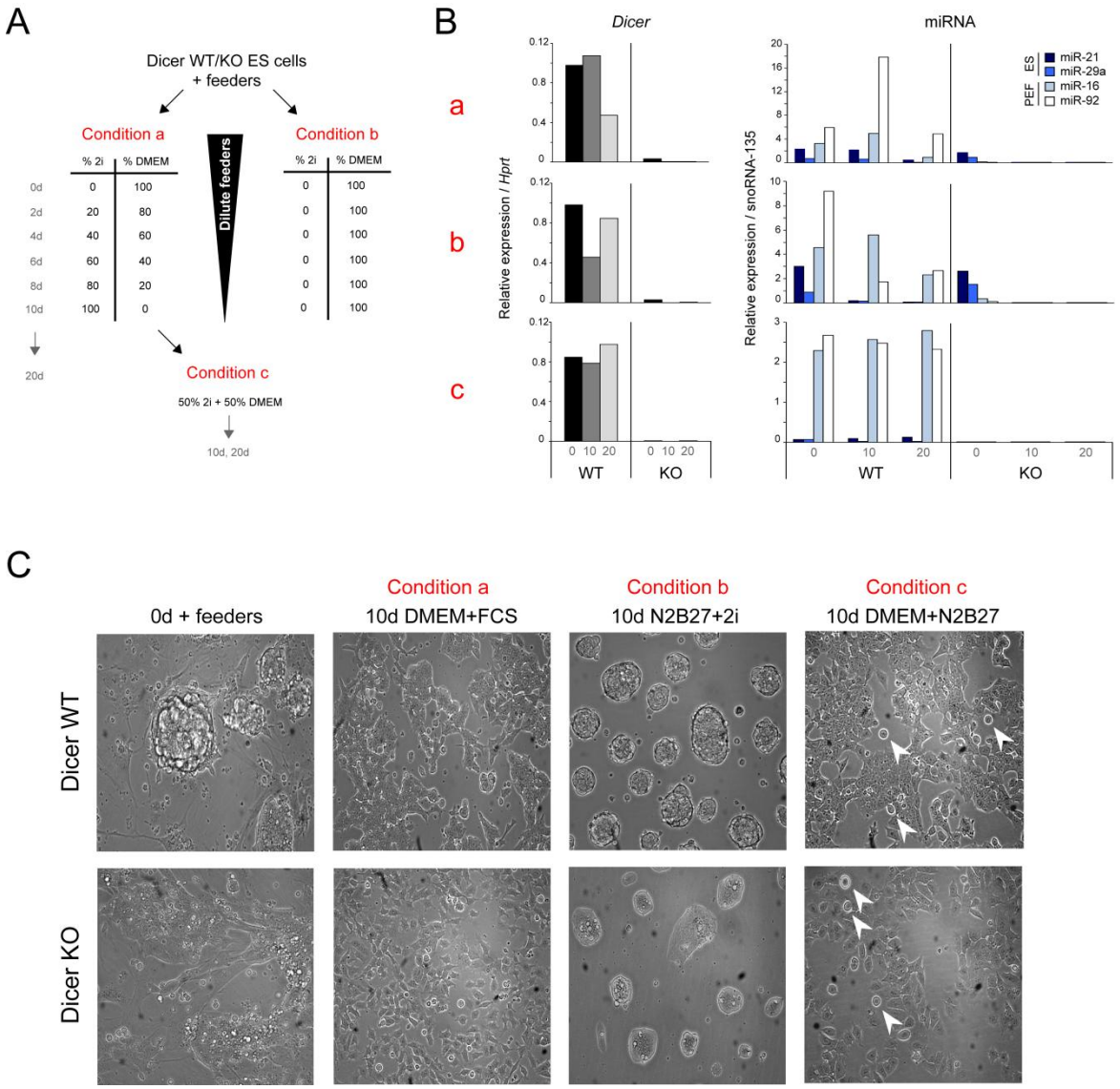



**Figure 4.2: Derivation of feeder-independent wildtype and microRNA-deficient mouse embryonic stem cells.**
(A) Schematic of protocol used to adapt feeder-dependent Dicer WT/KO ES cells to feeder-independent culture conditions. In Condition a: feeders were progressively diluted out per passage every 2 days, concurrent with a proportional increase in the ratio of N2B27+2i : DMEM+FCS. In Condition b: feeders were again diluted out per passage, but cells remained in culture with DMEM + FCS only. After 10 days in mixed media in both conditions during feeder depletion, cells were cultured for a further 10 days in each media alone without feeders before analysis. In Condition c: feeder-independent cells from Condition a were cultured in 50 : 50 N2B27+2i : DMEM + FCS for 20 days. Cells maintained in all conditions were cultured in the presence of LIF. (B) Verification of Dicer knockout and miRNA depletion in feeder-independent conditions. qPCR analysis of Dicer mRNA levels (left-hand panel) and miRNA expression (right-hand panel) in each condition after 0 days, 10 days and 20 days in culture as indicated. (C) Brightfield images illustrating morphological changes in cells cultured for 10 days in each condition described in (A). White arrows (Condition C: WT and KO) indicate mitotic cells characterised by rounded morphology.

However, it became apparent that when cultured in N2B27+2i, Dicer WT and KO cells form extremely rounded colonies which readily detach from the culture vessel (Fig. 4.2C). Whilst not necessarily an issue for standard culture and propogation of these lines, the protocols described above for isolation of mitotic populations both required the cells in question to be adherent. I observed during the process of derivation (progressive dilution of feeder cells concurrent with increasing amounts of 2i – Fig. 4.2A) that when cells were cultured in roughly proportional quantities of DMEM+FCS (standard ES cell culture media) and N2B27+2i, they formed colonies which were flatter and more like a monolayer than the highly rounded colonies seen in 2i conditions alone (Fig. 4.2C). As these conditions represent a morphology which would be more amenable to the mitotic arrest protocols, cells used in further experiments were cultured in 50% 2i and 50% DMEM.

Before optimising the mitotic arrest assay in these cells, it was first necessary to verify that the changes in culture conditions described above do not compromise the pluripotent state. To assess this, I analysed expression of key pluripotency markers *Oct4* and *Nanog* in both Dicer WT and KO ES cell lines after 10 and 20 days in the culture conditions illustrated in Figure 4.2. In addition, I also quantified expression of *Gata4* and *Sox1* in these cells, as expression of these genes is one of the early events in differentiation of ES cells to mesoderm (Kuo et al., 1997; Grépin et al., 1995) and ectoderm (Pevny et al., 1998) respectively. Upregulation of these genes is generally accepted to be an indication of *in vitro* differentiation in ES cell cultures.

It was observed that whilst expression of *Oct4* and *Nanog* was decreased in both Dicer WT and KO ES cells cultured in DMEM only without feeders, expression of these genes was maintained or in some instances increased in cells cultured in 2i only or the combined media (Fig. 4.3). This data suggests that without feeders, ES cells grown in DMEM begin to differentiate, despite the continued presence of LIF in the culture medium. This notion is supported by the fluctuation and slight increase in *Gata4* and *Sox1* in WT and KO cells grown in DMEM only. Cells cultured in 2i only or the combined media maintained low levels of expression of both differentiation markers, even after 20 days in culture (Fig. 4.3). These results indicate that culture of Dicer WT and KO ES cells in both N2B27+2i alone and the combined N2B27+DMEM media without feeders maintains the pluripotent state, whilst cells cultured in feeder-independent conditions in DMEM + FCS begin to differentiate.

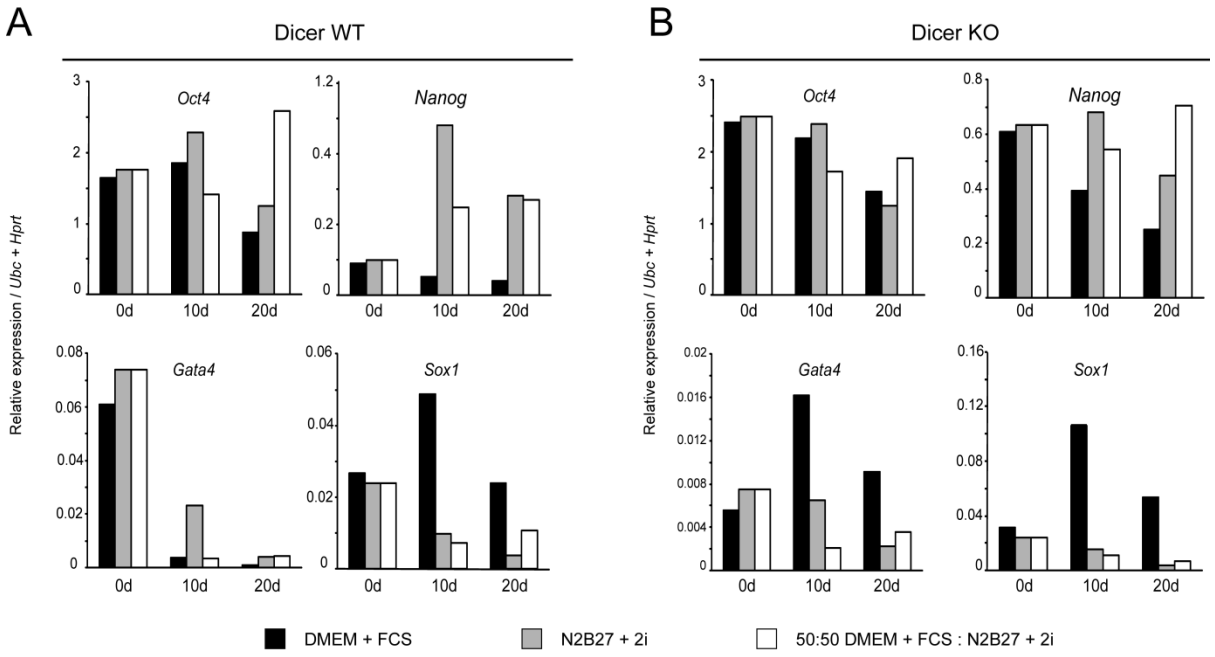



**Figure 4.3: Culture of Dicer WT and KO ES cells in feeder-independent conditions in combined media does not compromise the pluripotent state.**
qPCR analysis of pluripotency markers *Oct4* and *Nanog* and differentiation markers *Gata4* and *Sox1* in Dicer WT (A) and KO (B) ES cells in culture conditions described in Fig. 4.2.

Having established Dicer WT and KO ES cell lines with growth characteristics amenable to protocols for obtaining mitotic populations of cells, I then optimised a drug-mediated mitotic arrest assay in these lines to generate populations of cells for analysis of gene expression immediately post-cell division. Cells were cultured in the presence of 20ng/ml demecolcine for four hours, and then flasks were agitated gently to release mitotic cells. These cells were then washed in cold PBS, re-plated in demecolcine-free medium and put back in culture for three hours. Cells were collected at 30 minute intervals to assess progression from G2/M to G1.

Whilst both Dicer WT and KO lines yielded populations of cells highly enriched in G2-M after demecolcine-mediated arrest and mitotic shakeoff, the two populations showed different kinetics of release from mitosis to G1. After 2 hours in culture the percentage of cells in G1 in Dicer WT ES cells did not increase significantly above 58% (Fig. 4.4). If kept in culture for longer than 3 hours, the released population began to progress into S phase and eventually resumed a cell cycle profile reminiscent of an asynchronously growing culture (data not shown). In comparison, Dicer KO cells initially progressed slower from G2/M to G1 (21% at 1 hour release compared to 41% in WT) but then a greater proportion of cells was obtained in G1 after 2 hours (66% compared to 58%) and this difference increased by 3 hours release (78% compared to 59%). Whilst up to 20% of cells remained

in mitosis after 3 hours in culture, this population of cells is likely to be predominantly transcriptionally inactive (Hartl et al., 1993; Gottesfeld and Forbes, 1997) and therefore it was reasoned that analysis of G1-specific gene expression in populations of released cells is unlikely to be affected significantly by this residual population of mitotic cells.

It was additionally noted that the distinct cell cycle profile of asynchronous Dicer KO ES cells, characterised by accumulation of these cells in G1, is consistent with published findings by Wang *et al* (see 1.3.1.1).

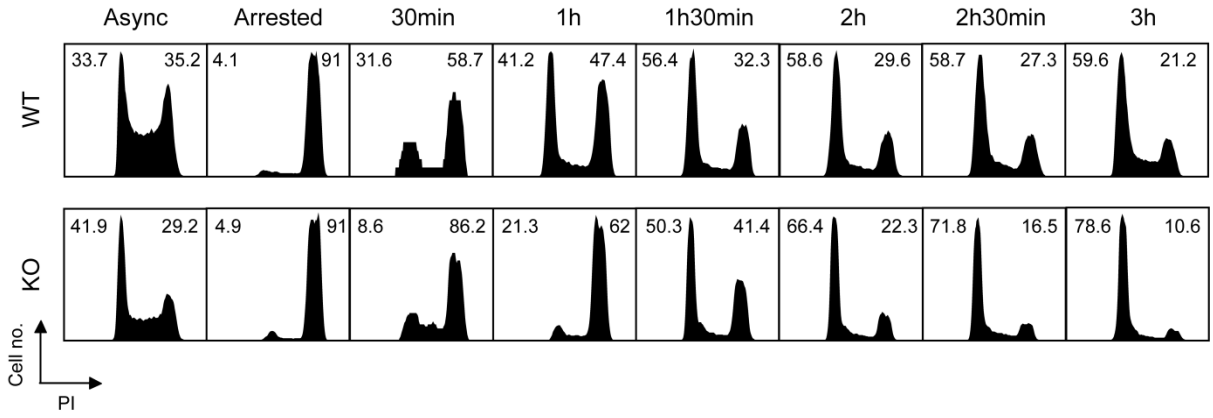



**Figure 4.4: Mitotic arrest and release yields populations of Dicer WT and KO ES cells in early G1.**
PI (propidium iodide) staining of populations obtained in mitotic arrest and release assay. Async – asynchronous cells. Arrested – cells treated for 4 hours with 20ng/ml demecolcine, followed by mitotic shakeoff. Timepoints – time post-release and culture of arrested cells in demecolcine-free media. Per graph: top left figure = % cells in G1; top right figure = % cells in G2/M. Gates not shown for clarity.

### 4.2.2 Genome-wide analysis of gene expression in asynchronous populations of Dicer WT/KO ES cells and in early G1

To ask whether lack of microRNAs through the G2-M to G1 transition results in deregulation of microRNA targets specifically in early G1, I analysed genome-wide gene expression changes in Dicer WT and KO ES cells immediately post-cell division. I prepared RNA samples for microarray analysis from three independent biological replicates of the mitotic arrest and release assay, using asynchronous populations as a control and cells two hours post-release from mitotic arrest as a population representative of cells in early G1 (Fig. 4.4). I considered that this timepoint represented a compromise between maximising the enrichment of released cells in G1 whilst simultaneously not allowing the cells to progress too far into the cell cycle.

Hybridisation and Principal Component Analysis (PCA) of the samples was carried out in the MRC CSC Genomics Laboratory. PCA clusters microarray replicates to illustrate the greatest sources of variance in the data, based on probe values and independent of sample identity. PCA is therefore a preliminary, unbiased visualisation of whether a) biological replicates are truly similar to each other and b) of whether there is a difference between experimental conditions. PCA analysis of either asynchronous or early G1 populations of Dicer WT and KO ES cells indicates that the primary sources of variance between samples are attributable to differences between the cell types (Dicer WT v KO - PCA #1), and to the difference in cell cycle stage (cell cycle – PCA #2) as anticipated. The plot also indicates that in the majority of cases, biological replicates of each condition cluster together, which suggests that the experimental replicates are reproducible and therefore likely to be of reliable quality. One sample, however (WT G1 Replicate 1), is a clear outlier (Fig. 4.5) which after re-assessing RNA quality post-extraction, was probably due to contamination of the RNA sample with phenol before labelling. This sample was excluded from further analysis. In general, clustering of samples using PCA indicates that there are differences in expression between Dicer WT and KO ES cells, in early G1 as well as in an asynchronous population.

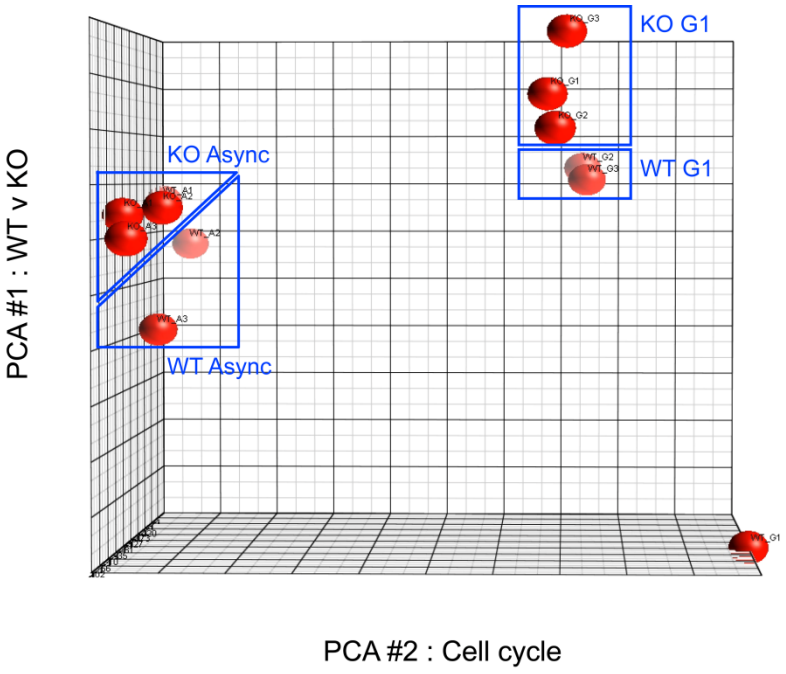



**Figure 4.5: Prinicipal Component Analysis (PCA) of Dicer WT/KO Async/G1 microarray samples.**
Visual representation of the most significant sources of variation in microarray data for Dicer WT/KO asynchronous/early G1 samples. Samples cluster according to presence or absence of miRNAs, and cell cycle stage. Clustering of biological replicates is delineated by blue boxes.

Initial processing of the raw data was carried out by Dr. Tom Carroll (MRC CSC, currently at CRUK, Cambridge Research Unit). I then analysed the data further to summarise general trends in changes in gene expression. Overall, a very similar number of genes were significantly deregulated (adjusted p-value < 0.05) in Dicer KO ES cells in an asynchronous population and in early G1 (Table 2). In terms of the number of genes deregulated in the asynchronous condition, these data are consistent with previously published microarray analysis in independently derived asynchronous Dicer WT/KO ES cells (Sinkonnen et al., 2008). A slightly greater percentage of genes were upregulated in Dicer KO cells in early G1 than in an asynchronous population (Table 2).

| |  Async |  Early G1 |
|---|---|---|
| % Deregulated genes | 15.9 | 16.3 |
| % Upregulated genes | 7.8 | 9.4 |
| % Downregulated genes | 8.1 | 6.9 |

**Table 2: Summary of Dicer KO vs WT microarray data in asynchronous cells and in early G1.**

It has previously been shown that the effect of a miRNA on expression level of its target mRNA is typically subtle (Lim et al., 2005; Baek et al., 2008), although there are rare examples of more dramatic miRNA-mediated changes in gene expression (Lee et al., 1993; Reinhart et al., 2000). Consistent with these findings, overall analysis of the fold change in expression in Dicer KO compared to WT cells in both cell cycle conditions showed that very few genes have an actual fold change of greater than 2 (log2FC of 1: black line on graph) (Fig. 4.6A). These general trends were still observed even when only those genes which changed significantly (adjusted p < 0.05) in Dicer KO cells were considered. The graphical representations in Fig. 4.6 also illustrate that most genes which are deregulated in Dicer KO cells change to a similar extent in early G1 as in an asynchronous population, and in the case of genes which are significantly deregulated in KO cells (Fig. 4.6B), always in the same direction.

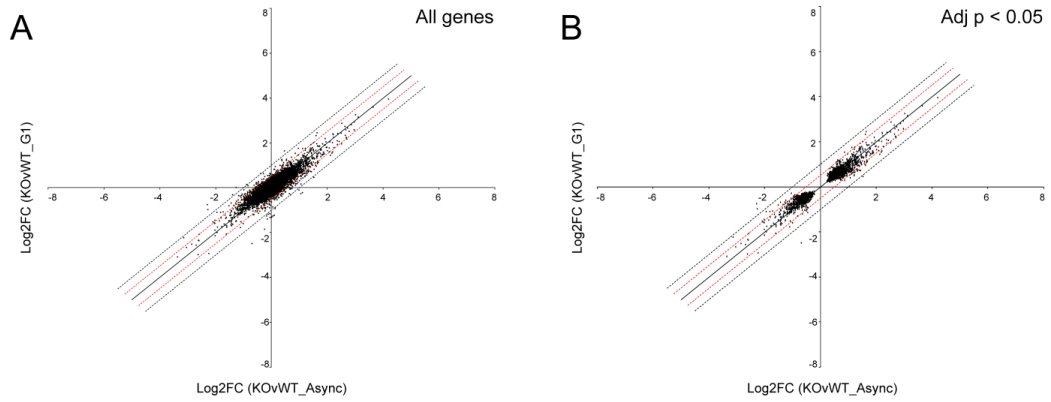



**Figure 4.6: Gene expression changes in Dicer KO cells are typically less than two-fold and similar in early G1 and asynchronous populations.**
(A) log2FC of Dicer KO compared to WT in asynchronous cells plotted against log2FC in Dicer KO compared to WT in early G1 – values for all genes in microarray dataset are shown. (B) As in (A), but only genes which are significantly deregulated in Dicer KO cells (adjusted p value < 0.05) are shown. In both graphs: black line = log2FC of 1 (actual fold change of 2); red line = log2FC of 0.5 (actual fold change of 1: genes below this line change to the same extent in both cell cycle conditions).

To ask whether the deregulated genes in Dicer KO cells are direct miRNA targets, the miReduce algorithm (described in 3.5) was applied to microarray data to identify motifs in the 3'UTR of genes which are significantly associated with either a positive or negative fold change in expression (analysis was carried out by A. Terry at the MRC Clinical Sciences Centre). This analysis identified a highly significant association of the gcacttt and gcactta motifs with upregulated genes in both asynchronous Dicer KO cells (Fig. 4.7A) and in early G1 (Fig. 4.7B). These motifs correspond to the binding sites for the miR-290 and miR-302 families respectively. The binding sites for these two families only differ by one base, and these miRNAs are the most highly expressed miRNAs in mouse (Marson et al., 2008) and human (Suh et al., 2004) embryonic stem cells respectively. Taken together with the subtle changes in expression of the majority of deregulated transcripts, this data suggests many of the genes upregulated in Dicer KO ES cells in both asynchronous cells and in early G1 are likely to be direct miRNA targets.

A

Async

| 3' UTR motif | p-value | No. of genes | Avg. expr. in Dicer KO | microRNA family |
|---|---|---|---|---|
| **gcacttt** | 0 | 1928 | 0.126 | miR-290 |
| **gcactta** | 0 | 1097 | 0.120 | miR-302 |
| **ctgagcc** | 0 | 2674 | 0.005 | miR-24 |
| **ttgtaaa** | $4.7 \times 10^{-10}$ | 3209 | 0.076 | — |
| **acactac** | $3.6 \times 10^{-7}$ | 1040 | 0.009 | miR-92 |

B

Early G1

| 3' UTR motif | p-value | No. of genes | Avg. expr. in Dicer KO | microRNA family |
|---|---|---|---|---|
| **gcacttt** | 0 | 1928 | 0.124 | miR-290 |
| **gcactta** | 0 | 1097 | 0.128 | miR-302 |
| **taaatca** | $1.9 \times 10^{-10}$ | 1668 | 0.088 | — |
| **attatgt** | $2.2 \times 10^{-6}$ | 1714 | 0.081 | — |
| **gtgctgg** | $2.0 \times 10^{-6}$ | 2757 | 0.021 | — |

**Figure 4.7: Motifs for ES-specific microRNA families are significantly associated with upregulated genes in Dicer KO ES cells.**
(A) Top five motifs located in the 3'UTR of upregulated genes in Dicer KO ES cells in an asynchronous population. (B) Top five motifs located in the 3'UTR of upregulated genes in Dicer KO ES cells in early G1.

#### 4.2.2.1 Microarray Analysis (1): G1-specific deregulation of histone genes in Dicer KO ES cells

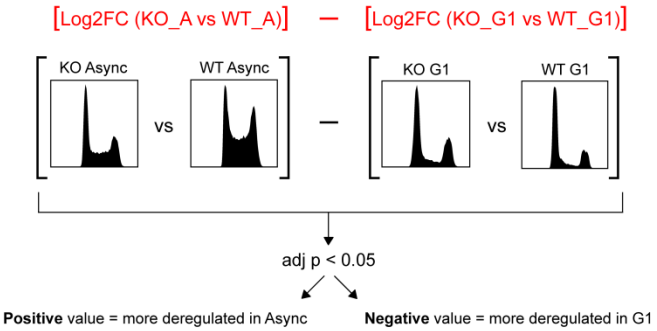
Next, in order to investigate a potential role for mitotically inherited miRNAs in regulation of gene expression immediately post-cell division, the microarray data were analysed in greater depth to assess whether any genes were deregulated specifically in Dicer KO ES cells in early G1. The four expression datasets are illustrated in Fig. 4.8A: asynchronous Dicer WT ES cells, asynchronous Dicer KO ES cells, early G1 Dicer WT ES cells, and early G1 Dicer KO ES cells. Changes in expression in asynchronous Dicer KO ES cells were identified by comparing this expression dataset to expression data in asynchronous control (WT) cells. The same approach was applied to identify expression changes in Dicer KO ES cells in early G1. The significance of these changes was assessed using a student's t-test and adjusted for multiple testing using the Bonferroni correction method. Only genes which were significantly deregulated (adjusted p-value < 0.05) were considered in further analyses.

Next, to identify genes with a significantly greater log2FC in early G1 compared to asynchronous cells, the value of log2FC of each significantly deregulated gene in Dicer KO ES cells in early G1 was subtracted from the log2FC in expression of the gene in an asynchronous population. If the log2FC of the gene in question was greater in G1, it would therefore yield a negative value using this approach. The significance of the difference in log2FC between asynchronous and early G1 Dicer KO ES cells was quantified using a student's t-test. This analysis was carried out by Dr. Tom Carroll.

Using this approach, only 12 genes were identified which changed specifically in Dicer KO cells in early G1 (adjusted p-value < 0.05), which were all transcripts for histone proteins (Fig. 8B). An additional 9 histone genes were also significantly deregulated in Dicer KO cells in G1 at the non-adjusted p-value level ($p < 0.05$). I validated expression of histone genes deregulated at both levels of statistical significance by qPCR (Fig. 4.8C). All qPCR validation of microarray data in this study was undertaken in three biological replicates of the mitotic arrest and release assay performed independently of experiments used for microarray analysis. This analysis confirms that transcripts from histone genes are significantly downregulated in Dicer KO ES cells specifically in early G1. The fact that this trend could be validated independently of the level of statistical significance in the arrays

(e.g. adjusted or non-adjusted p-value) illustrates that even subtle changes in expression identified in the microarray data can be confirmed.

A




B

| Gene | Difference in FC (calculated as in A) | Adjusted p-value |
|---|---|---|
| Hist2h2aa1 | -1.02 | 0.003 |
| Hist1h2af | -1.03 | 0.003 |
| Hist1h2ao | -1.06 | 0.003 |
| Hist1h3f | -1.32 | 0.003 |
| Hist1h1b | -1.80 | 0.003 |
| Hist2h4 | -1.06 | 0.003 |
| Hist2h3c1 | -1.33 | 0.004 |
| Hist1h2bq | -1.30 | 0.004 |
| Hist1h3g | -1.28 | 0.005 |
| Hist1h1a | -1.60 | 0.005 |
| Hist1h2be | -1.25 | 0.012 |
| Hist1h2bm | -1.10 | 0.013 |

C

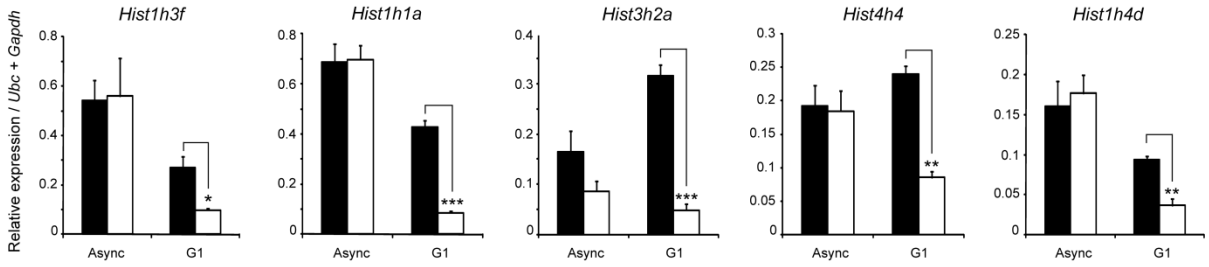



**Figure 4.8: Histone genes are significantly downregulated in Dicer KO cells specifically in early G1.**
(A) Approach used to identify genes specifically deregulated in Dicer KO cells in early G1. (B) List of 12 genes with a significantly greater fold change in Dicer KO cells in G1 compared to an asynchronous population. Values shown are calculated from probe values in array data. (C) qPCR validation of expression of histone genes in control (black) and KO (white) ES cells in an asynchronous population or in early G1. Error bars: s.t.d. of three independent biological replicates (see footnote). Asterisks indicate significance (T-test: * = $p < 0.05$; ** = $p < 0.005$; *** = $p < 0.0005$).

As histone gene expression is known to be regulated co-ordinately (Marzluff et al., 2008), and multiple transcripts are deregulated simultaneously (Fig. 4.8), it is likely that instead of these transcripts being direct miRNA targets themselves, an upstream regulatory mechanism is under miRNA-mediated control. In support of this, histone genes have a unique mRNA structure, characterised by a 3' stem loop in place of a canonical polyadenylated 3'UTR, and therefore are also likely to lack conventional miRNA binding sites. Deregulation of histone genes is therefore probably a secondary effect of loss of miRNAs (discussed further in 4.2.3.1).

#### 4.2.2.2 Microarray Analysis (2): Deregulation of bivalent genes in Dicer KO ES cells

Given the typically subtle effect of a given miRNA on a specific target mRNA, the approach described in Fig. 4.8A (in attempting to identify highly significant differences in fold change between conditions) was perhaps not the most appropriate when considering the expected phenotypic effects of loss of miRNA activity. As a result, I next used a different method of analysis to further probe the microarray data and attempt to identify G1-specific deregulated genes which are directly regulated by microRNAs.

