COMPUTATIONAL AND EXPERIMENTAL ANALYSIS OF MICRORNA REGULATION IN MOUSE EMBRYONIC STEM CELL PLURIPOTENCY AND DIFFERENTIATION

Rashmi Tripathi
Newnham College

Thesis presented for the degree of

Master of Philosophy

University of Cambridge
Wellcome Trust Sanger Institute

8th May 2008

PREFACE

This work is my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text as Acknowledegments. There are 21,566 words in this thesis, excluding Tables, Figures, Bibliography and Appendices.

ACKNOWLEGEMENTS

I would like to thank my supervisors, Dr. Anton Enright and Professor Allan Bradley for giving me the opportunity to carry out the work described in this thesis. I am especially grateful to them for their patient guidance. I would also like to thank all the members of the Enright and Bradley laboratories for providing a stimulating work environment and constant support during all times. I am very grateful to Dr. Pentao Liu and members of his laboratory, Peng Li, Dr. Wei Wang, Song Choon Lee and Dr. Dong Lu for their advice and help in providing plasmids and reagents for my recombineering experiments. I would like to thank Dr. Cordelia Langford, Dr. Peter Ellis, Dr. Robert Andrews and Dr. Gregory Lefebre for their advice and support in expression experiments. The Illumina data has been generated by Dr. Peter Ellis. Dr. Enright has also very kindly provided me with a script to look at microRNA seed enrichments in transcriptional networks. I am very grateful to Dr. Enright, Dr. Cei Abreu-Goodger and Siarhei for their enormous help in training me to perform Illumina and Exiqon data analysis on Bioconductor. Dr. Cei Abreu-Goodger has very kindly constructed Sylamer (program devised by Dr. Stijn van Dongen) plots for microRNA seed over-representations and sequence biases. Dr. Saini has enabled me to carry out random iterations for detecting microRNA binding sites on transcripts with her Perl script. Differential expression analysis and GO enrichment analysis for mRNA expression has been carried out by Siarhei Manakou. I would also like to thank Rafaella Rossi and Linda Rehaume for their help in using the Bioanalyzer.

I would like to thank Drs. George Vassiliou, Cei, Anton Enright, Charles Shaw-Smith and Professor Bradley for their comments on my thesis.

I would like to thank Joan Green and Andrew King for being absolutely wonderful in obtaining references and books and Sarah-Quakelaar Howard for making appointments to see Prof. Bradley. I would like to thank Dr. Christina Hedberg Delouka and Dr. Alex Bateman for their advice and support in general. I would like to thank all my fellow student peers for providing an enriching intellectual environment during my time in Cambridge. I am especially grateful to my tutor Prof. Harry Baker for providing a listening ear and advice outside my Department during trying times. I am very grateful to Cambridge Commonwealth Trust and the Wellcome Trust Sanger Institute for funding my studies. I am grateful to examiners for making suggestions and corrections for improvement of this work. I am extremely indebted to my family and for their unconditional love.

TABLE OF CONTENTS

	PREFACE	2
	ACKNOWLEDGEMENTS	3
	TABLE OF CONTENTS	4
	LIST OF FIGURES	6
	LIST OF TABLES	8
	LIST OF ABREVIATIONS	9
	SUMMARY	11
	CHAPTER ONE: LITERATURE REVIEW	13
1.1	Introduction	13
1.2	What are MicroRNAs?	14
1.3	MicroRNA Synthesis and Processing	14
1.4	Regulatory Effects of MicroRNAs	16
1.4.1	MicroRNA mediated Repression of Translation	16
1.4.2	MicroRNA mediated Activation of Translation	20
1.5	Clues regarding MicroRNA Functions in Mouse Embryonic Stem Cells	20
1.6	Determining MicroRNA-mRNA Interactions with miRanda,TargetScan, Pictar and Sylamer	22
	CHAPTER TWO: IN SILICO DETECTION OF MICRORNAS INTERACTING WITH KEY PLURIPOTENCY TRANSCRIPTS	25
2.1	Introduction: Generating Hypotheses for MicroRNA Functions in Stem Cells	25
2.2	Results	27
2.2.1	Increased Propensity of ES Cell MicroRNAs towards ES Cell Pluripotency Transcription Factor Transcripts	27
2.2.2	Is Enrichment of ES MicroRNA binding sites on 3'UTRs of Pluripotency Transcripts due to Chance?	30
2.2.3	Predicted Regulatory Networks for MicroRNAs in Mouse ES Cells	31
2.3	Summary	35
	CHAPTER THREE: MICRORNAS IMPACTING THE OCT4/NANOG TRANSCRIPTIONAL NETWORK	36
3.1	Introduction	36
3.2	Scanning most Up-regulated and Down-regulated Genes in the Oct4/Nanog Transcriptional Network	37
3.3	Results	41
3.3.1	Identification of MicroRNAs impacting the Nanog and Oct4 Transcriptional Network at FWER <0.001	41
3.3.2	Identification of MicroRNAs impacting the Nanog and Oct4 Transcriptional Network at FWER <0.01	43
3.3.3	Scanning entire Transcriptional Networks with Sylamer	45
3.4	Summary	47
	CHAPTER FOUR: ANALYSIS OF THE 'MIRNOME' AND 'TRANSCRIPTOME' IN DIFFERENTIATING ES CELLS	49
4.1	Introduction	49
4.2	Optimizing Embryoid Body formation	50
4.3	MicroRNA and Messenger RNA Analysis	50
4.3.1	Quality Control Assessment of mRNA and microRNA Expression Data	51
4.3.2	Assessing Positive and Negative Controls for Illumina Mouse mRNA and MicroRNA Chips	58
4.4	Results	61
4.4.1	Analysis of MicroRNA Differential Expression	61

4.4.2	Analysis of Gene Ontology Biological Process Enrichment in Differentially Expressed mRNAs	64
4.4.3	Sequence Bias and MicroRNA Seed Enrichments in 3'UTRs of Most Differentially Expressed mRNAs	66
4.4.4	mRNA Expression Analysis of Key Pluripotency Transcripts	70
4.4.5	Determining Correlations between Key Pluripotency Transcripts and their MicroRNA Regulators	72
4.5	Summary	77
	CHAPTER FIVE: MAKING TARGETING VECTORS FOR MICRORNA LOSS OF FUNCTION ANALYSIS IN ES CELLS	81
5.1	Introduction	81
5.2	Making Targeting Vectors by Recombineering	81
5.3	Results	83
5.3.1	First Step Recombineering	83
5.3.2	Second Step Recombineering	86
5.3.3	Third Step Recombineering	86
5.3.4	Sequence Verification of Targeting Vectors	88
5.4	Summary	89
	CHAPTER SIX: CONCLUSIONS AND DISCUSSION	90
	CHAPTER SEVEN: MATERIALS AND METHODS	94
7.1	Identification of MicroRNAs targeting Key Pluripotency Transcripts	94
7.2	Determining Significance of MicroRNA-mRNA Interactions	94
7.3	Identification of Up-regulated and Down-regulated Genes in the Oct4/Nanog Transcriptional Network	94
7.4	Identification of MicroRNA Targets in Different Gene Sets	95
7.5	ES Cell Culture	95
7.6	Embryoid Body Formation	96
7.7	Total RNA Isolation from ES Cells and Embryoid Bodies	96
7.8	MicroRNA Expression Analysis using miRcury Exiqon Microarrays	97
7.9	Microarray Data Analysis	98
7.10	Illumina Bead Arrays	99
7.10.1	Illumina MicroRNA Expression Profiling	99
7.10.2	Illumina Whole Genome Expression Profiling	103
7.10.3	Hybridisation, Washing and Signal Detection	105
7.11	Data Analysis of Illumina Bead Arrays	106
7.12	Using Sylamer	106
7.13	Making Vectors by Recombineering	106
	BIBLIOGRAPHY	112

LIST OF FIGURES

CHAPTER ONE

1.1	MicroRNA Processing, copied from Faller and Guo.	17
1.2	Different pathways of microRNA mediated degradation by RISC complex	18
1.3	Predicted target sites for miR-26a on SMAD-1 transcript	19
1.4	Three kinds of microRNA target sites	23

CHAPTER TWO

2.1	Star sequence	26
2.2	Increased Propensity of Stem Cell MicroRNAs towards core Pluripotency Transcription factor transcripts.	29
2.3	Average 3'UTR length of core set of pluripotency transcripts and control set.	30
2.4	Diagram depicting predicted binding sites for microRNAs on transcripts belonging to key pluripotency transcription factors	33
2.5	Diagram depicting predicted binding sites for microRNAs on Klf4 transcript	34
2.6	Diagram depicting predicted binding sites for microRNAs on cMyc transcript	34

CHAPTER THREE

3.1	Genes in the Oct4/Nanog Transcriptional Network	37
3.2	Graphical depiction of changes in gene-expression upon Oct4/Nanog knock-down in ES cells	38
3.3	MicroRNA binding patterns scanned in specific up-regulated and down-regulated gene sets, with mmu-miR-101a shown as an example	39
3.4	Flowchart showing the protocol for microRNA target identification in the Oct4/Nanog Transcriptional Network	40
3.5	MicroRNAs targeting the Nanog Transcriptional Network at FWER<0.001	42
3.6	MicroRNAs targeting the Oct4 Transcriptional Network at FWER<0.001	42
3.7	MicroRNA targeting the Nanog Transcriptional Network at FWER<0.01	44
3.8	MicroRNA targeting the Oct4 Transcriptional Network at FWER<0.01	45
3.9	Sylamer Plots for word and microRNA seed enrichments	46

CHAPTER FOUR

4.1	Box-plots of raw microRNA expression data	52
4.2	Box-plots of normalised microRNA expression data	53
4.3	Box-plots of raw mRNA expression data	54
4.4	Box-plots of normalised mRNA expression data	55
4.5	Cluster analysis of raw and normalised microRNA expression data	56
4.6	Cluster analysis of raw and normalised mRNA expression data	57
4.7	Box-plots of raw and normalised Exiqon data	59
4.8	Median Expression Values of Positive and Negative Control Probes on the Illumina mouse mRNA array	60
4.9	Median Expression Values of Positive and Negative Control Probes on the Illumina mouse microRNA array	60
4.10	Heatmap depicting differential microRNA expression.	62
4.11	MicroRNAs showing increased transient expression during differentiation of stem cells into embryoid bodies	63
4.12	Odds Ratios of Gene Ontology Biological Process Term Enrichments in genes which are most up and down-regulated during differentiation of embryonic stem cells into embryoid bodies	65
4.13	Sylamer Plot for sequence enrichment in 3'UTRs of genes most differentially expressed between day 0 and 3 of embryoid body formation	67
4.14	Sylamer Plots for maximum enrichments and depletions for microRNA seeds across all differentially expressed genes	69

4.15	Expression profiling of key pluripotency transcripts (Nanog, Pou5f1,Esrrb,Stat3,Tcf7,Sall4,Klf4,cMyc and Sox2) across three days of differentiation into embryoid bodies.	72
4.16	Plots depicting changes in mRNA and microRNA expression levels pluripotency transcripts and their predicted microRNA interacting partners	74
4.17	Plots depicting changes in mRNA and microRNA expression levels of pluripotency transcripts and their predicted microRNA interacting partners	76
	CHAPTER FIVE	
5.1	Recombineering Procedure	83
5.2	PCR diagnosis of successful recombineering events	85
5.3	Final targeting vector for miR-192	86
5.4	Final targeting vector for miR-345	87
5.5	Graphical representations of nucleotide substitutions and deletions detected in mir-345 and mir-192 vectors.	88

List of Tables

Table 2.1	p-values for ES MicroRNA Enrichments in Randomly Selected Transcript Sets	31
Table 4.1	Average Correlation between replicates in Illumina Arrays	52
Table 4.2	Frequency of AT/GC base-pairs in Seeds of Differentially expressed MicroRNAs	68
Table 4.3	Pearson's Correlation Values for Expression Values of Pluripotency Transcripts and their Predicted Interacting Partners.	75
Table 7.1	Media Composition	94
Table 7.2	Sample Volumes	96
Table 7.3	Sample Volumes	96
Table 7.4	Junction Primers	107
Table 7.5	Recombineering Primers	109
Table 7.6	Sequencing Primers	109

LIST OF ABBREVIATIONS

AmpR	Ampicillin Resistance
ARE	AU Rich Element
AT-GC	Adenine Thymine - Guanine Cytosine
AU	Adenylate Uridylate
BAC	Bacterial Artifical Chromosome
bic	B-cell integration cluster
BLAST	Basic Local Alignment and Search Tool
cDNA	complementary DNA
CIP	Calf-Intestinal Phosphatase
CO2	carbon dioxide
Dcp1	Decapping protein 1
Dcp2	Decapping protein 2
DGCR8	DiGeorge syndrome Critical Region 8
DNA	Deoxyribonucleic Acid
D-PCR	downstream-PCR
E.coli	*Escherichia coli*
EC	Embryonal Carcinoma
EGF	Epidermal Growth Factor
ES cell	Embryonic Stem cell
Esrrb	Estrogen related receptor beta
Exo	Exonuclease
FWER	Family Wise Error Rate
FXR1	Fragile X mental retardation
G6PD	Glucose 6 Phosphate Dehydrogenase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GO	Gene Ontology
GOBP	Gene Ontology and Biological Process
GSK3?	Glycogen synthase kinase 3-beta
GU	Guanine-Uracil
Klf4	Kruppel like factor 4
LB	Luria Broth
LIF	Leukemia Inhibitory Factor
LNA	Locked Nucleic Acid
Lrh1	Liver receptor homolog 1
miRNA	microRNA
mRNA	messenger Ribonucleic Acid
N2B27	Neurobasal B27 medium
NCBI	National Center of Biotechnology
Oct4	Octamer four transcription factor
P-bodies	Processing bodies
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
polyA	polyAdenine
pre-miRNA	precursor-microRNA
pri-miRNA	primary microRNA
p-value	probablity value
R2D2	dsRNA binding domain and Drosophila Dicer-2 binding domain
RanGTP	Ras related nuclear protein Guanosine Triphosphate
RecA	Recombination defective for A
RecBCD	Recombination defective for BCD
RISC	RNA Induced Silencing Complex
RNA	Ribonucleic acid
Sall4	Sal-like protein 4 transcription factor
Sox2	SRY(sex determining region Y)-box 2 transcription factor
ssDNA	single stranded DNA
Stat3	Signal transducers and activators of transcription factor

T58A	Threonine at position 58 replaced by Alanine
Tcf7	Transcription factor 7
Th2	Thymus lymphocyte helper 2
TNF-?	Tumour necrosis factor -?
TRBP	human immunodeficiency virus transactivating response RNA binding protein
TTP	tristetrapolin
UCSC	University of California Santa Cruz
U-PCR	upstream -PCR
UTR	Untranslated region
Xrn1	Exoribonuclease 1

SUMMARY

Gene-regulation is a key feature of all biological systems, which enables organisms to respond to various environmental stimuli. It can be mediated at the transcriptional and post-transcriptional level. It is also crucial in developmental processes like differentiation of embryonic stem cells into all lineages of a maturing embryo. Post-transcriptional silencing of messenger RNAs by small regulatory molecules called microRNAs is a type of a gene regulatory mechanism whereby multiple genes can be simultaneously down-regulated by a single microRNA entity. In order to understand how microRNAs might regulate pluripotency (the ability of stem cells to differentiate into any cell type) and homeostasis (steady state) in mouse embryonic stem cells, I used an *in silico* approach to build regulatory models of microRNA actions in stem cells. Transcripts and entire transcriptional networks belonging to key stem cell transcription factors were scanned for microRNA binding sites using miRanda on miRBase (John et al., 2004, Griffiths-Jones et al., 2006), TargetScan (Lewis et al., 2003, Lewis et al., 2005, Grimson et al., 2007), Pictar (Krek et al., 2005), a perl script (devised by Dr. Anton Enright) and Sylamer (a program devised by Dr. Stijn van Dongen). A number of stem cell microRNAs were predicted to interact with these transcripts and transcriptional networks hinting at a role of these microRNAs in regulating stem cell properties.

Furthermore, expression dynamics of both microRNA and mRNA molecules during stem cell differentiation into embryoid bodies was studied using Illumina bead arrays, which constitute a highly sensitive and robust expression measurement platform. Many microRNAs which are expressed in stem cells were found to be down-regulated during differentiation of stem cells indicating that these microRNAs might be essential in maintaining pluripotency and homeostasis in stem cells. MicroRNAs which are up-regulated during stem cell differentiation might be required for formation of new cell types. This work also uncovers a previously unseen sequence bias in AT-GC composition of 3'UTRs of genes that are most differentially expressed during differentiation. There is some evidence that this bias might facilitate the action of microRNAs which are also differentially expressed during stem cell differentiation. Correlation analysis of expression trends of pluripotency transcripts and their predicted microRNA interacting partners reveals both positive and negative correlations between the two. Analysis of Gene-Ontology terms in genes that are most differentially expressed, offers clues regarding the nature of gene-expression in day 3 embryoid bodies.

Lastly, targeting vectors for the purpose of making homozygous knock-outs of microRNAs in stem cells were constructed. Loss of function analysis in embryonic stem cells is expected to reveal further information regarding the role of microRNAs in maintaining stem cell pluripotency and causing differentiation. Expression or proteomic analysis of knock-out stem cell lines should also help classify microRNA targets in stem cells based upon whether they are degraded or in a state of translational inhibition.

CHAPTER ONE: LITERATURE REVIEW

1.1 Introduction

Stem cells are unique in their ability to divide continuously and differentiate into multiple cellular types upon receiving certain internally or externally mediated signals. These cells can be classified into embryonic, germ and somatic stem cell types according to their tissue of origin. Embryonic stem (ES) cells were first established in culture from the inner cell mass of the mouse blastocysts (Evans and Kaufman, 1981, Martin, 1981) and can give rise to all cell types found in the adult including germ-cells (Bradley et al., 1984) . This ability of stem cells to form all cell types is termed as 'pluripotency'. Over the years, these cells have proved to be of immense use in engineering genes in transgenesis experiments to understand gene function *in vivo*, obtaining animal models to study human diseases, and in generating tissues and specific cell types *in vitro.* Differentiation of stem cells into different lineages has been shown to be dependent upon expression of certain transcription factors (Pesce and Scholer, 2001, Reid, 1990). Recent advances involving reprogramming of somatic cells into pluripotent cells (Takahashi and Yamanaka, 2006, Takahashi, 2007) by introduction of specific factors like Oct4 (Octamer 4), Sox2 (SRY-box 2), Klf4 (Kruppel like factor 4), and cMyc (called 'Yamanaka factors') in mouse fibroblasts and Oct4, Sox2, Nanog and Lin28 in human somatic cells (Yu et al., 2007) has spurred interest in using these 'induced pluripotent stem cells' for transplantation in patients suffering from diverse degenerative disorders. Most of the identified molecules happen to be transcriptional regulators with the exception of Lin28 which has been implicated as a negative regulator of microRNA processing pathway (Viswanathan et al., 2008). In my thesis, I examine the role of a class of small regulatory RNA molecules called microRNAs in regulation of expression of stem cell pluripotency transcripts and transcriptional networks by computational methods. I also study changes in expression of these molecules and messenger mRNAs, including their potential targets, during the process of stem cell differentiation. Lastly, I describe the construction of two targeting vectors for microRNA loss of

function analysis in embryonic stem cells for discovery of specific modes of microRNA mediated messenger RNA regulation in embryonic stem cells and their potential role in embryonic stem cell differentiation.