Gene Set Enrichment Analysis (GSEA) is an algorithm designed to associate biological meaning to changes in expression from large datasets (Mootha et al., 2003; Subramanian et al., 2005), and is particularly applicable to differences which might not necessarily be statistically significant under usual parameters for analysis of microarray data. Using GSEA, all genes in the dataset are ranked by fold change (irrespective of the statistical significance of the fold change itself), and the GSEA algorithm asks whether a defined gene set is more significantly associated with genes which are upregulated or downregulated in the condition of interest. This is extremely useful for analysis of miRNA-mediated changes, as fold changes which are not necessarily statistically significant between Dicer WT and KO are those which are most likely to be directly mediated by microRNAs.

To analyse the microarray data discussed above using GSEA, it was first necessary to select gene sets of interest with which to interrogate the expression data. A list of predicted targets of the miR-290 and miR-302 families were used as a positive control for genes which should be associated with upregulated genes in Dicer KO ES cells compared to control; and the list of histone genes shown in Fig. 4.8B was used as a control for genes which should be downregulated. Then, a selection of gene sets associated with differentiation of specific lineages or terminally differentiated cell types were chosen (gene lists were obtained from GO term datasets) to ask whether genes associated with non-ES lineages were upregulated in Dicer KO cells, either globally (asynchronous cells) or in a cell cycle-dependent manner (early G1).

In addition to these gene sets, lists of genes marked by a bivalent chromatin signature were also included (Ku et al., 2008). Bivalent genes are conventionally defined by co-occurrence of both

trimethylation of lysine 4 on histone 3 (H3K4me3) and also trimethylation of lysine 27 on the same histone (H3K27me3) in the promoter region of the gene (Azuara et al., 2006; Bernstein et al., 2006). Genes marked by a bivalent chromatin signature are known to be key developmental regulators, and are some of the first genes to be expressed when an ES cell receives cues to differentiate. The presence of RNA polymerase II at the promoters of these genes (Stock et al., 2007) and evidence for low-level transcription despite the presence of H3K27me3 (De Gobbi et al., 2011; Brookes et al., 2012) suggests that bivalent genes are repressed by less stringent mechanisms than other genes associated with functions of terminally differentiated cells. It was hypothesised that therefore this set of genes may potentially be more at risk of being aberrantly transcribed than the non-ES lineage-inappropriate gene sets described in the paragraph above.

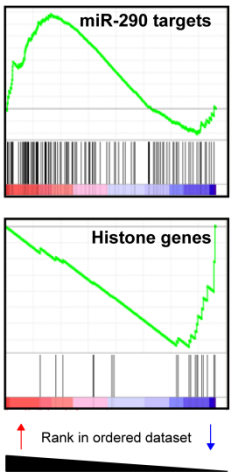
I ran the GSEA program on gene lists ranked by fold change in Dicer KO ES cells compared to control in both an asynchronous population and in early G1, using the gene sets described above. Figure 4.9A shows examples of GSEA output graphs illustrating the association of miR-290 targets and histone gene sets with upregulated and downregulated genes in Dicer KO ES cells compared to control. The gene list ranked by fold change in expression is represented along the x-axis; black bars represent where individual genes in a defined gene set (e.g. miR-290 targets) fall within the list of ranked genes. The false discovery rate (FDR) is used as an indication of the statistical significance of the association of a given gene set with either upregulated or downregulated genes. An FDR of less than 0.05 indicates that the association is likely to be identified in at least 95% of instances.

As anticipated, predicted targets of the miR-290/302 families were highly significantly associated with upregulated genes in both asynchronous cells and in early G1, and histone genes were significantly associated with downregulated genes only in Dicer KO cells in G1 (Fig 4.9B). These results indicate that GSEA algorithm is reliable and can detect genuine associations between gene sets and changes in gene expression.

None of the gene sets related to terminally differentiated cell types were statistically significantly associated with either upregulated or downregulated genes in Dicer KO cells, either in an asynchronous population or in early G1. However, genes marked by a bivalent chromatin signature (H3K4me3 + H3K27me3) were highly significantly associated with upregulated genes in asynchronous Dicer KO ES cells. Bivalent genes were also associated with upregulated genes in

Dicer KO ES cells in early G1, but to a less significant extent than in an asynchronous population (FDR < 20%).

A




B

| | ↑ in Dicer KO | | ↓ in Dicer KO | |
|---|---|---|---|---|
| | Async | G1 | Async | G1 |
| miR-290 targets | 0.00 | 0.00 | | |
| miR-302 targets | 0.02 | 0.05 | | |
| Histone genes | | | 0.12 | 0.00 |
| Bivalent genes (Ku 2008) | 0.02 | 0.19 | | |
| Immune system development | | | 1.00 | 0.96 |
| Synaptogenesis | | | 1.00 | 0.92 |
| Epithelial cell development | 0.14 | 0.07 | | |
| Muscle differentiation | 0.17 | 0.17 | | |
| Heart development | 0.33 | 0.30 | | |
| Axonogenesis | 0.45 | 0.71 | | |
| Brain development | 0.62 | 0.90 | | |
| Haemopoiesis | 0.97 | | | 0.78 |

**Figure 4.9: GSEA analysis of microarray data identifies global upregulation of bivalent genes in Dicer KO ES cells.**
(A) Representative GSEA output graphs for upregulated genes in Dicer KO ES cells (miR-290 targets) and downregulated genes (histone transcripts). (B) Summary of GSEA output assigning FDR values (shown) for association of defined gene set (left of table) with upregulated or downregulated genes in Dicer KO ES cells in given cell cycle conditions (top of table).

I validated the expression of a subset of bivalent genes in Dicer WT and KO ES cells by qPCR in both asynchronous cells and populations in early G1. Expression of previously validated miR-290 targets *Cdkn1a (p21), Rbl2* and *Lats2* (Wang et al., 2008) was analysed as a positive control; and quantification of non-miRNA targets *Oct4* and *Eed* used to assess secondary or non-specific effects of loss of miRNA activity in the Dicer KO cells. These data confirmed that bivalent genes are upregulated in Dicer KO ES cells in both an asynchronous population and in early G1.

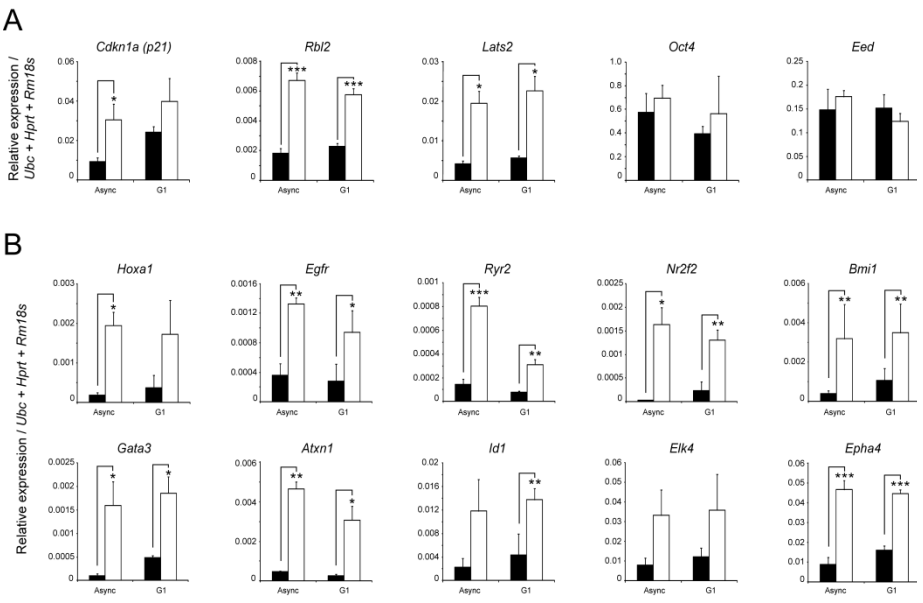



**Figure 4.10: Validation of microarray data by qPCR confirms upregulation of bivalent genes in Dicer KO ES cells.**
(A) qPCR analysis of validated miR-290 targets (*p21, Rbl2, Lats2*) and non-miRNA targets (*Oct4, Eed*) in Dicer WT (black) and KO (white) ES cell populations as indicated. (B) qPCR analysis of bivalent genes in Dicer WT/KO ES cell populations as indicated. Error bars: s.t.d. of three independent biological replicates. Asterisks indicate significance (T-test: * = p < 0.05; ** = p < 0.005; *** = p < 0.0005).

Whilst transcripts from bivalent genes are predicted to be miR-290 targets by both TargetScan and miReduce algorithms, I next sought to ask whether these genes are in fact direct targets of the miR-290 family. Until recently, identification of direct targets of miRNAs was challenging due to functional redundancy between miRNA family members with the same seed sequence. However, the recent development of tiny locked nucleic acid (LNA) inhibitors has facilitated inhibition of all family members simultaneously (Obad et al., 2011). These molecules bind specifically to the 7-8 nucleotide seed sequence of all members of a given miRNA family, providing a more informative and reliable indication of direct targets of miRNAs of interest.

Therefore, to identify whether bivalent genes are in fact direct targets of the miR-290 family in ES cells, I introduced an LNA inhibitor of the miR-290 family into Dicer WT ES cells (see Materials and Methods). If the genes in question are direct targets of these miRNAs, it would be expected that these transcripts would be upregulated in WT cells in which the activity of the miR-290 family has been specifically inhibited.

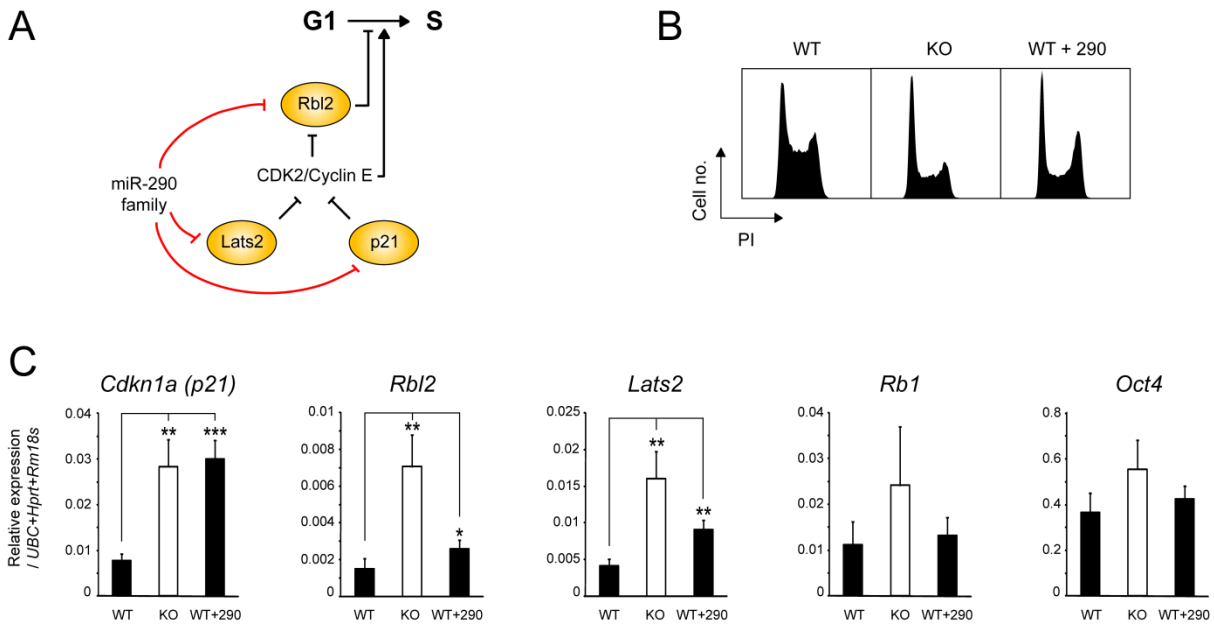



**Figure 4.11: Simultaneous inhibition of all members of miR-290 family in WT ES cells recapitulates cell cycle defect in Dicer KO ES cells and results in upregulation of validated miR-290 target genes.**
(A) Schematic adapted from (Wang et al., 2008) illustrating the role of the miR-290 family in regulation of the cell cycle in ES cells. (B) FACs analysis of cell cycle profiles of PI (propidium iodide) stained populations of asynchronous Dicer WT and KO ES cells, and WT ES cells transfected for 24 hours with a final concentration of 0.5μM LNA inhibitor of the miR-290 family (WT + 290). (C) qPCR analysis of validated miR-290 targets (*p21, Rbl2, Lats2*) and non-miR-290 targets (*Rb1, Oct4*) in cells transfected with control LNA (WT and KO) or 0.5μM miR-290 inhibitor (WT + 290) . Error bars: s.t.d. of three independent biological replicates. Asterisks indicate significance (T-test: * = $p < 0.05$; ** = $p < 0.005$; *** = $p < 0.0005$).

Previous studies (Wang et al., 2008) have inhibited individual members of the miR-290 family in WT ES cells and, using this strategy, identified key cell cycle regulators under the control of these RNAs (Fig. 4.11A). However, inhibition of specific members of the miRNA family independently did not recapitulate the characteristic cell cycle defect demonstrated by Dicer KO ES cells. The authors speculated that this was attributable to functional redundancy between family members. In support of this, transfection of WT ES cell lines used in this study with LNA oligos that simultaneously inhibit the entire miR-290 family successfully recapitulated a cell cycle profile in transfected WT cells that was reminiscent of KO cells (Fig. 4.11B). Upregulation of *p21* and *Rbl2* expression in LNA-transfected WT cells (and no effect on non-miR-290 target *Oct4*) provided additional confirmation that the LNAs were successfully inhibiting the miR-290 family.

I next asked whether loss of function of these miRNAs in Dicer WT ES cells had an effect on expression of bivalent genes. Almost all bivalent genes previously validated as upregulated in Dicer KO ES cells were similarly upregulated in LNA-290 transfected WT ES cells, suggesting that transcripts from bivalent genes are direct targets of the miR-290 family. The role of microRNAs in regulation of transcripts from bivalent genes is investigated in greater depth in Chapter 5.

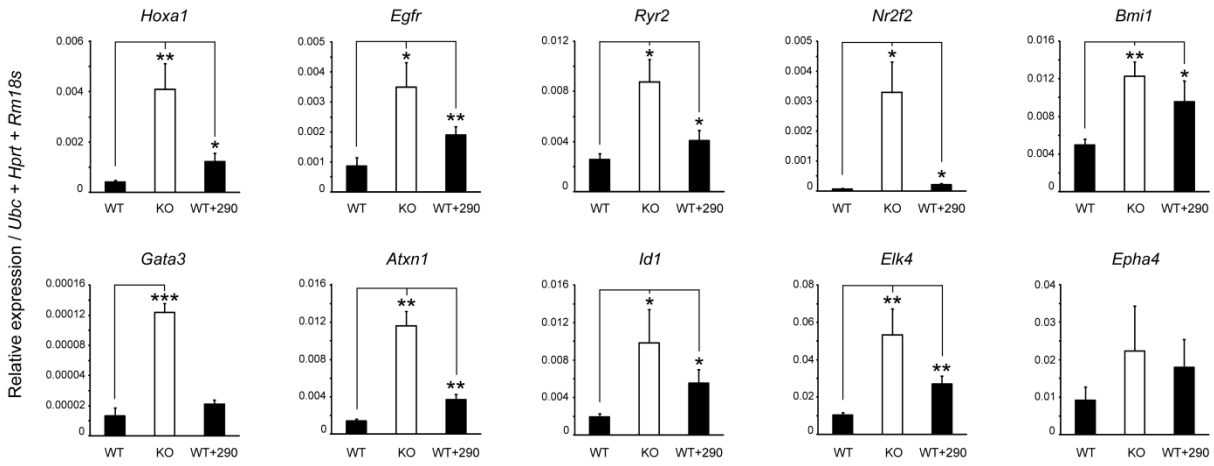


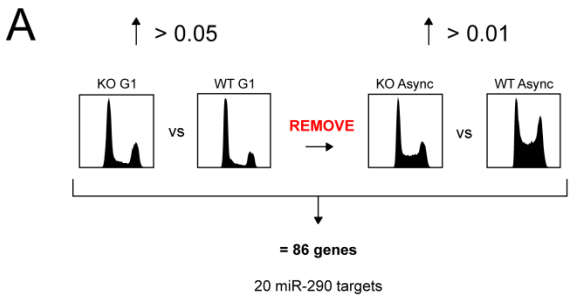

**Figure 4.12: Inhibition of miR-290 family in WT ES cells identifies bivalent genes as potential novel miRNA targets.** qPCR analysis of expression of bivalent genes in Dicer WT (WT) and KO (KO) ES cells transfected with 0.5μM control LNA, and in Dicer WT ES cells transfected with 0.5μM LNA inhibitor of the miR-290 family (WT + 290). Error bars: s.t.d. of three independent biological replicates. Asterisks indicate significance (T-test: * = $p < 0.05$; ** = $p < 0.005$; *** = $p < 0.0005$).

#### 4.2.2.3 Microarray Analysis (3): Identification of a subset of predicted miRNA targets deregulated to a greater extent in Dicer KO cells in early G1 compared to an asynchronous population

It is probable that the methods of analysis described in 4.2.2.1 and 4.2.2.2 failed to identify any genes deregulated specifically in Dicer KO ES cells in early G1 because gene expression changes in Dicer KO ES cells in early G1 are extremely similar to changes in expression in an asynchronous population (Fig. 4.6). Therefore, instead of looking for highly significant G1-specific changes in expression in early G1 (Fig. 4.8A) or the association of functionally related groups of genes with expression changes in this dataset (Fig. 4.9B), I next looked for genes that changed to a greater extent in Dicer KO ES cells in early G1 (as opposed to specifically in G1) when compared to the asynchronous population.

I first identified all genes which were upregulated in the KO cells microarray data, reasoning that this subset of genes were most likely to be directly regulated by miRNAs. Next, I identified genes with a log2FC greater than 0.5 in early G1, and then selected only those which were significantly upregulated in Dicer KO compared to control cells. From this subset of genes, I then removed genes which were also upregulated in asynchronous Dicer KO ES cells compared to control cells (Fig. 4.13A).

This approach identified 86 genes which were upregulated to a greater extent in Dicer KO ES cells in early G1 compared to an asynchronous population of cells, 20 of which were predicted to be targets of the miR-290 family by miReduce. The function of these genes was investigated using Gene Ontology (GO term) analysis, a method of assessing whether a given set of genes is associated with specific biological processes. GO terms involved in protein phosphorylation were associated with the 20 miR-290 targets which were upregulated to a greater extent in early G1, but was not statistically significant (Fig. 4.13B).




B

| Gene Ontology Term | Corrected p-value |
|---|---|
| Protein amino acid phosphorylation | 0.81 |
| Phosphorylation | 0.68 |
| Phosphorus metabolic process | 0.72 |
| Phosphate metabolic process | 0.72 |
| Response to wounding | 0.67 |

**Figure 4.13: Identification of predicted targets of the miR-290 family upregulated to a greater extent in Dicer KO ES cells in early G1.**
(A) Approach used to identify genes which are upregulated to a greater extent in Dicer KO ES cells in early G1 compared to an asynchronous population of cells. (B) GO terms associated with the 20 miR-290 target genes identified in (A). The relevant p-value for each GO term was corrected for multiple testing using the Bonferroni correction method.

I validated the expression of 12 of the 20 predicted miR-290 targets in the relevant populations of Dicer WT and KO ES cells by qPCR. Only two genes were significantly upregulated in Dicer KO cells in early G1 only (*Crk* and *Thap2*) (Fig. 4.14). An additional three genes (*Lrch3, Slc6a9* and *Hn1*) had a greater fold change in Dicer KO cells in G1 compared to an asynchronous population, although the change was not statistically significant (T-test: p value < 0.05). Apart from *Slc6a9*, which encodes the glycine transporter GLYT1 and is predominantly expressed in the brain (Hu et al., 2003), all of these genes are likely to be expressed in ES cells as they are either involved in development (*Hn1*) (Grabher et al., 2006), or ubiquitous cellular functions such as intracellular signalling (*Crk*) (Birge et al., 2009), DNA binding (*Thap2*) (Sabogal et al., 2010), or cytoskeletal regulation (*Lrch3*) (Foussard et al., 2010).

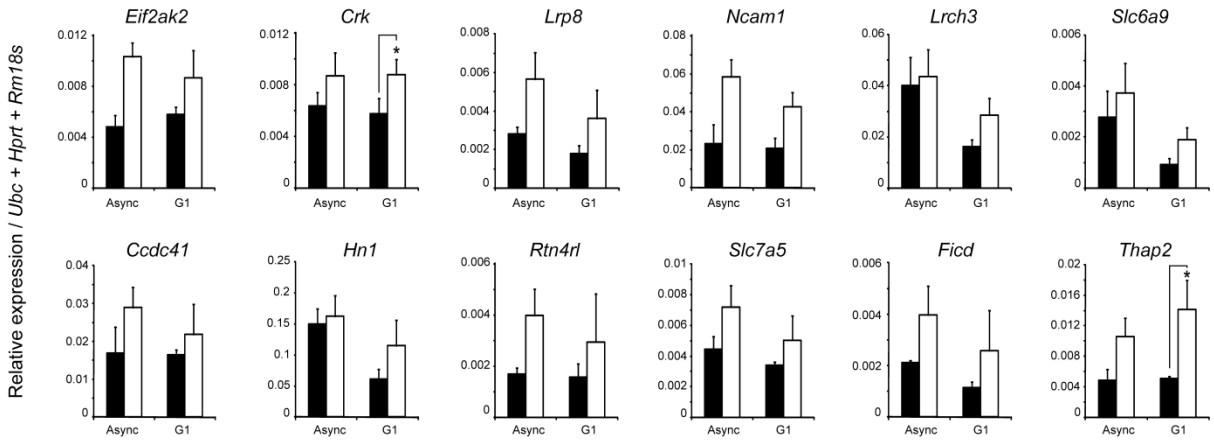



**Figure 4.14: Identification of predicted miR-290 targets which are deregulated to a greater extent in Dicer KO ES cells in early G1 compared to an asynchronous population.**
qPCR analysis of genes identified as changing to a greater extent in G1 compared to asynchronous Dicer KO ES cells in microarray data, based on parameters described in text. Black bars; Dicer WT cells. In all graphs black bars = WT cells; white bars = KO cells. Error bars: s.t.d. of three independent biological replicates. Asterisks indicate significance (T-test: * = p < 0.05).

I next addressed whether these genes are direct targets of the miR-290 family. Using the same assay as described in 4.2.2.2, I transfected WT ES cells with LNA inhibitors of the miR-290 family and analysed expression of *Crk, Lrch3, Slc6a9, Hn1* and *Thap2* by qPCR 24 hours later. Whilst *Hn1* and *Thap2* are upregulated in WT cells transfected with the miR-290 inhibitor, this change is only statistically significant in the case of *Thap2*; and expression of *Crk, Lrch3* and *SLc6a9* did not change (Fig. 4.15).

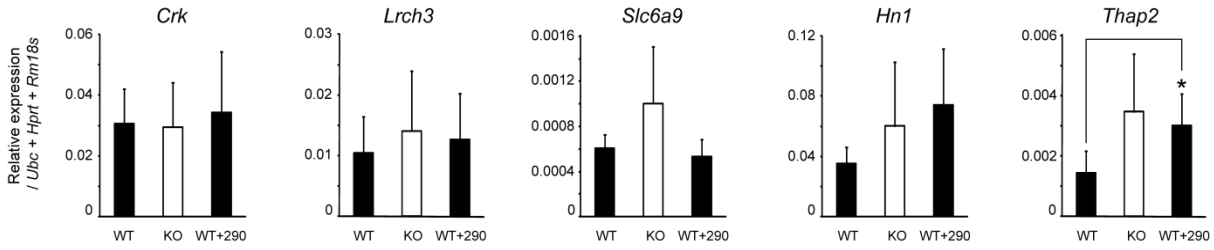



**Figure 4.15: Inhibition of the miR-290 family in WT ES cells identifies a subset of G1-deregulated genes as potential targets of ES-specific miRNAs.**
qPCR analysis of expression of G1-deregulated genes in Dicer WT and KO ES cells transfected with 0.5μM control LNA (WT and KO), and in Dicer WT ES cells transfected with 0.5μM LNA inhibitor of the miR-290 family (WT + 290). Error bars: s.t.d. of three independent biological replicates. Asterisks indicate significance (T-test: * = $p < 0.05$).

As all transfected populations in Fig. 4.15 are asynchronous, and these genes were specifically selected because they were not significantly upregulated in asynchronous Dicer KO ES cells it is therefore perhaps unsurprising that despite being predicted targets of the miR-290 family, expression of *Crk, Lrch3* and *SLc6a9* was unaffected by inhibition of the miR-290 family under these conditions. Further experiments are required to determine whether these genes are direct targets of the miR-290 family. This could be attempted by either looking at expression of these genes in LNA-transfected cells in early G1, or by using a 3' UTR assay (such as a luciferase reporter system) to identify whether miRNA binding sites in these genes are functional.

### 4.2.3 Discussion (1)

#### 4.2.3.1 Expression of histone gene expression is co-ordinately and specifically deregulated in Dicer KO ES cells in early G1

Histone gene expression is known to be regulated in a highly cell cycle-dependent manner, as rapid production of histone proteins is required specifically in S phase to package newly synthesised DNA. To facilitate stringent temporal regulation, histone mRNA transcripts have a unique 3'UTR conformation, characterised by a stem-loop structure (Marzluff et al., 2008). The presence of

the stem loop in place of a conventional polyadenylated 3'UTR precludes these transcripts from being directly targeted by miRNAs, and therefore co-ordinated downregulation of these genes in early G1 is likely to be attributable to miRNA-mediated control of upstream regulatory mechanisms governing expression of these genes.

One key player in transcriptional regulation of histone genes is the Nuclear Protein Ataxia-Telangiectasia Locus protein (NPAT), which physically associates with histone gene loci and stimulates transcription of multiple histone genes simultaneously (Gao et al., 2003). Upregulation of NPAT activity in G1 is required to induce rapid, co-ordinate expression of histone genes in late G1/early S phase. This is mediated at the transcriptional level by binding of the E2F family of transcription factors to histone genes (Gao et al., 2003), and at the post-transcriptional level by the CDK2/cyclin E complex, the activity of which is required for the G1/S transition (Zhao et al., 1998).

It is known that upregulation of p21 inhibits the activity of the CDK2/cyclin E complex, which consequently results in G1 arrest. Lack of progression from G1 to S phase also subsequently results in cytoplasmic retention of the E2F family of proteins, and therefore silencing of E2F target genes. Given that NPAT is positively regulated by both the CDK2/cyclin E complex and the E2F family of proteins, it is possible that loss of NPAT activity as a result of increased p21 expression is responsible for the co-ordinated downregulation of histone genes in early G1 in Dicer KO ES cells (Fig. 4. 16).

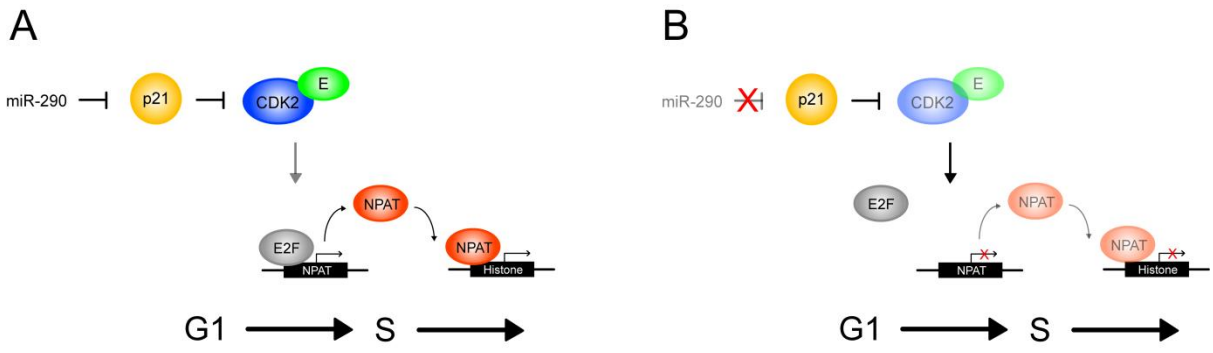



**Figure 4.16: Model for role of miRNA in G1-specific regulation of histone gene expression.**
In WT cells (A) p21 levels are attenuated by the activity of the miR-290 family. The CDK2/cyclinE complex is therefore active and facilitates progression from G1 to S, and subsequent activation of E2F targets, including NPAT. Expression of NPAT co-ordinately activates transcription of multiple histone genes at the onset of S phase. In Dicer KO cells (B), loss of miRNA activity leads to upregulation of p21, subsequent inhibition of CDK2/cyclinE function and G1 arrest. The E2F proteins therefore do not activate target genes, NPAT is not expressed, and transcription of histone genes is not induced.

Given that histone genes are not downregulated in asynchronous Dicer KO cells (Fig. 4.8) it is clear that activation of transcription of these genes is not entirely abolished in microRNA-deficient ES

cells. The above model is not, therefore, a comprehensive description of the role of miRNA in regulation of histone transcription, but may account for the G1-specific defect observed in this study.

#### 4.2.3.2 Identification of a potential role for miRNA in regulation of protein phosphorylation in early G1

Dynamic protein phosphorylation is a critical regulatory mechanism involved in mediating transitions between cell cycle stages. The association of GO terms related to protein phosphorylation with genes upregulated to a greater extent in Dicer KO ES cells in early G1 when compared to an asynchronous population (Fig. 4.13) could indicate that microRNAs play a role in co-ordinated regulation of multiple transcripts which are involved in this process immediately post-cell division. Further analysis of this observation could potentially add to current understanding of the role of miRNA in regulation of the cell cycle itself.

#### 4.2.3.3 Re-evaluation of mechanisms governing regulation of lineage-specific gene expression in pluripotent stem cells

Despite the range of approaches employed to analyse the microarray data generated in this study, eventually only a limited number of genes were identified which were predicted to be miRNA targets and which were deregulated to a greater extent in early G1 compared to an asynchronous population. A number of reasons proposed to account for this are discussed below.

1. Technical/assay design.

Microarray analyses can reliably detect changes in gene expression above a certain magnitude, and in addition, only actual fold changes of greater that 2 are usually considered as biologically relevant using this technique. However, due to the nature of miRNA activity, expression changes detected in these array samples are subtle (mostly less than 2-fold). Detecting statistically significant changes of this magnitude is challenging in the first instance, but then to add comparisons between conditions (in this case cell cycle stages) on top of this makes detection of changes which are specific to individual conditions difficult. Since this assay was designed, the more sensitive method of RNA profiling by deep sequencing (RNA-seq) has become far more widespread, and use of this technology could potentially be more informative in reliably detecting subtle changes of the magnitude described in this study.