1.2 What are MicroRNAs?

MicroRNAs (miRNAs) are short, single stranded 21-23nt family of RNAs which were first discovered in *Caenorhabditis elegans* (*C.elegans*). In pioneering experiments performed by Ambros and Wightman, the founding member of the microRNA family, lin-4, was shown to be involved in regulating the heterochronic gene lin-14 (Ambros and Horvitz, 1984, Lee et al., 1993, Wightman et al., 1993, Olsen and Ambros, 1999). MicroRNAs repress gene expression by targeting cognate messenger RNAs (mRNAs) for degradation or translational repression (Bartel, 2004). MicroRNA mediated post-transcriptional silencing is similar to the mechanism of RNA interference which was initially described in plants (Vanderkrol et al., 1990) and plant viruses (Baulcombe, 1996). However, recently it has been documented that under conditions of serum starvation and cell cycle arrest, some microRNAs may also up-regulate translation (Vasudevan et al., 2007).

1.3 MicroRNA Synthesis and Processing

MicroRNAs are primarily transcribed by RNA polymerase II (Lee et al., 2004) as monocistronic or polycistronic primary transcripts (pri-miRNAs). The transcription of microRNAs interspersed among Alu elements has also been shown to be dependent on RNA Polymerase III (Borchert et al., 2006). Primary transcripts (pri-miRNAs), which can be thousands of base pairs (bp) long are processed into 60-70bp precursors (pre-miRNAs) by the Microprocessor complex, which comprises of Drosha , a member of RNAase III family of nucleases and DGCR8 (Denli et al., 2004, Gregory and Shiekhattar, 2005), a double-stranded RNA binding protein. Directional cloning of pri-miRNA had initially revealed the presence of a stem structure of approximately 33 bp, with a

terminal loop and flanking segments (Lee et al., 2003). Recent studies have shown that the terminal loop may not be essential, whereas the flanking ssRNA segments are critical for processing (Han et al., 2006). The cleavage site on a pri-miRNA occurs at a distance of approximately 11bp from the stem-ssRNA junction. Purified DGCR8 interacts specifically with pri-miRNAs and the surrounding ssRNA segments are crucial for this binding to occur. Hence, DGCR8 has been termed as a "molecular-anchor" which performs the function of measuring the distance from the stem-ssRNA junction to initiate cleavage and processing of pri-miRNA molecules. There are also other features found on pri-miRNAs which include internal loops and bulges at specific positions and are conserved between microRNAs in different species (Han et al., 2006). These loops and bulges are, however, not essential for pri-miRNA processing (Saetrom et al., 2006).

Some microRNAs located within introns are processed by the spliceosome instead of the Drosha-DGCR8 Complex (Ruby et al., 2007, Okamura et al., 2007). Subsequent to splicing they are linearized by the debranching enzyme and folded into hairpin loops.

After initial processing of pri-miRNA into pre-miRNA in the nucleus by the Drosha complex or the spliceosome, pre- miRNAs are transported into the cytoplasm by Exportin 5 which is a carrier protein located in the nuclear membrane, and RanGTP (Bohnsack et al., 2004, Lund et al., 2004). In the cytoplasm the pre-miRNAs are processed by Dicer into short double-stranded RNA duplexes with 2 nucleotide 3'-overhangs (Hutvagner et al., 2001, Bernstein et al., 2003). These duplexes are then loaded onto the Ago1 (Argonaute1) or Ago2-RISC (RNA induced silencing complex) complexes depending on the degree of matches between the RNA molecules in the RNA duplex, as shown in *Drosophila* (Forstemann et al., 2007, Tomari et al., 2007). Perfectly matching duplexes like let-7 are loaded mostly onto Ago2-RISC complex while mismatches in the duplex shift the balance of loading towards Ago1-RISC (Tomari et al., 2007). After being loaded onto the RISC complex the double stranded RNA is unwound to a mature single stranded RNA (called guide strand) (Nykanen et al., 2001) and its complementary strand (the star sequence containing passenger

strand) is degraded by RISC (Gregory and Shiekhattar, 2005, Matranga et al., 2005, Rand et al., 2005, Leuschner et al., 2006). Recently it has been shown that the star sequence also has a chance to be incorporated into RISC and can be actively used in gene regulation (Okamura et al., 2008).

The functional human RISC complex (Gregory et al., 2007) is composed of three proteins Ago2, Dicer and TRBP and has been shown to cleave target RNA by using precursor microRNA (pre-miRNA) hairpin as the source to guide single stranded RNA to cognate RNA transcripts. RNA helicase A has been shown to interact with RISC by promoting association of double-stranded RNA with Ago2 in human cells (Robb and Rana, 2007). The entire processing pathway is depicted in Figure 1.1. The next section describes how microRNAs loaded in the functional RISC complex mediate their regulatory effects.

1.4 Regulatory Effects of MicroRNAs

1.4.1 MicroRNA mediated Repression of Translation

Although the exact mechanism of microRNA mediated gene silencing is not completely understood, various pathways have been elucidated for triggering this process. An mRNA molecule could be cleaved (Hutvagner and Zamore, 2002), translationally repressed at initiation or elongation steps (Olsen and Ambros, 1999, Seggerson et al., 2002) or rendered unstable by the cooperation of microRNAs and certain AU rich elements (AREs) (Jing et al., 2005) (Figure 1.2). Sufficient sequence complementarity between a microRNA and its target sequence is a requirement to initiate site-specific cleavage.

The formation of a continuous A-helix between the 2^{nd} and 12^{th} nucleotide of a microRNA (Chiu and Rana, 2003, Haley and Zamore, 2004) results in cleavage between the 10^{th} and 11^{th} nucleotides by the active site on Argonaute proteins (Song et al., 2004, Ma et al., 2005). The

degradation fragments released are then processed in the exosome (Mitchell et al., 1997, Orban and Izaurralde, 2005), a macromolecular complex responsible for RNA degradation (Figure 1.2A).

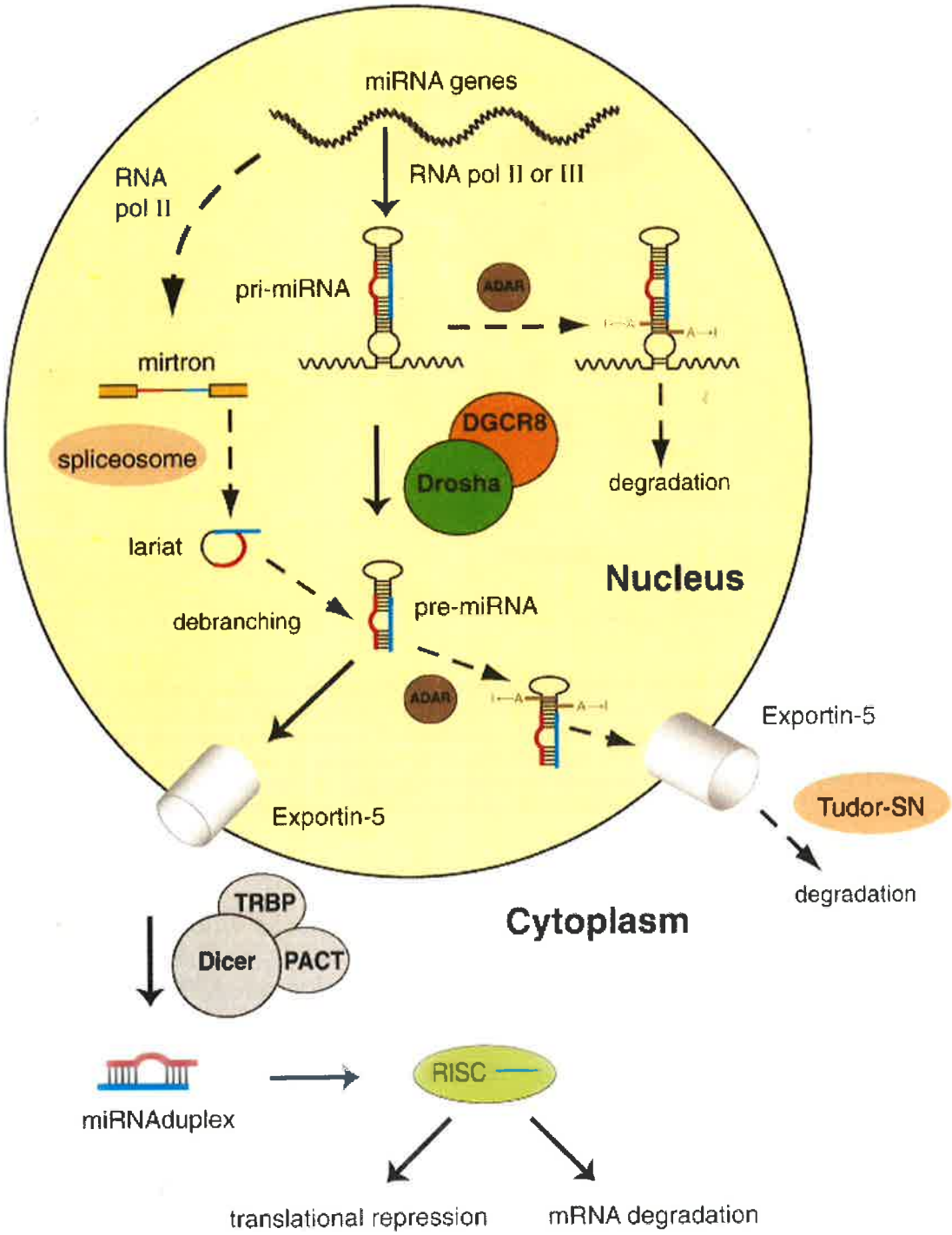


Figure 1.1: MicroRNA Processing, copied from Faller and Guo, 2008. MicroRNA genes are transcribed by RNA polymerase II/III to generate the primary transcripts, referred to as pri-miRNAs which can be either mono or polycistronic. The first-step processing mediated by DGCR8/Drosha results in pre-miRNAs of ~70 nt, which are recognized and exported into the cytoplasm by Exportin 5 complex into the cytoplasm. Upon export, Dicer participates in the second-step processing to yield an RNA duplex which is then incorporated into RISC (RNA induced silencing complex) as a mature single stranded RNA. The mature microRNA leads to translational repression or mRNA degradation depending upon complementarity to the target sequence. Alternatively, a class of microRNAs called mirtrons can be processed by the spliceosome machinery and the de-branching enzyme. The A-to-I modification of microRNAs by the process of RNA editing might further alter their specification. The non-processable microRNAs get degraded by nucleases such as Tudor-SN in the cytoplasm.

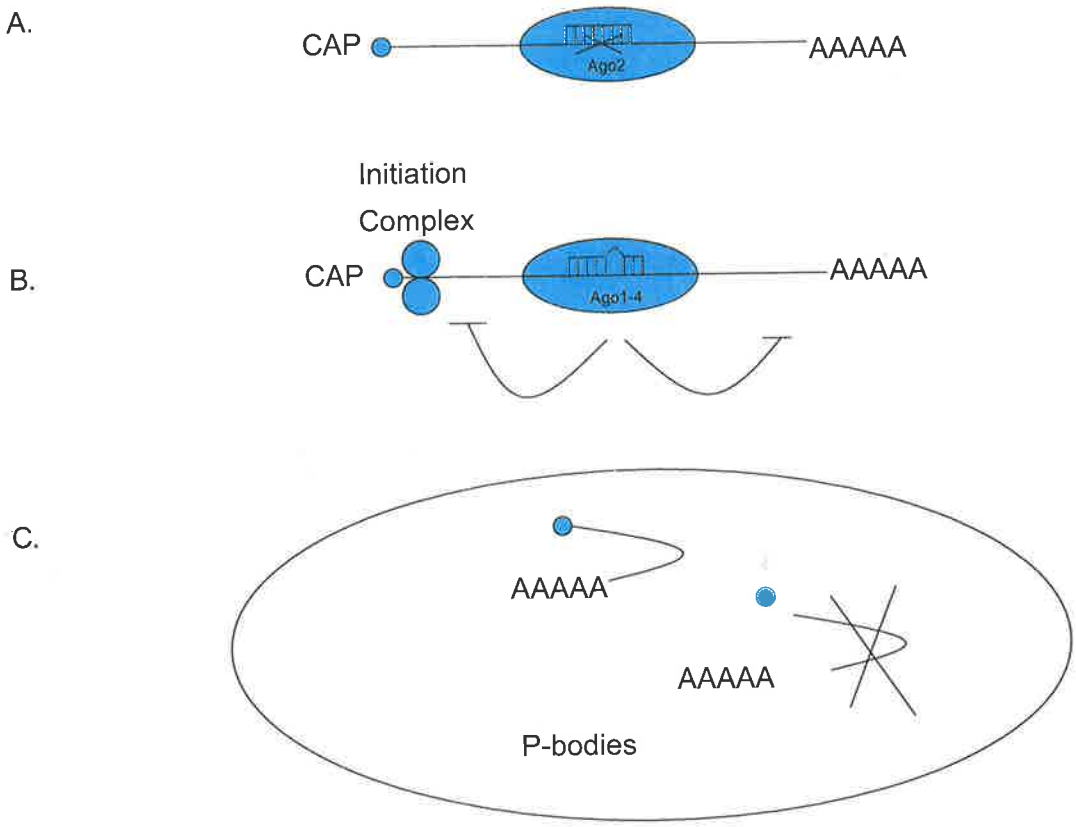


Figure 1.2: Different pathways of microRNA mediated degradation by the RISC complex. A. Sufficient complementarity between the microRNA and mRNA results in endonucleolytic cleavage by Argonaute protein, which is a part of the RISC complex and subsequent processing in the exosome. B. Insufficient complementarity could lead to translational inhibition of mRNA molecule by interference with initiation or elongation steps and ribosomal drop-off or C. where mRNAs could be transported to the cytoplasmic P-bodies and subjected to deadenylation-dependent decapping followed by degradation. Alternatively, mRNA molecules could be stored in these bodies temporarily and rescued later for translation.

The majority of microRNAs have limited complementarity to their known mRNA targets and the microRNA -target interaction is mostly restricted to the 5'end of the microRNA. However, even in these cases, a perfect match between nucleotides 2-8 of the microRNA (the so called "seed sequence" , See Figure 1.2A and 1.3) and the mRNA is important for recognition of the targets (Lewis et al., 2003, John et al., 2004, Kiriakidou et al., 2004).

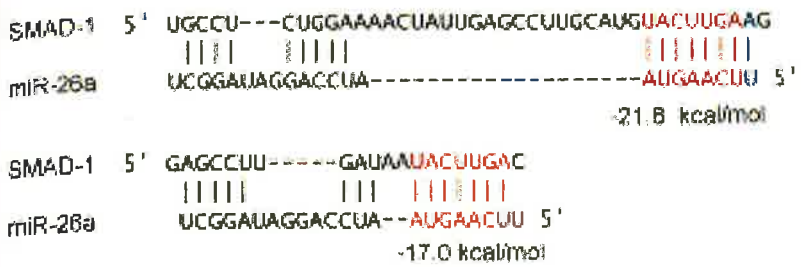


Figure 1.3: Predicted target sites for miR-26a on SMAD-1 transcript (from Lewis, Shih et al. 2003).

MicroRNA mediated inhibition of translation has been reported for a number of microRNAs (Olsen and Ambros, 1999, Seggerson et al., 2002, Nelson et al., 2004, Humphreys et al., 2005) where they associate with polysomes, silencing translation of the target at the initiation and elongation steps (Petersen et al., 2006) (Figure 1.2B and 1.2C) and causing ribosome drop off and accumulation of target transcripts in P-bodies (Liu et al., 2005, Sen and Blau, 2005). This localization is mediated via Argonaute proteins that interact with the proteins present within the P-bodies (Sen and Blau, 2005) where they are degraded by the RNA decay enzymes. Sometimes microRNA targeted, translationally repressed mRNA molecules might localize to P-bodies for storage purposes, from where they might be rescued later for translation (Kedersha et al., 2005, Pillai et al., 2005) (Figure 1.2C).

MicroRNA binding can also induce deadenylation-dependent decapping of mRNAs (Giraldez et al., 2006) followed by 5' to 3' decay (Wu et al., 2006). For some microRNAs like miR-16, binding to AU rich elements might recruit TTP (tristetraprolin) to AREs through members of the Argonaute family. TTP forms a complex with decapping enzymes like Dcp1, Dcp2 and Xrn1 in the P-bodies, thereby promoting mRNA decay (Jing et al., 2005, Kedersha et al., 2005, Lykke-Andersen and Wagner, 2005).

Hence among the various pathways available for microRNA mediated mRNA decay, the inherent stability of the mRNA molecule determines whether it will be either translationally repressed or subject to accelerated degradation (Valencia-Sanchez et al., 2006). Some of the factors

influencing this decision are sequence complementarity between the microRNA and mRNA molecule and AU composition of 3'-UTR of the mRNA, distance of the binding site from the center of long UTRs and cooperativity between microRNAs in eliciting repression (Grimson et al., 2007). However the triggering of the degradation pathway might be dependent upon certain developmental and environmental signals which are still not known.

1.4.2 MicroRNA mediated Activation of Translation

Previously it was demonstrated that tumor necrosis factor- α (TNFα) ARE can be translationally activated by serum starvation mediated cell–cycle arrest (Vasudevan and Steitz, 2007). This up-regulation in translation has been shown to be dependent upon microRNAs which facilitate recruitment of Argonaute (AGO) and fragile X mental retardation-related protein (FXR1) to AREs (Vasudevan et al., 2007). It has been proposed that microRNA mediated repression of translation is a property of proliferating cells, while cell cycle arrest results in translational up-regulation of AREs in cell lines (Vasudevan et al., 2007). These findings further expand the repertoire of regulatory activities mediated by microRNAs which might be important in their roles in development, differentiation and carcinogenesis.

1.5 Clues regarding MicroRNA Functions in Mouse Embryonic Stem Cells

The exact biological function of most individual microRNAs remains largely unknown. However, some clues regarding their role in embryonic development have been provided through studies loss of function analysis of Dicer and DGCR8, two key microRNA processing enzymes, in mice a nd stem cells. Dicer-/- mice die at embryonic day 7.5 and lack multipotent stem cells (Bernstein et al., 2003). Dicer deficient ES cells have a severe proliferation defect and they fail to differentiate (Murchison et al., 2005). However, Dicer plays an important role in the RNA interference pathway by controlling the processing of both microRNAs (Hutvagner et al., 2001,

Paddison et al., 2002) and other families of double stranded interfering RNAs. Recently it has been shown that many of the defects of Dicer-/- ES cells can be reversed by transfection of microRNAs belonging to the microRNA miR-290 cluster (Sinkkonen et al., 2008).

DGCR8 deficiency results in a global loss of all the microRNAs found in ES cells (Wang et al., 2007). In contrast to Dicer deficiency, DGCR8 deficiency results in delayed differentiation of ES cells. These results imply that the effect of Dicer loss in ES cells is independent of microRNAs (Wang et al., 2007) whereas DGCR8 deficiency phenotype is caused by the generalized loss of all microRNAs.

More direct evidence for the involvement of microRNAs in regulation of ES cell pluripotency has been provided by transfection of anti-miR-21 in ES cells which has resulted in reduction of stem cell pluripotency potentially through down-regulation of Sox2 and Nanog expression (Singh et al., 2008).

Additionally miR-134 has been shown to play a specific role in ectoderm differentiation of ES cells (Tay, 2007) by attenuating Nanog and Lrh-1 expression. These key findings have provided vital clues regarding the role of microRNAs in governing stem cell pluripotency. However, stem cell pluripotency is dependent on specific gene expression patterns governed by a consortium of pluripotency factors. In my thesis I have attempted to unravel the broader function of microRNAs in regulating this phenomenon using computational target prediction tools to extend my search for microRNA binding to a wide array of transcription factors and transcriptional networks and microRNA and mRNA expression analysis as described in subsequent chapters.