In addition, microarray analyses of gene expression in the defined conditions were performed two hours post-release from mitotic arrest. However, it could be that deregulation of some genes occurs to a greater extent at earlier or later timepoints post-mitotic exit. Therefore to analyse the role of microRNAs in the G2-M to G1 transition in the greatest depth possible, it would be informative to perform genome-wide expression analysis during a timecourse post-mitotic release.

Finally, throughout the analyses described here, expression changes in populations of Dicer WT and KO ES cells were compared to changes in asynchronous cells. However, given that a proportion of cells in an asynchronously growing population are also in G1, this was perhaps not an ideal control. Instead, it would be preferable to compare expression in early G1 with a population of cells depleted of G1 (e.g. S and G2/M only). Whilst a satisfactory method to obtain this population was not available at the start of this study, latter experiments using the elutriation system (see following section) to fractionate cells at distinct cell cycle stages could potentially be used to improve this assay.

2. Secondary effects in stable Dicer KO ES cells

Constitutive loss of Dicer and long-term *in vitro* culture in these conditions could plausibly result in secondary effects at the gene expression and phenotypic levels which are not directly attributable to loss of miRNA function, but more likely due compensatory mechanisms for loss of miRNA activity which have been induced over time. Repeating the assay described above in a conditional Dicer KO system would be more informative in terms of identification of changes directly related to and/or caused by loss of miRNA function.

3. Mitotic bookmarking maintains lineage-specific gene expression through cell division

As discussed in 1.3.2, the re-establishment of lineage-specific transcriptional profiles in early G1 has been proposed to be mediated in some circumstances by ‘mitotic bookmarks’; histone modifications and a subset of transcription factors thought to be stably associated with chromatin through the G2/M to G1 transition. Whilst evidence for this mechanism is disparate, it is possible that mitotic bookmarking serves to guide transcription factors and chromatin modifying proteins to relevant genomic sites in a highly specific manner in early G1, facilitating rapid and efficient establishment of cell-specific gene expression immediately post-cell division. Given the paucity of evidence presented

here for aberrant of lineage-inappropriate genes in early G1, this hypothesis is plausible; however, until more sensitive analyses of gene expression in Dicer KO cells in early G1 have been performed, a role for miRNAs in buffering cell-specific transcriptional profiles through this stage of the cell cycle should not be excluded.

4. Globally hyperactive genome in ES cells redefines concept of lineage-specific gene expression

Alternatively, whilst it is feasible that no global defects in lineage-specific gene expression were observed specifically in early G1 due to one or a combination of the reasons above, an alternative explanation could be considered. It has been proposed that ES cells exhibit a 'hyperactive' transcriptome, defined by low-level transcription from the majority of the genome (Guenther et al., 2007; Efroni et al., 2008; Meshorer and Misteli, 2006). Recent studies support this model by showing that even genes bound by PcG proteins are in fact transcribed (Brookes et al., 2012), some of which are also translated into protein. This plasticity is thought to contribute to the pluripotent state by facilitating rapid upregulation of key developmental regulators in response to cues to differentiate.

In light of these studies, one of the original hypotheses of this study, which considers derepression of lineage-inappropriate genes in early G1, is perhaps not as applicable to ES cells as it may be to somatic cell types. Whilst in terminally differentiated cell types, lineage-specific gene expression is maintained by stringent repression of lineage-inappropriate genes, in ES cells this may not be the case. Instead, if it is considered that the majority of the genome in pluripotent cells is transcribed to some extent, then the concept of lineage-appropriate gene expression is redefined by whether these transcripts are translated into protein, not whether they are transcriptionally active or silenced. Whether this balance is compromised through the G2-M to G1 transition cannot be addressed using data presented here due to the identification of insufficient G1-deregulated genes which are genuinely lineage-inappropriate in ES cells, and due to technical limitations of the assay. More sensitive transcriptome analyses at more frequent timepoints post-mitotic release, as described above, would be required to probe this idea further.

Considering the idea of differential mechanisms for maintenance of cell-specific transcriptional states in somatic and pluripotent cells, it would be informative to analyse gene expression immediately post-cell division in a terminally differentiated cell type, and assess these

results in parallel with those from the ES study described above. In pursuit of this aim, a Dicer KO 3T3 fibroblast line was derived (Appendix I). Unfortunately, these cells proved to not be amenable to the mitotic arrest and release assay, and therefore experiments on these cells were not continued in the interests of consistency between experimental systems. Optimisation of an alternative system to explore this idea further would be an attractive avenue for future work.

Whilst the concept of a globally hyperactive genome in ES cells perhaps negates the original hypothesis, in another sense it introduces a fundamental role for miRNA-mediated gene regulation in maintenance of the pluripotent state. Low-level transcription of the majority of the genome – including multiple lineage-inappropriate genes – is thought to be required to facilitate differentiation. However, the cell must ensure that transcripts from these genes are not translated into protein in order to maintain the pluripotent state: therefore co-ordinated regulation of multiple transcripts by ES-specific miRNAs could plausibly contribute to this. This mechanism is supported by global upregulation of bivalent genes in Dicer KO ES cells; a concept which is investigated further in the following chapter.

In summary, data presented in this chapter supports a role for miRNA in regulation of lineage-inappropriate transcripts in pluripotent cells as exemplified by global upregulation of bivalent genes predicted to be direct targets of the ES-specific miR-290 family. Given that this deregulation is not restricted to or greater in early G1, I next addressed the question of whether expression of these genes is specifically deregulated at any other stage in the cell cycle.

## 4.3 Analysis of gene expression in Dicer WT/KO ES cells throughout the cell cycle

As discussed in Chapter 1, the G2-M to G1 transition is just one example of a point during the cell cycle which presents a challenge to gene regulatory mechanisms. In order to investigate whether lineage-specific transcriptional profiles are compromised through any other transitional phases during the cell cycle in microRNA-deficient ES cells, a method was required to isolate cell populations representative of all cell cycle stages.

### 4.3.1 Optimisation of elutriation system

Elutriation is a drug-free method of isolating cells in distinct stages of the cell cycle from an asynchronous population (Banfalvi, 2008). The technique is dependent on the progressive increase in cell size and density throughout the cell cycle, facilitating separation of distinct populations of cells via

centrifugation. An asynchronous population of cells is loaded into a specifically designed chamber inside a centrifuge. The application of centrifugal force to the chamber results in distribution of cells of distinct densities in a gradient within the chamber (Fig. 4.17A). By gradually increasing the flow rate of liquid pumped into the chamber whilst the centrifuge is in operation, cells are gradually eluted out of the chamber in proportion with their size (smallest cells are eluted first). Thus, a continuum of populations representative of all cell cycle stages can be obtained using the same method. Cell-cycle profiles of populations representative of all stages of the cell cycle in Dicer WT and KO ES cells obtained using this technique are illustrated in Fig.4.17B.

Consistent with previously published data (Wang et al., 2008) and observations in this study (Fig. 4.4) the accumulation of Dicer KO cells in early G1 (and subsequent deficiency of S phase cells) is reflected in populations of elutriated cells (Fig. 4.17B). Despite many attempts to optimise the conditions used for the elutriation assay in order to obtain a population of Dicer KO ES cells enriched in S phase, this proved to be extremely challenging. At best, populations of S phase cells in the Dicer KO line contained residual populations of either G1 or G2/M. These data indicate that a very small proportion of asynchronously growing Dicer KO ES cells exist in S phase, strengthening the evidence for the previously described miRNA-mediated G1-S transition which is defective in these cells.

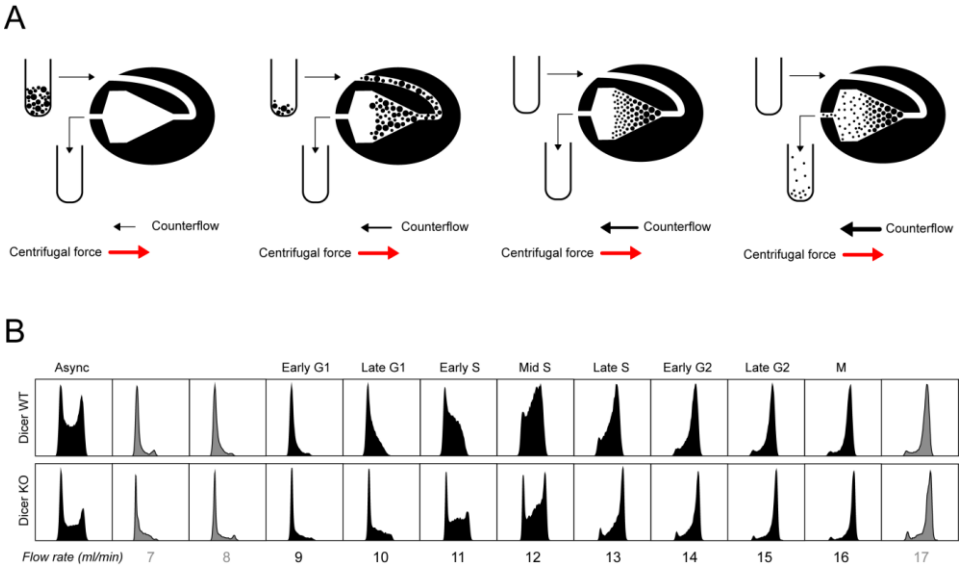



**Figure 4.17: Isolation of cells at distinct stages of the cell cycle using elutriation.**
(A) Schematic of elutriation system. An asynchronous population of cells are loaded into a chamber, which is subject to centrifugal force. This results in distribution of the cells within the chamber according to size. By increasing the flow rate of liquid into and out of the chamber, cells can be eluted according to size, corresponding to progressive stages of the cell cycle. (B) Cell cycle profiles (FACs analysis of propidium iodide (PI) stained cells) of elutriated populations of Dicer WT and KO ES cells. The flow rate required to elute this populations is indicated below each profile. Grey profiles indicate populations excluded from further analysis due to high proportion of cell debris (7/8ml/min) and doublets (17ml/min) in these fractions. Cell cycle stages which eluted populations represent are indicated above the relevant profiles.

Cells used in the elutriation system described in this section were also grown in the combined media of N2B27 + 2i and DMEM + 10% FCS (50:50 ratio), identical to conditions used for the mitotic arrest and release assay. This was to minimise variation between conditions used in the two systems and ensure that data obtained by each method was comparable.

### 4.3.2 Quantification of gene expression through the cell cycle

As illustrated above, the elutriation system is dependent on the progressive increase in cell size that occurs gradually throughout the cell cycle. This phenomenon has important implications for quantification of mRNA levels in representative populations from each cell cycle stage. As the cell increases in size, its total RNA content would also be expected to increase proportionally. This was confirmed by quantification of total RNA (in nanograms) extracted from one million cells isolated by elutriation from distinct cell cycle stages (Fig 4.18B).

By definition, as housekeeping genes are required for ubiquitous and essential cellular functions, these genes would be expected to be transcribed at a rate proportional with the increase in cell size, and therefore the amount of transcripts from these genes would increase throughout the cell cycle. Therefore, if normalised to mRNA levels of housekeeping genes as in standard gene expression analyses, transcript levels from any genes which are not transcribed at a similarly increased rate would appear to decrease through the cell cycle; not because this is a real biological trend, but because levels of the housekeeping gene are increasing (Fig. 4.18A).

To identify an alternative means of normalisation for gene expression and thus avoid this issue, I introduced a defined quantity of a synthetic RNA with no known homology to the mouse genome into lysed populations of one million cells from each elutriated fraction. RNA was then extracted from each sample, and levels of genes of interest were normalised to the exogenously introduced RNA as well as endogenously transcribed housekeeping genes. By doing so, it was reasoned that absolute (as opposed to relative) values of the mRNA of interest could be quantified at defined stages of the cell cycle.

To test this system, I quantified total levels of *Hprt* and the 'spike' RNA in each elutriated fraction as determined by C(t) value per one million cells. As predicted, whilst levels of *Hprt*

consistently increased throughout the cell cycle (Fig. 4.18C), in concordance with an increase in cell size and total RNA (Fig. 4.18B), levels of the exogenous spike RNA remained constant (Fig 4.18D).

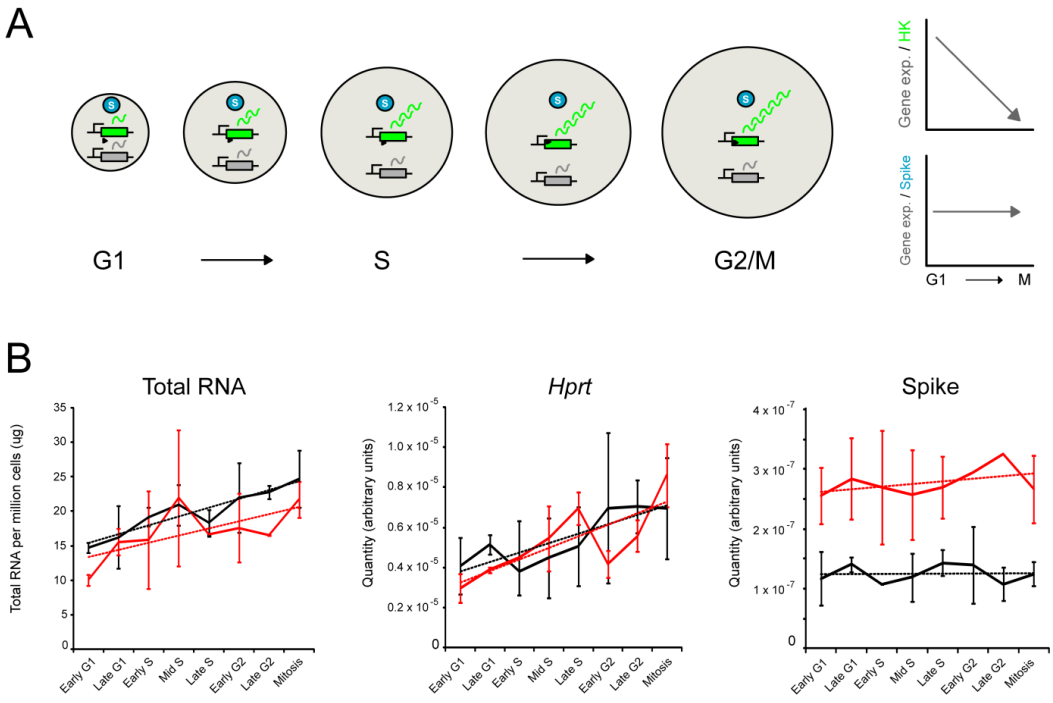



**Figure 4.18: Comparison of endogenous and exogenous RNAs for normalisation of gene expression through the cell cycle.**
(A) Representation of hypothetical changes in gene expression through the cell cycle. Expression of housekeeping genes (green) increases through the cell cycle in proportion with cell size. Genes which are not transcriptionally upregulated through the cell cycle are shown in grey. Introduction of the spike RNA (S) gives a constant parameter for normalisation of levels of both subsets of genes through the cell cycle. Schematic graphs (right) indicate apparent decrease in level of basally transcribed genes when normalised relative to housekeeping genes (HK); and the constant actual level of expression when normalised to the spike RNA. (B) Quantification of total RNA (ng) per 1x10^6 Dicer WT (black) and KO (red) ES cells in each elutriated fraction. Quantification of *Hprt* (C) and spike RNA (D) per 1x10^6 cells based on C(t) values. Error bars: s.t.d of three independent biological replicates.

However, when total levels of the spike RNA was quantified per one million cells in WT and KO cells at different stages of the cell cycle, significantly more spike RNA was consistently detected in KO cells at the same cell cycle stage when compared to WT cells (Fig. 4.18D). These results meant that mRNA levels normalised to spike RNA could not be compared between WT and KO cells at defined cell cycle stages. The system could still be used, however, to quantify changes in the amount of RNA in WT/KO cells in G2-M relative to the quantity in G1, in order to measure relative accumulation or degradation of the mRNA through the cell cycle. Therefore the above methods of normalisation provide two alternative means of quantification of mRNA levels in Dicer WT/KO ES cells through the cell cycle; firstly, proportionally to the biological state of the cell (normalised to *Hprt*) and secondly, relative to the amount of transcript in G1 (normalised to spike RNA).

### 4.3.3 Kinetics of histone gene expression in elutriated populations

As a positive control to assess whether the elutriation system itself and the above methods of gene expression analysis could reliably detect cell cycle-dependent changes in gene expression, I quantified levels of transcripts from histone genes in elutriated populations of WT ES cells. As previously discussed in 4.2.3.1, transcription from histone genes is highly cell cycle regulated, with mRNA production rapidly increasing through G1 and peaking in S phase, followed by degradation of histone mRNA transcripts in G2/M (Marzluff et al., 2008). This expression pattern could be detected in mRNA levels of the *Hist1h1a* and *Hist4h4* genes when normalised to either the exogenously introduced spike RNA (Fig. 4.19A) or *Hprt* (Fig. 4.19B). These data illustrate that elutriation can be used to detect cell-cycle dependent changes in gene expression.

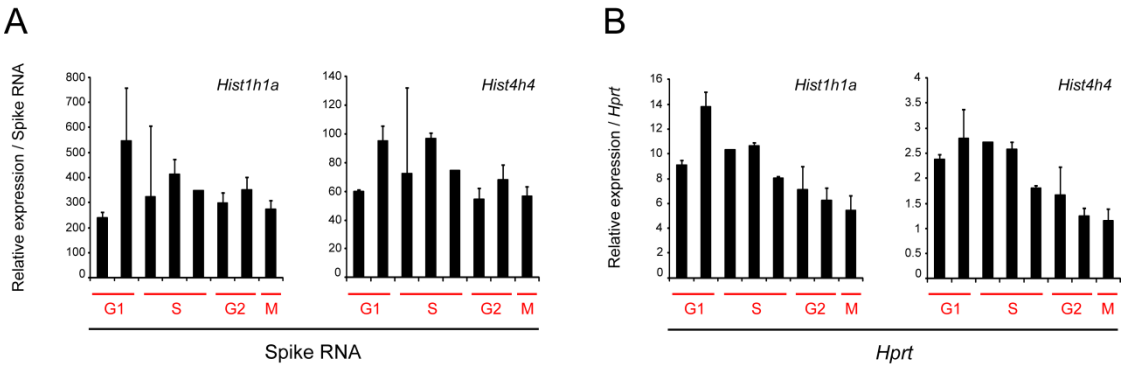



**Figure 4.19: Cell cycle-dependent changes in histone gene expression can be detected in elutriated populations of WT ES cells.**
qPCR analysis of *Hist1h1a* and *Hist4h4* expression in WT elutriated populations normalised to (A) spike RNA and (B) *Hprt*. Cell cycle stages represented by the populations analysed are indicated in red. Error bars: s.t.d. of three independent biological replicates.

### 4.3.4 Expression of validated microRNA targets through the cell cycle in Dicer KO ES cells

Next, I was interested to investigate whether genes known to be directly regulated by miRNAs exhibit cell-cycle dependent patterns of expression in Dicer WT or KO ES cells. In addition to the previously discussed miR-290 targets *Cdkn1a (p21), Rbl2* and *Lats2, Casp2* (Zheng et al., 2011) and *Arid4a* (Ciaudo et al., 2009) have also been validated as targets of the miR-290 family in mouse ES cells. *Casp2* plays a role in regulation of apoptosis under conditions of genotoxic stress, and *Arid4a* (previously known as retinoblastoma-binding protein 1) is involved in recruitment of histone modifying complexes to the promoters of genes repressed by Rb (retinoblastoma 1). Expression of *Oct4* (which is not targeted by ES-specific miRNAs) was assessed as a control.

Cell-cycle dependent expression of these genes in Dicer WT and KO ES cells was analysed in populations of cells obtained by elutriation and normalised to either endogenously expressed *Hprt* or exogenously introduced spike RNA (Fig. 4.20).

Expression of *Oct4* showed no difference between Dicer WT and KO ES cells at any stage of the cell cycle when normalised to *Hprt*, and normalisation to the spike RNA in G1 and G2-M showed that transcripts from this gene accumulate to a similar extent throughout the cell cycle in both Dicer WT and KO ES cells. *Cdkn1a* was significantly upregulated in WT cells through the cell cycle, corresponding to an almost 2-fold accumulation in G2-M when compared to G1. This trend was amplified in Dicer KO cells, resulting in a 2.6-fold accumulation of transcript in G2-M. A similar trend was observed in expression of both *Rbl2* and *Lats2*, suggesting that miRNAs may play a role in attenuating accumulation of target mRNAs through the cell cycle as the cell grows and divides.

Conversely, whilst *Arid4a* and *Casp2* exhibited patterns of expression through the cell cycle in Dicer WT ES cells similar to those observed in *Cdkn1a, Rbl2* and *Lats2*, in Dicer KO cells, transcripts from these genes appear to be downregulated from G1 to G2-M (as indicated by both elutriation data and by quantification to the spike RNA).

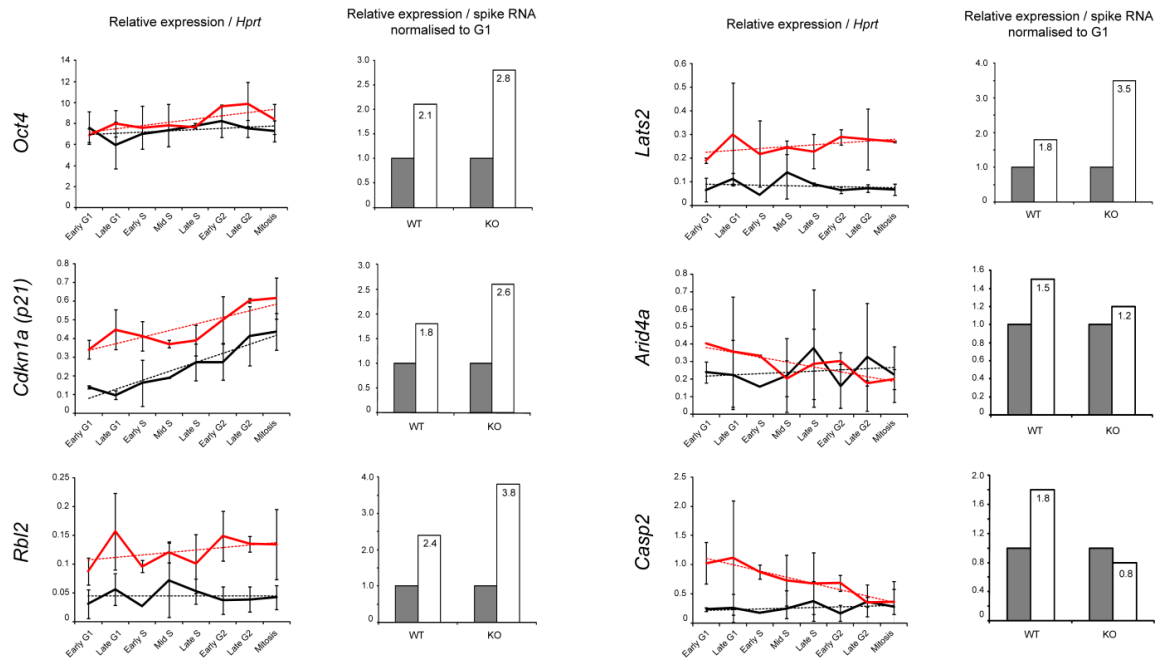



**Figure 4.20: Transcripts from validated miR-290 targets *Cdkn1a*, *Rbl2* and *Lats2* accumulate through the cell cycle to a greater extent in Dicer KO ES cells compared to WT.**
Line graphs show qPCR analysis of gene expression in elutriated populations of Dicer WT (black) and KO (red) ES cells normalised to *Hprt.* Dashed line illustrates trend in expression through the cell cycle. Error bars: s.t.d of three independent biological replicates. Bar graphs illustrate quantity of mRNA normalised to spike RNA in G1 (grey bars) and G2/M (white bars) in Dicer WT and KO ES cells as indicated. Numbers within white bars show fold change in G2/M compared to G1.

### 4.3.5 Expression of bivalent genes through the cell cycle in Dicer KO ES cells

Having identified that transcripts from validated targets of the miR-290 family appear to accumulate through the cell cycle in Dicer KO ES cells, I was next interested to assess whether transcripts from bivalent genes (as potential novel targets of miR-290) similarly exhibit any cell-cycle dependent expression patterns. I analysed the expression of a subset of bivalent genes (previously shown to respond to LNA-mediated inhibition of the miR-290 family in WT ES cells) in elutriated populations of WT and KO ES cells, and normalised the levels of these transcripts to both *Hprt* and the spike RNA. It was observed that as in the case of validated miRNA targets, transcripts from bivalent genes accumulate to a greater extent in G2/M in Dicer KO ES cells compared to WT (Fig. 4.21).

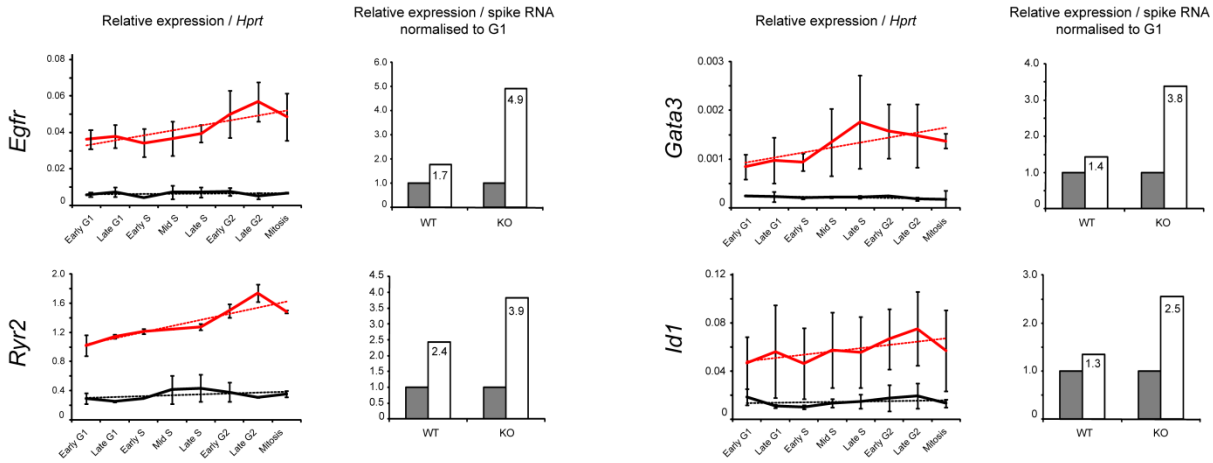



**Figure 4.21: Transcripts from bivalent genes predicted to be targets of the miR-290 family accumulate through the cell cycle to a greater extent in Dicer KO ES cells compared to WT.**
Line graphs show qPCR analysis of gene expression in elutriated populations of Dicer WT (black) and KO (red) ES cells normalised to *Hprt.* Dashed line illustrates trend in expression through the cell cycle. Error bars: s.t.d of three independent biological replicates. Bar graphs illustrate quantity of mRNA normalised to spike RNA in G1 (grey bars) and G2/M (white bars) in Dicer WT and KO ES cells as indicated. Numbers within white bars show fold change in G2/M compared to G1.

Taken together, these data suggest that miRNAs may play a role in attenuating accumulation of both ES-expressed and lineage-inappropriate mRNA transcripts through the cell cycle, as global transcription and therefore total RNA increases in proportion with cell size.

### 4.3.6 Discussion

#### 4.3.6.1 Use of elutriated populations to detect trends in gene expression through the cell cycle

The elutriation system has many advantages, in terms of being entirely independent of a requirement for chemical agents for cell cycle fractionation which may induce secondary, non-specific effects, and in the capacity of the assay to isolate populations of cells representative of all stages of the cell cycle using the same method. However, as a result of the continuous elution of cells from the gradient established inside the chamber (Fig. 4.17A) it is difficult to use elutriated populations to identify highly cell-cycle stage specific changes in gene expression; this assay is more appropriate to detecting trends in expression through the cell cycle.

This is exemplified by analysis of the expression of histone genes *Hist1h1a* and *Hist4h4* in Fig. 4.19. Expression of these genes is known to be almost entirely shut down after S phase, and only reactivated after the G2/M to G1 transition (see 4.2.3.1). Whilst a trend reflecting this regulation is detected in elutriated populations, the changes in expression are less dramatic than would be anticipated given the stringency with which mRNA levels of these genes are known to be regulated (Marzluff et al., 2008). This is likely due to the presence of cells from previous and subsequent fractions present in a given elutriated population.

In addition, the fact that absolute depletion of histone transcripts in G2/M was not observed in this study could also be attributable to the fact that analyses presented here quantify steady-state mRNA levels. As a result, residual transcripts produced in previous cell cycle stages could also contribute to the mRNA level quantified in mitotic cells, even if transcription of these genes has been repressed at this point. More informative analyses could be obtained by analysing levels of nascent RNA in elutriated fractions, or by assessing RNA production using nuclear-run on assays.

#### 4.3.6.2 Accumulation of microRNA targets through the cell cycle in Dicer KO ES cells

The use of differential systems of normalisation in populations of cells obtained by elutriation allows analysis of cell cycle dependent trends in gene expression to be assessed either relative to the growth in cell size and total RNA through the cell cycle (relative to *Hprt*), or in terms of absolute levels

of a given transcript at one stage of the cell cycle compared to another (relative to exogenously introduced spike RNA).

Using these two methods, I analysed expression of validated miR-290 targets *Cdkn1a (p21), Rbl2* and *Lats2* through the cell cycle. Transcript levels of *Cdkn1a* appear to increase through the cell cycle relative to *Hprt*, indicating that expression of this gene increases at a rate greater than *Hprt* in WT ES cells. Expression of *Rbl2* and *Lats2* appears to not change through the cell cycle relative to *Hprt*, suggesting that these genes are transcribed at a similar rate. However, normalisation of all three genes to the spike RNA in G1 and G2/M suggests that loss of miRNA-mediated regulation of these transcripts results in their accumulation in G2/M to a greater extent in Dicer KO ES cells compared to WT.

Conversely, validated miRNA targets *Arid4a* and *Casp2* did not follow this trend. In WT ES cells, transcript levels appear to not change relative to *Hprt*, and accumulate slightly in G2/M when normalised to spike RNA – suggesting that like *Rbl2* and *Lats2*, expression of these genes increases in WT cells through the cell cycle in proportion with housekeeping genes and total RNA. However, in Dicer KO ES cells, these genes are upregulated in G1 but downregulated through the cell cycle. This pattern could possibly be attributed to upregulation of a negative regulator of these genes during late S phase and early G2 which is also a miRNA target.

When the pattern of expression of bivalent genes was analysed in the same system, it was observed that in WT cells, transcript levels do not change through the cell cycle when normalised to *Hprt,* and accumulate in G2/M to a similar extent as the actively expressed miRNA targets discussed above. This data suggests that despite being transcribed at low levels, transcripts from these genes are subject to the same proportional changes in global transcription as actively expressed genes. As with the patterns in expression of validated miRNA targets through the cell cycle, transcripts from bivalent genes appear to accumulate to a greater extent in Dicer KO ES cells in G2/M than in WT cells.

It was noted that the degree of upregulation in bivalent genes in Dicer KO ES cells through the cell cycle was greater than would be anticipated if attributable directly to loss of post-transcriptional regulation by miRNAs, and yet the fact that these genes respond as direct targets in the LNA-mediated miRNA inhibition assay suggests that they are directly regulated by the miR-290

family. It was then considered that this phenotype could potentially be caused by combination of deregulation at both the transcriptional and post-transcriptional levels, the cumulative effect of which could possibly account for the magnitude of deregulation observed. This question is addressed in the following chapter.

Taken together, these data support a role for miRNA in regulation of both lineage-specific and inappropriate transcripts through the cell cycle. These results support the notion that lineage-inappropriate genes are actively transcribed in ES cells, and suggest that a subset of these transcripts is post-transcriptionally regulated by miRNAs. Expression analysis in cell populations obtained by elutriation indicates that both ES-expressed and lineage-inappropriate miRNA targets accumulate to a greater extent in Dicer KO ES cells throughout the cell cycle, suggesting that miRNAs may play a role in facilitating a gradual, global increase in transcription as the cell grows in size whilst simultaneously maintaining transcript levels within a threshold tolerable to the cell.

With regard to future experiments, it would be valuable to perform the above analyses in cells conditionally deleted for the Dicer gene. When discussing the possibility of accumulation of transcripts which are direct miRNA targets, it should be acknowledged that these cells have been cultured for sustained periods of time without miRNAs, and therefore the levels of accumulation and upregulation of direct targets could potentially be partially attributable to this. Quantification of transcript accumulation through the cell cycle may therefore be more informative if analysed in cells which have not undergone many cell cycles in a miRNA-deficient state, and would present an interesting avenue for future investigation.

### 4.4 Summary and concluding remarks

In conclusion, the experimental systems described in this chapter were designed to investigate the biological role of mitotically inherited microRNAs in mammalian cells. It was hypothesised that as the G2/M to G1 transition could compromise chromatin-based mechanisms of gene regulation, the presence of microRNAs through this transition could stabilise transmission of cell-specific transcriptional profiles. However, data presented in this chapter indicates that lineage-specific gene expression is not perturbed Dicer KO ES cells in Dicer KO cells immediately post-cell division. This could potentially be attributable to global low-level transcription of the majority of the

genome in pluripotent cells; so that 'lineage-inappropriate' transcripts are not actually transcriptionally repressed. The hypothesis that chromatin-based mechanisms of gene regulation may be compromised during the G2/M to G1 transition may be more applicable to terminally differentiated cell types, where lineage-specific gene expression is regulated by more stringent mechanisms at the transcriptional level.

Despite the lack of genes specifically deregulated in early G1 in Dicer KO ES cells, a role for microRNAs in regulation of lineage-specific gene expression was identified in the discovery of a substantial number of bivalent genes predicted to be directly targeted by members of the ES-specific miR-290 family. Analysis of expression of bivalent genes throughout the cell cycle suggests that loss of miRNA-mediated regulation results in accumulation of these transcripts in G2/M in a manner similar to ES-expressed target genes. These data support a model whereby in pluripotent cells, miRNAs contribute towards maintenance of lineage-specific transcriptional profiles by maintaining transcript levels of both expressed and lineage-inappropriate genes within a threshold tolerable to the cell which is concurrent with the global increase in total RNA as the cell grows and divides. Whilst this function is not restricted or specific to a given stage in the cell cycle as originally hypothesised, these findings support a role for miRNA in regulation of lineage-specific gene expression through the cell cycle.

# 5.0 MicroRNA-dependent transcriptional and post-transcriptional regulation of bivalent genes

## 5.1 Introduction

In the previous chapter, I identified global upregulation of transcripts from genes marked by the bivalent chromatin signature of H3K4me3 and H3K27me3 in microRNA-deficient ES cells. The transcription of bivalent genes is repressed by Polycomb proteins (PcG) (Stock et al., 2007; Margueron and Reinberg, 2011; Klymenko et al., 2006; Ku et al., 2008). In this chapter I examine whether the deregulation of these transcripts in microRNA-deficient ES cells is attributable to loss of post-transcriptional regulation, transcriptional regulation, or a combination of both.

## 5.2 Evidence for direct post-transcriptional regulation of bivalent genes by microRNAs

A direct role for miRNAs in post-transcriptional regulation of transcripts from bivalent genes is strongly supported by the finding that these genes are upregulated in WT ES cells transfected with an LNA inhibitor of the miR-290 family (Fig. 4.11). To substantiate these results, I carried out GSEA analysis of published microarray data for another Dicer KO ES cell line into which the miR-290 family had been reintroduced (Sinkonnen et al., 2008). My analysis showed a highly significant association of bivalent genes with downregulated transcripts in miR-290 transfected cells (Fig. 5.1A). This result, using an independent dataset, provides strong support for a role of the miR-290 miRNAs in the regulation of bivalent genes. I further confirmed this by transfection of a miR-291a-3p mimic into the Dicer KO ES cells. This miRNA was selected as it is the most highly expressed expressed member of the miR-290 family (Houbaviy et al., 2003). qPCR analysis of validated miR-290 targets (Fig. 5.1B) and bivalent genes (Fig. 5.1C) showed that, these genes are downregulated on reintroduction of miR-291a-3p. Taken together, these results provide strong evidence that transcripts from bivalent genes are directly regulated by ES-specific miRNAs.

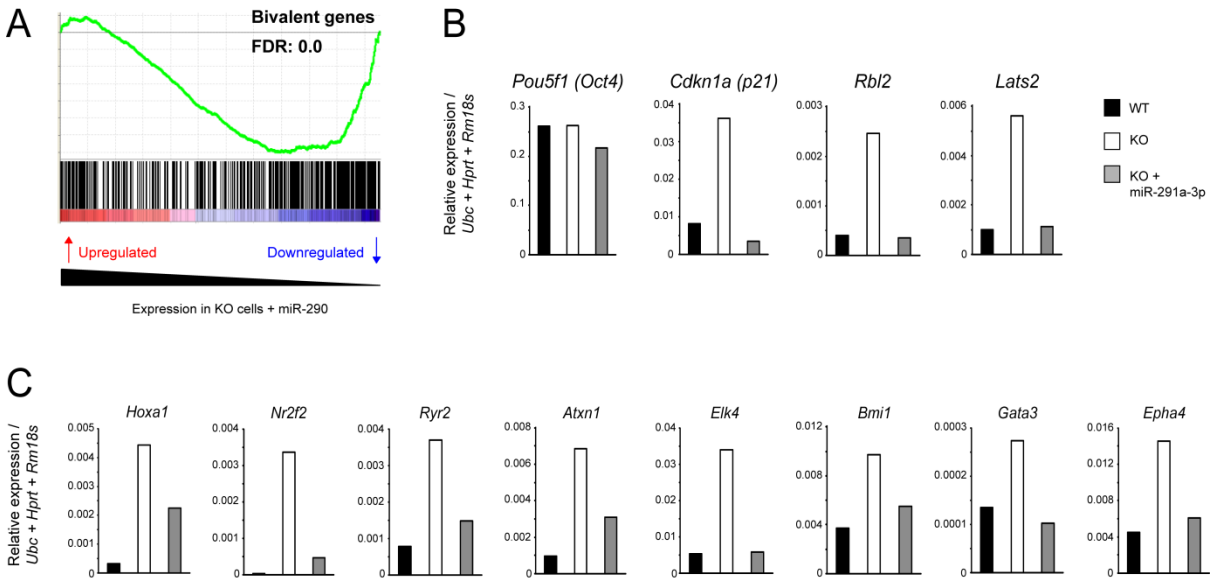



**Figure 5.1: Reintroduction of the miR-290 family into Dicer KO ES cells indicates that transcripts from bivalent genes are directly regulated by ES-specific miRNAs.**
(A) GSEA analysis of microarray gene expression data for Dicer KO cells transfected with all members of miR-290 cluster compared to Dicer KO cells transfected with a control RNA (Sinkonnen et al., 2008). The associated gene set is a list of genes marked by H3K4me3 and H3K27me3 as defined in (Ku et al., 2008). Note that these genes are downregulated on reintroduction of the miR-290 family into Dicer KO ES cells, suggesting that transcripts from bivalent genes are direct targets of the miR-290 family. FDR: false discovery rate. qPCR analysis of expression of (B) control genes and (C) bivalent genes in Dicer WT (black) and KO (white) cells transfected with a control RNA, and Dicer KO cells transected with a final concentration of 0.1μM miR-291a-3p mimic for 24 hours (grey). *Oct4:* non-miR-290 target. *P21/Rbl2/Lats2:* validated targets of the miR-290 family.

In addition, there is a highly significant enrichment for miR-290 binding sites in the 3'UTR of bivalent genes when compared to the rest of the genome (odds ratio 1.6; p value < 0.005). Given that miRNAs can target multiple mRNAs, it is tempting to speculate that members of the miR-290 family play a role in co-ordinated post-transcriptional regulation of transcripts from promoters marked by H3K4me3 and H3K27me3 in ES cells. This regulatory network could poise the cell for differentiation by simultaneously targeting multiple transcripts from key developmentally regulated genes, allowing low level transcription whilst ensuring that the mRNAs produced are not translated and therefore do not compromise the pluripotent state.

## 5.3 Expression of PRC2 components is not altered in Dicer KO ES cells

As discussed in Chapter 4, the degree of upregulation of bivalent genes in Dicer KO ES cells is greater than would be expected if caused by loss of miRNA activity alone. I therefore examined whether deregulation at the transcriptional level, as a result of defective PcG function, could contribute to the magnitude of change observed.

To address this possibility, I first analysed RNA and protein levels of core members of the Polycomb repressive complex 2 (PRC2) and global levels of H3K4me3 and H3K27me3 in Dicer WT and KO ES cells. Quantification of RNA levels of key PRC2 components by qPCR indicates that only Jarid2 is significantly deregulated in Dicer KO ES cells at the RNA level, but this is not reflected at the protein level (Fig. 5.2B). Expression of all other PRC2 components analysed does not change in Dicer KO ES cells; and likewise global levels of H3K4me3 and H3K27me3 is not altered.

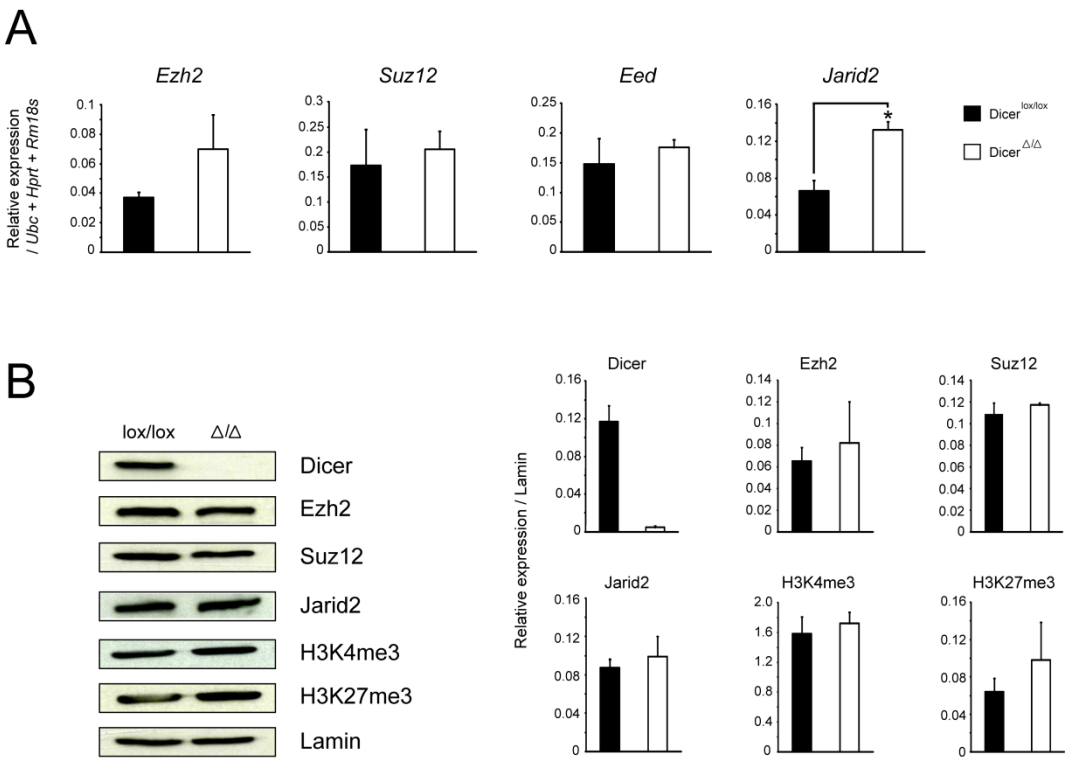



**Figure 5.2: Expression of core PRC2 components and global levels of histone modifications H3K4me3 and H3K27me3 do not change in Dicer KO ES cells.**
(A) qPCR analysis of PcG mRNA levels in Dicer WT (black) and KO (white) ES cells. Asterisks indicate significance (T-test): * = $p < 0.05$. (B) Western blot analysis of protein expression of Dicer, PcG proteins Ezh2 and Suz12, and global levels of H3K4me3 and H3K27me3 in Dicer WT and KO ES cells. Lamin was used a loading control. (C) Quantification of protein levels in (B) using Odyssey Image Studio software (Li-COR). Error bars: s.t.d, of three biological replicates.

## 5.4 Levels of Ezh2 binding and H3K27me3 are reduced at bivalent genes in Dicer KO ES cells

Although core PRC2 components are not globally deregulated in Dicer KO ES cells, I considered the possibility that binding of PRC2 proteins to specific target genes could be defective. Decreased PRC2 binding at bivalent genes could result in loss or reduction in H3K27me3, and subsequently alter the transcriptional state of these genes. To address this, I carried out chromatin immunoprecipitation (ChIP) assays in Dicer WT and KO ES cells. Binding of Ezh2 and levels of

H3K4me3 and H3K27me3 was analysed at the promoters of bivalent genes that were deregulated in Dicer KO ES cells (Fig. 4.10). The levels of both Ezh2 binding and H3K27me3 are reduced at deregulated bivalent genes, but levels of H3K4me3 are not altered. As discussed in Chapter 1, the mechanisms underyling recruitment of the TrxG proteins which catalyse H3K4me3 are not fully characterised; data presented here suggests that loss of miRNA activity impacts PRC2 binding and subsequently levels of H3K27me3, but not the deposition of H3K4me3.

Taking into account that global levels of Ezh2 and H3K27me3 do not change in Dicer KO cells (Fig. 5.2), these results suggest that either PRC2 recruitment to bivalent genes may be impaired in these cells, or alternatively, PRC2 is redistributed to other genomic sites via an as yet undefined mechanism. These possibilities are discussed in 5.6.1.

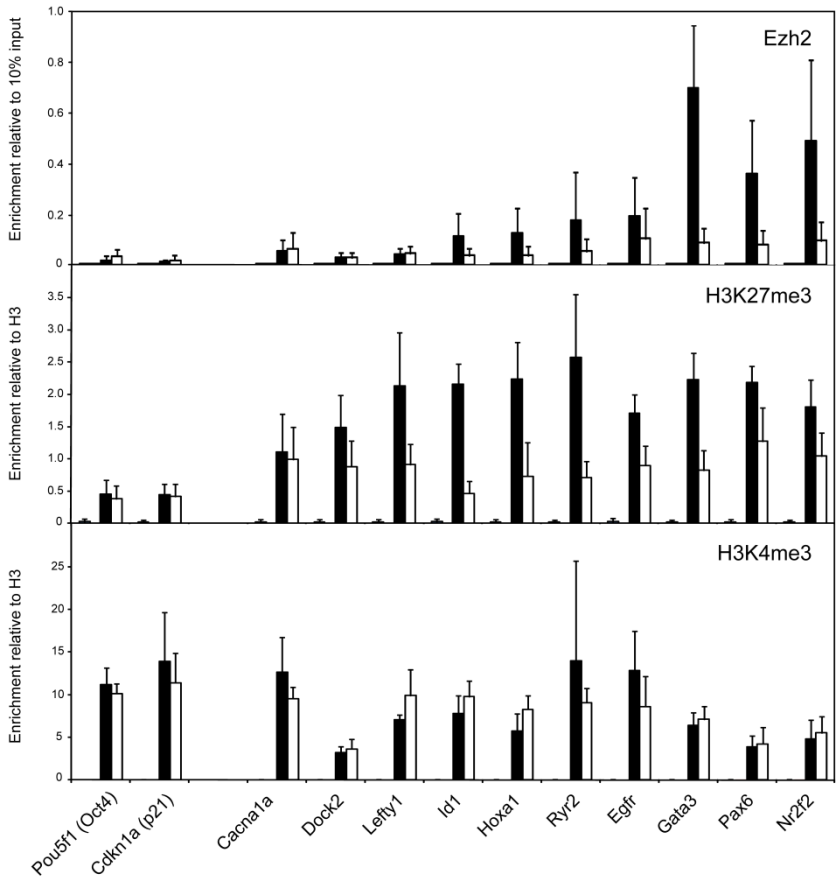



**Figure 5.3: Loss of Ezh2 binding at deregulated bivalent genes in Dicer KO ES cells correlates with reduced levels of H3K27me3 but not H3K4me3 at target genes.**
Quantification of Ezh2 (top panel) H3K27me3 (middle panel) and H3K4me3 (bottom panel) by ChIP in Dicer WT (black) and KO (white) ES cells. *Nanog* and *Oct4* were used as negative controls for Ezh2 and H3K27me3, and as a positive control for H3K4me3. Binding was analysed by qPCR of immunoprecipitated DNA from three independent biological replicates (error bars = s.t.d. of these experiments). Ezh2 binding is normalised relative to 10% input, whereas H3K27me3/H3K4me3 are normalised relative to H3.

## 5.5 Production of full length nascent RNA from bivalent genes is increased in Dicer KO ES cells

Given the observation that both Ezh2 binding and H3K27me3 are decreased at the promoters of bivalent genes in Dicer KO ES cells, I wanted to address whether this phenotype correlates with transcriptional deregulation. I approached this problem by quantification of nascent RNA from PRC2 bound genes in Dicer WT and KO ES cells, to ascertain whether upregulation of bivalent genes in Dicer KO ES cells is partially attributable to transcriptional deregulation as a result of an increase in full length nascent RNA transcripts.

I designed sets of qPCR primers to span exon-intron boundaries at both the 5' and 3' end of the unspliced RNA transcript. Due to the presence of primers in the intronic regions of the gene, this method should therefore only detect nascent unspliced RNA. Using this system, data presented in Fig. 5.4 indicate that levels of nascent RNA from *Oct4* and *p21,* which are not bound by Ezh2 or marked by H3K27me3, do not change significantly in Dicer KO ES cells. Interestingly, quantification of nascent RNA from both the 5' and 3' ends of bivalent genes *Id1, Hoxa1, Egfr* and *Nr2f2* is increased in miRNA-deficient ES cells. This suggests that loss of Ezh2 binding and H3K27me3 correlates with an increase in full-length, nascent RNA from bivalent genes in Dicer KO ES cells. It is therefore likely that transcriptional deregulation of PRC2-bound genes contributes to the global increase in mature mRNA from these genes.

To assess whether transcriptional deregulation of these genes is dependent on ES-expressed miRNAs in Dicer WT and KO ES cells, 5' and 3' nascent RNA from the genes described above was quantified in WT cells transfected with an LNA inhibitor of the miR-290 family (see 4.2.2.2). In three out of four of the bivalent genes analysed, inhibition of the miR-290 family in WT ES cells resulted in upregulation of full-length nascent RNA after 24 hours, indicating that transcriptional deregulation of these genes is miRNA-dependent.

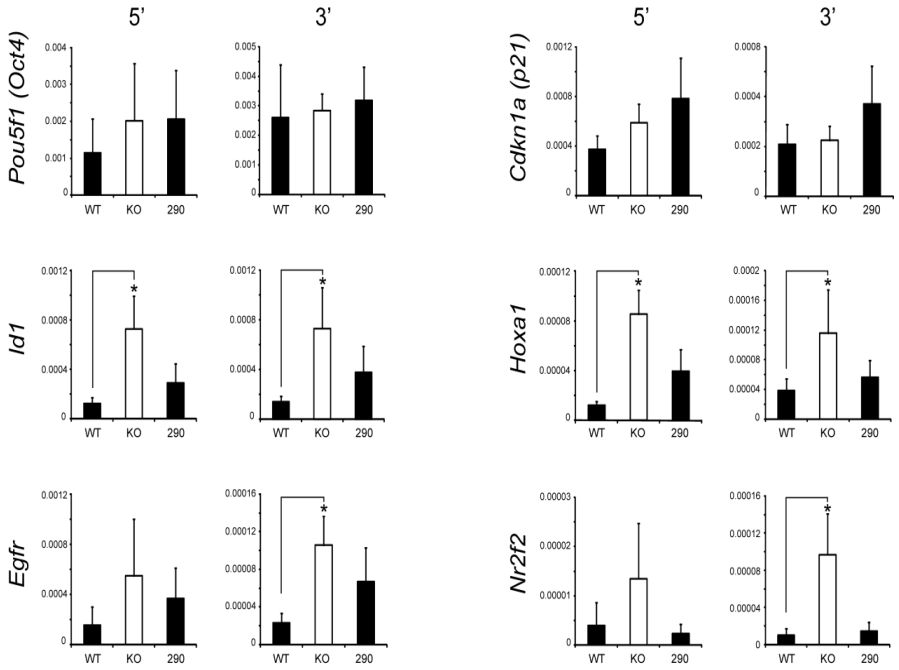



**Figure 5.4: Loss of PcG-binding and H3K27me3 at bivalent genes in Dicer KO ES cells correlates with increased full-length nascent RNA from target genes in a miR-290 dependent manner.** qPCR quantification of nascent RNA from the 5' and 3' end of control genes (*Oct4, p21*) and bivalent genes (*Id1, Hoxa1, Egfr, Nr2f2*) in Dicer WT, KO, or WT cells transfected with LNA inhibitor of the miR-290 family (as indicated). Expression of all genes is normalised to *Ubc, Hprt* and *Rm18s.* Error bars: s.t.d. of three independent biological replicates. Asterisks indicate significance (T-test): * = $p < 0.05$.

## 5.6 Kinetics of expression of mature and nascent RNA from bivalent genes in conditional Dicer KO ES cells suggests that miRNAs function at both the transcriptional and post-transcriptional level in regulation of PRC2 bound genes

Data presented above indicates that bivalent genes are deregulated transcriptionally as well as post-transcriptionally in Dicer-deficient ES cells. To further investigate the contribution of miRNAs to regulation of these genes at the transcriptional and post-transcriptional level, a previously established conditional Dicer KO ES cell system (Nesterova et al., 2008) was employed. I reasoned that if miRNAs function independently at both the transcriptional and post-transcriptional level in regulation of bivalent genes, the kinetics of deregulation of nascent and mature RNA would be distinct in cells acutely depleted of miRNA. If transcriptional deregulation is dependent on miRNAs and caused by loss of PRC2 binding and concurrent reduction in H3K27me3, it could take some time (e.g. several cell cycles) for this phenotype to be apparent. However, if miRNAs directly regulate mature mRNAs from bivalent genes at the post-transcriptional level, it would be expected that this phenotype would be more immediately discernible. Thus, deregulation of levels of mature mRNA would be evident before deregulation of nascent RNA. On the other hand, if upregulation of mature mRNA from bivalent genes is a direct consequence of transcriptional deregulation, it would be expected that levels of nascent RNA would increase prior to, or concurrent with, levels of mature mRNA.

As previously mentioned (4.2.1) the 'wildtype' cells used in this study are derived from ERT2Cre Dicer$^{lox/lox}$ mice, and therefore treatment of these cells with 4-hydroxytamoxifen (4-OHT) results in activation of the Cre recombinase and excision of the RNaseIII domain of the Dicer gene (see schematic of locus (Fig. 3.5A) and description of Cre/LoxP system in Appendix I). As this system was originally characterised by Nesterova *et al* in feeder-dependent culture conditions, it was first necessary to optimise the conditions required to fully delete the Dicer locus and subsequently efficiently deplete cells of miRNAs in the feeder-independent system used in this study.

ERT2Cre Dicer$^{lox/lox}$ ('WT') ES cells were cultured in the presence of either 4-OHT or the carrier (ethanol) at a final concentration of 80nM. Samples for analysis of genomic DNA, mRNA, miRNA and protein were taken at two day intervals. Deletion of the floxed region at the genomic level was observed after two days in culture with 4-OHT (Fig 5.5A), followed by loss of Dicer mRNA (Fig. 5.5B) and protein (Fig. 5.5C) by four days. An 80-90% reduction in miR-291a-3p and miR-92 (two of the most highly expressed miRNAs in ES cells (Houbaviy et al., 2003)) was observed after six days in culture with 4-OHT (Fig.5.5D).

Levels of Dicer mRNA (and miRNA) began to show a slight increase in floxed cells after 10 days in culture with 4-OHT. These data suggest that a subpopulation of cells which escape deletion gain a growth advantage over the miRNA-deficient cells. For this reason, further experiments using this system were limited to a timecourse of 10 days in 4-OHT or EtOH.

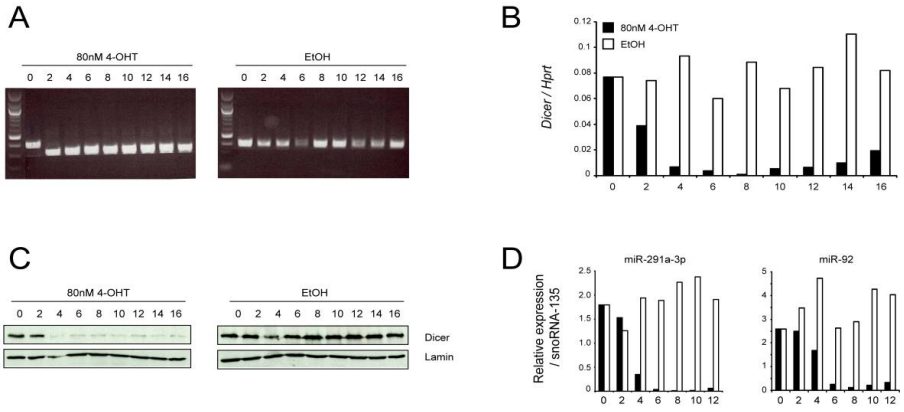



**Figure 5.5: Characterisation of feeder-independent conditional Dicer KO ES cell system.**
(A) Genomic PCR of ERT2Cre Dicer$^{lox/lox}$ ES cells cultured for 16 days in 80nM 4-OHT or EtOH as indicated. Floxed amplicon = 390bp; deleted amplicon = 300bp. (B) qPCR analysis of Dicer mRNA levels in the conditions described in (A). (C) Western blot analysis of Dicer protein levels in the conditions described in (A). Lamin was used as a loading control. (D) qPCR analysis of miRNA levels in ERT2Cre Dicer$^{lox/lox}$ ES cells cultured for 16 days in 80nM 4-OHT (black) or EtOH (white). Analysis was only possible up to 12 days due to insufficient RNA for small RNA extraction/amplification obtained from later timepoints.

Using this ES cell system, I could assess whether mature and nascent RNA levels from bivalent genes show similar or distinct kinetics of deregulation on acute depletion of miRNA activity (Fig. 5.6). The extent of transcriptional deregulation of bivalent genes after 10 days in culture with 4-OHT (assessed by the fold change in levels of nascent RNA at 10 days compared to 0 days) varied considerably amongst the bivalent genes analysed, and levels of nascent RNA only increased more than 3-fold in one of the eight genes analysed. On the other hand, all bivalent genes analysed were consistently upregulated at the post-transcriptional level (as quantified by levels of mature RNA after 10 days culture in the presence of 4-OHT relative to 0 days), with the fold increase in levels of mature RNA exceeding the upregulation of nascent RNA in the case of all bivalent genes assessed. Interestingly, the degree of transcriptional deregulation appeared to correlate with the level of Ezh2 binding (Fig. 5.6).

Data presented in Fig. 5.6 suggests that not only are bivalent genes transcribed in WT ES cells, in accordance with published data discussed in the previous chapter, but that the degree of transcriptional activity varies significantly even within genes commonly marked by a bivalent chromatin signature. Whereas the effect of loss of miRNA activity on levels of nascent RNA from these genes appears to be correlated with PRC2 binding in WT cells, and therefore level of transcription, upregulation of mature mRNA from bivalent genes appears to be independent of PRC2 activity and transcriptional state. Whilst these patterns are preliminary observations based on a few candidate genes, these results support differential and distinct functions for miRNA in regulation of bivalent genes at the transcriptional and post-transcriptional level in ES cells.