1.6 Determining MicroRNA-mRNA Interactions with miRanda, TargetScan, Pictar and Sylamer

Interactions between microRNAs and their target genes were first identified in a genetic screen carried out in *C.elegans (Lee et al., 1993)*. Identification of microRNA targets computationally through specific base-pairing rules, thermodynamic parameters using classical RNA folding programs and cross-species conservation scores offers a useful approach for discovering potential microRNA-3'UTR interactions (Sethupathy et al., 2006). Indeed, there have been some application of *in silico* microRNA target prediction methods for narrowing down potential candidates for experimental validation of targets in various biological processes ranging from cancer (Koscianska et al., 2007) to viral infection (Skalsky et al., 2007). Current experimental approaches for microRNA-target verifications rely on performing luciferase assays (Rodriguez et al., 2007b), co-expression analysis of microRNAs and their targets *in vivo* by tagging microRNAs and their targets with specific fluorescent proteins (Stark et al., 2003), quantitative PCR, *in situ* hybridisation and western blot analysis (Kuhn et al., 2008).

MicroRNA target sites have been classified into three categories (Sethupathy et al., 2006): (i) 5'-dominant canonical, which have perfect base pairing to the 'seed' sequence at the 5'end of the microRNA as well as towards the 3'end, (ii) 5'-dominant seed only, which have perfect base pairing to the 5'seed only and limited base-pairing to the 3'end and (iii) 3'-compensatory sites, which have extensive base pairing to the 3'end of the microRNA to compensate for imperfect or a shorter stretch of base pairing to the 5'end of the microRNA (Figure 1.4).

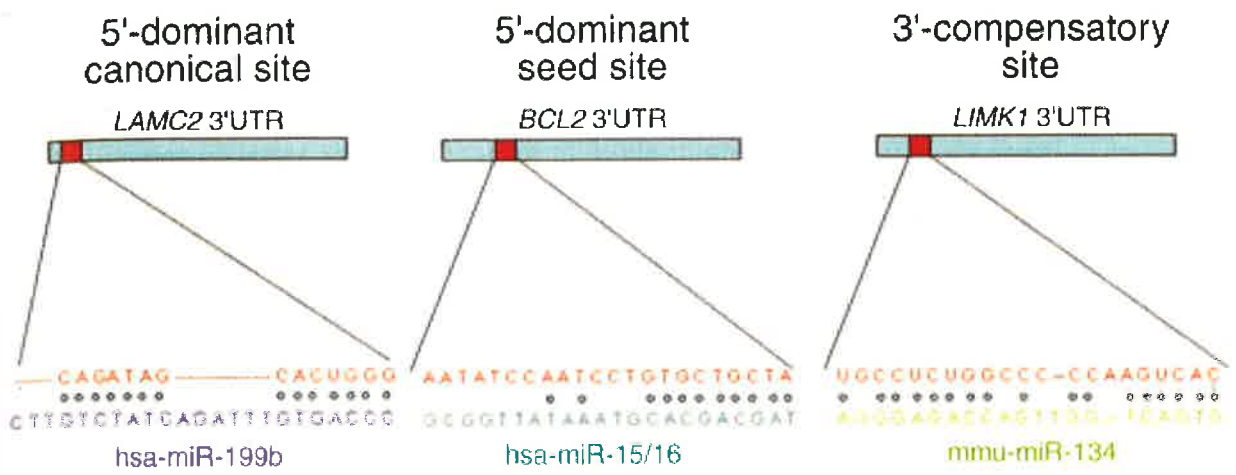


Figure 1.4: Three kinds of microRNA target sites (from (Sethupathy et al., 2006)).

miRanda is a target prediction algorithm that optimizes sequence complementarity using position specific rules and interspecies conservation (John et al., 2004) to predict target sites on 3'UTRs of mRNAs. This algorithm allows flexibility for (i) non-uniform distribution of a number of sequence-complementary sites for different microRNAs (ii) 5'-3' asymmetry, with higher scores for matches at the 5'end, compared to the 3'end, but at the same time allowing matches at the 3'end to compensate for low complementarity at the 5'end and (iii) G:U wobbles, which are detrimental to repression (John et al., 2005). These computational rules have been deduced from a wide range of experiments (Brennecke et al., 2003, Johnston and Hobert, 2003, Lin et al., 2003, Vella et al., 2004, Doench and Sharp, 2004). This program has also been found to be most sensitive for predicting 3'-compensatory target sites (Sethupathy et al., 2006) compared to other algorithms like TargetScan (Lewis et al., 2005) and PicTar (Krek et al., 2005) which have been found to be more appropriate for prediction of 5'-dominant sites (Sethupathy et al., 2006). TargetScan also takes into account certain other parameters like AU-content, distance from the center of 3'UTR and cooperativity of microRNA binding sites (Grimson et al., 2007). For, miRanda I have filtered targets from miRBase based upon *org p-values* (Griffiths-Jones et al., 2006, Griffiths-Jones et al., 2008) (which denote the probability of the same microRNA family binding to multiple transcripts within the same orthologous group) to minimize false positives.

While TargetScan and miRanda are programs that help identify all the targets of a single microRNA, Pictar (Krek et al., 2005), is another program that relies on probabilistic identification of combinations of target sites (PicTar) for multiple microRNAs on a single target. The statistical model applied by Pictar relies on synergistic effects of microRNAs binding in a cooperative manner to a single target molecule.

In my thesis I have applied these three programs to identify crucial microRNA-transcription factor transcript interactions that might be critical in the maintenance of stem cell pluripotency. Additionally I have also applied 'Sylamer', a program developed by Stijn van Dongen (paper submitted) which ranks genes in order of decreasing expression, comparing two different conditions, and calculates cumulative enrichment of microRNA sequence 'words' or motifs in the entire spectrum of differentially expressed genes. Sylamer plots can help detect specific enrichments of sequence motifs and microRNA seeds in the 3'UTRs of entire transcriptomes.

CHAPTER TWO: *IN SILICO* DETECTION OF MICRORNAS INTERACTING WITH KEY PLURIPOTENCY TRANSCRIPTS

2.1 Introduction: Generating Hypotheses for MicroRNA Functions in Mouse Embryonic Stem Cells

Embryonic stem cells have the potential to continuously renew themselves in culture and an ability to differentiate into multiple cell types upon manipulation of culture conditions. This ability of stem cells to form any cell is termed as 'pluripotency'. Major evidence for the potential role of microRNAs in regulating stem cell pluripotency has been provided by knocking out DGCR8 in stem cells, one of the key proteins involved in binding to the stem-loop structure of microRNA precursors in the microRNA production pathway (Wang et al., 2007). As mentioned earlier, DGCR8 deficiency results in delayed stem cell differentiation and failure of these cells to down-regulate expression of certain transcription factors like Oct4, Nanog and Sox2 during this process (Wang et al., 2007). Hence, one can hypothesize that microRNAs expressed in stem cells may function in regulating stem cell pluripotency and differentiation by the following mechanisms:

1. MicroRNAs directly bind to key stem cell specific transcripts belonging to the Yamanaka factors (Takahashi and Yamanaka, 2006) and Nanog (Mitsui et al., 2003), regulating their expression in stem cells.
2. MicroRNAs bind to transcripts of co-regulators (Zhou et al., 2007) of Oct4, Nanog and Sox2, regulating their expression.
3. MicroRNAs might bind to 3'UTRs of the downstream targets of Oct4 and Nanog transcription factors (Loh et al., 2006) which would constitute entire transcriptional networks (Chapter three).

In order to test the first two hypotheses, the 3'UTRs of transcripts belonging to Yamanaka transcription factors, Nanog (Mitsui et al., 2003) and co-regulators of Oct4, Nanog and Sox2, mainly Esrrb (estrogen related receptor beta), Stat3, Tcf7 (Transcription factor 7), Sall4 (Sal-like protein 4) and Lrh1 (liver receptor homolog 1) (Zhou et al., 2007), were scanned for microRNA binding sites using a microRNA target prediction algorithm, miRanda (John et al., 2004) on miRBASE (Griffiths-Jones et al., 2006), a web-based resource of all known microRNAs and targets. This algorithm reports predicted microRNA-target interactions based upon sequence complementarity between a mature microRNA and a target site in the 3'UTR of a transcript, binding energy of the microRNA-target duplex and evolutionary conservation of the target site sequence. Only microRNAs with predicted target sites having *org p-value* (Griffiths-Jones et al., 2006, miRBase Targets) less than 0.01 were taken. All star sequences (Figure 2.1), corresponding to the part of the precursor sequences, complementary to the mature microRNAs were excluded from the analysis. Further, all the microRNAs identified to bind to these transcripts were filtered based upon their expression signatures in stem cells. The identity of microRNAs expressed in stem cells was determined from microRNA.org (see Materials and Methods). I have also used TargetScan (Lewis et al., 2003, Lewis et al., 2005, Grimson et al., 2007) and Pictar to detect interactions between these transcripts and stem cell microRNAs.

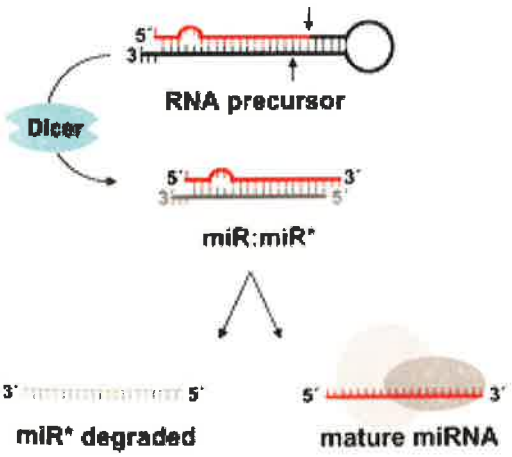


Figure 2.1: Star sequence. Asymmetric processing of microRNA hairpin loops by Dicer results in mature microRNAs and the complementary star sequences which are subject to degradation. Figure taken from http://www.sciencemag.org/feature/data/prizes/ge/2006/schwarz.dtl

2.2 Results

2.2.1 Increased Propensity of ES Cell MicroRNAs towards ES Cell Pluripotency Transcription Factor Transcripts

In order to test the first two hypotheses which predict that microRNAs might directly bind to 3'UTRs of transcripts belonging to 'Yamanaka factors' (Takahashi and Yamanaka, 2006), Nanog (Chambers et al., 2003), and a number of additional transcription factors like Esrrb, Stat3, Tcf7, Sall4 and Lrh1 implicated in regulating stem cell pluripotency (Zhou et al., 2007), I searched for binding sites on these transcripts corresponding to microRNAs found to be expressed in ES cells (from microRNA.org, See Materials and Methods). These searches were performed using miRanda (on miRBase) (Griffiths-Jones et al., 2006), TargetScan (Grimson et al., 2007) and Pictar (Krek et al., 2005). As a control set I used ten additional transcription factor transcripts found to be expressed in ES cells and down-regulated during differentiation (derived from mRNA expression Illumina data, Chapter four).

As seen in Figure 2.1, there are greater number of binding sites observed on the ten core pluripotency transcription factor transcripts as compared to the control set (p-value : 0.002). Binding sites corresponding to ES microRNAs are predicted for Oct4 (Pou5f1), Nanog, Sox2, Klf4, cMyc, Stat3, Tcf7, Sall4 and Lrh1 transcripts. There is considerable overlap between predictions for microRNA binding sites found on Sox2, cMyc, Stat3, Tcf7 and Sall4 transcripts obtained with these three prediction programs. There are also a number of binding sites detected exclusively by individual programs which reflect the differences in the statistical procedures applied by them to generate microRNA target predictions (Sethupathy et al., 2006). The greatest enrichment of sites is observed on Stat3 and Sall4 transcripts.

Analysis of microRNAs interacting with these transcripts reveals that multiple transcripts could be potentially regulated by a single microRNA entity. For example, microRNA mmu-miR-21 has been shown recently to regulate stem cell pluripotency by affecting Nanog and Sox2 expression (Singh et al., 2008). In addition to Nanog and Sox2, I also find binding sites corresponding to this microRNA on Tcf7 and Stat3 transcripts (Figure 2.1). MicroRNA mmu-mir-24 might potentially co-regulate Oct4 and Tcf7 transcripts. Further, one also observes that microRNAs belonging to miR-290 and miR-93 cluster families might co-regulate Stat3 expression (Figure 2.6) indicating common functionality of these microRNAs in regulating stem cell pluripotency. Overall these results suggest that microRNAs expressed in ES cells have greater propensity of binding to key stem pluripotency transcription factor transcripts.

Transcripts having longer UTR length might be expected to possess greater number of sites by chance alone. In order to estimate whether these pluripotency transcripts have longer 3'UTRs than the control set, I compared the average length of the 3'UTRs and found that the 3'UTRs belonging to pluripotency transcripts are slightly longer than the control transcripts, although the difference between them is not statistically significant (Figure 2.3). It is likely that the UTR length has very little effect on these results for ES cell microRNA binding site enrichment on stem cell transcripts. However one cannot rule out the possibility that these transcripts might have evolved slightly longer UTRs to accommodate microRNA regulatory functions in determining ES cell pluripotency.

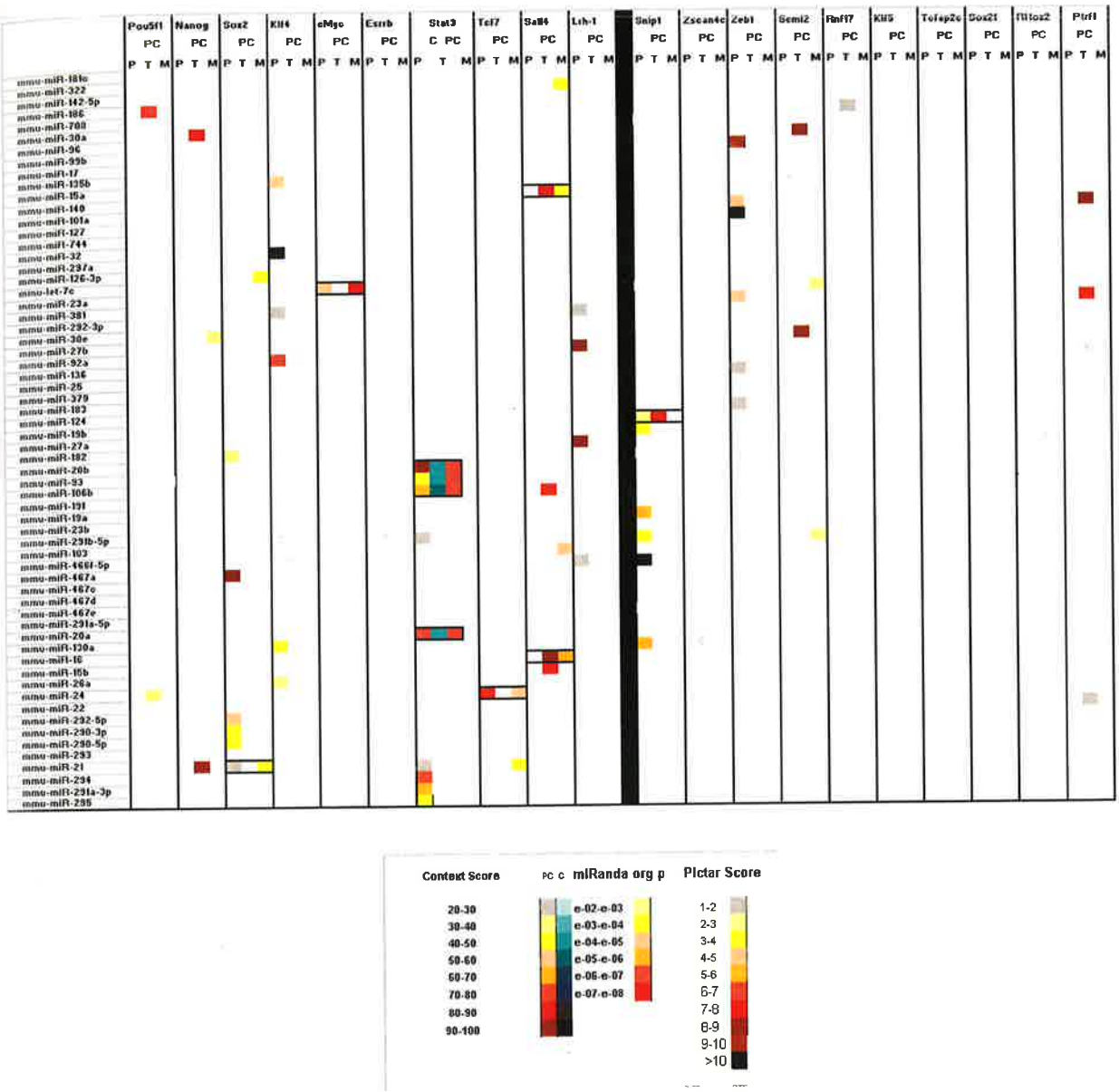


Figure 2.2: Increased Propensity of Stem Cell MicroRNAs towards Core Pluripotency Transcription Factor Transcripts. Ten core pluripotency transcription factor transcripts (left panel) and ten control transcription factor transcripts (right panel) were screened for microRNA binding sites using Pictar, Targetscan and miRanda. MicroRNAs have been derived from microRNA.org based upon their expression in ES cells. Context score (TargetScan) with distinct poorly conserved (PC) and conserved (C) scores, org p-value (miRanda) and Pictar score ranges are depicted below the chart. One observes that core set of pluripotency transcription factors have greater number of binding sites compared to control set (p-value: 0.002, single tailed t-test).

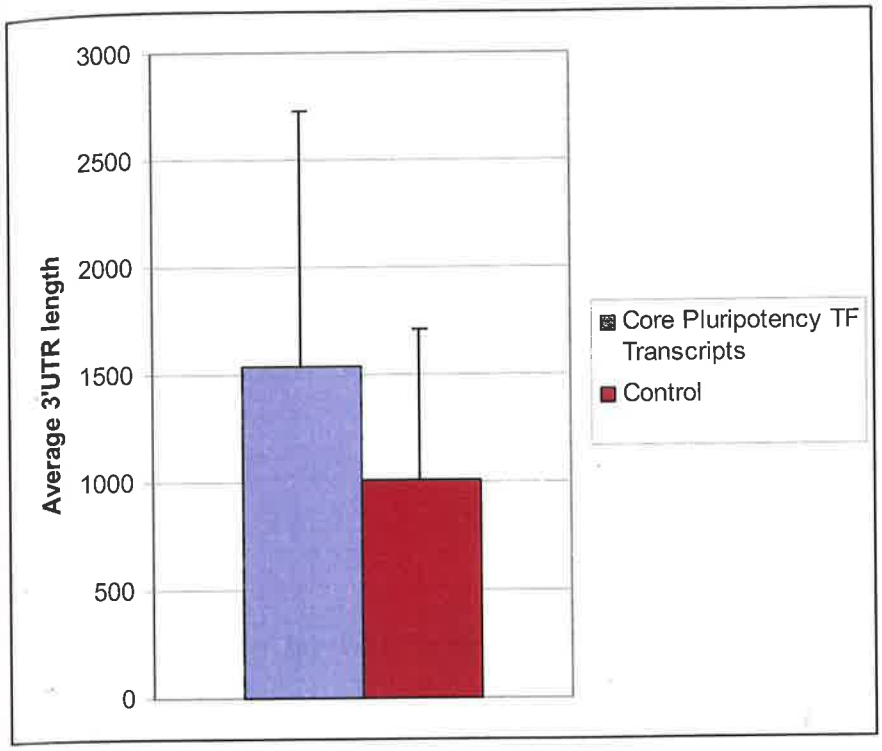


Figure 2.3: Average 3'UTR length of core set of pluripotency transcripts and control set. Also shown are error bars (n=10, s.d). The difference between the pluripotency transcript set and the control set is not statistically significant (p: 0.23).