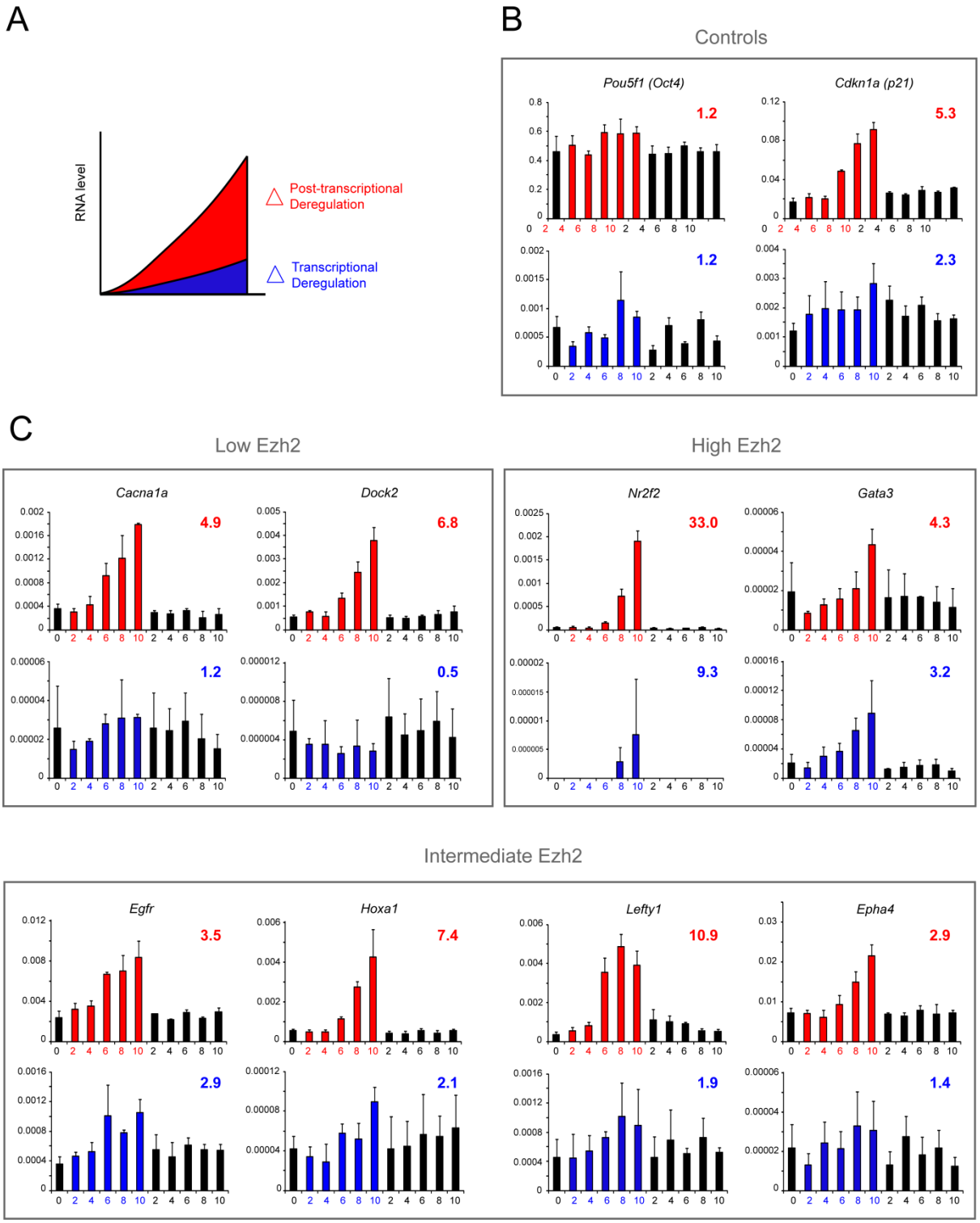



**Figure 5.6: Kinetics of expression of mature and nascent RNA from bivalent genes after acute miRNA depletion.**
(A) Schematic indicating dissociation of post-transcriptional and transcriptional contribution to miRNA-dependent upregulation of bivalent genes using the conditional Dicer KO ES cell system. The fold change in levels of mature (red) and nascent (blue) RNA at 10 days post-miRNA depletion compared to wildtype levels (0d) is shown on each graph in (B) and (C). Where the Δ in mature RNA is greater than Δ in nascent RNA, this strongly suggests that miRNAs play a role primarily in post-transcriptional regulation of mature transcripts, and that miRNA-dependent transcriptional deregulation is a secondary, indirect effect of loss of Dicer function. (B) and (C) qPCR analysis of levels of mature (red) and nascent (blue) RNA in ERT2Cre Dicer$^{lox/lox}$ ES cells treated with 80nM 4-OHT for 10 days. In the case of both forms of RNA, levels of expression of indicated genes in the EtOH control are shown in black. Numbers underneath each graph correspond to the number of days in culture. Genes analysed are grouped according to PcG levels as indicated in Fig.5.3. (B) *Pou5f1 (Oct4):* non-miRNA target. *Cdkn1a (p21):* validated miRNA target. (C) Bivalent genes. Expression of all genes is normalised to *Ubc, Hprt* and *Rm18s.* Error bars: s.t.d. of three independent biological replicates.

## 5.7 Discussion

The aim of this chapter was to establish whether global upregulation of bivalent genes in Dicer KO ES cells is attributable to loss of direct post-transcriptional regulation of these genes by miRNAs, or due to deregulation of the repressive activity of PcG proteins at the transcriptional level. Data presented here suggests that in fact ES-expressed miRNAs function in regulation of bivalent genes at both levels; directly targeting mature mRNA transcripts, and playing an as yet undefined role in PRC2 binding to target genes.

### 5.7.1 Loss of PcG binding at bivalent genes in Dicer KO ES cells results in transcriptional deregulation of genes marked by both H3K4me3 and H3K27me3

My experiments show that global levels of PRC2 components remain unchanged in Dicer KO ES cells, but Ezh2 binding and levels of H3K27me3 are reduced at the promoter regions of bivalent genes in these cells, and production of full-length nascent RNA from candidate genes is increased. Interestingly, the miR-290 family appears to be important in transcriptional as well as post-transcriptional regulation of bivalent genes, as inhibition of these miRNAs in WT ES cells results in an increase in nascent RNA from candidate PRC2 bound genes (Fig. 5.4). These results indicate that transcriptional deregulation of bivalent genes in microRNA-deficient ES cells could contribute to the upregulation of mature RNA levels detected by microarray, as opposed to this phenotype being entirely attributable to loss of post-transcriptional regulation of these genes.

However, the significant enrichment for miR-290 target sites within the 3'UTR of bivalent genes, combined with the fact that these transcripts respond in miRNA inhibition and re-introduction assays, suggests that bivalent genes are direct miRNA targets in ES cells. Whilst additional 3'UTR assays would be informative in further investigating the post-transcriptional role of miRNA in regulation of these genes, data presented in this chapter therefore indicate that miRNAs play a role in regulation of bivalent genes at both the transcriptional and post-transcriptional level in pluripotent cells.

The complexity of the pluripotent state is gradually being uncovered by genome-wide datasets correlating expression of mature and nascent RNA with maps of transcription factor binding and epigenetic modifications. It is now becoming apparent that fully elucidating the regulatory

mechanisms defining and governing bivalency requires consideration of multiple layers of regulation. In the experiments described in this study, genes were defined as 'bivalent' by identification of both H3K4me3 and H3K27me3 chromatin marks in the promoter region of the gene (Ku et al., 2008). A recent study by Brookes *et al* has shown that within the broad classification of bivalency a range of subclasses can be defined by correlating PcG binding, RNA expression level and status of RNA polymerase II (RNAPII) (Brookes et al., 2012). Similar observations have also been made by others using a candidate based approach (De Gobbi et al., 2011). Results shown in this chapter support the ideas purported by these authors, identifying that detectable levels of mature and nascent RNA varies significantly between genes commonly marked by both H3K4me3 and H3K27me3 in WT ES cells. Data presented in this thesis suggests that co-ordinated post-transcriptional regulation of transcripts from bivalent genes by the same highly expressed miRNA family prevents these transcripts from being translated, independent of the variable extent of expression of these genes at the transcriptional level.

Leading on from this, analysis of the kinetics of mature and nascent RNA expression in the conditional Dicer KO system revealed a potential correlation between miRNA activity, PRC2 binding and the transcriptional state of the candidate genes in question. *Nr2f2* and *Gata3* are bound by high levels of Ezh2 (Fig. 5.3), which correlates with a striking upregulation of nascent RNA on acute depletion of miRNA activity (Fig. 5.6). On the other hand, *Cacna1a* and *Dock2* are bound by low levels of Ezh2 despite being marked by both H3K4me3 and H3K27me3, and are not significantly deregulated at the transcriptional level on loss Ezh2 binding. Genes which have an intermediate level of PRC2 binding - *Egfr, Hoxa1, Lefty1* and *Epha4* – exhibit minor deregulation at the transcriptional level when miRNA activity is lost. It should be noted that as the changes in levels of nascent RNA from control genes *Pouf51 (Oct4)* and *Cdkn1a (p21)* are similar to changes in candidate genes bound by intermediate levels of PRC2, further assessment of an additional panel of both control and candidate genes would be required to conclusively demonstrate differences in levels of nascent RNA on acute miRNA depletion between these two groups.

These results indicate that transcriptional deregulation of bivalent genes in Dicer KO ES cells appears to be dependent on the level of PRC2 binding and therefore transcriptional activity of the gene in question, suggesting that the role of miRNA in regulation of transcription from bivalent genes

is indirect and not necessarily globally applicable to all genes marked by H3K4me3 and H3K27me3. On the other hand, upregulation of mature RNA on loss of miRNA activity appears to be common to all bivalent genes analysed, independently of PRC2 binding and transcriptional state. Taken together, these findings suggest that miRNAs play a role in regulation of bivalent genes at both the transcriptional and post-transcriptional level in ES cells.

The complex nature of the correlations between these phenotypes highlights the intricate regulatory mechanisms which maintain the pluripotent state; and suggest that the definition of bivalency as used in this study is simplistic. Whilst the data shown here strongly implicates miRNAs in regulation of bivalent genes at multiple levels, one limitation of the experiments presented in this chapter is the candidate-based approach of the analysis – it would be informative to carry out similar studies genome-wide using RNA-seq to quantify levels of nascent RNA in Dicer KO ES cells, and ChIP-seq to determine genome-wide patterns of PcG-binding and histone modifications. Using these techniques, potential correlations between the degree of transcriptional deregulation, transcription factor binding and histone modifications hinted at by data in Fig. 5.7 could be more thoroughly explored. It could also be useful to correlate data on RNAPII status and DNA methylation to these analyses, as additional key mechanisms involved in transcriptional regulation.

In addition, the multi-faceted role of miRNA in the regulation of bivalent genes exemplifies the difficulty of dissecting phenotypes associated with the Dicer KO system. As discussed in the previous chapter, constitutive loss of all miRNAs in stable Dicer KO ES cells inevitably results in deregulation of a multitude of cellular processes, making identification of directly regulated miRNA processes challenging. The benefit of using the conditional KO system instead is the advantage it holds in being able to uncover primary effect of loss of miRNA activity, as illustrated by the subtle changes in expression shown in Fig. 5.7. Any further analysis of the role of miRNAs in regulation of these genes would benefit from parallel analyses being undertaken in the conditional Dicer KO system to avoid misleading phenotypes attributed to secondary effects of long-term culture without miRNAs.

As data accumulates regarding the multitude of different complexes bound to promoters of bivalent genes in ES cells, population heterogeneity, and the apparently dynamic nature of gene regulation in ES cells which is thought to be governed by fluctuating levels in key pluripotency factors (Chambers et al., 2007; Kalmar et al., 2009) it is clear that gene regulation in ES cells is complex and

sensitive to perturbation by extrinsic and intrinsic factors. The ability of miRNAs to co-ordinately regulate multiple genes confers robustness to biological systems in light of perturbations in the inter- and extracellular environment (Ebert and Sharp, 2012). Data presented in this study provides evidence for co-ordinated regulation of key developmentally regulated genes by ES-expressed miRNAs, suggesting that miRNAs could play a role in stabilising pluripotency within highly heterogenous populations of ES cells, and incorporating both direct and indirect effects of miRNA activity into the regulatory networks governing the pluripotent state.

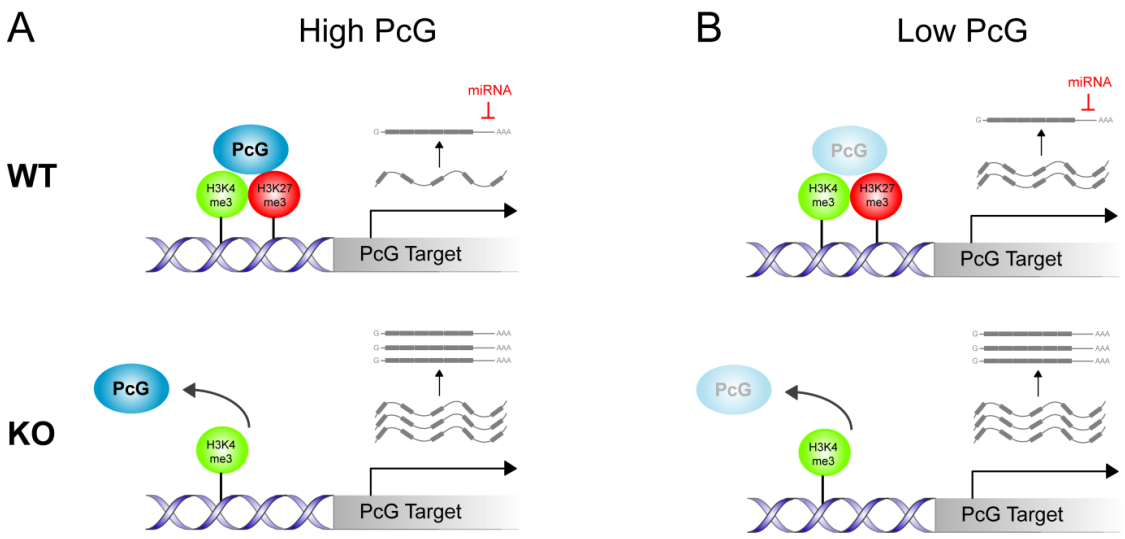



**Figure 5.7: Incorporation of post-transcriptional regulation by miRNA into regulatory networks governing expression of bivalent genes in pluripotent cells.**
(A) In WT cells, bivalent genes bound by high levels of PcG are transcribed at almost undetectable levels. Low levels of mature RNA produced from nascent RNA are negatively regulated by miRNAs. In KO cells, loss of PcG binding results in a significant increase in nascent RNA from bivalent genes, and a combination of increased nascent RNA and loss of post-transcriptional regulation by miRNAs also results in a concurrent increase in levels of mature mRNA. (B) In WT cells, genes bound by low levels of PcG produce slightly more nascent RNA than genes bound by high levels of PcG (A), however mature transcripts from these genes are similarly negatively regulated post-transcriptionally by miRNAs. In KO cells, loss of PcG binding results in a subtle increase in nascent RNA, but the difference between WT and KO is less striking due to these genes being transcribed to some extent in WT cells, as opposed to stringently repressed. However, a similar increase in mature RNA is observed from genes bound by both high and low levels of PcG due to loss of post-transcription regulation by miRNAs.

Although these models illustrate how miRNA-mediated regulation of bivalent genes can be separated into two distinct roles at the transcriptional and post-transcriptional level, they do not provide an explanation for aberrant PcG binding in Dicer KO ES cells. When considering possible mechanisms underlying this phenotype, it is important to note that expression of Ezh2 and H3K27me3 at the population level is not altered in microRNA-deficient cells. As a result, loss of Ezh2 binding at bivalent genes (and associated reduction in H3K27me3) cannot be explained by global downregulation of key PcG proteins. Instead, these data support a model whereby PcG recruitment is defective in Dicer KO ES cells. The mechanistic basis of this phenotype could not be pursued

experimentally in this study due to time constraints. However, potential mechanisms which may explain this phenotype are discussed below.

#### 5.7.1.1 A direct role for miRNA in transcriptional gene silencing?

As discussed in Chapter 1, the role of the RNAi pathway in heterochromatin formation in yeast is well established. In addition, recent studies have extended this mechanism to expressed genomic regions in *S. pombe*, indicating that processing of dsRNAs from nascent RNA transcripts by the RNAi machinery is involved in silencing of active gene loci in lower organisms (Bühler et al., 2006). Studies in human cells have indicated that a similar mechanism of RNA-mediated transcriptional gene silencing (TGS) may persist in mammalian cells; but these observations are based on introduction of exogenous RNAs, and the studies do not resolve whether RNA-mediated TGS is dependent (Morris et al., 2004) or independent (Ting et al., 2005) of DNA methylation of the target gene promoter. Whilst a role for the RNAi pathway in TGS in mammalian cells is not yet well understood; in the absence of conclusive evidence that this mechanism does not persist in higher organisms, it remains plausible that the loss of Dicer in ES cells could potentially directly contribute to upregulation of transcription from bivalent genes. Future studies investigating RNA-mediated TGS in mammalian systems will be informative in establishing whether this mechanism may be involved in the phenotype observed in this study.

#### 5.7.1.2 Does dephosphorylation of Ezh2 abolish ncRNA-mediated PcG recruitment to bivalent genes?

The role of non-coding RNA in Polycomb recruitment to target genes is well documented. In particular, the long non-coding RNAs (lncRNA) HOTAIR (Rinn et al., 2007) and RepA (Zhao et al., 2008) have been shown to mediate PRC2 recruitment to the Hox gene loci and inactive X chromosome respectively, mediating silencing of these regions. Genome-wide analyses suggest that the function of lncRNA may extend beyond these well characterised examples (Guttman et al., 2009; Kapranov et al., 2007), although functionality is yet to be assigned to the majority of these transcripts. Recent studies using RNA- or DNA-protein crosslinking followed by high-throughput sequencing suggests that multiple lncRNAs from different loci interact with PRC2 (Zhao et al., 2010) and that ncRNA-mediated PRC2 binding is sequence-specific, and acts to nucleate broad regions of PcG

binding (Chu et al., 2011). Together, these studies provide evidence that ncRNA-mediated PRC2 recruitment is an important component of gene regulation by PcG.

Recently, two independent groups have shown that phosphorylation of Ezh2 is required for binding of PRC2 to target loci (Chen et al., 2010; Kaneko et al., 2010). In one study, the authors claim that this phenotype is due to increased affinity of the phosphorylated form of Ezh2 for non-coding RNA, enhancing PRC2 recruitment to target genes (Kaneko et al., 2010). Both studies show that in both mouse and human cells, Ezh2 phosphorylation is mediated by CDK1 and CDK2.

Cdkn1a (p21) is a well-documented inhibitor of the kinase activity of both CDK1 and CDK2. As demonstrated by Wang *et al* and independently validated in Dicer KO cells used in this study (Fig. 4.11), *p21* is a direct target of the miR-290 family of miRNAs, and is consequently upregulated in Dicer KO ES cells. This suggests that activity of CDK1 and CDK2 could be downregulated in Dicer KO ES cells. If the above studies are correct, upregulation of p21 and subsequent loss of CDK1 and CDK2 activity in Dicer KO ES cells could therefore plausibly result in reduction of Ezh2 phosphorylation and defective PRC2 recruitment to target genes.

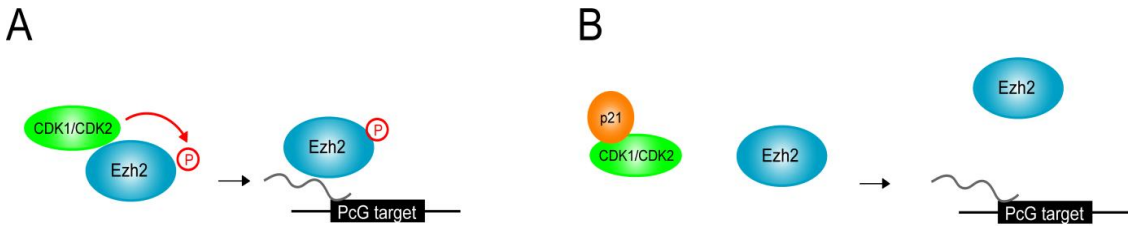



**Figure 5.8: Model (1): Inhibition of CDK1/CDK2 activity in Dicer KO ES cells results in loss of Ezh2 phosphorylation and abolishes ncRNA-mediated recruitment of PRC2 to target genes.**
In WT cells (A) CDK1/CDK2 phosphorylates Ezh2, increasing affinity of the PRC2 complex for ncRNA and subsequently mediating recruitment of PcG to target genes which are efficiently silenced. In Dicer KO cells (B) p21 is upregulated, inhibiting the kinase activity of CDK1/CDK2. Ezh2 is therefore not phosphorylated, does not interact with ncRNA, and is not recruited to target genes; resulting in transcriptional upregulation of these targets.

To investigate this model further, the phosphorylation state of Ezh2 would first need to be tested in Dicer KO ES cells. Currently there is no commercially available antibody for this modification, so this analysis would involve either obtaining the custom derived antibody generated by Kaneko *et al*, or producing a new antibody independently. If phosphorylation of Ezh2 is indeed reduced in Dicer KO ES cells, *in vitro* assays determining the affinity of Ezh2 to ncRNA in WT and KO ES cells would be required; and finally, inhibition of p21 in Dicer KO ES cells to determine if downregulation of p21 activity could restore Ezh2 phosphorylation and subsequent binding to target genes.

#### 5.7.1.3 Is Polycomb redistributed to genomic regions that become hypomethylated in cells with reduced DNA methyltransferase expression?

Alternatively, a number of studies have also suggested that Polycomb complexes are directly recruited to hypomethylated CpG islands. In 2007, Tanay *et al* used a comparative genomics approach to identify over 200 highly conserved genomic regions corresponding to CpG domains and PRC2 binding (Tanay et al., 2007). More recently, two different groups have shown that CpG island chromatin is sufficient to recruit PRC2, independent of the surrounding context of the region (Mendenhall et al., 2010; Lynch et al., 2012). Genome-wide support for this hypothesis has been provided by a study which developed a technique called ChIP-bisulfite-sequencing (ChIP-BS-seq), which can be used to generate high resolution methylation data for regions of the genome associated with a given histone modification or protein of interest (Brinkman et al., 2012). The authors found that H3K27me3 co-localises with DNA methylation genome-wide, except for at CpG islands. In ES cells deficient for the enzymes which catalyse DNA methylation, peaks of H3K27me3 were lost and replaced by broad local enrichments (BLOCs) of the modification associated with hypomethylated DNA; suggesting that domains of CpG-rich chromatin (when aberrantly hypomethylated) can recruit PcG, subsequently resulting in deposition of H3K27me3.

Interestingly, aberrant hypomethylation of several genomic sites in Dicer KO ES cells has been identified previously (Benetti et al., 2008; Nesterova et al., 2008; Sinkonnen et al., 2008), and has been shown to be attributable to downregulation of key enzymes involved in establishment and maintenance of DNA methylation. There are four key DNA methyltransferase enzymes in mammalian cells: the catalytically active Dnmt1, Dnmt3A and Dnmt3B, and one regulatory protein, Dnmt3L (Denis et al., 2011). Dnmt1 is responsible for maintenance of DNA methylation through DNA replication, via recognition of hemi-methylated DNA strands; whilst Dnmt3A/B are required for *de novo* methylation of specific genomic sites. Analysis of the mRNA expression levels of the Dnmts in Dicer KO cells is inconsistent between studies; although whilst different groups do not concur on whether expression of Dnmt1 and Dnmt3B changes in microRNA-deficient ES cells, Dnmt3A and Dnmt3L are consistently observed to be significantly downregulated. This defect corresponds with hypomethylation of the *Xist* promoter (Nesterova et al., 2008) subtelomeric sites and repetitive elements (Benetti et al., 2008) and the *Oct4* promoter (when differentiation is induced) (Sinkonnen et al., 2008) in Dicer KO ES cells.

This phenotype is intriguing in the context of the recent studies discussed above which describe recruitment of PcG to hypomethylated CpG islands. It is plausible that in ES cells constitutively depleted of microRNAs, large regions of the genome are hypomethylated due to downregulation of key DNA methyltransferases in these cells. These loci could subsequently act as novel PcG binding sites, recruiting PcG proteins away from conventional binding sites (such as the promoters of bivalent genes). Thus, peaks of H3K27me3 are reduced in Dicer KO ES cells (Fig. 5.3), whilst global levels of the modification do not change (Fig. 5.2).

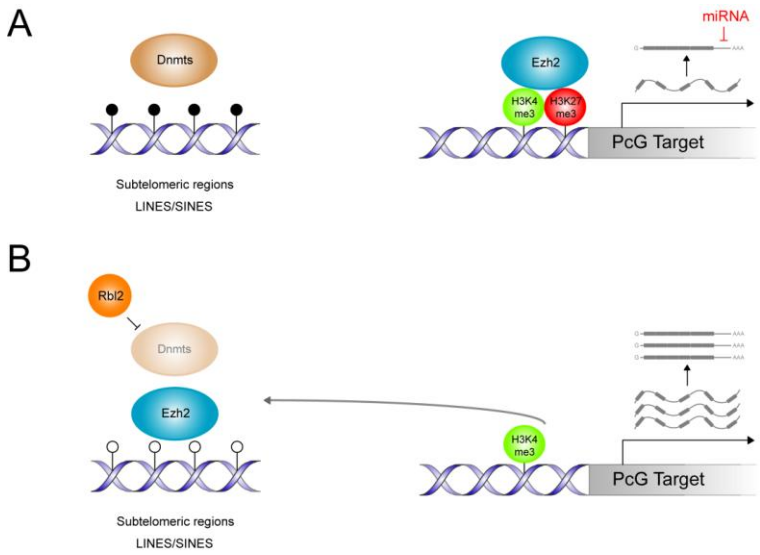



**Figure 5.9: Model (2): Downregulation of Dnmts in Dicer KO ES cells creates novel hypomethylated sites which recruit PcG away from canonical target genes.**
(A) In WT cells, DNA methyltransferases (Dnmts) methylate GC-rich regions such as subtelomeric sites and repetitive elements. PcG remains bound to target genes, characterised by low-level production of RNAs which are post-transcriptionally regulated by ES-specific miRNAs. In KO cells (B) downregulation of key Dnmts (potentially due to upregulation of Rbl2) results in hypomethylation of these regions, which then recruit PcG away from target genes. Loss of PcG at these regions results in reduction of H3K27me3, and increased production of both nascent and processed RNA, in the latter case due to loss of post-transcriptional regulation by miRNAs.

To test this hypothesis, binding of PcG proteins to potential novel hypomethylated sites in Dicer KO ES cells could be determined by ChIP. Dependency of this phenotype on Dnmts could be evaluated by reintroduction of Dnmts into Dicer KO cells and identifying whether binding of PcG to bivalent genes is restored.

Consideration of this mechanism requires a number of additional factors to be taken into account. Firstly, as mentioned previously, published data describing the extent of deregulation of individual Dnmts in different Dicer KO ES cell lines is not always consistent despite hypomethylation being a commonly described phenotype in these cells. Determining which of the Dnmts is

mechanistically responsible for hypomethylation in Dicer KO ES cells is critical to determining the precise role which miRNA may play in this process.

Secondly, the mechanism described above suggests that upregulation of PcG-bound genes in Dicer KO ES cells is indirectly associated with hypomethylation of large regions of the genome in these cells. Given the well-documented function of DNA methylation of promoter regions in gene silencing, a potential direct effect on gene expression due to loss of DNA methylation should also be considered. Genome-wide analysis of 5mC in WT ES cells suggest that around 20% of bivalent genes are marked by some degree of promoter methylation (Fouse et al., 2008) therefore it is plausible that upregulation of at least a subset of bivalent genes in Dicer KO ES cells could therefore be directly contributed to by loss of promoter methylation. Bisulfite sequencing of the promoter regions of deregulated bivalent genes in Dicer KO ES cells could be used to address this.

Following on from this, the Ten-eleven translocation (Tet) family of proteins have been recently identified as responsible for the oxidation of 5mC to 5hmC in mammalian cells (Tahiliani et al., 2009). Several datasets have been independently generated which map the genome-wide distribution of 5mC and 5hmC in pluripotent and differentiating cells (Ficz et al., 2011; Pastor et al., 2011; Wu et al., 2011; Williams et al., 2011). Localisation of 5hmC at the TSS of genes and with exons of expressed genes is consistently detected, despite the range of different methods used to generate the datasets. In addition, 5hmC is shown to be more frequently associated with euchromatin than heterochromatin and/or repetitive sequences, which are known to be marked by 5mC. Significantly, 5hmC (but not 5mC) is also shown to be associated with promoters of bivalent genes more often than would be expected by chance (Pastor et al., 2011; Williams et al., 2011). The precise role of 5hmC in gene regulation is not yet fully understood; correlation of 5hmC with gene expression data in the above studies produces conflicting results. However, if large regions of the genome are hypomethylated in Dicer KO ES cells as a result of downregulation of the Dnmt family of proteins, the subsequent effect this may have on levels of 5hmC at bivalent gene promoters – and as a result, gene expression – should be considered and possibly experimentally investigated.

#### 5.7.1.4 A role for RNAi in long-range interactions between PcG-repressed genes?

Early investigations into the molecular function of PcG proteins identified that these complexes are found in distinct nuclear foci in both *Drosophila* (Buchenau et al., 1998) and mammals

(Saurin et al., 1998). It is has since been shown that these foci represent complex higher order chromatin structures containing PcG proteins, where physical looping of chromatin domains bound by PcG is involved in silencing of the genes within these regions (Lanzuolo et al., 2007). It has been suggested that in *Drosophila* the insulator protein CTCF is involved in mediating these chromatin loops (Ferraiuolo et al., 2010) although the mechanisms underlying the establishment of these domains remains to be elucidated. Of relevance to this study is the finding that in *Drosophila,* RNAi components are required for long-range interactions between PcG bound genes (Grimaud et al., 2006). However, the authors show that RNAi components are not required for PcG recruitment; and therefore whilst chromatin looping may prove to be a functional aspect of PcG-mediated gene regulation in mammals, it is likely that this is independent or downstream of miRNA-dependent PcG recruitment to target genes.

It is important to note that all the potential mechanisms discussed above are not necessarily mutually exclusive. As acknowledged by Zhao *et al*, for example, not all known PRC2 binding sites are associated with PRC2-interacting long ncRNAs, indicating that at other genomic sites it is probable that alternative mechanisms mediate this interaction (Zhao et al., 2010). In light of this, it could be that the role of microRNA in regulation of PcG recruitment, either directly or indirectly, is not comparable for all PcG bound genes. What does appear to be consistent, however, is the function of ES-specific microRNAs in post-transcriptional regulation of transcripts produced from PcG bound genes; strengthening the argument for distinct roles for microRNAs in regulation of bivalent genes at both the transcriptional and post-transcriptional level in ES cells.

In summary, data presented in this chapter presents evidence for regulation of lineage-inappropriate genes in ES cells by miRNAs. These results indicate that post-transcriptional regulation of key developmentally regulated genes attributes a novel function to ES-specific miRNAs in maintenance of the pluripotent state. Deregulation of a subset of bivalent genes at the transcriptional level is a secondary effect of loss of miRNA activity attributable to defective binding of PRC2 in these cells, and therefore it can be concluded that miRNAs function independently at the both transcriptional and post-transcriptional levels to regulate expression of lineage-inappropriate genes in ES cells.

# 6.0 General Discussion and Future Perspectives

The propagation of cell-specific gene expression through mitosis is essential to the preservation of highly specialised lineages in multicellular organisms, yet the mechanisms which facilitate this process are not fully understood. The aim of this study was to determine whether miRNAs play a role in maintenance of lineage-specific gene expression through transitions in the cell cycle which potentially compromise chromatin-dependent gene regulatory mechanisms.

## 6.1 Mammalian microRNAs are inherited through cell division and are functional in daughter cells

Identification of the functional transmission of miRNAs between cells has been well documented in plants and lower organisms, but evidence for a similar phenomenon in mammals is scarce. In this study, I present evidence that exogenously introduced microRNAs are incorporated into endogenous mammalian silencing complexes and are also subject to intracellular degradation processes, indicating that despite their synthetic nature, these RNAs function as endogenous miRNAs inside transfected cells. Transfected miRNA mimics are transmitted through cell division from parent cells to progeny, and are capable of downregulating expression of specific mRNA transcripts in daughter cells. In future work, it would be valuable to pursue emerging technologies capable of tagging and tracking endogenous RNAs to unambiguously investigate whether endogenously derived microRNAs are similarly inherited through mitosis.