2.2.2 Is Enrichment of ES Cell MicroRNA Binding Sites on 3'UTRs of Pluripotency Transcripts due to Chance?

After binding sites corresponding to key ES cell microRNAs were detected on 3'UTRs of key pluripotency transcripts, it was decided to test whether this enrichment of binding sites is caused by 'chance'. In order to test this hypothesis, ten thousand iterations were performed for detecting binding sites corresponding to all microRNAs expressed in stem cells on 3'UTRs of four out of ten randomly selected transcripts using miRanda. Transcript sets with greater than 0 hits for microRNAs were counted and the p-value was calculated to be 0.24. A very strict cut-off of greater than 0 hits for microRNAs binding to randomly selected transcripts could be responsible for this high p-value. As Sall4 possesses binding sites corresponding to 4 microRNAs, it was decided to

repeat this test of enrichment for greater than 1, 2 and 3 microRNA binding sites in 3'UTRs of these randomly selected transcripts. A reduction in p-value was observed with an increase in microRNA hits from 0 to 3 (Table 2.1). These results imply that there might be a slight enrichment of ES microRNA binding sites on Sall4 transcript as compared to other randomly selected transcripts, although transcripts containing greater than 0 microRNA binding sites, may be identified by chance alone (p=0.24).

Table 2.1: p-values for ES MicroRNA Enrichments in Randomly Selected Transcript Sets (Acknowledgements: Dr. Harpreet Saini)

ES MicroRNA Hits in 4/10 Randomly Selected Transcripts	p-value
>0	0.24
>1	0.19
>2	0.078
>3	0.015

2.2.3 Predicted Regulatory Networks for MicroRNAs in Mouse ES Cells

These *in silico* results suggest that ES cell specific microRNAs might directly regulate the expression of Oct4, Nanog, Sox2, Klf4, cMyc, Stat3, Sall4, Tcf7 and Lrh-1 transcripts. Using these results, hypothetical regulatory networks were built to graphically depict interactions between microRNAs and these key stem cell specific transcripts in relation to stem cell pluripotency. Figures 2.4-2.6 depict these regulatory networks. It is known that knock-down of Oct4 results in differentiation of stem cells into the endoderm and trophoblast lineages (Hay et al., 2004). ES cells containing homozygous deletions for Nanog lose their pluripotency and differentiate into the extra-embryonic endoderm lineage (Mitsui et al., 2003, Yuan et al., 1995). Sox2 cooperates with Oct4 in regulating gene expression and is expressed in pre-implantation embryos and in ES and EC

(embryonic carcinoma) cells in a similar manner as Oct4 (Yuan et al., 1995). These transcription factors form a core network regulating pluripotency in stem cells (Figure 2.4). Sall4 and Tcf7 have been implicated in the extended regulatory network of transcription factors in stem cells, composed of Oct4, Nanog and Sox2, by co-regulating the expression of these transcripts (Loh et al., 2006). The overall model demonstrating the putative impact of microRNAs on these transcription factor transcripts is shown in Figure 2.4.

Klf4 or Kruppel-like-factor-4 has been identified as a binding partner for Oct4 in regulating the expression of an ES cell specific gene *Lefty*1 which has been shown to be important for left-right axis determination during early embryogenesis (Nakatake et al., 2006). Although Klf4 null mice show no abnormalities during early embryogenesis, over-expression of Klf4 has been shown to inhibit embryoid body formation during differentiation indicating that this gene is essential for maintenance of stem cells in the pluripotent state (Li et al., 2005). Figure 2.5 highlights the model, whereby stem cell specific microRNAs might negatively regulate Klf4 transcript thereby regulating the expression of *Lefty*1.

The key player of LIF (leukaemia inhibitory factor)dependent self-renewal pathway in embryonic stem cells, Stat3, has been shown to regulate the expression of Myc transcription factor (Cartwright et al., 2005). Maintained expression of unphosphorylated stable Myc (T58A) is sufficient for stem cell pluripotency, even in the absence of LIF, while the dominant negative form of Myc promotes premature differentiation. Phosphorylation of T58A promotes GSK3β mediated degradation of Myc (Sears et al., 2000, Cartwright et al., 2005).

Figure 2.6 depicts the overall model for microRNAs negatively regulating the expression of cMyc transcript. Hence it is possible that microRNAs might provide an additional layer of regulation to maintain the expression level of these key transcription factors in a state of homeostasis.

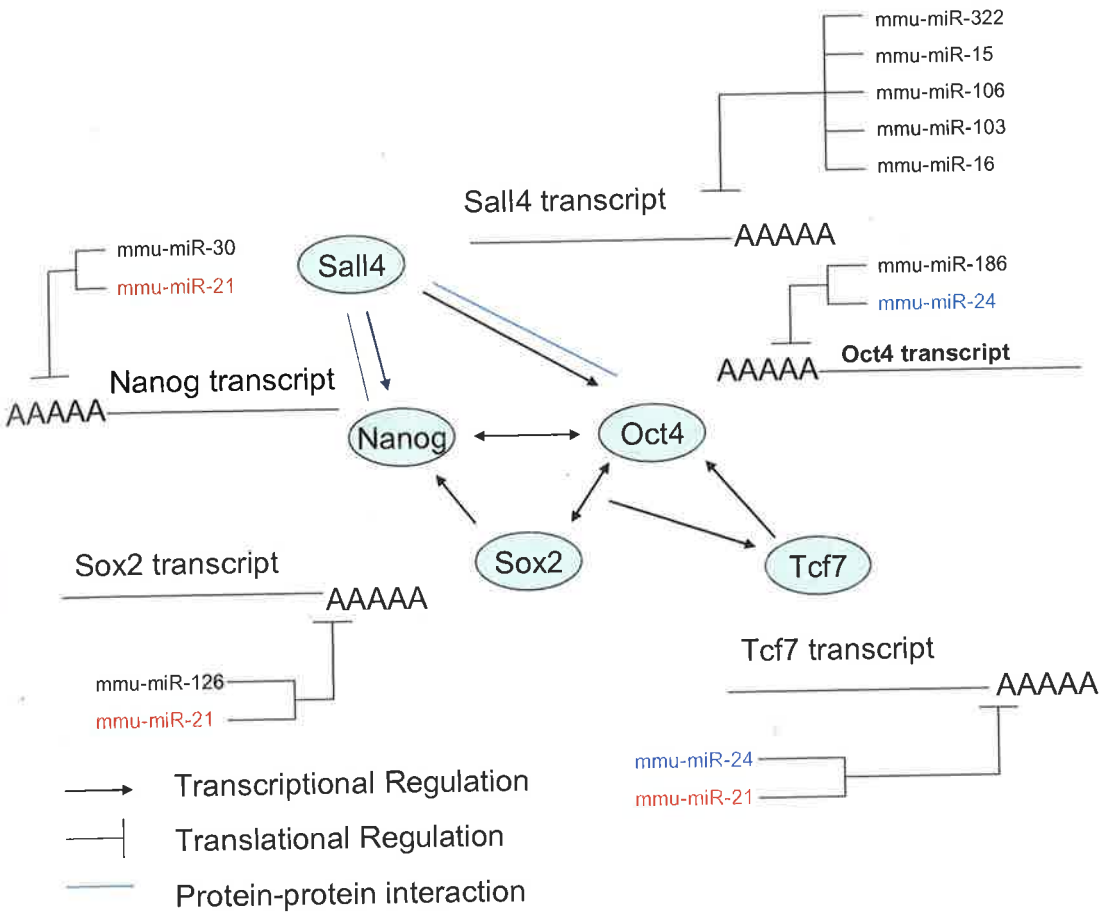


Figure 2.4: Diagram depicting predicted binding sites for microRNAs on transcripts belonging to key pluripotency transcription factor transcripts. Oct4, Nanog and Sox2 form the core of this regulatory network, with Sall4 and Tcf7 playing the role of regulators of these core transcription factors. There exists a feed-back relationship between these core set of regulators and Sall4 and Tcf7. MicroRNAs mmu-miR-21 and mmu-mir-24 have been highlighted in red and blue text respectively as they are predicted to target more than one set of transcripts. Diagram partially adapted from Zhou *et al.*, 2007. (Legend shown in bottom left of the Figure).

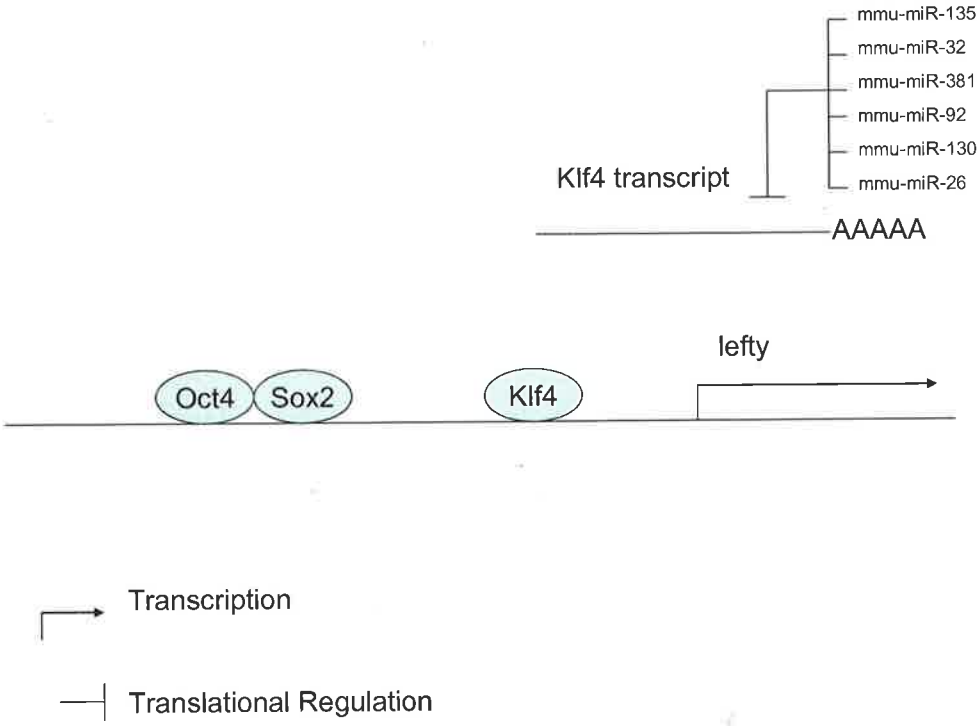


Figure 2.5: Diagram depicting predicted binding sites for microRNAs on Klf4 transcript. Klf4 cooperates with Oct4 and Sox2 to activate the promoter for *Lefty* in embryonic stem cells. Inhibition of Klf4 by microRNAs is predicted to reduce lefty expression and stem cell pluripotency. Legend is shown in the bottom left of the figure.

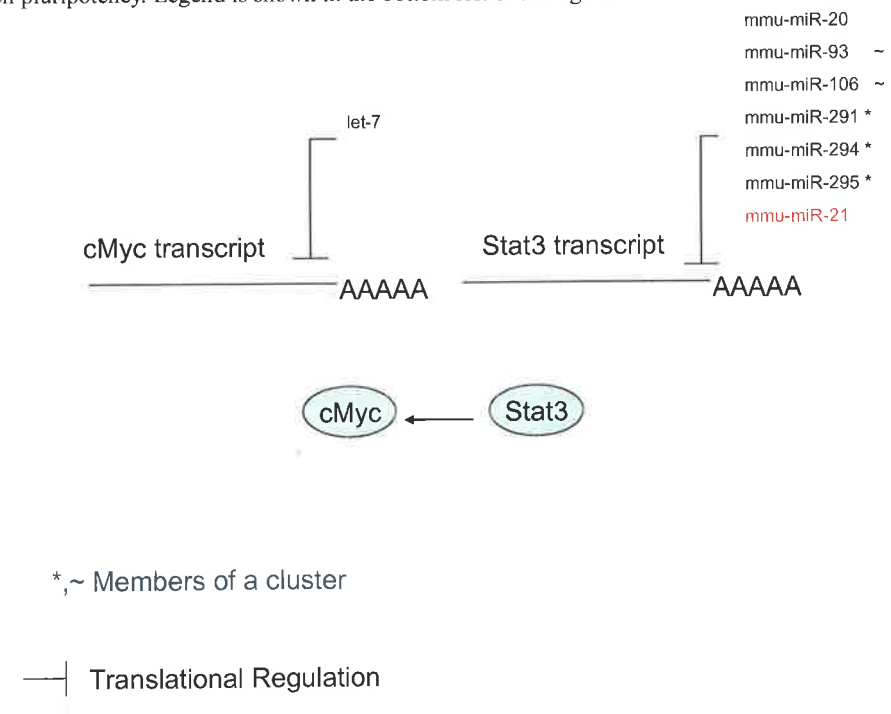


Figure 2.6: Diagram depicting predicted binding sites for microRNAs on the cMyc and Stat3 transcripts. Inhibition of these two transcripts is predicted to reduce stem cell pluripotency. MicroRNAs mmu-miR-291, mmu-miR-294 and mmu-miR-295 belong to the same cluster, while mmu-miR-93 and mmu-miR-106 belong to a separate cluster. MicroRNA miR-21 is highlighted in red as it has been shown to regulate a number of other pluripotency transcripts like Sall4, Sox2 and Tcf7. Legend is shown in the bottom left of the figure.

2.3 Summary

Scanning 3'UTRs of key stem cell pluripotency transcripts like Oct4, Nanog, Klf4, Sox2, Stat3,Sall4, cMyc, Tcf7 and Lrh1 reveals the presence of binding sites for numerous microRNAs also expressed in stem cells (Houbaviy et al., 2003, Thomson et al., 2004). Moreover, I find that microRNAs expressed in ES cells show increased propensity towards these transcription factor transcripts as compared to other randomly selected transcripts. It is also observed that multiple transcripts might be regulated by a single microRNA entity. Additionally, a number of microRNAs belonging to the same cluster family are predicted to interact with Stat3 transcription factor transcript indicating common functionality of family members in regulating stem cell pluripotency. Overall, these results hint at a potential function of these microRNAs in regulating the expression of pluripotency transcripts in stem cells.

CHAPTER THREE: *IN SILICO* DETECTION OF MICRORNAS IMPACTING THE OCT4/NANOG TRANSCRIPTIONAL NETWORK

3.1 Introduction

The properties of self-renewal and pluripotency of embryonic stem cells depend upon specific gene expression patterns which in turn are determined by transcription factors. Oct4 and Nanog constitute a key set of transcription factors that regulate pluripotency and self-renewal in stem cells. In order to study the microRNAs impacting the downstream network of genes affected by a combination of these stem cell transcription factors, the microarray data obtained from experiments involving knock-down of Oct4 and Nanog by RNA interference in embryonic stem cells (Loh et al., 2006), was downloaded from GEO DataSets at NCBI as CEL files and processed in GEPAS (Herrero et al., 2003) for log transformation and normalisation (See Materials and Methods). Figure 3.1 shows the entire set of differentially expressed genes in the Oct4 and Nanog transcriptional network. Genes expressed differentially in wild type and knock-down cells were ranked according to their t-statistics and scrutinized for motif and microRNA binding site enrichments using two different computational approaches. In the first approach a perl script (provided by Dr. Anton Enright) was used to detect microRNA binding site enrichments in genes that were most up and down-regulated in stem cells upon Oct4/Nanog knock-down compared to wild-type cells. In the second approach entire transcriptional networks downstream of these transcription factors were scanned for microRNA binding site enrichments using 'Sylamer' (program devised by Dr. Stijn van Dongen, paper submitted).

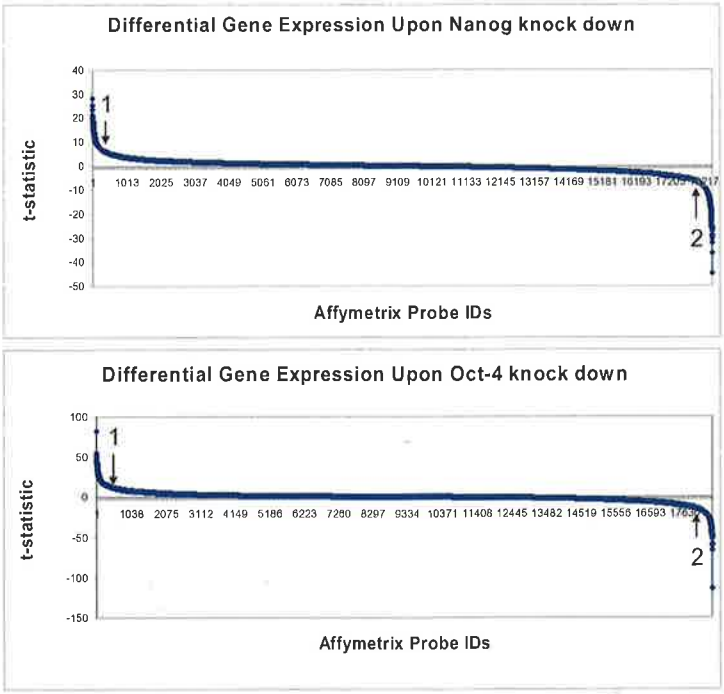


Figure 3.1: Genes in the Oct4/Nanog Transcriptional Network (Loh et al., 2006). A. Genes either down-regulated, 1, or up-regulated, 2, upon Nanog or Oct4 knock-down by RNA interference in mouse embryonic stem cells. The y-axis depicts the t-statistic values where positive values indicate genes which are most highly expressed in the control (mouse embryonic stem cells not subjected to RNA interference for either Nanog or Oct4) versus experimental cells (subjected to knock-down by RNA interference for Nanog or Oct4). The genes were filtered according to a family wise error rate (FWER) <0.001 and 0.01. There were 11 and 30 genes in the down-regulated and up-regulated sets (upon Nanog knock-down) respectively which passed this filtering in the Nanog transcriptional network at p< 0.001. Similarly, 76 and 123 of the down-regulated and up-regulated genes respectively (upon Oct4 knock down), passed this filtering for the Oct4 transcriptional network at p<0.001. At p<0.01 there were 40 genes which were down-regulated and 66 genes which were up-regulated in the Nanog transcriptional network. For the Oct4 transcriptional network, there were 209 and 328 genes in the down-regulated and up-regulated genes respectively.

3.2 Scanning most Up-regulated and Down-regulated Genes in the Oct4/Nanog Transcriptional Network

The most up-regulated or down-regulated genes were identified by filtering out genes based on their t-statistics and FWER (Family wise error rate< 0.001 and <0.01) (see Materials and Methods). Figure 3.2 describes a model for the effect observed in ES cells upon knock-down of Oct4 and Nanog and the resultant changes in gene expression.