## 6.2 MicroRNAs regulate expression of lineage-appropriate genes through the cell cycle in ES cells

The analysis of microarray data described in this study, which assesses gene expression changes in both asynchronous and G1-synchronised populations of microRNA-deficient ES cells, illustrates the difficulty in dissecting the Dicer KO phenotype - multiple pathways and processes are deregulated simultaneously, and physical changes in expression are commonly of an extremely low magnitude and often weak statistical significance. Despite these challenges, confidence in the biological significance of this data could be ascertained from the consistency in overall gene expression changes with previously published datasets generated from independently derived Dicer

KO ES cell lines (Sinkonnen et al., 2008). Analysis of genome-wide expression changes in Dicer KO cells immediately post-cell division identified G1-specific protein phosphorylation and histone gene expression as potential novel mechanisms not previously described to be under miRNA mediated control.

In this study, I intended to test the hypothesis that the mitotic inheritance of microRNAs could serve to stabilise transmission of lineage-specific transcriptional profiles from parent cells to progeny. Whilst very few lineage-inappropriate genes were identified which were likely to be regulated by miRNAs and specifically deregulated in early G1, this finding could potentially be explained by studies suggesting that the majority of the genome is in fact transcribed to some extent in ES cells (Efroni et al., 2008; Guenther et al., 2007). Given this concept, the definition of 'lineage-inappropriate' in ES cells should be reconsidered in the context of the original hypothesis, as genes not required for ES-cell activity are not stringently transcriptionally repressed in these cells. The idea that chromatin-based mechanisms of gene regulation may be compromised through the G2/M to G1 transition, thus resulting in aberrant production of lineage-inappropriate transcripts, is therefore arguably less relevant in pluripotent cells. Future studies in somatic cell lines, such as the Dicer KO 3T3 line derived in this study (Appendix I), would be informative in further investigating whether miRNAs have a specific function through this transition in terminally differentiated cell types. Such experiments would require more sensitive methods of transcriptome analysis in order to detect changes of the magnitude predicted (such as RNA-sequencing or single-cell expression profiling).

Despite the lack of evidence to support a role for miRNA in regulation of gene expression specifically in early G1, data presented in this study demonstrates that without miRNAs, transcripts from genes which are known to be directly regulated by miRNAs accumulate through the cell cycle. These results suggest that miRNAs may play a role in maintaining levels of target mRNA transcripts within a threshold tolerable to the cell concurrent with a global increase in total RNA as the cell grows and prepares for cell division. The pattern of expression and accumulation observed in actively transcribed miRNA targets was also observed in expression of transcripts from bivalent genes, indicating that microRNAs play a role in regulation of both ES-expressed and lineage-inappropriate genes throughout the cell cycle.

## 6.3 Incorporation of combinatorial miRNA-mediated regulation of bivalent genes into regulatory networks governing the pluripotent state

In terminally differentiated cells, lineage-specific gene expression is established and maintained by both sequence-specific and chromatin-dependent mechanisms which silence genes not required for the activity of that particular cell. Whilst recent evidence suggests that this process can in fact be reversed, and somatic cells can be reprogrammed to re-activate silenced genes (Plath and Lowry, 2011) it has been shown that somatic cells can retain an 'epigenetic memory' of their differentiated state, creating a barrier to fully re-establishing pluripotency – highlighting the stringency of mechanisms regulating lineage specification in differentiated cells.

On the other hand, pluripotent stem cells must, by definition, remain in an inherently dynamic transcriptional state, capable of responding to cues to differentiate by rapidly upregulating or silencing relevant panels of genes simultaneously. Maintenance of this 'poised' transcriptional state requires co-ordinated, multi-layer regulation. The ability of a single microRNA to target multiple mRNA transcripts makes these regulatory molecules ideal candidates for mediating transitions between phenotypes which require co-ordinated changes in a large number of genes. In accordance with this, reciprocal expression of the miR-290 and let-7 miRNA families in ES cells has been shown to be involved in the transition from a pluripotent to differentiated state (Melton et al., 2010), and over-expression of ES-specific miRNAs in human somatic cells has been shown to significantly increase the efficiency of reprogramming (Anokye-Danso et al., 2011).

Data presented in this study provides a mechanistic basis for the contribution of ES-expressed miRNAs to maintenance of the pluripotent state. The observation that transcripts from bivalent genes are globally upregulated in Dicer KO ES cells, coupled with reciprocal experiments inhibiting or reintroducing specific miRNAs in Dicer WT and KO ES cells respectively, implies that transcripts from these genes are directly targeted by ES-specific miRNAs. I also found that bivalent genes are deregulated at the transcriptional level in miRNA-deficient ES cells as a result of reduced PcG binding to these genes. Further investigation of the mechanistic basis for loss of PcG binding at bivalent genes in microRNA-deficient ES cells remains an interesting direction for future studies.

In summary, data presented in this thesis incorporates post-transcriptional regulation by microRNAs into the regulatory mechanisms which facilitate the global increase in total cellular RNA

throughout the cell cycle in preparation for cell division. Future experiments using more sensitive assays to fractionate the cell cycle in a more defined manner and gene expression analysis at the single cell level will be required to attempt to identify subtle fluctuations in cell-cycle dependent gene expression, and to uncover the mechanisms which regulate this.

This study also highlights that lineage-specific gene expression in pluripotent cells is defined and regulated in a manner distinct to the activating and repressive mechanisms predominantly governing cell identity in somatic cells. Given that non-ES expressed genes are in fact transcribed in ES cells, this study indicates that combinatorial regulation of bivalent genes by microRNAs at both the transcriptional and post-transcriptional level represents a novel component of the regulatory networks governing the pluripotent state.

In conclusion, data presented in this thesis contributes to current understanding of mechanisms which regulate two fundamental cellular processes. Firstly, cell growth and division is a universal phenomenon applicable to almost all mammalian cell types, and therefore understanding the mechanisms which regulate this process is essential to our understanding of cell biology. Secondly, elucidating the mechanisms which maintain the pluripotent state (and how these change in order to facilitate differentiation) is also critical in terms of current research seeking to generate somatic cells from pluripotent cells *in vitro* for use in cell replacement therapy. Whilst much remains to be established in terms of fully elucidating the intricate molecular networks which regulate the cell cycle and the pluripotent state, incorporation of miRNA-mediated regulation into these systems could aid future studies in both fields.

# 7.0 References


Akbar, A. N., Vukmanovic-Stejic, M., Taams, L. S., and Macallan, D. C. (2007). The dynamic co-evolution of memory and regulatory CD4+ T cells in the periphery. Nature Reviews Immunology *7*, 231–237.

Alleman, M., Sidorenko, L., McGinnis, K., Seshadri, V., Dorweiler, J. E., White, J., Sikkink, K., and Chandler, V. L. (2006). An RNA-dependent RNA polymerase is required for paramutation in maize. Nature *442*, 295–298.

Anokye-Danso, F., Trivedi, C. M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P. J., Epstein, J. a, et al. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell *8*, 376–388.

Arasu, P., Wightman, B., and Ruvkun, G. (1991). Temporal regulation of lin-14 by the antagonistic action of two other heterochronic genes, lin-4 and lin-28. Genes & Development *5*, 1825–1833.

Arroyo, J. D., Chevillet, J. R., Kroh, E. M., Ruf, I. K., Pritchard, C. C., Gibson, D. F., Mitchell, P. S., Bennett, C. F., Pogosova-Agadjanyan, E. L., Stirewalt, D. L., et al. (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proceedings of the National Academy of Sciences of the United States of America *108*, 5003–5008.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nature Cell Biology *8*, 532–538.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. Nature *455*, 64–71.

Banfalvi, G. (2008). Cell cycle synchronization of animal cells and nuclei by centrifugal elutriation. Nature Protocols *3*, 663–673.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215–233.

Bazzini, A. A., Lee, M. T., and Giraldez, A. J. (2012). Ribosome Profiling Shows That miR-430 Reduces Translation Before causing mRNA decay in Zebrafish. Science *336*, 233–236.

Baù, D., Sanyal, A., Lajoie, B. R., Capriotti, E., Byron, M., Lawrence, J. B., Dekker, J., and Marti-Renom, M. a (2011). The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. Nature Structural & Molecular Biology *18*, 107–114.

Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P., et al. (2008). A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. Nature Structural & Molecular Biology *15*, 268–279.

Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E., and Lai, E. C. (2007). Mammalian mirtron genes. Molecular Cell *28*, 328–336.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Bernstein, B. W., and Bamburg, J. R. (2010). ADF/Cofilin: a functional node in cell biology. Trends in Cell Biology *20*, 187–195.

Bernstein, E., Caudy, a a, Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Bernstein, E., Kim, S. Y., Carmell, M. a, Murchison, E. P., Alcorn, H., Li, M. Z., Mills, A. a, Elledge, S. J., Anderson, K. V., and Hannon, G. J. (2003). Dicer is essential for mouse development. Nature Genetics *35*, 215–217.

Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I., and Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. Cell *125*, 1111–1124.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209–213.

Birge, R. B., Kalodimos, C., Inagaki, F., and Tanaka, S. (2009). Crk and CrkL adaptor proteins: networks for physiological and pathological signaling. Cell Communication and Signaling *7*, 13.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Branco, M. R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biology *4*, e138.

Braun, J. E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Molecular Cell *44*, 120–133.

Brenner, S., Jabob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature *190*, 576–581.

Brinkman, a. B., Gu, H., Bartels, S. J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, a., Meissner, a., et al. (2012). Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Research. *22,* 1-11.

Brookes, E., de Santiago, I., Hebenstreit, D., Morris, K. J., Carroll, T., Xie, S. Q., Stock, J. K., Heidemann, M., Eick, D., Nozaki, N., et al. (2012). Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. Cell Stem Cell *10*, 157–170.

Buchenau, P., Hodgson, J., Strutt, H., and Arndt-jovin, D. J. (1998). The Distribution of Polycomb-Group Proteins During Cell Division and Development in. Cell *141*, 469–481.

Burns, D. M., D'Ambrogio, A., Nottrott, S., and Richter, J. D. (2011). CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation. Nature *473*, 105–108.

Bühler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873–886.

Campos, E. I., and Reinberg, D. (2009). Histones: Annotating Chromatin. Annual Review of Genetics *43*, 559–599.

Carroll, N. A. L. O., Erhardt, S., Pagani, M., and Barton, S. C. (2001). The Polycomb-Group Gene Ezh2 Is Required for Early Mouse Development. Molecular and Cellular Biology, *21*, 4330–4336.

Chalfie, M., Horvitz, H. R., and Sulston, J. E. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. Cell *24*, 59–69.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Chang, T.-C., Zeitels, L. R., Hwang, H.-W., Chivukula, R. R., Wentzel, E. a, Dews, M., Jung, J., Gao, P., Dang, C. V., Beer, M. a, et al. (2009). Lin-28B transactivation is necessary for Myc-mediated let-7 repression and proliferation. Proceedings of the National Academy of Sciences of the United States of America *106*, 3384–3389.

Chatterjee, S., and Grosshans, H. (2009). Active turnover modulates mature microRNA activity in Caenorhabditis elegans. Nature *461*, 546–549.

Chekulaeva, M., Mathys, H., Zipprich, J. T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nature Structural & Molecular Biology *18*, 1218–1226.

Cheloufi, S., Dos Santos, C. O., Chong, M. M. W., and Hannon, G. J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584–589.

Chen, K., and Rajewsky, N. (2006). Deep conservation of microRNA-target relationships and 3'UTR motifs in vertebrates, flies, and nematodes. Cold Spring Harbor symposia on quantitative biology *71*, 149–156.

Chen, S., Bohrer, L. R., Rai, A. N., Pan, Y., Gan, L., Zhou, X., Bagchi, A., Simon, J. A., and Huang, H. (2010). Cyclin-dependent kinases regulate epigenetic gene silencing through phosphorylation of Ezh2. Nature Cell Biology *12*, 1108–1114.

Chen, T., and Li, E. (2004). Structure and Function of Eukaryotic DNA Methyltransferases. Current Topics in Developmental Biology *60*, 55–89.

Chen, Z.-xia, and Riggs, A. D. (2011). DNA methylation and demethylation in mammals. The Journal of Biological Chemistry *286*, 18347–18353.

Chi, S. W., Hannon, G. J., and Darnell, R. B. (2012). An alternative mode of microRNA target recognition. Nature Structural and Molecular Biology *19*, 321–328.

Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479–486.

Chiang, H. R., Schoenfeld, L. W., Ruby, J. G., Auyeung, V. C., Spies, N., Baek, D., Johnston, W. K., Russ, C., Luo, S., Babiarz, J. E., et al. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes & Development *24*, 992–1009.

Chu, C., Qu, K., Zhong, F. L., Artandi, S. E., and Chang, H. Y. (2011). Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Molecular Cell *44*, 1–12.

Ciaudo, C., Servant, N., Cognat, V., Sarazin, A., Kieffer, E., Viville, S., Colot, V., Barillot, E., Heard, E., and Voinnet, O. (2009). Highly dynamic and sex-specific expression of microRNAs during early ES cell differentiation. PLoS Genetics *5*, e1000620.

Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G. J., Lawson, N. D., et al. (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694–1698.

Cobb, B. S., Nesterova, T. B., Thompson, E., Hertweck, A., O'Connor, E., Godwin, J., Wilson, C. B., Brockdorff, N., Fisher, A. G., Smale, S. T., et al. (2005). T cell lineage choice and differentiation

in the absence of the RNase III enzyme Dicer. The Journal of Experimental Medicine *201*, 1367–1373.

Cooper, K. (1959). Cytogenetic analysis of the major heterochomatic elements (especially Xh and Y) in Drosophila melanogaster, and the theory of “heterochromatin”. Chromosoma *10*, 535–588.

Crick, F. (1970). Central Dogma of Molecular Biology. Nature *227*, 561–563.

Czech, B., and Hannon, G. J. (2011). Small RNA sorting: matchmaking for Argonautes. Nature reviews. Genetics *12*, 19–31.

Davis, B. N., Hilyard, A. C., Lagna, G., and Hata, A. (2008). SMAD proteins control DROSHA-mediated microRNA maturation. Nature *454*, 56–61.

Dawe, H. R., Minamide, L. S., Bamburg, J. R., and Cramer, L. P. (2003). ADF/cofilin controls cell polarity during fibroblast migration. Current Biology *13*, 252–257.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Denis, H., Ndlovu, M. N., and Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO Reports *12*, 647–656.

Djupedal, I., and Ekwall, K. (2009). Epigenetics: heterochromatin meets RNAi. Cell Research *19*, 282–295.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay. Science *336*, 237–240.

Dodge, J. E., Okano, M., Dick, F., Tsujimoto, N., Chen, T., Wang, S., Ueda, Y., Dyson, N., and Li, E. (2005). Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. The Journal of Biological Chemistry *280*, 17986–17991.

Dostie, J., Richmond, T. a, Arnaout, R. a, Selzer, R. R., Lee, W. L., Honan, T. a, Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Research *16*, 1299–1309.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A., and Noble, W. S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363–367.

Ebert, M. S., and Sharp, P. A. (2012). Roles for MicroRNAs in Conferring Robustness to Biological Processes. Cell *149*, 515–524.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., Bazett-Jones, D. P., Le Grice, S., McKay, R. D. G., Buetow, K. H., et al. (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437–447.

Egli, D., Birkhoff, G., and Eggan, K. (2008). Mediators of reprogramming: transcription factors and transitions through mitosis. Nature Reviews Molecular Cell Biology *9*, 505–516.

Egli, D., Rosains, J., Birkhoff, G., and Eggan, K. (2007). Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. Nature *447*, 679–685.

Ehrlich, M., A-gama-sosa, M., Huang, L.-hsiang, Midgett, R. M., Kenneth, C., Mccune, R. A., and Gehrke, C. (1982). Nucleic Acids Research. *10*, 2709–2722.

Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.-F., Doerks, T., Dorner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. Genes & Development *21*, 2558–2570.

Fabian, M. R., Cieplak, M. K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T. F., et al. (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nature Structural & Molecular Biology *18*, 1211–1217.

Farthing, C. R., Ficz, G., Ng, R. K., Chan, C.-F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS Genetics *4*, e1000116.

Faust, C., Lawson, K. a, Schork, N. J., Thiel, B., and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. Development *125*, 4495–4506.

Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proceedings of the National Academy of Sciences of the United States of America *93*, 10887–10890.

Ferraiuolo, M. a, Rousseau, M., Miyamoto, C., Shenker, S., Wang, X. Q. D., Nadler, M., Blanchette, M., and Dostie, J. (2010). The three-dimensional architecture of Hox cluster silencing. Nucleic Acids Research *38*, 7472–7484.

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. a, Marques, C. J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398–402.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. a, Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116–1122.

Flynt, A. S., and Lai, E. C. (2008). Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nature Reviews Genetics, 831–842.

Fouse, S. D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell *2*, 160–169.

Foussard, H., Ferrer, P., Valenti, P., Polesello, C., Carreno, S., and Payre, F. (2010). LRCH proteins: a novel family of cytoskeletal regulators. PloS One *5*, e12257.

Fox, M. H. (2003). Methods for Synchronizing Mammalian Cells. Methods in Molecular Biology *241*, 11–16.

Galgano, A., Forrer, M., Jaskiewicz, L., Kanitz, A., Zavolan, M., and Gerber, A. P. (2008). Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. PloS One *3*, e3164.

Ganier, O., Bocquet, S., Peiffer, I., Brochard, V., Arnaud, P., Puy, A., Jouneau, A., Feil, R., Renard, J.-P., and Méchali, M. (2011). Synergic reprogramming of mammalian cells by combined exposure to mitotic Xenopus egg extracts and transcription factors. Proceedings of the National Academy of Sciences of the United States of America *108*, 17331–17336.

Gao, G., Bracken, A. P., Burkard, K., Pasini, D., Classon, M., Attwooll, C., Sagara, M., Imai, T., Helin, K., and Zhao, J. (2003). NPAT expression is regulated by E2F and is essential for cell cycle progression. Molecular and Cellular Biology *23*, 2821–2833.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. Molecular Cell *45*, 344–356.

De Gobbi, M., Garrick, D., Lynch, M., Vernimmen, D., Hughes, J. R., Goardon, N., Luc, S., Lower, K. M., Sloane-Stanley, J. a, Pina, C., et al. (2011). Generation of bivalent chromatin domains during cell fate decisions. Epigenetics & Chromatin *4*, 9.

Gottesfeld, J. N., and Forbes, D. (1997). Mitotic repression of the transcriptional machinery. Trends in Biochemical Sciences *22*, 197–202.

Gottwein, E., Cai, X., and Cullen, B. R. (2006). A novel assay for viral microRNA function identifies a single nucleotide polymorphism that affects Drosha processing. Journal of Virology *80*, 5321–5326.

Grabher, C., von Boehmer, H., and Look, a T. (2006). Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. Nature Reviews Cancer *6*, 347–359.

Grewal, S. I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. Current Opinion in Genetics and Development *20*, 134–141.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. Cell *124*, 957–971.

Grépin, C., Robitaille, L., Antakly, T., and Nemer, M. (1995). Inhibition of transcription factor GATA-4 expression blocks in vitro cardiac muscle differentiation. Molecular and Cellular Biology *15*, 4095–4102.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., Eussen, B. H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Guenther, M. G., Levine, S. S., Boyer, L. a, Jaenisch, R., and Young, R. a (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77–88.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., Huarte, M., Zuk, O., Carey, B. W., Cassady, J. P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Han, J., Pedersen, J. S., Kwon, S. C., Belair, C. D., Kim, Y.-K., Yeom, K.-H., Yang, W.-Y., Haussler, D., Blelloch, R., and Kim, V. N. (2009). Posttranscriptional crossregulation between Drosha and DGCR8. Cell *136*, 75–84.

Hartl, P., Gottesfeld, J., and Forbes, D. J. (1993). Mitotic repression of transcription in vitro. The Journal of Cell Biology *120*, 613–624.

He, L., He, X., Lim, L. P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D., et al. (2007). A microRNA component of the p53 tumour suppressor network. Nature *447*, 1130–1134.

He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. Cell Research *21*, 442–465.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Molecular Cell *32*, 276–284.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. The Journal of Biological Chemistry *279*, 48350–48359.

Hochegger, H., Takeda, S., and Hunt, T. (2008). Cyclin-dependent kinases and cell-cycle transitions: does one fit all? Nature Reviews Molecular Cell Biology *9*, 910–916.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A. (1965). Structure of a Ribonucleic Acid. Science *147*, 1462–1465.

Houbaviy, H. B., Murray, M. F., and Sharp, P. a (2003). Embryonic stem cell-specific MicroRNAs. Developmental Cell *5*, 351–358.

Hsieh, C.-S., Lee, H.-M., and Lio, C.-W. J. (2012). Selection of regulatory T cells in the thymus. Nature Reviews Immunology *12*, 157–167.

Hu, S., Ohno, K., Eulenburg, V., Szo, K., Richter, D., Betz, H., and Gene, G. (2003). Inactivation of the Glycine Transporter 1 Gene Discloses Vital Role of Glial Glycine Uptake in Glycinergic Inhibition. *40*, 785–796.

Huang, B., Zhao, J., Lei, Z., Shen, S., Li, D., Shen, G.-X., Zhang, G.-M., and Feng, Z.-H. (2009). miR-142-3p restricts cAMP production in CD4+CD25- T cells and CD4+CD25+ TREG cells by targeting AC9 mRNA. EMBO Reports *10*, 180–185.

Hutvagner, G., McLachlan, J., Pasquinelli, A. E., Balint, E., Tuschl, T., and Zamore, P. D. (2001). A Cellular Function for the RNA-Interference Enzyme Dicer in the Maturation of the let-7 Small Temporal RNA. Science *293*, 834–838.

Ingham, P. W. (1983). Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. Nature, 591–593.

Jackman, J., and O'Conner, P. M. (2001). Methods for Synchronizing Cells at Specific Stages of the Cell Cycle. In Current Protocols in Cell Biology, pp. 8.3.1–8.3.20.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins†. Journal of Molecular Biology *3*, 318–356.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074–1080.

Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. (2009). Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biology *7*, e1000149.

Kaneko, S., Li, G., Son, J., Xu, C.-F., Margueron, R., Neubert, T. A., and Reinberg, D. (2010). Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes & Development *24*, 2615–2620.

Kanellopoulou, C., Muljo, S. a, Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D. M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes & Development *19*, 489–501.

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermuller, J., Hofacker, I. L., et al. (2007). RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. Science *316*, 1484–1488.

Kastan, M. B., and Bartek, J. (2004). Cell-cycle checkpoints and cancer. Nature *432*, 316–323.

Katoh, T., Sakaguchi, Y., Miyauchi, K., Suzuki, T., Kashiwabara, S.-I., Baba, T., and Suzuki, T. (2009). Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes & Development *23*, 433–438.

Kawahara, Y., Zinshteyn, B., Chendrimada, T. P., Shiekhattar, R., and Nishikura, K. (2007a). RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Reports *8*, 763–769.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., and Nishikura, K. (2007b). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137–1140.

Kedde, M., Strasser, M. J., Boldajipour, B., Oude Vrielink, J. a F., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P. M., van Duijse, J., Ørom, U. A., et al. (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273–1286.

Kim, H. H., Kuwano, Y., Srikantan, S., Lee, E. K., Martindale, J. L., and Gorospe, M. (2009). HuR recruits let-7/RISC to repress c-Myc expression. Genes & Development *23*, 1743–1748.

Kim, K., Doi, a, Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M. J., Ji, H., Ehrlich, L. I. R., et al. (2010). Epigenetic memory in induced pluripotent stem cells. Nature *467*, 285–290.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes & Development *20*, 1110–1122.

Knight, S. W., and Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science *293*, 2269–2271.

Kosak, S. T., Skok, J. a, Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science *296*, 158–162.

Kowalczyk, M. S., Hughes, J. R., Garrick, D., Lynch, M. D., Sharpe, J. a, Sloane-Stanley, J. a, McGowan, S. J., De Gobbi, M., Hosseini, M., Vernimmen, D., et al. (2012). Intragenic enhancers act as alternative promoters. Molecular Cell *45*, 447–458.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M. B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H. J., Schübeler, D., et al. (2010a). Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. Cell *141*, 618–631.

Krol, J., Loedige, I., and Filipowicz, W. (2010b). The widespread regulation of microRNA biogenesis, function and decay. Nature Reviews Genetics *11*, 597–601.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genetics *4*, e1000242.

Kuo, C. T., Morrisey, E. E., Anandappa, R., Sigrist, K., Lu, M. M., Parmacek, M. S., Soudais, C., and Leiden, J. M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. Genes & Development *11*, 1048–1060.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Current Biology *12*, 735–739.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nature Cell Biology *9*, 1167–1174.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., and Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. The EMBO Journal *23*, 4051–4060.

Leffak, M. I., Grainger, R., and Weintraub, H. (1977). Conservative Assembly and Segregation of Nucleosomal Histones. Cell *12*, 837–845.

Lei, H., Oh, S. P., Okano, M., Jüttermann, R., Goss, K. a, Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195–3205.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Current Biology *15*, 1501–1507.

Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M., and Cheung, V. G. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. Science *333*, 53–58.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769–773.

Liu, Y., Zhang, P., Li, J., Kulkarni, A. B., Perruche, S., and Chen, W. (2008). A critical function for TGF-beta signaling in the development of natural CD4+CD25+Foxp3+ regulatory T cells. Nature Immunology *9*, 632–640.

Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053–2056.

Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., and Cui, Q. (2008). An analysis of human microRNA and disease associations. PloS One *3*, e3420.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95–98.

Lynch, M. D., Smith, A. J., De Gobbi, M., Flenley, M., Hughes, J. R., Vernimmen, D., Ayyub, H., Sharpe, J. A., Sloane-Stanley, J. A., Sutherland, L., et al. (2012). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. The EMBO Journal *31*, 317–329.

Lyons, a B. (2000). Analysing cell division in vivo and in vitro using flow cytometric measurement of CFSE dye dilution. Journal of Immunological Methods *243*, 147–154.

Lyons, a B., and Parish, C. R. (1994). Determination of lymphocyte division by flow cytometry. Journal of Immunological Methods *171*, 131–137.

Maison, C., Bailly, D., Peters, A. H. F. M., Quivy, J.-P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nature Genetics *30*, 329–334.

Mantei, A., Rutz, S., Janke, M., Kirchhoff, D., Jung, U., Patzel, V., Vogel, U., Rudel, T., Andreou, I., Weber, M., et al. (2008). siRNA stabilization prolongs gene knockdown in primary T lymphocytes. European Journal of Immunology *38*, 2616–2625.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Marson, A., Levine, S. S., Cole, M. F., Frampton, G. M., Brambrink, T., Johnstone, S., Guenther, M. G., Johnston, W. K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell *134*, 521–533.

Martínez-Balbás, M. a, Dey, a, Rabindran, S. K., Ozato, K., and Wu, C. (1995). Displacement of sequence-specific transcription factors from mitotic chromatin. Cell *83*, 29–38.

Marzluff, W. F., Wagner, E. J., and Duronio, R. J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nature Reviews Genetics *9*, 843–854.

Mayr, C., and Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell *138*, 673–684.

Mazzarella, L., Jørgensen, H. F., Soza-Ried, J., Terry, A. V., Pearson, S., Lacaud, G., Kouskoff, V., Merkenschlager, M., and Fisher, A. G. (2011). Embryonic stem cell-derived hemangioblasts remain epigenetically plastic and require PRC1 to prevent neural gene expression. Blood *117*, 83–87.

Melo, S. a, Ropero, S., Moutinho, C., Aaltonen, L. a, Yamamoto, H., Calin, G. a, Rossi, S., Fernandez, A. F., Carneiro, F., Oliveira, C., et al. (2009). A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. Nature Genetics *41*, 365–370.

Melton, C., Judson, R. L., and Blelloch, R. (2010). Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature *463*, 621–626.

Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., Ku, M., and Bernstein, B. E. (2010). GC-Rich Sequence Elements Recruit PRC2 in Mammalian ES Cells. PLoS Genetics *6*, 1–10.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nature reviews. Molecular Cell Biology *7*, 540–546.

Milne, T. a, Kim, J., Wang, G. G., Stadler, S. C., Basrur, V., Whitcomb, S. J., Wang, Z., Ruthenburg, A. J., Elenitoba-Johnson, K. S. J., Roeder, R. G., et al. (2010). Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. Molecular Cell *38*, 853–863.

Miska, E. a, Alvarez-Saavedra, E., Abbott, A. L., Lau, N. C., Hellman, A. B., McGonagle, S. M., Bartel, D. P., Ambros, V. R., and Horvitz, H. R. (2007). Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. PLoS Genetics *3*, e215.

Molnar, A., Melnyk, C. W., Bassett, A., Hardcastle, T. J., Dunn, R., and Baulcombe, D. C. (2010). Small Silencing RNAs in Plants Are Mobile and Direct Epigenetic Modification in Recipient Cells. Science *328*, 872–875.

Mootha, V. K., Lindgren, C. M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature Genetics *34*, 267–273.

Morris, K. V., Chan, S. W.-L., Jacobsen, S. E., and Looney, D. J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science *305*, 1289–1292.

Mseka, T., Bamburg, J. R., and Cramer, L. P. (2007). ADF/cofilin family proteins control formation of oriented actin-filament bundles in the cell body to trigger fibroblast polarization. Journal of Cell Science *120*, 4332–4344.

Muntean, A. G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J. a, Basrur, V., Elenitoba-Johnson, K. S. J., and Hess, J. L. (2010). The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. Cancer Cell *17*, 609–621.

Nesterova, T. B., Popova, B. C., Cobb, B. S., Norton, S., Senner, C. E., Tang, Y. A., Spruce, T., Rodriguez, T. a, Sado, T., Merkenschlager, M., et al. (2008). Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. Epigenetics & Chromatin *1*, 2.

Neumann, P., Yan, H., and Jiang, J. (2007). The centromeric retrotransposons of rice are transcribed and differentially processed by RNA interference. Genetics *176*, 749–761.

Ng, R. K., and Gurdon, J. B. (2008). Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nature Cell Biology *10*, 102–109.

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. Annual Review of Biochemistry *79*, 321–349.

Nolde, M. J., Saka, N., Reinert, K. L., and Slack, F. J. (2007). The Caenorhabditis elegans pumilio homolog, puf-9, is required for the 3'UTR-mediated repression of the let-7 microRNA target gene, hbl-1. Developmental Biology *305*, 551–563.