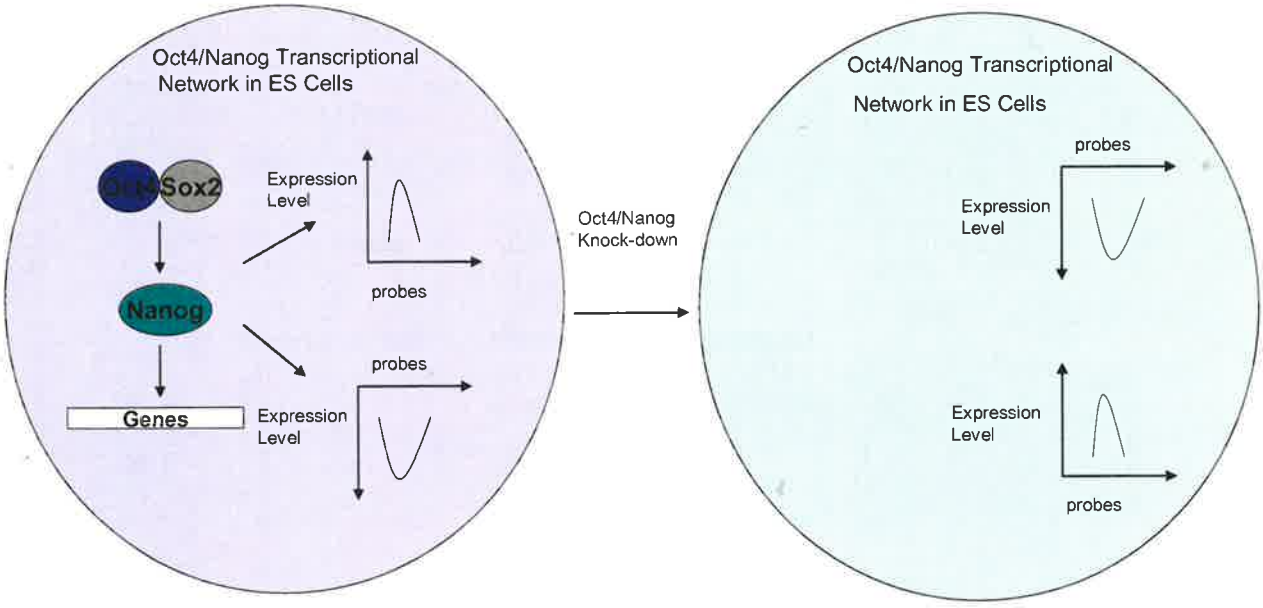


Figure 3.2: Graphical depiction of changes in gene-expression upon Oct4/Nanog knock-down in ES cells. Knocking down either Oct4 or Nanog causes ES cells to differentiate and also results in dramatic changes in the genes in the entire network. Genes previously up-regulated in ES cells are down-regulated upon Oct4/Nanog knock-down, while genes down-regulated in ES cells are subsequently up-regulated.

Entire sets of up-regulated and down-regulated genes, were scanned for the presence of 6-mer, 7-mer, 8-mer and double 6-mer microRNA binding sites. The 6-mer sequence corresponds to the 'seed' sequence which has been shown to be important for microRNA binding. The 7-mer and 8-mer sequences are extensions of this seed sequence. MicroRNA mmu-miR-101a is shown as an example (Figure 3.3). The over-representation cut-off scores selected for microRNA enrichment in specific gene sets are listed in the table in Figure 3.3. These scores were calculated as described in the Materials and Methods Section.

A. mmu-miR-101a Seed Patterns for finding Overrepresented Targets in the 3'UTRs of a Gene Set

```
6'mer          CAUGACAU
                |||||||
               GTACTGTA

               CAUGACAU
5'7mer         ||||||||
               GTACTGTA

3'7mer         CAUGACAU
               ||||||||
               GTACTGTA

8-mer          CAUGACAU
               ||||||||
               GTACTGTA

Double 6'mer   CAUGACAU..... CAUGACAU
               |||||||       |||||||
               GTACTGTA..... GTACTGTA
```

Identification of MicroRNAs Overrepresented in Up-regulated/Down-regulated Genes Set					
	6-mer	5'-7-mer	3'-7mer	8-mer	Double 6-mer
Overrepresentation Filtering Scores	>1.0	>1.2	>1.2	>1.5	>1.0

Figure 3.3: MicroRNA binding patterns scanned in specific up-regulated and down-regulated gene sets, with mmu-miR-101a shown as an example (A). Targets obtained were subjected to filtering for seed pattern overrepresentation scores according to the figures shown in the table.

The flowchart shown in Figure 3.4 describes the entire protocol for target identification. MicroRNA targets were discovered for most up-regulated and down-regulated gene sets in the Nanog and Oct4 (Figures 3.1) transcriptional networks. Genes downstream of these transcription factors are important in the maintenance of stem cell pluripotency (Loh et al., 2006).

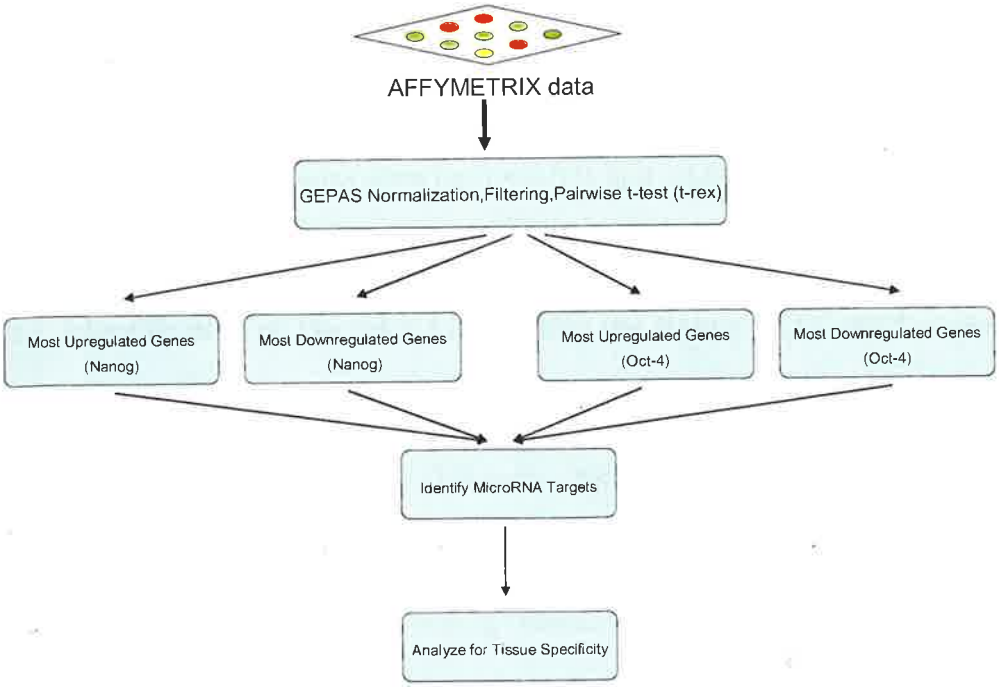


Figure 3.4: Flowchart showing the protocol for microRNA target identification in the Oct4 /Nanog Transcriptional Network.

According to the hypothesis of 'selective avoidance' proposed by Farh *et al.* (Farh et al., 2005), microRNAs and their targets localise in a different space-time during development. Hence, one would expect microRNAs and their targets (which are normally expressed in stem cells and down-regulated upon Oct4 and Nanog knock-down) to localise in different tissues or cell types during development. Hence, in order to investigate if computational results obtained from this analysis are in agreement with 'avoidance hypothesis', information regarding tissue specific localisation of microRNAs predicted to interact with these genes was also determined from expression data obtained by Landgraf *et al.* (Landgraf et al., 2007).

3.3 Results

MicroRNAs impacting genes most up and down-regulated upon Oct4 and Nanog knock-down at FWER (Family wise error rate) < 0.001 and <0.01 were identified using a perl script.

3.3.1 Identification of MicroRNAs targeting the Nanog and Oct4 Transcriptional Network at FWER<0.001

There were 11 transcripts down-regulated upon Nanog knock-down at this FWER cut-off value of 0.001. As can be observed in Figure 3.5, mmu-miR-345 is predicted to target the most down-regulated genes in the Nanog transcriptional network. mmu-miR-345 is specifically expressed during mammalian brain development (Miska et al., 2004). It is also expressed in the adult brain, kidney and placenta (Landgraf et al., 2007). Most of the high affinity (8-mer) binding sites for these microRNAs are concentrated in 2 of 11 transcripts in this gene set. There is a greater than 8 fold enrichment of the 8-mer binding sequence in this gene set compared to a random set of UTRs obtained from the mouse genome. These results are in agreement with the 'avoidance hypothesis' described previously according to which target expression often precedes microRNA expression during development (Farh et al., 2005). In this case, the expression of genes downstream of Nanog transcription factor in embryonic stem cells precedes the expression of miR-345 in the mammalian brain during development (Miska et al., 2004).

In the Oct4 transcriptional network, one detects 76 transcripts downregulated upon Oct4 knock-down in ES cells at this FWER cut-off value. MicroRNA mmu-miR-101a is seen to target the most down-regulated set of genes (Figure 3.6). This microRNA shows widespread expression in differentiated tissues like brain, skin, adipocytes, kidney, T-cell, B-cell, pancreas, testis, ovary, placenta and embryo (Landgraf et al., 2007). This microRNA is also found to localise in embryonic stem cells and potentially targets 4 of 76 transcripts in this gene set with high specificity corresponding to 8-mer target sequence. There is a greater than 3-fold enrichment for the 8-mer

binding sequence in this gene set compared to a random set of 3'UTRs obtained from the mouse genome.

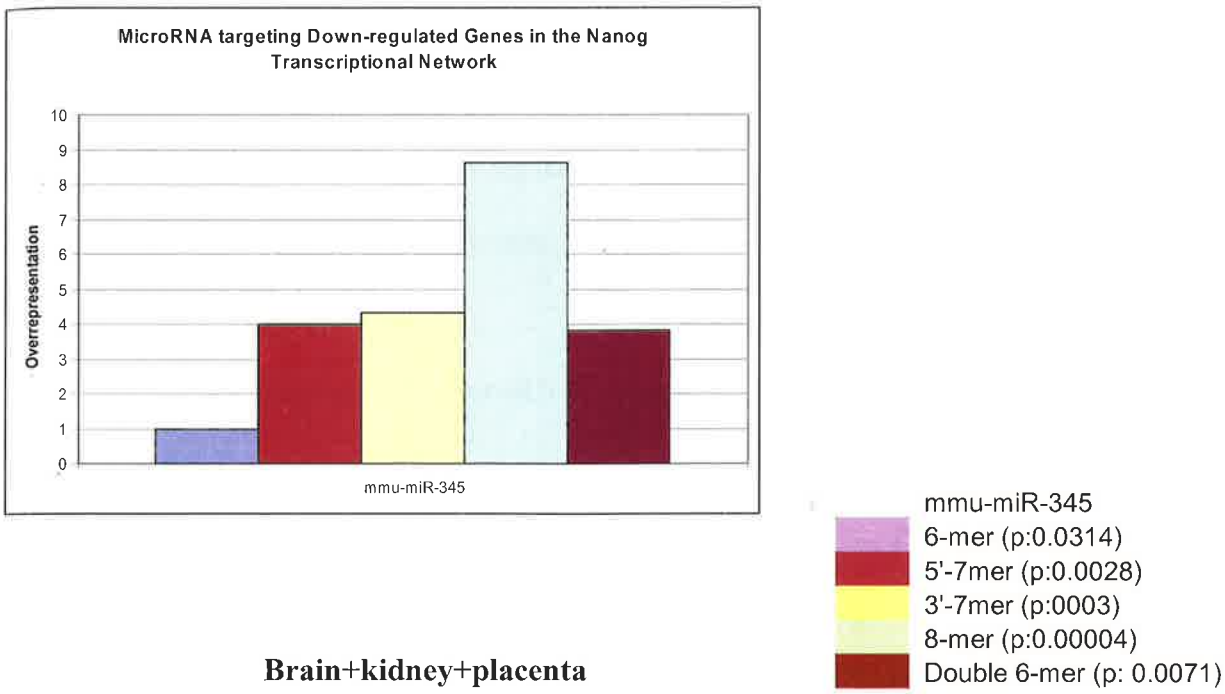


Figure 3.5: A. MicroRNAs targeting the Nanog Transcriptional Network at FWER<0.001. A. mmu-miR-345, expressed in brain, kidney and placenta, targets the most down-regulated genes in the Nanog transcriptional network upon Nanog knock-down in ES cells compared to Nanog expressing ES cells. The p-values for individual binding sites are also shown.

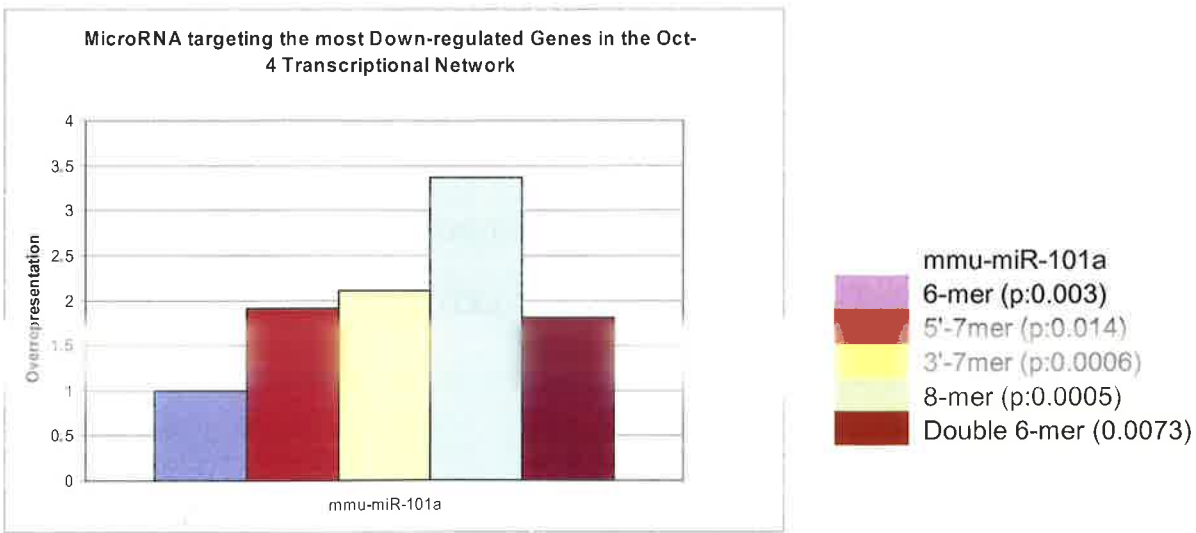


Figure 3.6: MicroRNA targeting the Oct4 transcriptional network at FWER<0.001. mmu-miR-101a targets the most down-regulated genes in the Oct4 transcriptional network upon Oct4 knock-down in ES cells compared to ES cells expressing Oct4. This microRNA is expressed in ES cells, brain, skin, adipocytes, kidney, T-cell, B-cell, pancreas, testis, ovary, placenta and embryo. The p-values for individual microRNAs are shown.

There is no clear agreement of these results with the 'avoidance hypothesis' (Farh et al., 2005) as mmu-miR-101a shows overlap in its expression pattern with its target genes in embryonic stem cells.

No microRNAs were found to be enriched in the up-regulated gene sets at this cut-off. Hence, it was decided to screen up-regulated and down-regulated genes at lower FWER cut-off values of 0.01 for microRNA binding sites.

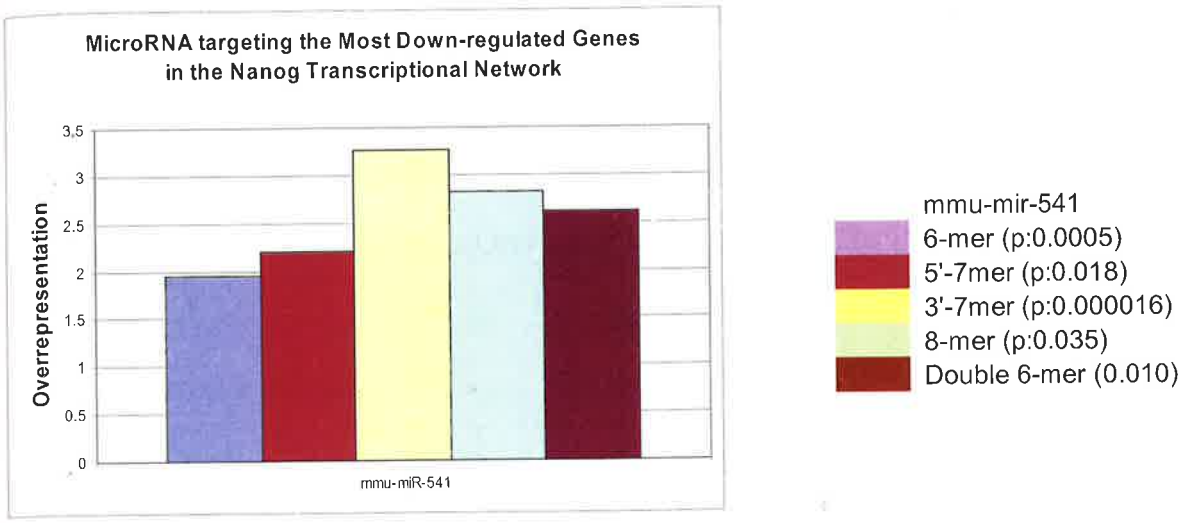
3.3.2 Identification of MicroRNAs targeting the Nanog and Oct4 Transcriptional Network at FWER<0.01.

At a higher FWER cut-off threshold of 0.01, there were 40 genes which were down-regulated and 66 genes which were up-regulated in the Nanog transcriptional network. For the Oct4 transcriptional network, there were 209 and 328 genes in the down-regulated and up-regulated sets respectively. In the Nanog transcriptional network, mmu-miR-541 was found to be enriched in the most down-regulated genes. This microRNA shows widespread expression in various differentiated tissues like the brain, kidney, lung, pancreas, ovary, embryo and stem cells (Landgraf et al., 2007). 14 of 40 transcripts show high affinity for the 8-mer binding site on this microRNA. There is a greater than two-fold enrichment observed for the 8-mer binding sequence in this set of genes compared to a random 3'UTRs obtained from the mouse genome (Figure 3.7(A)). Again, these results are contradictory to the 'avoidance hypothesis' (Farh et al., 2005).

mmu-miR-300 was found to be enriched in up-regulated genes in the Nanog transcriptional network. 19 of 66 transcripts showed high affinity of binding of the 8-mer sequence (3.7(B)). This 8-mer sequence is overrepresented more than three fold in these 3'UTRs as compared to randomized set from the whole mouse genome. Furthermore, mmu-miR-300 is expressed in embryonic stem cells (Landgraf et al., 2007, Houbaviy et al., 2003). Here, one detects a reverse

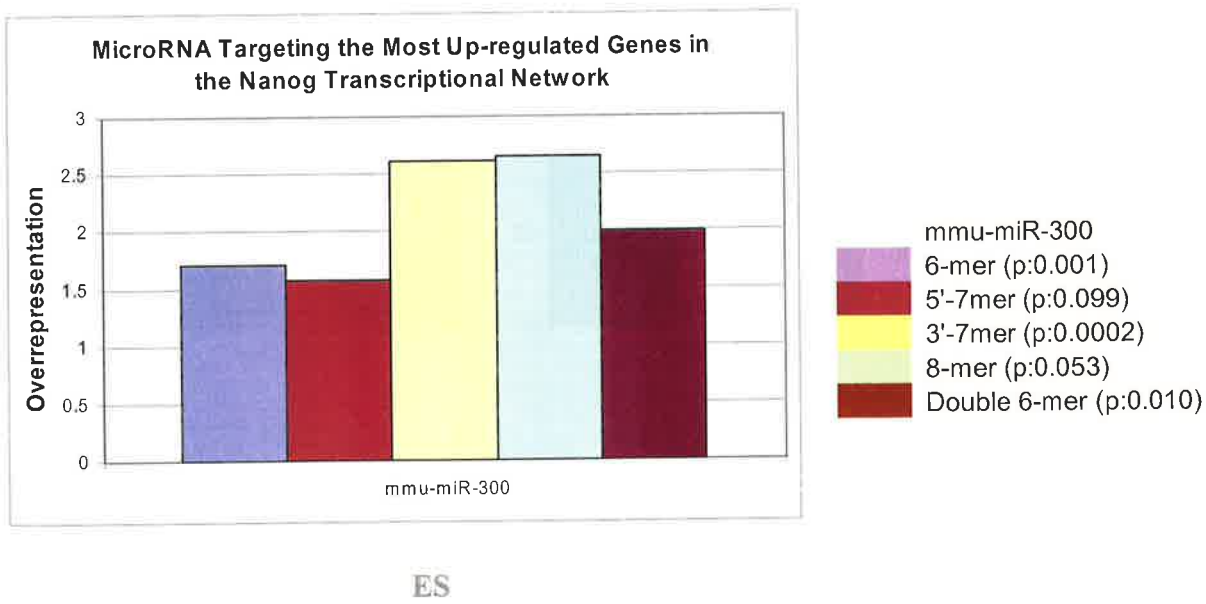
scenario of the 'avoidance hypothesis' where microRNA expression is seen to precede expression of target genes most up-regulated upon Nanog knock-down in stem cells.

A.



ES+brain+kidney+lung+pancreas+ovary+embryo

B.



ES

Figure 3.7 MicroRNA targeting the Nanog Transcriptional Network at FWER<0.01. A. mmu-miR-541 targets the most down-regulated genes in the Nanog transcriptional network upon Nanog knock-down in ES cells compared to wild type ES cells. This microRNA is expressed in ES cells and differentiated tissues including brain, kidney, lung, pancreas, ovary and embryo. B. mmu-miR-300 targets the most up-regulated genes in the network. This microRNA is expressed in stem cells. The p-values for individual binding sites are also shown.

For the Oct4 transcriptional network, mmu-miR-130a and mmu-miR-291a display enrichment in the most up-regulated genes. 94 of 328 and 97 of 328 transcripts show high enrichment of the 8-mer binding sequence for mmu-miR-130a and mmu-miR-291a respectively.

Further, there was a greater than two-fold and a greater than 1.5-fold over-representation of the 8-mer binding sequence for mmu-miR-130a and mmu-miR-291a respectively, in the 3'UTRs of the up-regulated gene sets compared to a randomized set of mouse 3'UTRs (Figure 3.8). Both these microRNAs are expressed exclusively in stem cells. There was no significant microRNA enrichment observed for the down- regulated genes. In these results, one also detects a reverse scenario of the 'avoidance hypothesis' where microRNA expression is seen to precede expression of target genes most up-regulated upon Oct4 knock-down in stem cells.

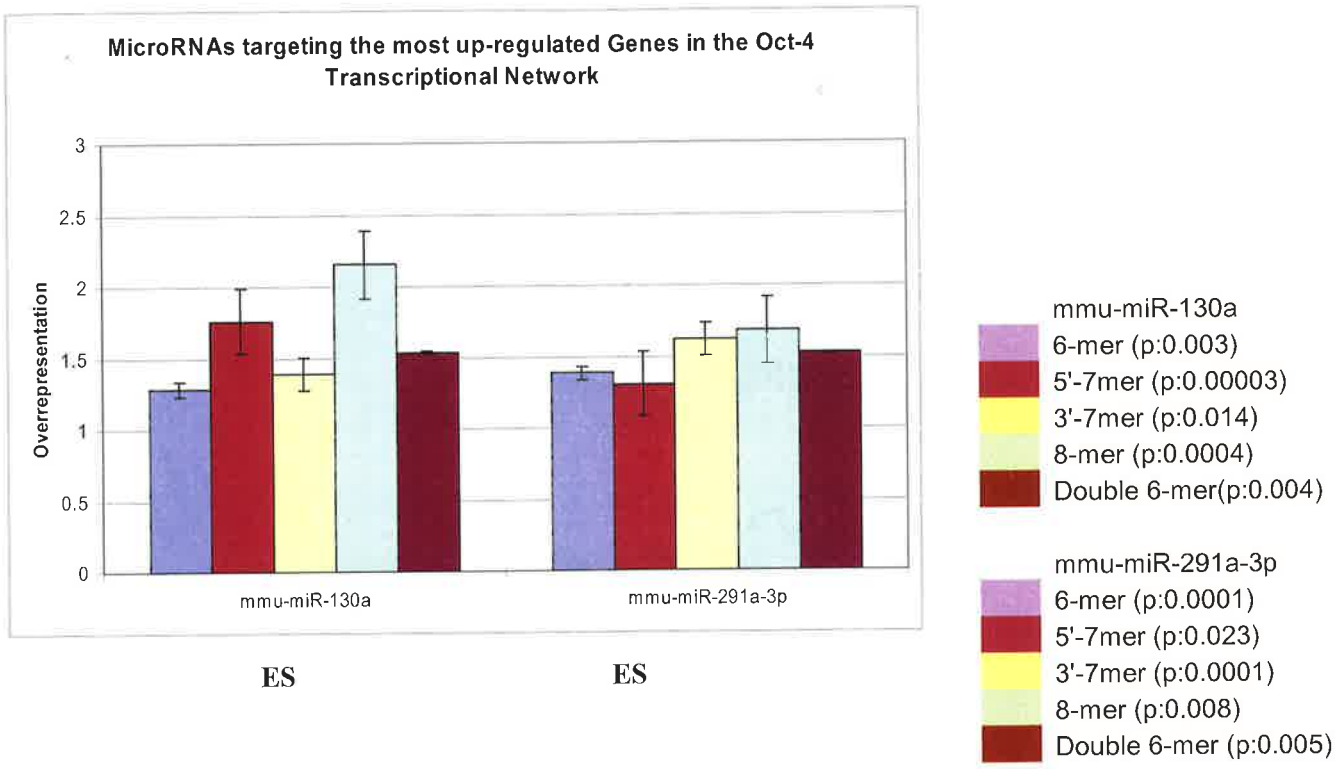


Figure 3.8. MicroRNA targeting the Oct4 Transcriptional Network at FWER<0.01. mmu-miR-130a and mmu-miR-291a-3p target the most up-regulated genes in the Oct4 transcriptional network upon Oct4 knock-down in ES cells compared to wild type ES cells. Both microRNAs are expressed in ES cells. The p-values are depicted adjacent to the chart.

3.3.3 Scanning entire transcriptional networks with 'Sylamer'

'Sylamer' is a program developed by Stijn van Dongen which ranks genes according to increasing fold change values across different conditions, and calculates cumulative enrichment of sequence 'words' or motifs in the entire spectrum of differentially expressed genes. Sylamer plots can help

detect specific enrichments of sequence motifs and microRNA seeds in the 3'UTRs of entire transcriptomes. Oct4 and Nanog transcriptional networks were scanned for specific microRNA seed enrichments using this program.

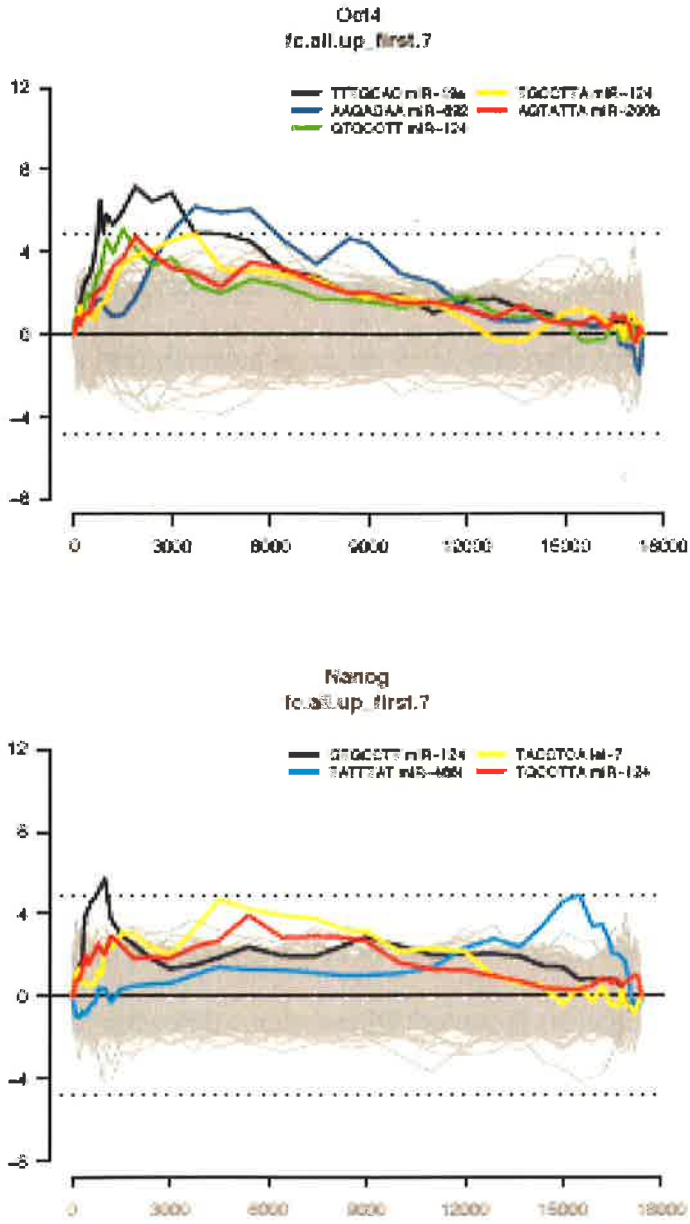


Figure 3.9: Sylamer Plots for microRNA seed enrichments in the Oct4 and the Nanog transcriptional networks. Genes have been ranked according to their fold-change from left to right. (Acknowledgements: Dr. Cei Abreu-Goodger). Each grey line depicts enrichment for a particular word or microRNA seed across ranked genes, while seeds or words showing maximum enrichments and depletions have been coloured. The y-axis represents log of p-values of cumulative hyper-geometric test performed for detecting enrichments or depletions. The x-axis consists of probes ranked according to their t-statistics which have been scanned for seed or word enrichments in windows consisting of 100 genes. The y-axis represents the log10 transformed significance p-value (positive for enrichment and negative for depletion). The error line depicts an e-value (Bonferonni corrected p-value) of 0.01.

One detects specific enrichments for mmu-miR-19a, mmu-miR-692 and mmu-miR-124 seed sequences in genes up-regulated upon Oct4 knock-down in ES cells. Seeds belonging to miR-124 and let-7 are overrepresented in genes up-regulated after Nanog knock-down in ES cells. MicroRNAs miR-19a and miR-124 are known to be expressed in stem cells and differentiated tissues ((Landgraf et al., 2007). MicroRNA miR-692 has been isolated from developing mouse embryos (Mineno et al., 2006). These results suggest that these microRNAs might regulate genes down-stream of these transcription factors which might be down-regulated during differentiation of stem cells. There is an enrichment of seeds belonging to miR-466l in genes expressed in ES cells. This microRNA has been isolated from ES cells, and cells treated with 5-aza-deoxycytidine which is a DNA-methyl transferase inhibitor (Calabrese et al., 2007) and might be involved in regulation of DNA methylation pathway. It is possible that genes downstream of Nanog transcription factor in ES cells might also be involved in regulating this pathway in ES cells.

The two different computational approaches applied, yield different results for microRNA enrichments. The first method analyzes microRNA seed overrepresentation scores in selected gene sets compared to the whole mouse genome, while Sylamer calculates cumulative enrichment of seeds across ranked gene sets and helps determine more global enrichments and depletions across entire transcriptional networks composed of thousands of genes. The first approach is useful if one is interested specifically in small gene sets, like those consisting of genes most differentially expressed between two conditions, while the second approach detects microRNAs with a more wide-spread effect on gene expression.

Summary

Scanning entire transcriptional networks downstream of Oct4 and Nanog transcription factors reveals enrichment of binding sites of stem cell microRNAs in genes which are most up-regulated upon knock-down of either of these transcription factors in stem cells. One also observes

an enrichment of binding sites corresponding to microRNAs expressed in differentiated tissues among transcripts which are most down-regulated upon knock-down of these transcription factors in stem cells. Sylamer analysis has resulted in further identification of binding sites corresponding to microRNAs expressed in stem cells in these transcription networks indicating a function of these microRNAs in specifically regulating genes involved in pluripotency pathways. These results show that microRNAs in stem cells might help regulate entire transcriptional networks down-stream of Oct4 and Nanog forming an additional layer of regulation in maintenance of stem cell homeostasis. Some of the results obtained in this chapter are in agreement with the 'avoidance hypothesis' (Farh et al., 2005) while others are contradictory. These computational predictions could be further verified by experiments such as luciferase assays.

CHAPTER FOUR: ANALYSIS OF THE 'MIRNOME' AND 'TRANSCRIPTOME' IN DIFFERENTIATING ES CELLS

4.1 Introduction

In this chapter, I study microRNA and mRNA expression dynamics during the process of stem cell differentiation, during which there is a systematic loss of stem cell pluripotency and differentiation of mouse embryonic stem cells into ectodermal and endodermal layers of embryoid bodies. Embryoid body formation recapitulates many aspects of embryogenesis *in vitro* enabling one to study of stem cell differentiation in a well defined system (Martin et al., 1977). However, one must note that overall it is an artificial system that cannot completely represent the phenomenon of *in vivo* differentiation.

The following questions were addressed by the microRNA and mRNA expression analyses carried out in this chapter:

1. How do expression levels of microRNAs (constituting the 'mirnome') change during stem cell differentiation into day three embryoid bodies?
2. What are the characteristics of messenger RNAs (constituting the 'transcriptome') most differentially expressed during stem cell differentiation in terms of Gene Ontology (GO) enrichment scores, general sequence composition and microRNA seed over-representations?
3. How do expression levels of transcripts belonging to key pluripotency transcription factors, discussed in Chapter two, change during differentiation of stem cells into day three embryoid bodies?
4. What is the correlation between temporal expression trends of pluripotency transcripts and microRNAs (predicted to interact with them using miRanda (John et al., 2004)) ,TargetScan (Grimson et al., 2007) and Pictar (Krek et al., 2005) during differentiation of

stem cells into day three embryoid bodies?

4.2 Optimizing Embryoid Body Formation

Initially mouse embryonic stem cells were plated in feeder free conditions and monitored for embryoid body formation. It was observed that ES cells tended to aggregate randomly resulting in embryoid bodies of various sizes. It was also observed that larger embryoid bodies tended to differentiate into ectodermal and endoermal layers faster than smaller embryoid bodies. In order to overcome this problem of asynchronous stem cell differentiation, it was decided to utilize the hanging drop method to synchronise differentiating stem cells.

Mouse embryonic stem cells were plated in hanging drops to form embryoid bodies at a density of 1000 cells per drop (Kurosawa, 2007) using a previously described method (Guan et al., 1999). The three dimensional environment provided to stem cells allowed them to form uniform aggregates resulting in the formation of a single embryoid body per drop. The hanging drop method enables cells to differentiate in a synchronised manner compared to other differentiation procedures which lead to random aggregation of cells in culture (Kurosawa, 2007).

4.3 MicroRNA and Messenger RNA Expression Analysis

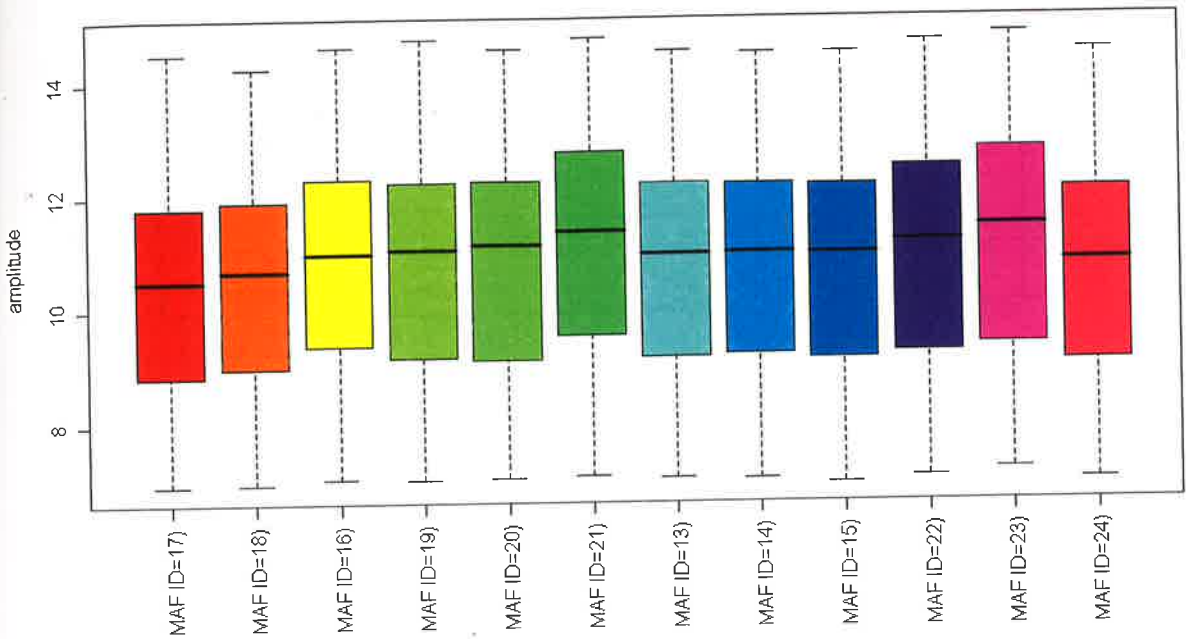
MicroRNA and mRNA expression was measured using Illumina Mouse-6 BeadChip arrays at days 0, 1, 2 and 3 of embryonic stem cell differentiation (See Materials and Methods). Advantages of this platform include greater sensitivity, high feature density allowing smaller reagent and sample volumes and standardized hybridization set ups for reproducible data generation. MicroRNA expression was also measured using an alternative Exiqon microarray platform (See Materials and Methods). The mRNA and microRNA Illumina expression experiments were performed by the Sanger Institute Microarray Facility Staff.

4.3.1 Quality Control Assessment of mRNA and MicroRNA Expression Data

Quality Control assessment of microarrays relies on looking at systematic differences between arrays and spatial artefacts within individual arrays. Prior to data processing it was ensured that all arrays contained uniform density of spots. Images were subject to background correction and raw data was transformed and normalised using lumi (Du et al., 2008) on R/Bioconductor (Team, 2007), a package for statistical computing. The normalisation procedure applied for both, is quantile, which makes the distribution of intensities equivalent across all samples. This method proved to be most efficient in reducing variation between microRNA arrays. For of consistency, this method was also applied to the mRNA data. Box-plots, density plots, pair-wise correlation plots, pair-wise MA plot between microarrays and sample-relation plots were plotted for both raw and normalised data. For convenience, only box-plots of microRNA and mRNA data, before and after normalisation, are depicted in Figures 4.1 -4.4. As can be observed, there is variation in expression levels between microRNA and mRNA chips, which is minimized after transformation and normalisation procedures. Figures 4.5 and 4.6 depict the sample relations based on hierarchial clustering of individual replicates at various time points. Replicates at various time-points cluster well together, even before normalisation for microRNA data (Figure 4.5), whereas, normalisation significantly improves sample clustering for mRNA data (Figure 4.6). There is a high degree of correlation between replicates of normalised mRNA and microRNA data (Table 4.1).