Obad, S., dos Santos, C. O., Petri, A., Heidenblad, M., Broom, O., Ruse, C., Fu, C., Lindow, M., Stenvang, J., Straarup, E. M., et al. (2011). Silencing of microRNA families by seed-targeting tiny LNAs. Nature Genetics *43*, 371–378.

Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., and Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell *130*, 89–100.

Olins, A. L., and Olins, D. E. (1974). Spheroid Chromatin Units (v Bodies). Science *183*, 330–332.

Ono, S., Minami, N., Abe, H., and Obinata, T. (1994). Characterization of a novel cofilin isoform that is predominantly expressed in mammalian skeletal muscle. The Journal of Biological Chemistry *269*, 15280–15286.

Orom, U. A., Nielsen, F. C., and Lund, A. H. (2008). MicroRNA-10a Binds the 5'UTR of Ribosomal Protein mRNAs and Enhances Their Translation. Molecular Cell *30*, 460–471.

Paige, J. S., Nguyen-Duc, T., Song, W., and Jaffrey, S. R. (2012). Fluorescence imaging of cellular metabolites with RNA. Science *335*, 1194.

Paige, J. S., Wu, K. Y., and Jaffrey, S. R. (2011). RNA mimics of green fluorescent protein. Science *333*, 642–646.

Pardee, A. (1989). G1 events and regulation of cell proliferation. Science *246*, 603–608.

Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E., and Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. The EMBO Journal *23*, 4061–4071.

Pasquinelli, a E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Müller, P., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86–89.

Pasquinelli, A. E. (2012). MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nature Reviews Genetics *13*, 271–282.

Pastor, W. a, Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., McLoughlin, E. M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394–397.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W. M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R. M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Molecular Cell *38*, 603–613.

Peters, A. H. F. M., Mermoud, J. E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., and Jenuwein, T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nature Genetics *30*, 77–80.

Petersen, C. P., Bordeleau, M.-E., Pelletier, J., and Sharp, P. a (2006). Short RNAs repress translation after initiation in mammalian cells. Molecular Cell *21*, 533–542.

Pevny, L. H., Sockanathan, S., Placzek, M., and Lovell-Badge, R. (1998). A role for SOX1 in neural determination. Development *125*, 1967–1978.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research *29*, e45.

Pillai, R. S., Bhattacharyya, S. N., Artus, C. G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. Science *309*, 1573–1576.

Plath, K., and Lowry, W. E. (2011). Progress in understanding reprogramming to the induced pluripotent state. Nature Reviews Genetics *12*, 253–265.

Polager, S., and Ginsberg, D. (2008). E2F - at the crossroads of life and death. Trends in Cell Biology *18*, 528–535.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nature Biotechnology *28*, 1057–1068.

Probst, A. V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nature Reviews Molecular Cell Biology *10*, 192–206.

Ramachandran, V., and Chen, X. (2008). Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. Science *321*, 1490–1492.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469–474.

Raver-Shapira, N., Marciano, E., Meiri, E., Spector, Y., Rosenfeld, N., Moskovits, N., Bentwich, Z., and Oren, M. (2007). Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. Molecular Cell *26*, 731–743.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, a E., Bettinger, J. C., Rougvie, a E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901–906.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. a, Goodnough, L. H., Helms, J. a, Farnham, P. J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Roberts, A. P. E., Lewis, A. P., and Jopling, C. L. (2011). miR-122 activates hepatitis C virus translation by a specialized mechanism requiring particular RNA components. Nucleic Acids Research *39*, 7716–7729.

Ruby, J. G., Jan, C. H., and Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83–86.

Ryan, J. A. (Corning I. L. S. Cell Cloning by Serial Dilution in 96 Well Plates Protocol. Available at: http://www.level.com.tw/html/ezcatfiles/vipweb20/img/img/34963/3-2Single_cell_cloning_protocol.pdf [Accessed January 12, 2010].

Sabogal, A., Lyubimov, A. Y., Corn, J. E., Berger, J. M., and Rio, D. C. (2010). THAP proteins target specific DNA sites through bipartite recognition of adjacent major and minor grooves. Nature Structural and Molecular Biology *17*, 117–123.

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., and Burge, C. B. (2008). Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. Science *320*, 1643–1647.

Satyanarayana, A., and Kaldis, P. (2009). Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. Oncogene *28*, 2925–2939.

Sauer, B. (1987). Functional Expression of the cre-lox Site-Specific Recombination System in the Yeast Saccharomyces cerevisiae. Molecular and Cellular Biology *7* (6). 2087-2096.

Saurin, a J., Shiels, C., Williamson, J., Satijn, D. P., Otte, a P., Sheer, D., and Freemont, P. S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. The Journal of Cell Biology *142*, 887–898.

Savatier, P., Huang, S., Wiman, K. G., and Samarut, J. (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. Oncogene *9*, 809–818.

Schrider, D. R., Gout, J.-F., and Hahn, M. W. (2011). Very few RNA and DNA sequence differences in the human transcriptome. PloS One *6*, e25842.

Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nature Reviews Molecular Cell Biology *12*, 799–814.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63.

Sinkonnen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C. G., Zavolan, M., Svoboda, P., and Filipowicz, W. (2008). MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. Nature Structural and Molecular Biology *15*, 259–267.

Sood, P., Krek, A., Zavolan, M., Macino, G., and Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. Proceedings of the National Academy of Sciences of the United States of America *103*, 2746–2751.

Soutschek, J., Akinc, A., Bramlage, B., Charisse, K., Constien, R., Donoghue, M., Elbashir, S., Geick, A., Hadwiger, P., Harborth, J., et al. (2004). Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. Nature *432*, 173–178.

Stack, S. M., Brown, D. B., and Dewey, W. C. (1977). Visualization of interphase chromosomes. Journal of Cell Science *26*, 281–299.

Stam, M., Belele, C., Dorweiler, J. E., and Chandler, V. L. (2002). Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes & Development *16*, 1906–1918.

Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*, 1133–1146.

Stead, E., White, J., Faast, R., Conn, S., Goldstone, S., Rathjen, J., Dhingra, U., Rathjen, P., Walker, D., and Dalton, S. (2002). Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. Oncogene *21*, 8320–8333.

Stein, R., Gruenbaum, Y., Pollack, Y., and Razin, A. (1982). Clonal inheritance of the pattern of DNA methylation in mouse cells Biochemistry : Biochemistry *79*, 61–65.

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nature Cell Biology *9*, 1428–1435.

Strahl, B. D., and Allis, D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., and Ebert, B. L. (2005). Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide. October.

Suh, M.-R., Lee, Y., Kim, J. Y., Kim, S.-K., Moon, S.-H., Lee, J. Y., Cha, K.-Y., Chung, H. M., Yoon, H. S., Moon, S. Y., et al. (2004). Human embryonic stem cells express a unique set of microRNAs. Developmental Biology *270*, 488–498.

Sulston, J. E., and Horvitz, H. R. (1981). Abnormal cell lineages in mutants of the nematode Caenorhabditis elegans. Developmental Biology *82*, 41–55.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. a, Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Tanay, A., O'Donnell, A. H., Damelin, M., and Bestor, T. H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proceedings of the National Academy of Sciences of the United States of America *104*, 5521–5526.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J. R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K.-ichi (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Research *38*, 8164–8177.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. Cell *148*, 664–678.

Tay, Y., Zhang, J., Thomson, A. M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124–1128.

Terasima, T., and Tolmach, L. J. (1963). Growth and nucleic acid synthesis in sychronously dividing populations. Experimental Cell Research *30*, 344–362.

Ting, A. H., Schuebel, K. E., Herman, J. G., and Baylin, S. B. (2005). Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. Nature Genetics *37*, 906–910.

Todaro, G. J., and Green, H. (1963). Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. The Journal of Cell Biology *17*, 299–313.

Tritschler, F., Huntzinger, E., and Izaurralde, E. (2010). Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of deja vu. Nature Reviews Molecular Cell Biology *11*, 379–384.

Tsuchiya, S., Oku, M., Imanaka, Y., Kunimoto, R., Okuno, Y., Terasawa, K., Sato, F., Tsujimoto, G., and Shimizu, K. (2009). MicroRNA-338-3p and microRNA-451 contribute to the formation of basolateral polarity in epithelial cells. Nucleic Acids Research *37*, 3821–3827.

Vartiainen, M. K., Mustonen, T., Mattila, P. K., Ojala, P. J., Thesleff, I., Partanen, J., and Lappalainen, P. (2002). The Three Mouse Actin-depolymerizing Factor / Cofilins Evolved to Fulfill Cell-Type – specific Requirements for Actin Dynamics. Molecular Biology of the Cell *13*, 183–194.

Voinnet, O., and Baulcombe, D. C. (1997). Systemic signalling in gene silencing. Nature *389*, 553.

Voinnet, O., Vain, P., Angell, S., and Baulcombe, D. C. (1998). Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. Cell *95*, 177–187.

Wang, F., Koyama, N., Nishida, H., Haraguchi, T., Reith, W., and Tsukamoto, T. (2006). The assembly and maintenance of heterochromatin initiated by transgene repeats are independent of the RNA interference pathway in mammalian cells. Molecular and Cellular Biology *26*, 4028–4040.

Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. a, Gupta, R. a, et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J. E., Baehner, L., and Blelloch, R. (2008). Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. Nature Genetics *40*, 1478–1483.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nature Genetics *39*, 380–385.

White, J., and Dalton, S. (2005). Cell Cycle Control of Embryonic Stem Cells. Stem Cell Reviews *1*, 131–138.

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. a C., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343–348.

Winston, W. M., Molodowitch, C., and Hunter, C. P. (2002). Systemic RNAi in C. elegans requires the putative transmembrane protein SID-1. Science *295*, 2456–2459.

Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E., and Zhang, Y. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes & Development *25*, 679–684.

Xu, C.-R., Cole, P. A., Meyers, D. J., Kormish, J., Dent, S., and Zaret, K. S. (2011). Chromatin "Prepattern" and Histone Modifiers in a Fate Choice for Liver and Pancreas. Science *332*, 963–966.

Yamagata, K., Fujiyama, S., Ito, S., Ueda, T., Murata, T., Naitou, M., Takeyama, K.-I., Minami, Y., O'Malley, B. W., and Kato, S. (2009). Maturation of microRNA is hormonally regulated by a nuclear receptor. Molecular Cell *36*, 340–347.

Yang, J.-S., and Lai, E. C. (2011). Alternative miRNA Biogenesis Pathways and the Interpretation of Core miRNA Pathway Mutants. Molecular Cell *43*, 892–903.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Young, D. W., Hassan, M. Q., Pratap, J., Galindo, M., Zaidi, S. K., Lee, S.-hee, Yang, X., Xie, R., Javed, A., Underwood, J. M., et al. (2007). Mitotic occupancy and lineage-specific transcriptional control of rRNA genes by Runx2. Nature *445*, 442–446.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Steward, R., and Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. Science *307*, 932–935.

Yu, M., Mazor, T., Huang, H., Huang, H.-T., Kathrein, K. L., Woo, A. J., Chouinard, C. R., Labadorf, A., Akie, T. E., Moran, T. B., et al. (2012). Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors. Molecular Cell *45*, 330–343.

Zaidi, S. K., Young, D. W., Montecino, M., van Wijnen, A. J., Stein, J. L., Lian, J. B., and Stein, G. S. (2011). Bookmarking the Genome: Maintenance of Epigenetic Information. The Journal of Biological Chemistry *286*, 18355–18361.

Zhang, H., Hao, Y., Yang, J., Zhou, Y., Li, J., Yin, S., Sun, C., Ma, M., Huang, Y., and Xi, J. J. (2011). Genome-wide functional screening of miR-23b as a pleiotropic modulator suppressing cancer metastasis. Nature Communications *2*, 554.

Zhao, J., Dynlacht, B., Imai, T., Hori, T., and Harlow, E. (1998). Expression of NPAT, a novel substrate of cyclin E-CDK2, promotes S-phase entry. Genes & Development *12*, 456–461.

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., Kingston, R. E., Borowsky, M., and Lee, J. T. (2010). Genome-wide Identification of Polycomb-associated RNAs by RIP-Seq. Molecular Cell *40*, 939–953.

Zhao, J., Sun, B. K., Erwin, J. a, Song, J.-J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756.

Zheng, G. X. Y., Ravi, A., Calabrese, J. M., Medeiros, L. a, Kirak, O., Dennis, L. M., Jaenisch, R., Burge, C. B., and Sharp, P. a (2011). A latent pro-survival function for the mir-290-295 cluster in mouse embryonic stem cells. PLoS Genetics *7*, e1002054.

Zhou, V. W., Goren, A., and Bernstein, B. E. (2011). Charting histone modifications and the functional organisation of mammalian genomes. Nature Reviews Genetics *12*, 7–18.

Zilberman, D., Cao, X., and Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

# 8.0 Appendices

## I. Derivation of stable and conditional Dicer KO 3T3 fibroblasts

Due to the unique properties of embryonic stem cells, in their capacity to self-renew and differentiate, I considered that it would be informative to conduct parallel investigations of the role of microRNAs in gene expression through specific stages of the cell cycle in a terminally differentiated cell type. To this end, I derived stable and conditional Dicer KO 3T3 fibroblasts from the same mice used by Nesterova *et al* to establish the ES cell lines described in the main body of this work.

### a. Derivation of conditional Dicer KO 3T3 fibroblasts

To establish a 3T3 fibroblast line from ERT2Cre x $Dicer^{lox/lox}$ mice (Cobb et al., 2005b), I followed a previously described protocol for immortalisation of embryonic fibroblasts (Todaro and Green, 1963). Briefly, I disaggregated 17-19 day old embryos from ERT2Cre x $Dicer^{lox/lox}$ and ERT2Cre x $Dicer^{wt/wt}$ mice, and plated each embryo in a 60mm dish. Every three days, I trypsinised, counted, and replated cells at a density of $3 \times 10^5$ per 60mm dish. Several dishes were kept in culture at all times, so that if the number of cells counted per dish fell below $3 \times 10^5$ when trypsinised, dishes could be pooled so that the number of cells re-plated was always consistent.

As illustrated in Fig. I.I, the growth rate of $Dicer^{wt/wt}$ and $Dicer^{lox/lox}$ 3T3 was approximately equivalent between lines throughout the derivation process. After 20-25 days in culture, both populations of cells underwent a crisis point whereby the cells ceased to proliferate. However, this was anticipated (Todaro and Green, 1963), and after continually trypsinising and replating these cells at the same density for an additional two weeks, the growth rate of both populations began to increase. After the three months in culture illustrated in Fig. 8.1, the growth rate of both $Dicer^{wt/wt}$ and $Dicer^{lox/lox}$ 3T3 stabilised and remained consistent at approximately tripling of the plated population every three days (data not shown).

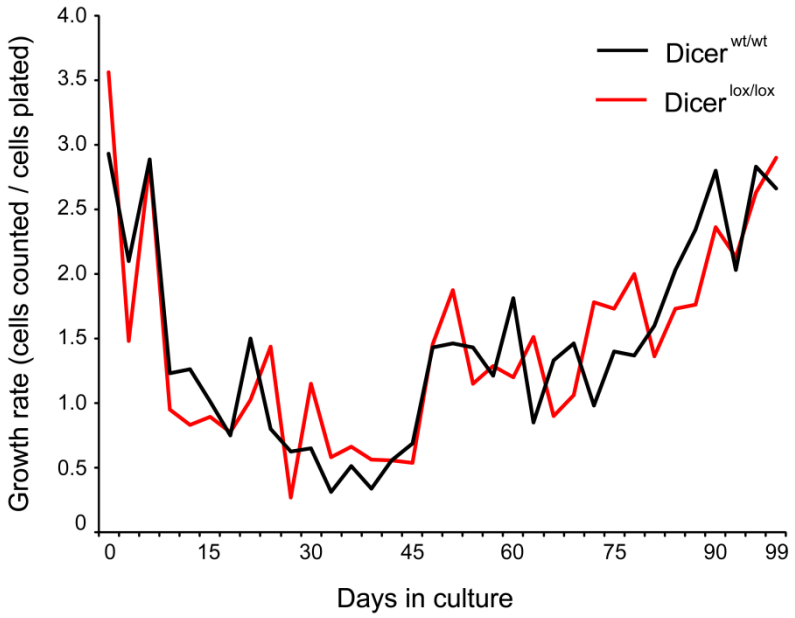



**Figure 8.1: Immortalisation of Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 fibroblasts.**
Growth rate (no. of cells counted after 3 days growth / no. of cells plated) of Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 fibroblasts from 0 to 99 days in culture.

Due to the requirement to pool cultures from different embryos in order to maintain viability of the cultures during immortalisation, the cell populations of Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 described were highly heterogeneous, which could make analysis of data obtained using these cells in later experiments difficult to interpret. To alleviate this problem, I sub-cloned both Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 lines using a serial dilution method (Ryan), and expanded surviving clonal populations for further analysis.

Next, it was necessary to optimise the conditions required to acutely deplete these cells of microRNAs using the tamoxifen-mediated Cre-loxP system. The Cre enzyme is a site-specific recombinase which specifically excises DNA flanked by the loxP sequences (Sauer, 1987). In the mice used to derive the 3T3 lines described here (Cobb et al., 2005b) the catalytic RNase III domain required for Dicer-dependent RNA processing is flanked by loxP sites (Fig. 3.5A). Fusion of the Cre enzyme to the ligand-binding domain of the human oestrogen receptor (ER), which is activated by binding of tamoxifen, generates a conditional, inducible system to activate Cre (Feil et al., 1996). Therefore, to induce Cre in the 3T3 lines derived in this study, I next optimised the amount of tamoxifen required to mediate deletion of the RNase III domain whilst maximising cell viability.

Based on conditions used by Nesterova *et al* to activate Cre in the ES cell lines derived from the same mice,I cultured Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 in final concentrations of 800nM and 80nM of 4-

hydroxytamoxifen (4-OHT) for 16 days. As shown in Fig. A2, culturing both Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 in 80nM 4-OHT appeared to mediate deletion of the RNase III domain as effectively as 800nM 4-OHT; by 4 days in culture, the flanked region could no longer be detected at the genomic level in both conditions (Fig. 8.2A). The WT allele remained unaffected despite culture in either concentration of 4-OHT. Equivalently, after 6 days in culture, Dicer mRNA levels were reduced by 95% in 80nM 4-OHT, and 97% in 800nM 4-OHT (Fig. 8.2B).

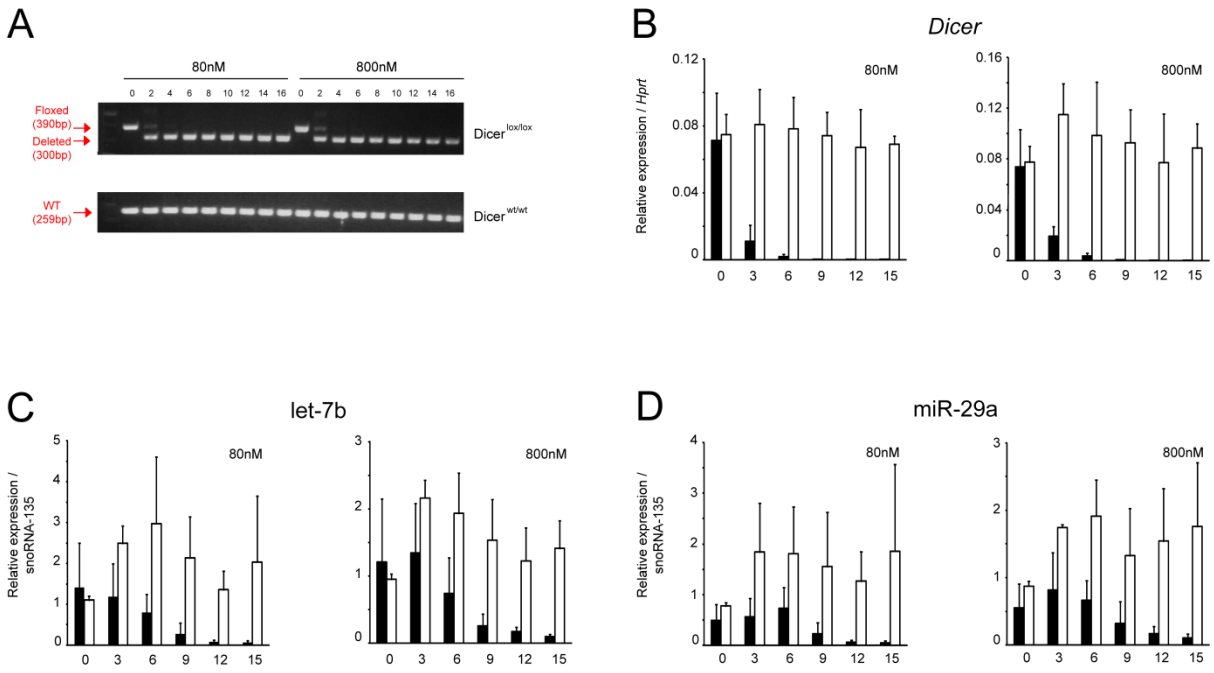



**Figure 8.2: Optimisation of tamoxifen-mediated acute depletion of miRNAs in ERT2Cre Dicer$^{lox/lox}$ 3T3 fibroblasts.**
(A) Genomic PCR of RNase III domain in ERT2Cre x Dicer$^{wt/wt}$ and ERT2Cre x Dicer$^{lox/lox}$ 3T3 fibroblasts treated with 80nM or 800nM 4-OHT as indicated. Numbers above the images indicate the number of days in culture. qPCR analysis of Dicer (B), let-7b (C) and miR-29a (D) levels in conditions described in (A). Numbers along the x axis indicate days in culture. Black bars: Dicer$^{lox/lox}$. White bars: Dicer$^{wt/wt}$. Error bars: s.t.d. of two independent biological replicates.

### b. Derivation of stable Dicer KO 3T3 fibroblasts

Having successfully derived a conditional Dicer KO 3T3 fibroblast line (and corresponding WT control) I then wanted to establish a stable Dicer KO cell line, in order to obtain a comparable system in terminally differentiated cells to the constitutive Dicer KO ES cell line used in the main text of this study. To do this, I cultured ERT2Cre x Dicer$^{lox/lox}$ 3T3 cells in 80nM 4-OHT for 11 days. After this time, I subcloned surviving cells were by serial dilution as above. Only three clones survived expansion from a single cell. I analysed these populations for expression of *Dicer* and three miRNAs highly expressed in fibroblasts (Fig. 8.3).

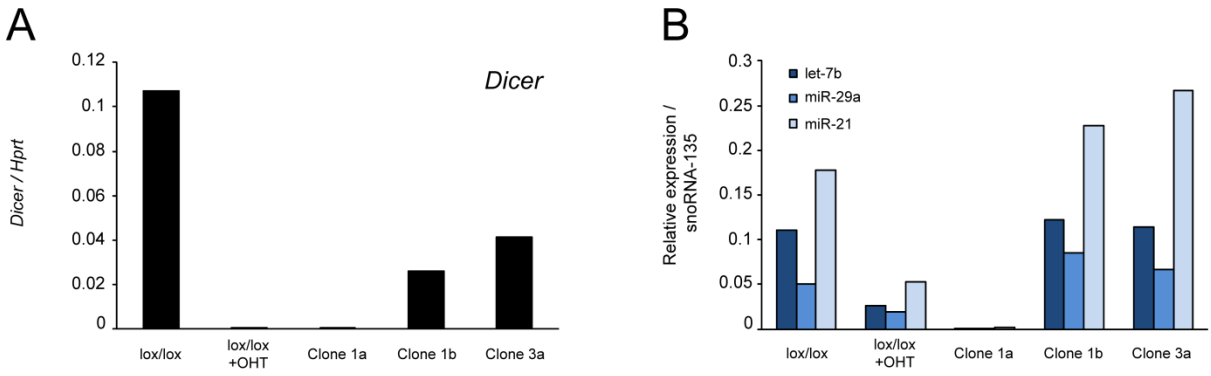



**Figure 8.3: Derivation of stable Dicer KO 3T3 fibroblasts.**
(A) qPCR analysis of Dicer expression in ERT2Cre x Dicer$^{lox/lox}$ 3T3 cells before 4-OHT treatment (lox/lox); after 11 days culture in 80nM 4-OHT (lox/lox +OHT) and in clones which survived subcloning and expansion from the parental tamoxifen treated population shown here. (B) qPCR analysis of miRNA levels in the populations described in (A).

This analysis showed that only one of the three clones had in fact efficiently and stably deleted Dicer and was subsequently miRNA-deficient. This is unsurprising for a number of reasons. Firstly; embryonic lethality of Dicer KO mice illustrates the fundamental biological role of these RNAs (Bernstein et al., 2003), therefore it is perhaps to be expected that cells entirely deficient of miRNAs are predominantly not viable. Secondly; the difficulty in deriving constitutive Dicer KO cells is reflected in the problems also encountered by Nesterova *et al* during establishment of the Dicer KO ES cells (Nesterova et al., 2008) suggesting that the difficulties observed here are not specific to the method used to derive these cell lines, but to the biology of microRNA-deficient cells. Finally; during optimisation of a similar conditional Dicer KO system in the ES cells, it was observed that if kept in culture for prolonged periods of time, subpopulations of ERT2Cre x Dicer$^{lox/lox}$ ES cells which had escaped deletion began to emerge (Fig. 5.5B,D). The fact that clones 1b and 3a still express miRNAs is likely to be attributable to this phenomenon.

### c. Optimisation of mitotic arrest and release assay in Dicer WT and KO 3T3 fibroblasts

Next, I attempted the mitotic arrest and release assay (as described in 4.2.1) in the conditional Dicer KO 3T3 cell line. I cultured Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 fibroblasts in 80nM 4-OHT for 10 days, then treated the cells with 20ng/ml demecolcine for 4 hours. After this point, flasks were gently agitated to release mitotic cells. I then washed the mitotic cells in cold PBS and replated them in demecolcine-free media for 2 hours, recapitulating conditions used in the ES cell system. However, I found that after two hours, only 40-50% of both Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ cells cells had progressed into G1 (Fig. 8.4). I then repeated the experiment, this time allowing cells 3 – 4 hours in culture post-

mitotic arrest. Whilst this did increase the proportion of cells in early G1, I then doubted whether these populations could be considered as representative of cells immediately post-cell division. As populations of cells in early G1 were specifically required for comparative assays to be carried out 3T3 fibroblasts in parallel with the ES cell studies, these populations were therefore not used for further analysis.

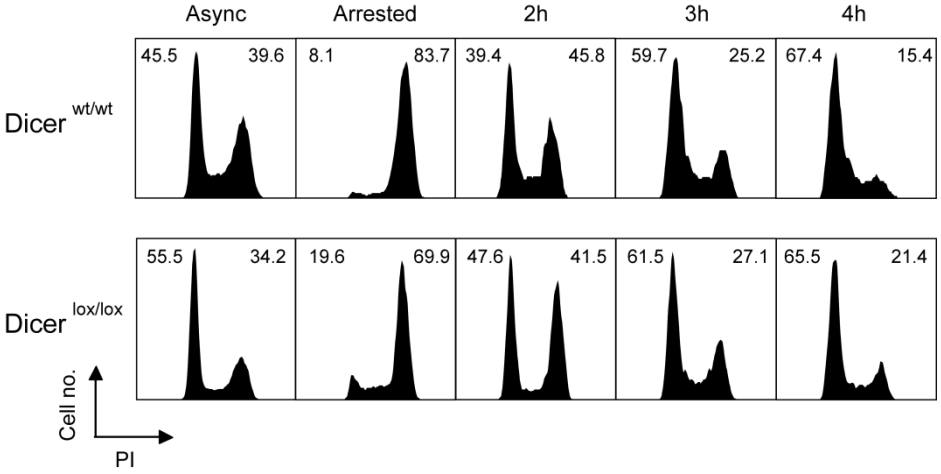



**Figure 8.4: The mitotic shakeoff and release assay is not suitable for obtaining populations of Dicer KO 3T3 fibroblasts immediately post-cell division.**
FACs analysis of cell cycle profiles of populations of Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 cells stained with propidium iodide (PI) obtained using the mitotic arrest and release assay. The numbers in the top left and right-hand corners of each graph indicate the percentage of live cells in G1 and G2/M respectively. Gates are not shown for clarity.

It would have been useful, if time were available, to attempt alternative cell cycle synchronisation methods in Dicer$^{wt/wt}$ and Dicer$^{lox/lox}$ 3T3 fibroblasts, enabling equivalent experiments to be performed in parallel with the Dicer KO ES cells described in the main text. This was not possible during the course of this study, but remains a key direction for future work.

## II. MicroRNA-mediated regulation of the actin cytoskeleton

During the process of expanding the stable Dicer KO 3T3 cell line (clone 1a) described in Appendix I, I observed that these cells exhibited a distinct morphology compared to WT cells. When polymerised actin was visualised in these cells by Immunofluorescence (Fig. 8.5), I saw that this morphology was characterised by more disorganised internal actin structure, and reduced cell-cell adhesion. Additionally, I noticed that the morphological defects identified in the Dicer KO 3T3 cells were reminiscent of loss of colony formation seen in Dicer KO ES cells grown in feeder-independent conditions (Fig. 4.2C). Based on these observations, and the fact that data presented in this study identified *Cfl2* (a highly conserved protein involved actin depolymerisation) as a novel miRNA target in T cells (Fig. 3.8C), I reasoned that miRNAs might play a role in regulation of the actin dynamics across multiple cell types.

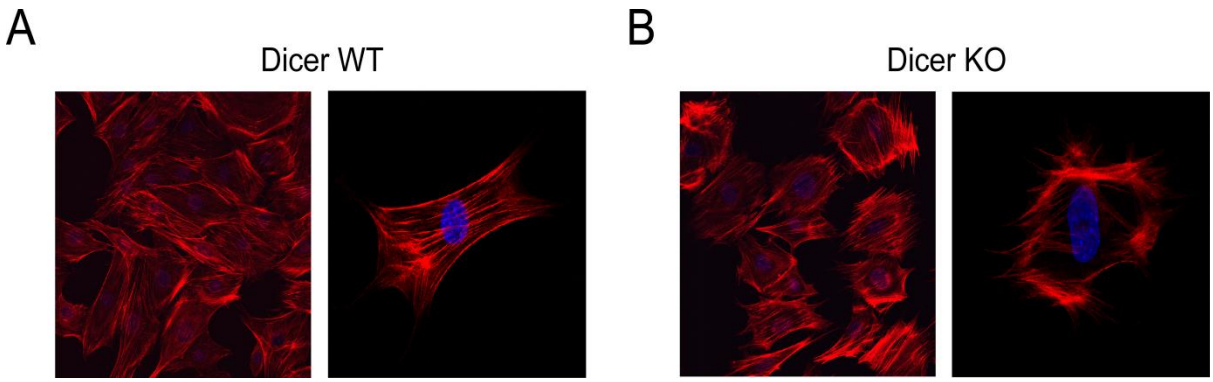



**Figure 8.5: Morphological differences in Dicer KO 3T3 fibroblasts compared to WT cells.**
Immunofluorescence analysis of $Dicer^{wt/wt}$ and $Dicer^{lox/lox}$ 3T3 fibroblasts stained with phalloidin (red) to identify polymerised actin filaments and DAPI (blue) to show the cell nucleus.