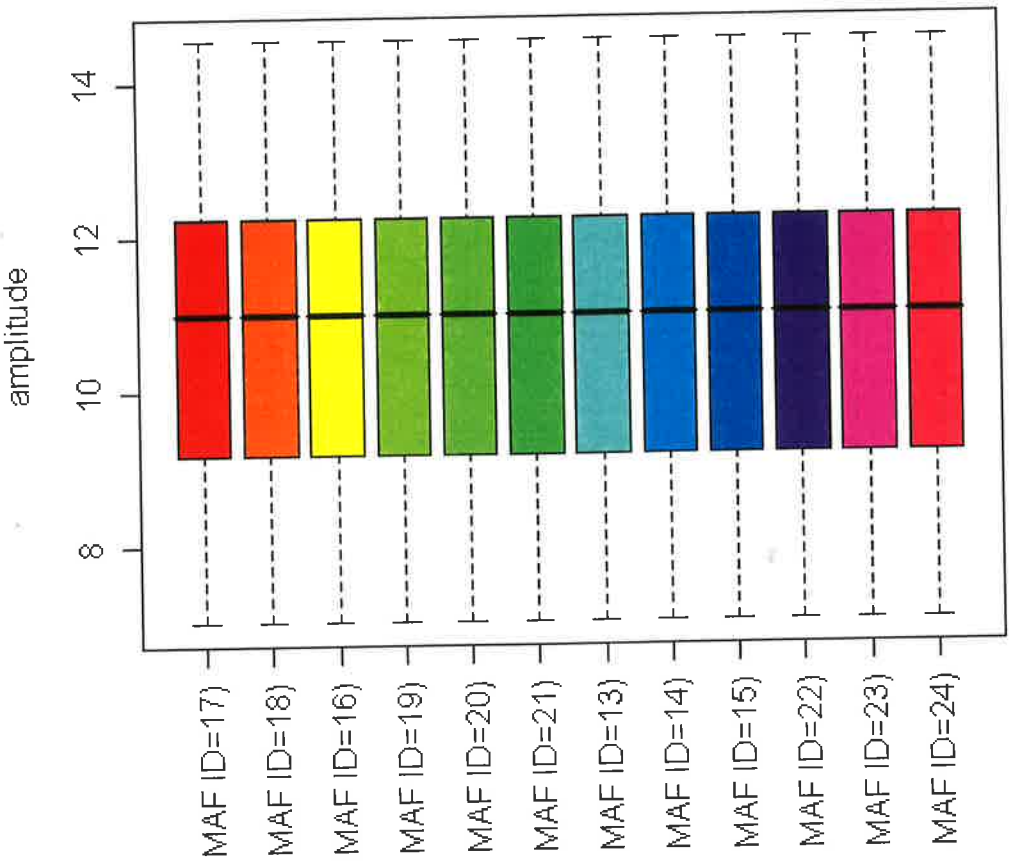
MicroRNA expression analysis was also performed using Exiqon as an alternative platform. Total RNA isolated from cells at day 0, day 1, day 2 and day 3 was labelled with Hy3 and Hy5 labelling dyes and hybridised onto microarray slides, with corresponding dye swaps for each sample.

Table 4.1: Average Correlation between Replicates in Illumina arrays		
	mRNA	miRNA
Day 0	0.98	0.99
Day 1	0.99	0.99
Day 2	0.99	0.99
Day 3	0.99	0.99



Day 1 R1 **Day 1 R2** **Day 1 R3** **Day 2 R1** **Day 2 R2** **Day 2 R3** **Day 0 R1** **Day 0 R2** **Day 0 R3** **Day 3 R1** **Day 3 R2** **Day 3 R3**

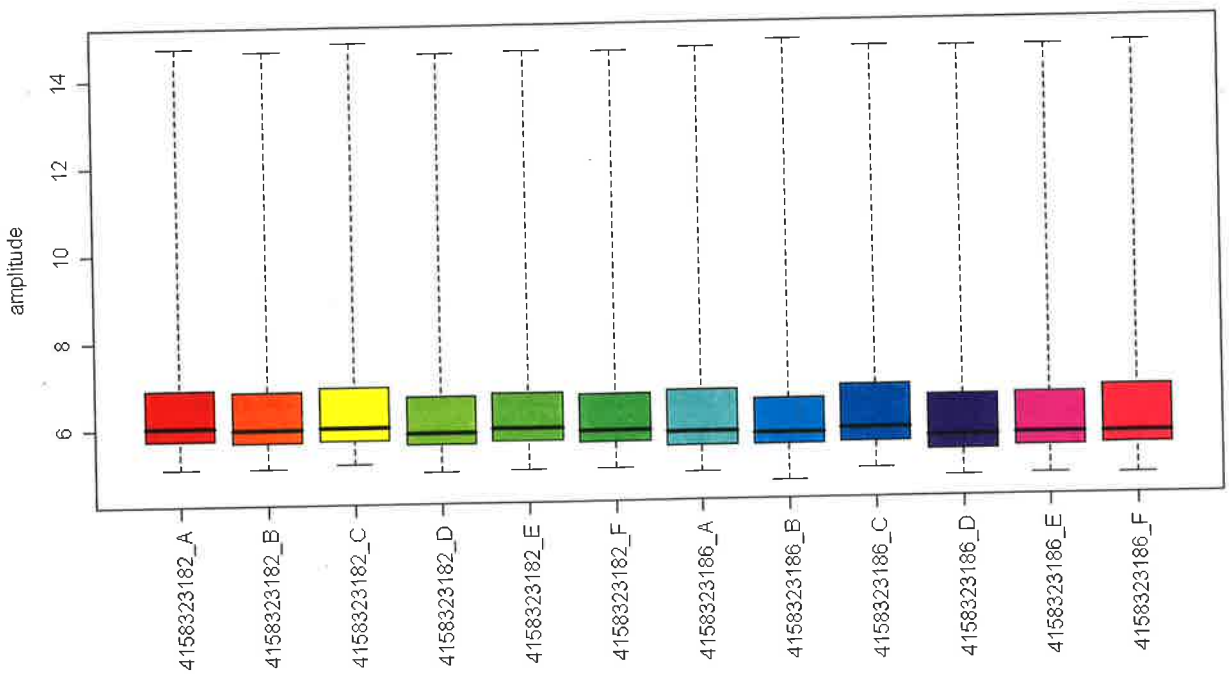
Figure 4.1: Box-plots of raw microRNA expression data obtained from time points 0, 1, 2 and 3 of stem cell differentiation into embryoid bodies. (Acknowledgement: Dr. Cei Abreu-Goodger). The x-axis shows MAF Ids. which represent the Illumina chip Ids used. (MAF IDs 13-15 represent Day 0 replicates, MAF IDs 16-18 represent Day 1 replicates. MAF IDs 19-21 represent Day 2 replicates and MAF IDs 22-24 represent Day 3 replicates)
R1, R2,R3: Replicates 1,2 and 3.



D1	D1	D1	D2	D2	D2	D0	D0	D0	D3	D3	D3
R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3

Figure 4.2: Box-plots of normalised microRNA expression data (obtained from time points 0,1,2 and 3 of stem cell differentiation into embryoid bodies). (Acknowledgement: Dr. Cei Abreu-Goodger). The x-axis shows MAF Ids, which represent the Illumina chip Ids used. (MAF IDs 13-15 represent Day 0 replicates, MAF IDs 16-18 represent Day 1 replicates, MAF IDs 19-21 represent Day 2 replicates and MAF IDs 22-24 represent Day 3 replicates)

D0,D1,D2,D3: Days 0,1,2 and 3.
R1, R2,R3: Replicates 1,2 and 3.



D0R1 D0R2 D0R3 D1R1 D1R2 D1R3 D2R1 D2R2 D2R3 D3R1 D3R2 D3R3

Figure 4.3: Box-plots of raw mRNA expression data (obtained from time points 0,1,2 and 3 of stem cell differentiation into embryoid bodies)(Acknowledgement: Dr. Cei Abreu-Goodger). The Illumina chip IDs are indicated (eg. 4158323182_A) on the x-axis.

D0,D1,D2,D3: Days 0,1,2 and 3
R1, R2,R3: Replicates 1,2 and 3.

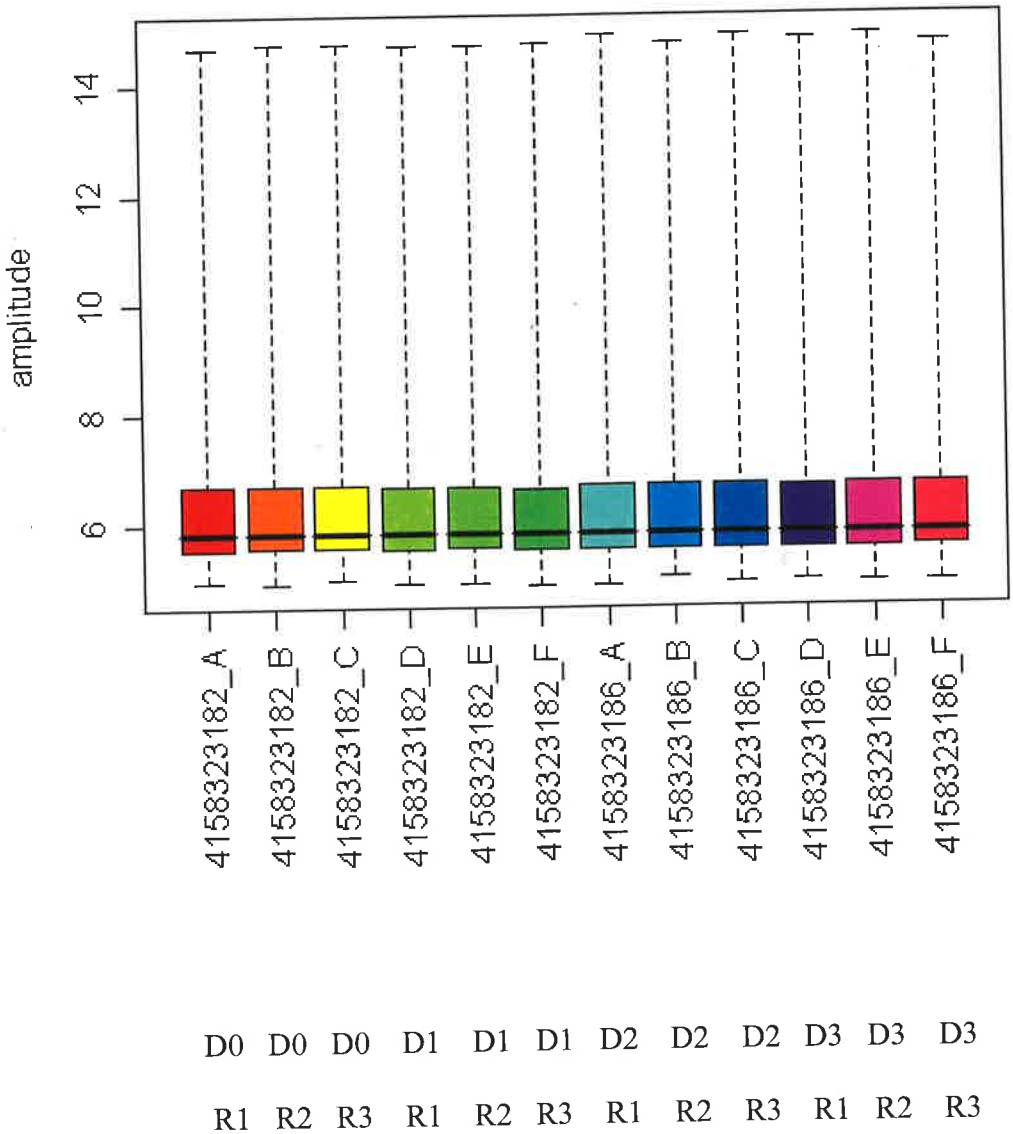
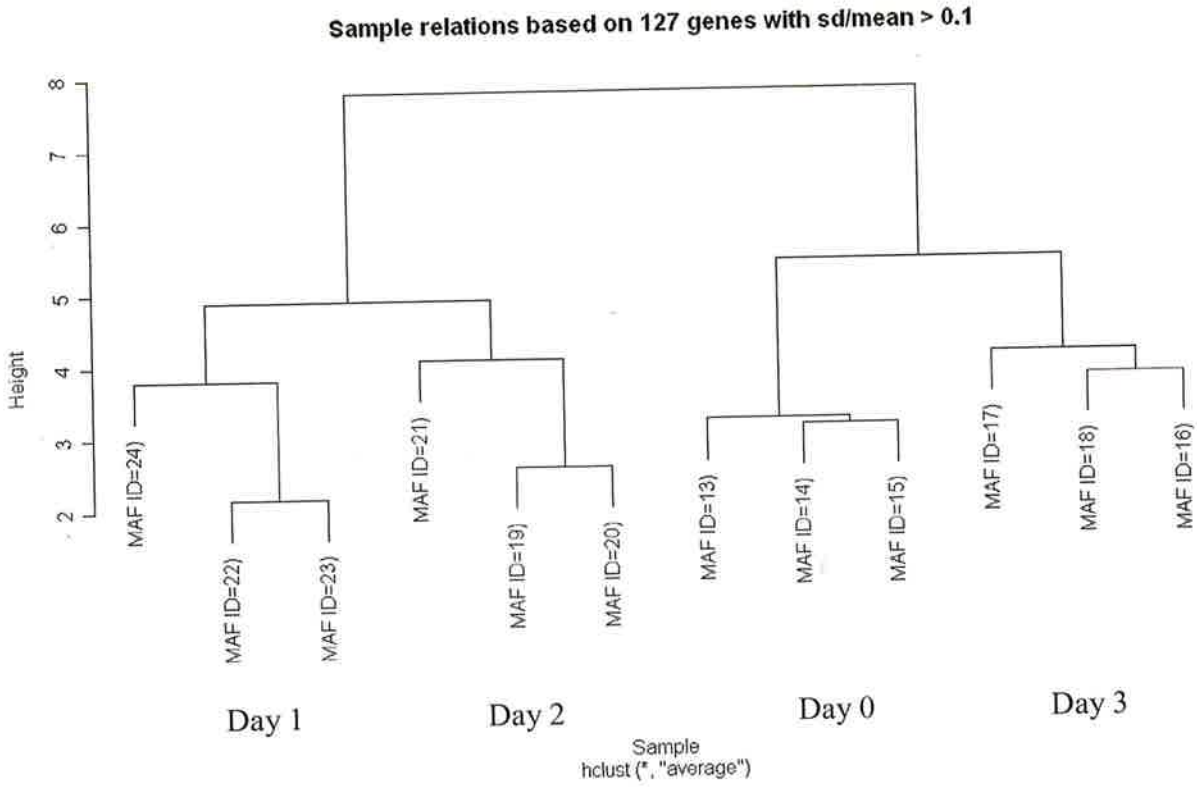


Figure 4.4: Box-plots of normalised mRNA expression data (obtained from time points 0,1,2 and 3 of stem cell differentiation into embryoid bodies) (Acknowledegment: Dr. Cei Abreu-Goodger). The Illumina chip IDs are indicated (eg. 4158323182_A) on the x-axis.

D0,D1,D2,D3: Days 0,1,2 and 3
R1, R2,R3:Replicates 1,2 and 3.

A.



B.

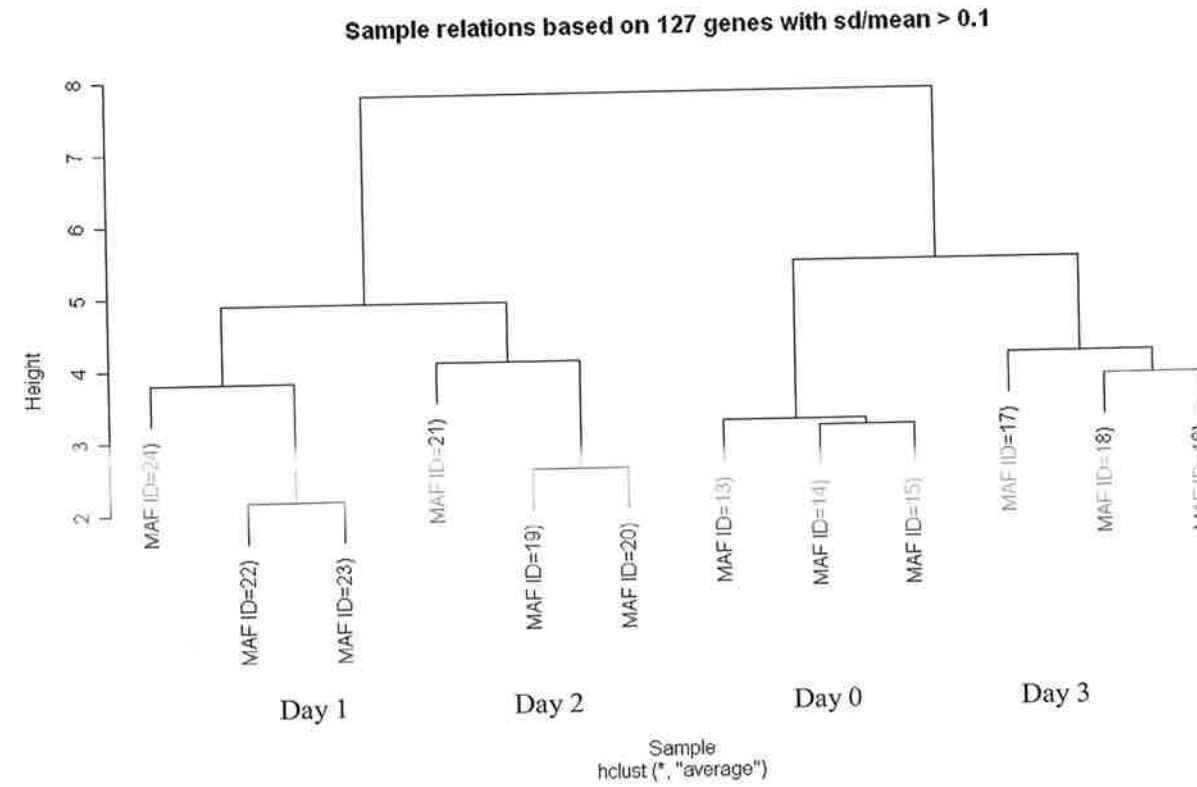
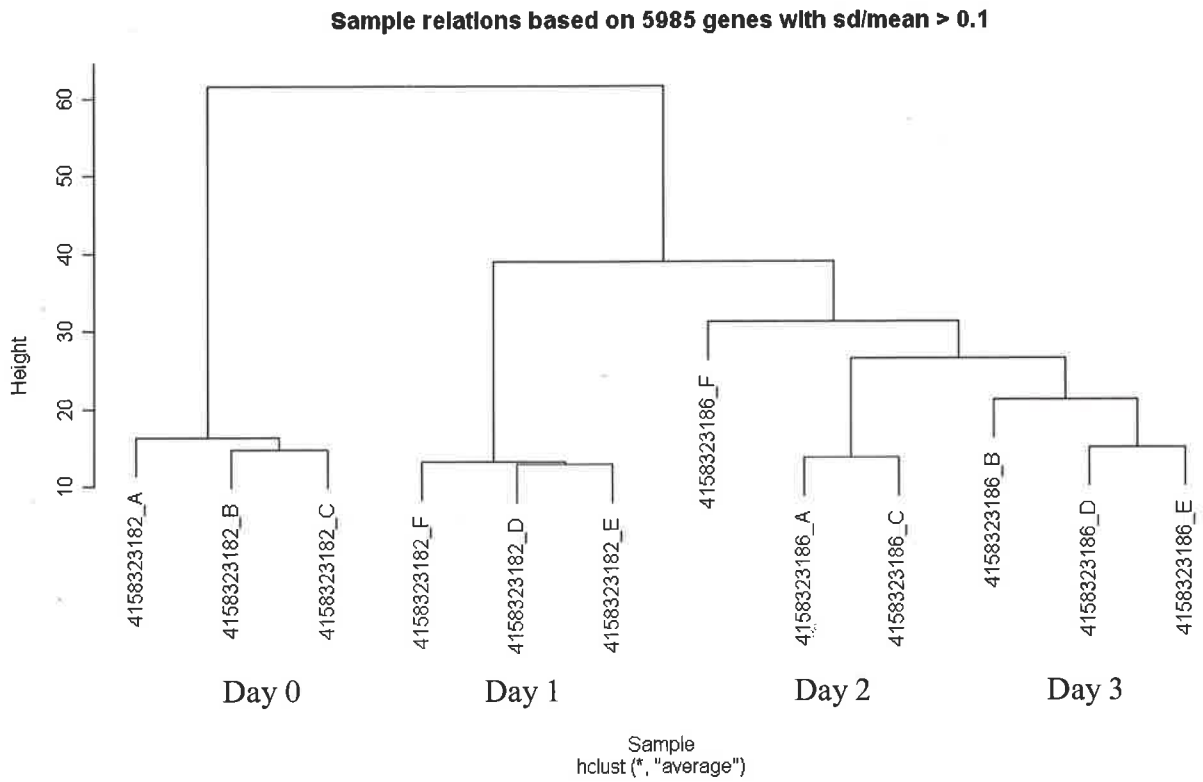


Figure 4.5: Cluster analysis of raw (A) and normalised (B) microRNA expression data. (Days 0,1,2 and 3 MAF IDs 13-15 represent Day 0 replicates, MAF IDs 16-18 represent Day 1 replicates, MAF IDs 19-21 represent Day 2 replicates and MAF IDs 22-24 represent Day 3 replicates) (Acknowledegment: Dr. Cei Abreu-Goodger). MAF IDs represent the Illumina chip Ids used.