The cofilin family of proteins have been shown to play a fundamental role in the process of cell polarisation (Dawe et al., 2003; Mseka et al., 2007). As these proteins are predicted to be under miRNA-mediated control (see Chapter 3), it is plausible that polarisation might be defective in microRNA-deficient cells. Indeed, studies have shown that downregulation of specific miRNAs in epithelial cells abolishes their polarity (Tsuchiya et al., 2009). Additionally, it has been well established that cell polarisation is a pre-requisite for cell migration; a process in which a number of microRNAs have been implicated (Zhang et al., 2011).

To investigate whether microRNA-deficient 3T3 fibroblasts exhibit defects in cell polarisation, a quantitative method of assessing this phenotype was required. CYTOO chips™ are coverslips

coated with fibronectin in distinct patterns, which ensure that adherent cells adopt specific shapes on attachment to the surface. As hundreds of these micropatterns are present on each chip, the slide can then be scanned automatically and multiple images compiled to generate meaningful statistics regarding actin distribution, avoiding the subjective analysis of images of single cells.

To assess whether the morphological abnormalities observed in Dicer KO 3T3 under standard culture conditions (Fig. 8.5) result in an impaired ability of these cells to polarise, I plated Dicer WT and KO (the constitutive KO; Clone 1a – see Fig. 8.3) onto CYTOO chips with either circular or crossbow micropatterns. The crossbow shape is designed to mimic an environment which induces cell polarisation, whereas the circular micropattern is used as a control. I allowed four hours for cells to attach to the micropatterns before fixing and staining cells with phalloidin and DAPI. Examples of the uniform distribution of cells on each pattern are shown in Fig.8.6.

Using the circular micropattern (Fig. 8.6A) it was observed that Dicer WT cells adopt a relatively uniform shape and distribution of cortical actin (with the exception of the cell in the top right hand corner, which, due to the presence of two nuclei, appears to be undergoing mitosis). On the other hand, Dicer KO cells have highly variable morphologies, characterised by apparently randomly distributed actin filaments, focal spots of actin staining, and membrane protrusions (Fig. 8.6A). Dicer WT 3T3 cells plated onto the crossbow pattern show distinctive patterns of similarly uniform actin staining, with intense staining of cortical actin on the leading edge of the cell (the region bound to the curved section of the crossbow) and longitudinal internal filaments orientated in the direction of polarisation (in this case, from right to left). Again, in Dicer KO cells, these patterns of staining seem to be deregulated. Intense staining of cortical actin is observed around the entire perimeter of microRNA-deficient cells, not just at the leading edge. Similarly, internal actin filaments are shorter and not uniformly orientated, as in the WT cells.

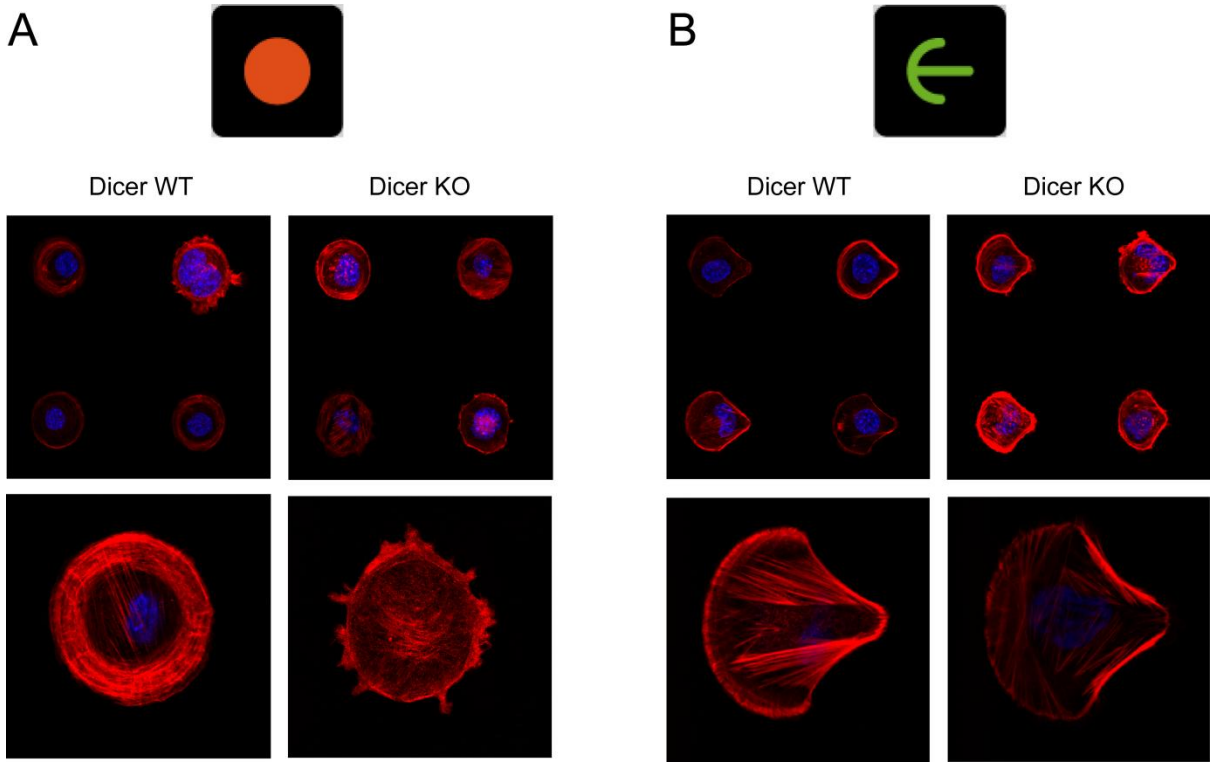



**Figure 8.6: Systematic analysis of actin distribution in Dicer WT and KO 3T3 fibroblasts.**
Images of Dicer WT and KO 3T3 fibroblasts plated on (A) circular and (B) crossbow shaped fibronectin micropatterns stained with phalloidin (red) to identify polymerised actin filaments and DAPI (blue) to show the cell nucleus. Upper panels illustrate representative sections of chips which contain hundreds of similar, uniformly distributed micropatterns.

These findings are intriguing, given that correct orientation of actin filaments in the direction of cell migration is required for cell polarisation, and has been shown to be dependent on the cofilin proteins (Mseka et al., 2007). In order to quantify the defect observed in Fig. 8.6B, CYTOO chips depicted in Fig. 8.6 were automatically scanned, so that hundreds of images of both Dicer WT and KO 3T3 fibroblasts adhered to the micropatterns were collected. In the case of the crossbow pattern, the cell was subdivided into defined regions spanning from the leading edge to the rear of the cell. Using data collated from all scanned images, the average intensity of phalloidin staining in these regions was quantified (Fig. 8.7A). This data therefore represents the average intracellular distribution of actin in total populations of Dicer WT and KO 3T3 fibroblasts which have been induced to polarise[2].

This data shows that in WT 3T3 fibroblasts induced to polarise, cortical actin at the leading edge of the cell is increased (shown by intense staining in Fig. 8.7A, quantified in Fig. 8.7B). A significant reduction in actin content throughout the cell body is then observed, with a slight peak in the staining at the rear of the cell. In Dicer KO 3T3, actin distribution in the polarised cell is far more

[2] Automated data collection from CYTOO chips and sub-regional quantification of actin intensity was carried out by Dr Dirk Dormann in the Microscopy Laboratory at the MRC Clinical Sciences Centre.

uniform (Fig. 8.7A). A slight peak in intensity can be observed at the leading edge of the cell (Fig. 8.7B) but this is not as pronounced as in WT cells. Whilst preliminary, these data suggest that Dicer KO 3T3 fibroblasts may be defective in cell polarisation due to their inability to redistribute intracellular actin when polarisation is induced.

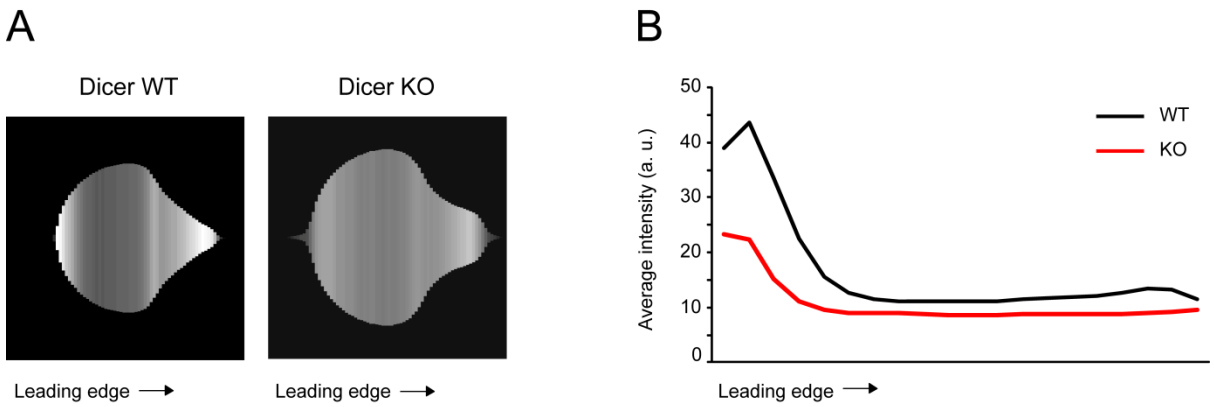



**Figure 8.7: Polarised actin distribution is impaired in Dicer KO 3T3 fibroblasts.**
(A) Representation of average intensity of phalloidin stain in defined longitudinal sections of the cell from leading edge to rear. Intensity is based on the average value for the defined region collated from 81 images of cells adhered to the crossbow micropattern in each genotype. (B) Quantification of the average intensity in each defined region in Dicer WT and KO 3T3 fibroblasts. A. u.: arbitrary units.

Data shown in Fig. 8.7A also indicates that despite the constraints of the uniformly sized fibronectin pattern, Dicer KO cells appear to be larger than WT cells. To assess this, average images were generated of Dicer WT and KO 3T3 fibroblasts using all data collected from scanning the CYTOO chips. These images were overlaid to assess the size difference between cells adhered to the same micropattern (Fig. 8.8). This analysis indicated that indeed Dicer KO cells are, based on a population level average, bigger than Dicer WT cells.

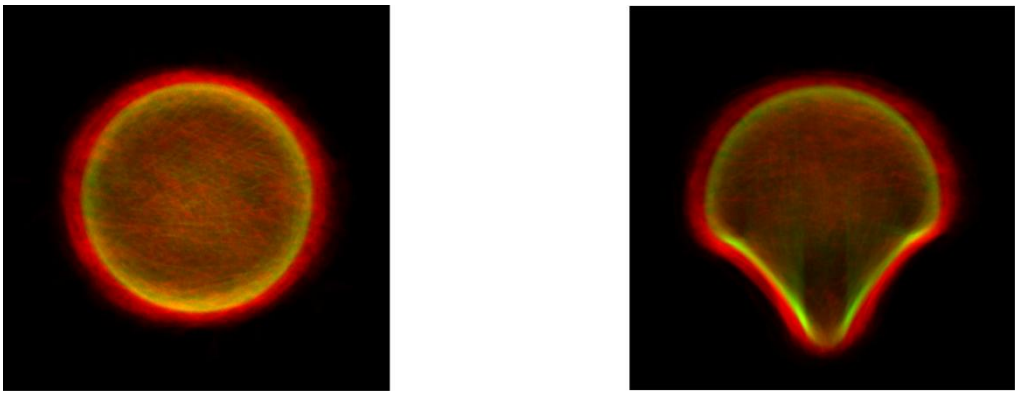


**Figure 8.8: Dicer KO 3T3 cells are generally greater in size than Dicer WT 3T3, independent of cell morphology.**
Overlay of average projections of actin distribution in Dicer WT (yellow) and KO (red) 3T3 fibroblasts. Images are generated from the average of 81 individual images of micropatterns on CYTOO chips.

In investigating a potential function for miRNA in regulation of the actin cytoskeleton, I was then interested to ascertain whether the increased cell size of Dicer KO cells correlates with increased total actin content. As anticipated, this was shown to be the case in both adherent cells (as quantified by cell area and total intensity of phalloidin staining of cells bound to the circular micropattern – Fig. 8.9A) and suspension cells (as quantified by FACs – Fig. 8.9B).

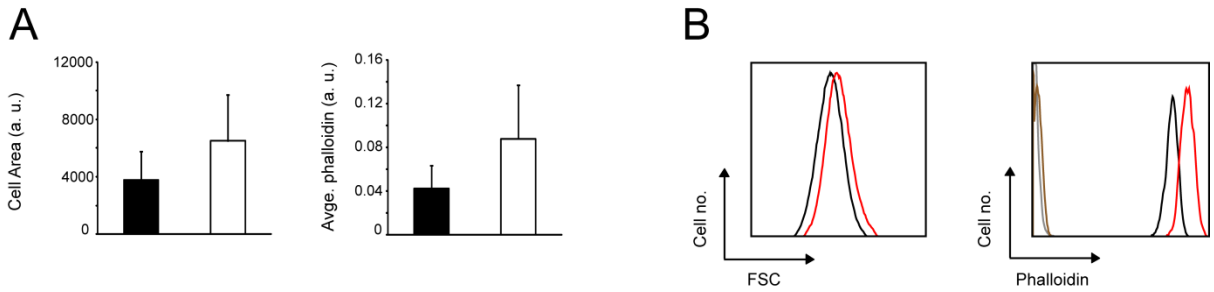



**Figure 8.9: Increased cell size in Dicer KO 3T3 fibroblasts correlates with increase in total actin content.**
(A) Quantification of average cell area and intensity of phalloidin stain in Dicer WT (black) and KO (white) 3T3 fibroblasts based on 81 images of cells adherent to the circular micropattern. Error bars: s.t.d. of values for 81 cells. (B) FACs analysis of cell size (left-hand panel, based on forward scatter; FSC) of Dicer WT (black) and KO (red) 3T3 fibroblasts in a non-adherent state. Total phalloidin content per population was also quantified (left-hand panel; unstained WT cells (grey) and unstained KO cells (brown)).

Taken together, the analyses presented here show that microRNA-deficient 3T3 fibroblasts are significantly different in morphology to Dicer WT cells. One component of this difference is the size of the two cells types; Dicer KO cells are significantly larger than Dicer WT cells, which correlates with a greater total actin content. In addition to this, Dicer KO 3T3 fibroblasts fail to systematically orientate actin filaments when induced to polarise, suggesting that this process may be impaired in microRNA-deficient cells.

Mechanistically, it is possible that this defect is in part due to deregulation of the cofilin family of proteins. To assess whether expression of genes coding for key family members *Cfl1, Cfl2* and *ADF* is deregulated in Dicer KO 3T3 fibroblasts, I quantified mRNA levels of these genes by qPCR in Dicer WT and KO 3T3 cells (Fig. 8.10). This data shows that whilst expression of *ADF* does not change in Dicer KO cells, *Cfl2* is significantly upregulated (consistent with data in Dicer KO $CD4^+$ T cells). However, the most highly expressed of the three genes, *Cfl1*, is significantly downregulated in Dicer KO 3T3 cells. *Cfl1* is the predominant form of the protein expressed in non-muscle tissues (Bernstein and Bamburg, 2010a) and therefore downregulation of this protein is more likely to be phenotypically relevant than upregulation of the miRNA-targeted *Cfl2.*

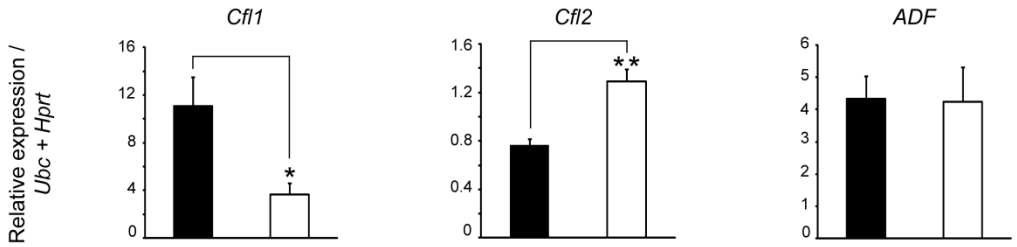
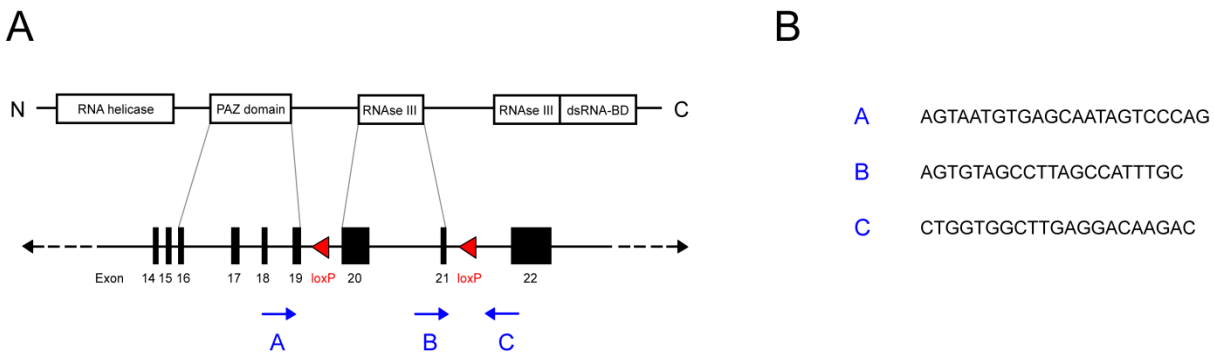



**Figure 8.10: mRNA levels of key members of the actin regulatory proteins are deregulated in Dicer KO 3T3 fibroblasts.** qPCR analysis of mRNA levels of key members of the cofilin family of proteins in Dicer WT (black) and KO (white) 3T3 fibroblasts. Error bars: s.t.d. of three independent biological replicates. Asterisks indicate significance (T-test): * = $p < 0.05$; ** = p , 0.005).

Whilst it is plausible that miRNAs mediate an upstream mechanism regulating *Cfl1* expression, the fact that this gene is not likely to be directly targeted by miRNAs made further investigation into the mechanistic basis of the actin defect in Dicer KO 3T3 too complex for the scope of this study. Given the known role of miRNAs in metastatic tumours, however, future analysis into the precise role of miRNAs in regulation of actin dynamics would be an interesting line of future research.

## III. Primer sequences

**Figure 8.11: Location and sequences of primers used for analysis of genomic deletion of RNase III domain in cells derived from ERT2Cre x Dicer$^{lox/lox}$ mice.**




(A) Schematic of region of modified Dicer gene containing the floxed RNase III domain in ERT2Cre x Dicer$^{lox/lox}$ mice. Location of primers used for genotyping are indicated in blue (A, B and C). WT alleles which do not contain the loxP sites are amplified by primers B and C and produce a PRC product of 259bp. Floxed alleles containing the loxP sites are amplified by primers B and C and produce a PCR product of 390bp. Deleted alleles (from a floxed parental line) are amplified by primers A and C and produce a PCR product of 309bp. (B) Sequences of primers shown in (A) used for genotyping (5' to 3').

**Table 3: Applied Biosystems (AB) Assay Ids for TaqMan® miRNA primer/probe pairs used for qPCR quantification of miRNAs.**

| miRNA | AB Assay ID |
|---|---|
| snoRNA-135 | 1230 |
| miR-124a | 1182 |
| miR-16 | 391 |
| miR-155 | 2571 |
| miR-142-3p | 1189 |
| miR-21 | 397 |
| miR-291a-3p | 2592 |
| miR-92 | 430 |
| let-7b | 2619 |
| miR-29a | 1189 |

**Table 4: Primer sequences for amplification of mRNA transcripts by qPCR.**

All primer sequences are shown 5' to 3'.

| Gene | Forward | Reverse |
|---|---|---|
| *Ac9* | CAACAGCGTGAGGGTCAAGAT | CATGGAGTCGAATTTGGGGTC |
| *ADF* | TGTATGACGCCAGCTTTGAG | AGCCACCTAGCTTTTCAGCA |
| *Arid4a* | ACACAACACAGTTGGTGCAAG | AGAGTTCTCTCATCACCGTCA |

| Gene | Forward | Reverse |
|---|---|---|
| *Ash1l* | CCTCGGTGGACTAAAGTGGTG | CGCTGGCTCAGAACTATTTGA |
| *Atxn1* | GTGTCTCCTGTGCAGTACGC | GTGACCGGGTTGCCTGATG |
| *Bmi1* | ATCCCCACTTAATGTGTGTCCT | CTTGCTGGTCTCCAAGTAACG |
| *Casp2* | CTCCCACCGTTGAGCTGTG | CCGTAGCATCTGTGGATAGGC |
| *Ccdc41* | TTAGAAGCTCCAATGCGAGAAC | AGCTGGTTATGTAGCTCCTCTT |
| *Cd4* | CCATCTCTCTTAGGCGCTT | ATTCCCCTTCCTTCCCCA |
| *Cd69* | TTGTACGAGAAGTTGGAATCATCTG | GAATATCGCTTCAGAAACGTCATGTCC |
| *Cdkn1a* | CGAGAACGGTGGAACTTTGAC | CAGGGCTCAGGTAGACCTTG |
| *Cfl1* | CCCTACACCACTTTTGTCAAGAT | ACCTTCTCGTAGCAGTTAGCTT |
| *Cfl2* | GCATCTGGAGTTACAGTGAATGA | CACCAATGTCACCCACCAAGA |
| *Cpeb2* | TTGGAGCAACCATCAGAACAG | TACCCGAGAACGGCTTCTTCA |
| *Crk* | GGCTCCGAATAGGAGATCAAGA | CCCGCACATACTCTGCCTC |
| *Dicer* | CGATATTCGATCCTCCTGTGA | GCCAGCAAGCAGTCTTTTGT |
| *Eed* | CAAATGTGTGAACAGCCTCAAGGAA | GCATCAGCATCTACATAGGACTGCAA |
| *Egfr* | GCCATCTGGGCCAAAGATACC | GTCTTCGCATGAATAGGCCAAT |
| *Eif2ak2* | TGGATTGTCACACGAGTGCAT | TACCAACGAACAAGCCTTGCT |
| *Elk4* | GAGCCCTGCGATACTACTACG | ACCGTCAGTGGGTCCATTTTC |
| *Epha4* | TGGAATTTGCGACGCTGTCA | CACTTCCTCCCACCCTCCTT |
| *Ezh2* | TCAGGATGAAGCAGACAGAAGAGG | CCTATCCTGTGGTCACCATTAACC |
| *Ficd* | CAGTCAAGACTACGGCGTCTC | AACTCATTCAGTGCGTCTACAAA |
| *Gapdh* | TGCACCACCAACTGCTTAGC | GGCATGGACTGTGGTCATGAG |
| *Gata3* | CCCTCCGGCTTCATCCTCT | CTGCACCTGATACTTGAGGC |
| *Gata4* | GAGGCTCAGCCGCAGTTGCAG | CGGCTAAAGAAGCCTAGTCCTTGCTT |
| *Hist1h1a* | AATAAAGGCACACTGGTGCAG | CCCCAGCAGTCTTCTTAGGTT |
| *Hist1h3f* | GTGAAGAAACCTCACCGCTAC | GTCCTCAAACAGACCCACAAG |
| *Hist1h4d* | TCTTGGTAAGGGTGGTGCTAA | ACGTTCTCCAGGAACACCTTC |
| *Hist3h2a* | AAGCTCGTGCAAAAGCGAAG | GCAGTCAGATATTCTAGCACAGC |
| *Hist4h4* | GAAGCGCATCTCGGGTCTC | CATAGCCGTAACCGTCTTGC |
| *Hn1* | TGGACCCTAACAGCAGGAACA | GCCTAATGAAAAATTGGACCCAC |
| *Hoxa1* | AAACCCTCCCAAAACAGGGAA | CGCGCTCGTGTAAGGTACTT |
| *Hprt* | TCAGTCAACGGGGGACATAAA | GGGGCTGTACTGCTTAACCAG |
| *Id1* | AGGTGAACGTCCTGCTCTAC | GTCCCGACTTCAGACTCCG |
| *Jarid2* | CCCAGGCTGCAAGCACAAA | CCTCGAAGGCAGAGAAAGGTAAGA |
| *Lats2* | GAGGTGCTTCTCCGCAAA | AGCATCTCAAAGAGAATCACACC |

| Gene | Forward | Reverse |
|---|---|---|
| *Lrch3* | CAAACTGTGCCTTCCCAAATTG | CGAGAAGTCGAGCCGTATCAAA |
| *Lrp8* | TCCTGCCGAGAAGTTAAGCTG | AAGAACGCAAGTCCCATCCC |
| *Nanog* | GCATCTTCTGCTTCCTGGCAA | GAACTATTCTTGCTTACAAGGGTCTGC |
| *Ncam1* | AGAAATCAGCGTTGGAGAGTCC | TCGTCATCATTCCACACCACT |
| *Nr2f2* | CCTCAAGGCCATAGTCCTGTT | GGCTCCTAACGTACTCTTCCAA |
| *Pou5f1* | CCCAAGGTGATCCTCTTCTGCTT | GAGAAGGTGGAACCAACTCCCG |
| *Rb1* | TGCATCTTTATCGCAGCAGTT | GTTCACACGTCCGTTCTAATTTG |
| *Rbl2* | GGGAGACATGGATTTATCTGGT | GCCAAGAGTGACCTGTGGA |
| *Rm18s* | GTAACCCGTTGAACCCCATT | CCATCCAATCGGTAGTAGCG |
| *Rtn4rl* | GCACCCGAGACCTTCCAAG | TGGAGGTACTCGATATGGTTGT |
| *Ryr2* | ACGGCGACCATCCACAAAG | CGGGGGAACATTCTTGGAATTG |
| *Slc6a9* | TACCTCTGCTATCGCAACGG | CAAACTGGCCGAAGGAAAGC |
| *Slc7a5* | TGTGGTAGCACATTTCTCCTACA | ATGCAGATAAAAACCCGTTCTCT |
| *Sox1* | TGAACGCCTTCATGGTGTGGTC | GCGCGGCCGGTACTTGTAAT |
| *Suz12* | TTTGAGACATTTTCTCAGGGACCTACA | TGTAGTCAGCGTCTCCTTAACAGCAAT |
| *Tgfbr1* | TCCCAACTACAGGACCTTTTTCA | GCAGTGGTAAACCTGATCCAGA |
| *Thap2* | GGAACCCTATGGAGGCGAAAA | TTGATCCACCTTCGAGTTGCT |
| *Ubc* | AGGAGGCTGATGAAGGAGCTTGA | TGGTTTGAATGGATACTCTGCTGGA |
| *Wasl* | ATGCTCCAAATGGTCCCAATC | CTTGGTGTTCCAATATCTGCCT |

**Table 5: Primer sequences for amplification of specific genomic regions in chromatin immunoprecipitation analyses.**

All primer sequences are shown 5' to 3'.

| Gene | Forward | Reverse |
|---|---|---|
| *Pou5f1* | TGGAGACTTTGCAGCCTGAG | TCTCCAACTTCACGGCATTG |
| *Cdkn1a* | AATTGGAGTCAGGCGCAGAT | GCATTGCTACGGGGAAGAAC |
| *Cacna1a* | GCAGAGTGAGCCTGCGAGTA | GAAGAGCCCCCATCAGTGTC |
| *Dock2* | CCCTTCACCTGCTTTTCCTG | TGGGTTCATGGCACAGAAAG |
| *Lefty1* | GGCTGATCATGCCTTCCTTC | CCTCTAACATCCGCCTCACC |
| *Id1* | AACCGCAAAGTGAGCAAGGT | GGTGGTCCCGACTTCAGACT |
| *Hoxa1* | GCGCAGGATTGGAAAGTTGT | ATACCCCATCCTTGGCAGTG |
| *Ryr2* | GCTGTCGTCACGAAAACCAG | CCCCCAAGTCGAGCAGATAC |
| *Egfr* | CTCCTCTTCTGGCTCCCAAA | AGCCCAGTCTCCCTTCCTTC |

| | | |
|---|---|---|
| *Gata3* | CGTAGCCCTGACGGAGTTTC | TATCCGCTGACGGAAGAGGT |
| *Pax6* | TACTGGATGGCCCCTGAACT | GCCTCTGGCTGGTATTGTCC |
| *Nr2f2* | CCAGAGCAGGCTAGCCAAAC | CTCAAAGTGGGCATGAGACG |

**Table 6: Primer sequences for amplification of the 5' end of nascent RNA transcripts by qPCR.**

All primer sequences are shown 5’ to 3’.

| Gene | Forward | Reverse |
|---|---|---|
| *Pou5f1* | TGGAGACTTTGCAGCCTGAG | TCCCAATTCCCTTCACTGCT |
| *Cdkn1a* | GTCGCTGTCTTGCACTCTGG | TCTCCTCTCCAGGCCTCTTG |
| *Id1* | TCATCGTGCCTCCTGGAGTA | GGCTGCGGTAGTGTCTTTCC |
| *Hoxa1* | CTTCTCCAGCGCAGACCTTT | CTTTTGCCTCTGCACACCTG |
| *Egfr* | AGCAGCCTCCCTCCTCTTCT | TCTCTGACCGGGAGAGGTTG |
| *Nr2f2* | CACCCGCCAAACTAAAGGAG | CGAGGCATTGCATACACGTT |

**Table 7: Primer sequences for amplification of the 3' end of nascent RNA transcripts by qPCR.**

All primer sequences are shown 5’ to 3’.

| Gene | Forward | Reverse |
|---|---|---|
| *Pou5f1* | GCCCTGCAGAAGGAGCTAGA | CCCCTTGCGCCTACATTAAG |
| *Cdkn1a* | TCCCTCTGACCAGGAGGTGT | TCCTGACCCACAGCAGAAGA |
| *Id1* | CCTCTCCCCTCCTTGTCCTT | GGCTGCGGTAGTGTCTTTCC |
| *Hoxa1* | TTCCCGATAGGGTTTGTGCT | TCTGTGAGCTGCTTGGTGGT |
| *Egfr* | AAATCGCTTATGCCCATTGC | TGGCCTCTTGGGAACAGATT |
| *Nr2f2* | CCTCCTCCCCCTTGAAGTTT | TTTCCTGCAAGCTTTCCACA |