A.



B.

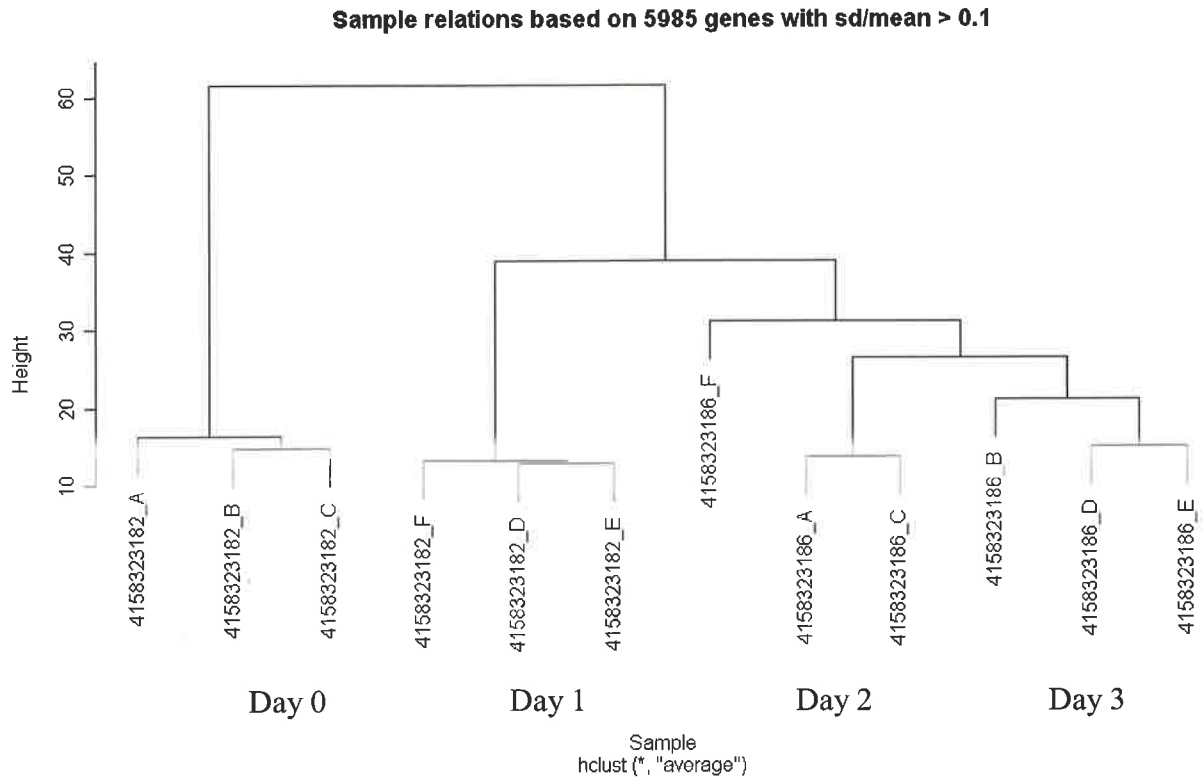


Figure 4.6: Cluster analysis of raw (A) and normalised (B) mRNA expression data (corresponding to days 0,1,2 and 3) (Acknowledegment:Dr. Cei Abreu-Goodger).

The slides were scanned using ScanArray Express and the images were quantified in the EasyQuant mode. All images were subject to background correction. The data thus obtained was normalized for all the arrays using the print tip loess method within arrays and aquantile normalisation between arrays (Figure 4.7). The average correlation of M-values between individual replicates was determined to be 0.17, which is poor. This could be due to possible technical variations in the sample preparation and processing stages or spotting errors on arrays. As the data from the Illumina platform was reproducible, it was decided to rely on this for all further downstream analysis.

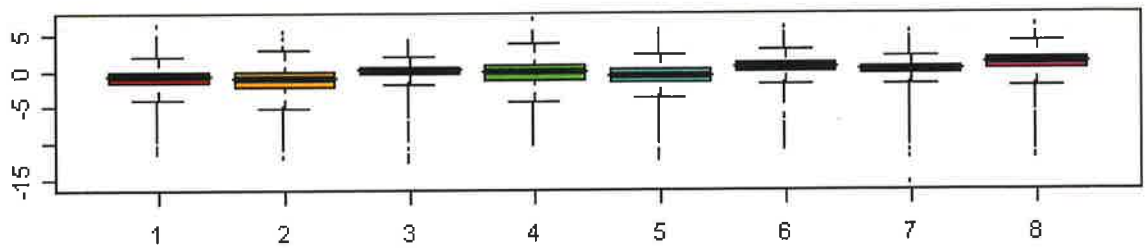
4.3.2 Assessing Positive and Negative Controls for Illumina Mouse mRNA and MicroRNA Chips

Microarray data is characterised by a lot of noise and hence it becomes important to distinguish actual biological signals from non-specific intensity values to extract meaningful data. Illumina bead-arrays have certain negative control probes incorporated in their chips which help in setting filtering thresholds for selecting genes for further downstream analysis.

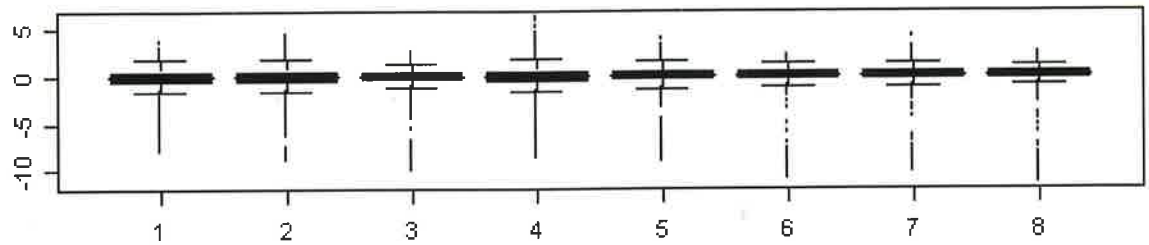
For mRNA arrays, the negative control corresponding to GFP, which is not present in the mouse was selected. I manually picked ubiquitously expressed GAPDH (glyceraldehyde 3-phosphate dehydrogenase) G6PD as a positive control.

Similarly median expression values for negative control for microRNAs were also plotted along with mmu-miR-15a which has been previously shown to be expressed in stem cells (Houbaviy et al., 2003, Thomson et al., 2004). The negative control was also used for filtering out probes in lists containing differentially expressed microRNAs.

Raw Data



Normalised (Print-tip-loess)



Normalised (Aquantile)

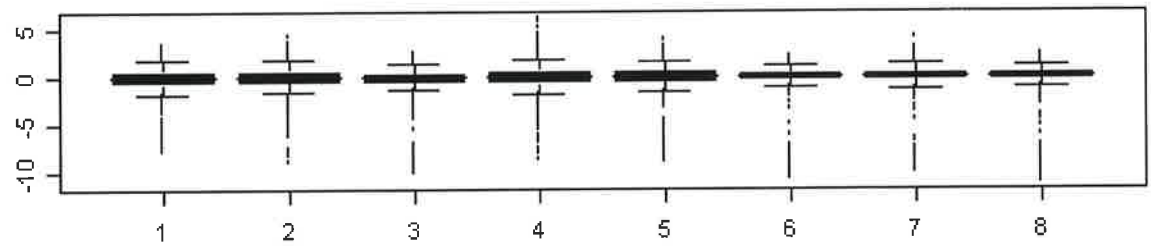


Figure 4.7: Box-plots of raw and normalised Exiqon data. Normalization was performed within arrays (print-tip loess) and between arrays (aquantile). The x axis represents 8 different arrays, 'Day 0': 1 and 2(Dye swap); 'Day 1': 3 and 4(Dye swap); 'Day 2': 5 and 6 (Dye swap); 'Day 3': 7 and 8 (Dye swap). The y axis represents the raw intensity values for each array. Acknowledgement: Dr. Anton Enright.

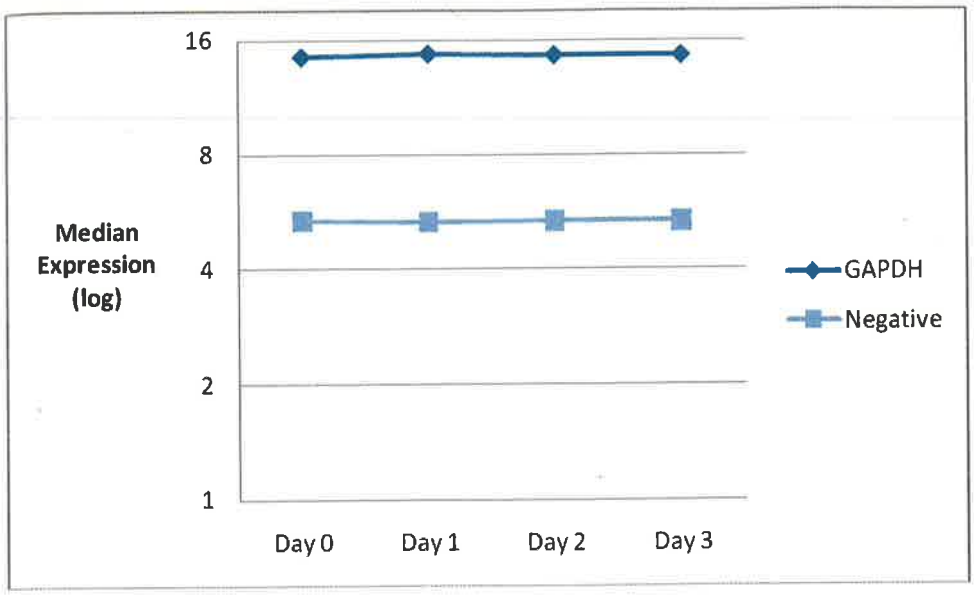


Figure 4.8: Median Expression Values of Positive and Negative Control Probes on the Illumina mouse mRNA array. GFP (Green Fluorescent Protein) was used as negative control. A cut-off of 6 (average median-value across all time-points) was chosen as the detection limit after normalisation of data for further expression analysis. GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) was used as positive control probe. (n=3, sd error bars are very small)

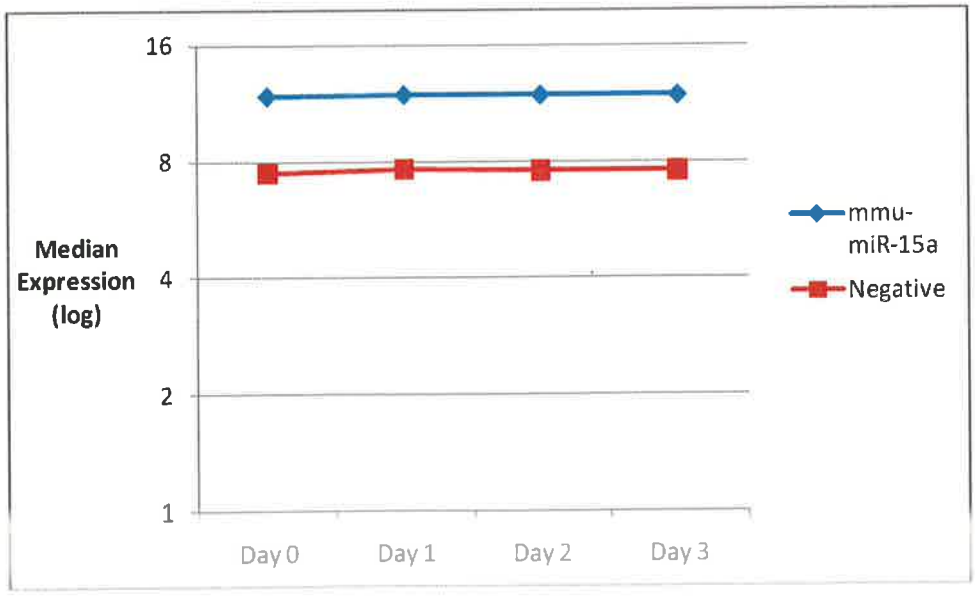


Figure 4.9: Median Expression Values of Positive and Negative Control Probes on the Illumina mouse microRNA array. A cut-off of 8 (average median-value across all replicates) was chosen as the detection limit after normalisation of data for further expression analysis. MicroRNA mmu-mir-15a was selected as positive controls. (n=3, sd error bars are very small).

4.4 Results

4.4 .1 Analysis of MicroRNA Differential Expression

Out of a total of 380 microRNAs probed for expression in mouse embryonic stem cells and differentiated states, there were 322 microRNAs detected on day 0 and 312 in day 3 embryoid bodies after log transformation and normalisation at a threshold p-value less than 0.01. There were 305 microRNAs expressed at both time points (Appendix Table 1).

Differential expression of microRNAs was determined between days 1-0, 2-0, 3-0, 2-1 and 3-2 by the eBayes method (Smyth, 2004) in the limma package in R. Top tables with the most up-regulated and down-regulated microRNAs at adjusted p-values <0.01 and log fold change >1 were constructed (See Supplementary Table 1). Figure 4.10B shows expression changes of microRNAs observed during first three days of differentiation of stem cells into embryoid bodies observed at p-value cut-off of 0.0001 (after Benjamini Hochberg Correction (Benjamini and Hochberg, 1995)). A number of microRNAs previously shown to be expressed in stem cells (Houbaviy et al., 2003, Thomson et al., 2004) were simultaneously down-regulated, coinciding with differentiation of stem cells and loss of pluripotency. These results hint at a potential role of these microRNAs in maintenance of stem cell self-renewal capacity and pluripotency. Other microRNAs which are up-regulated during differentiation might be required for formation of new cell types. There have been numerous examples where microRNA expression co-incides with appearance of new cell types, like the photoreceptor cells in *Drosophila* (Li and Carthew, 2005), Th2 cells in mice (Rodriguez et al., 2007a) and muscle cells in Xenopus laevis (Chen et al., 2006).

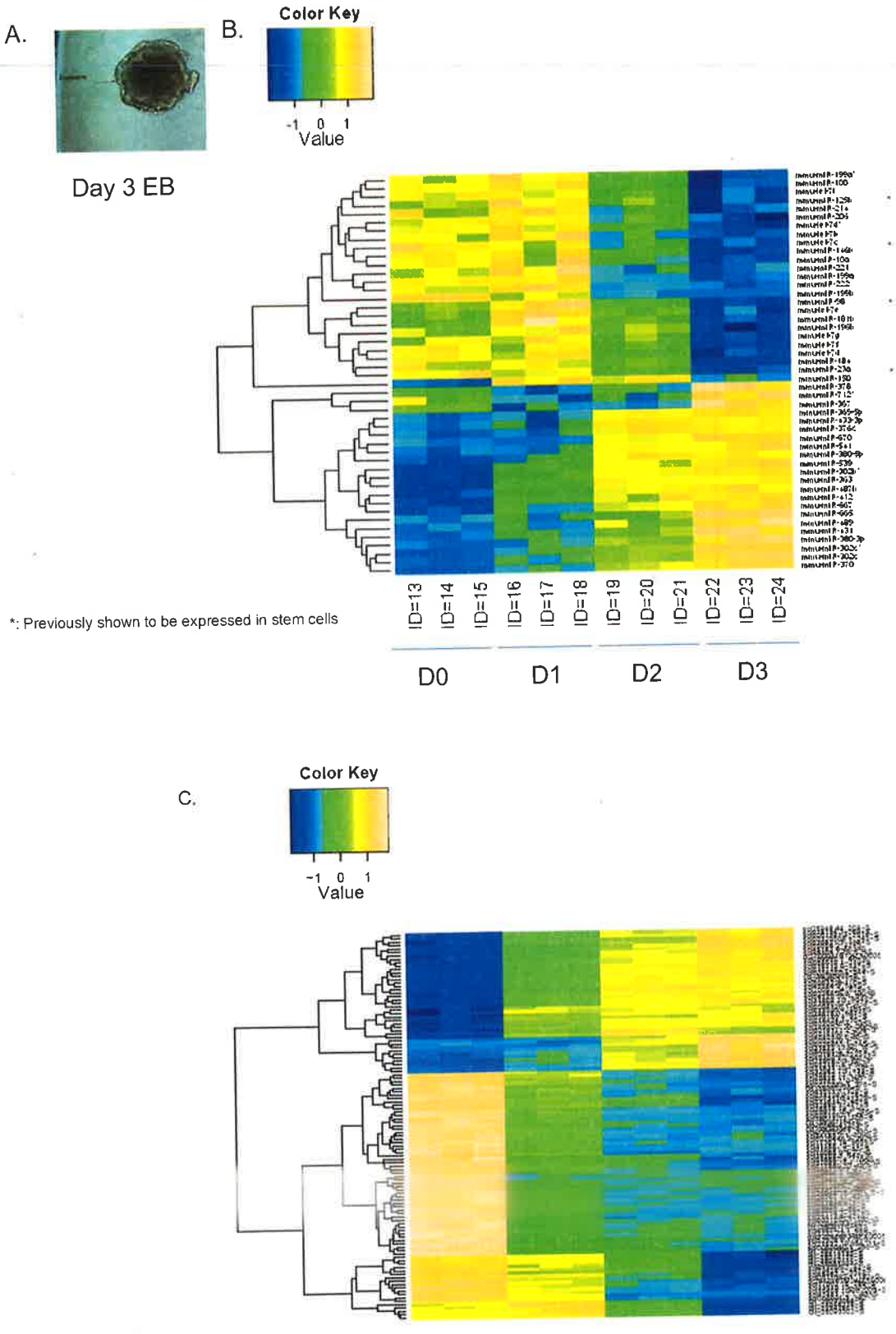


Figure 4.10: Heatmap depicting changes microRNA and mRNA expression during synchronised differentiation of stem cells into embryoid bodies. A. Day 3 embryoid body. B. Heat-map demonstrating microRNA expression changes. Color key depicts row-wise z-scores calculated for drawing a heat map representing microRNAs expressed differentially from day 0 to day 3 of embryoid body formation. A number of microRNAs previously shown to be expressed in ES cells (*) are down-regulated between these time points. The heatmap depicts microRNAs expressed differentially at adjusted p-value cut-off of 0.0001 (derived after Benjamini Hochberg correction) between day 0 and 3 of ES cell differentiation. C. Heatmap depicting mRNA expression changes between day 0 and day 3 of embryoid body formation at adjusted p-value cut-off of 10^{-13} (derived after Benjamini Hochberg correction). Acknowledgements for microRNA plot: Dr. Cei Abreu Goodger.

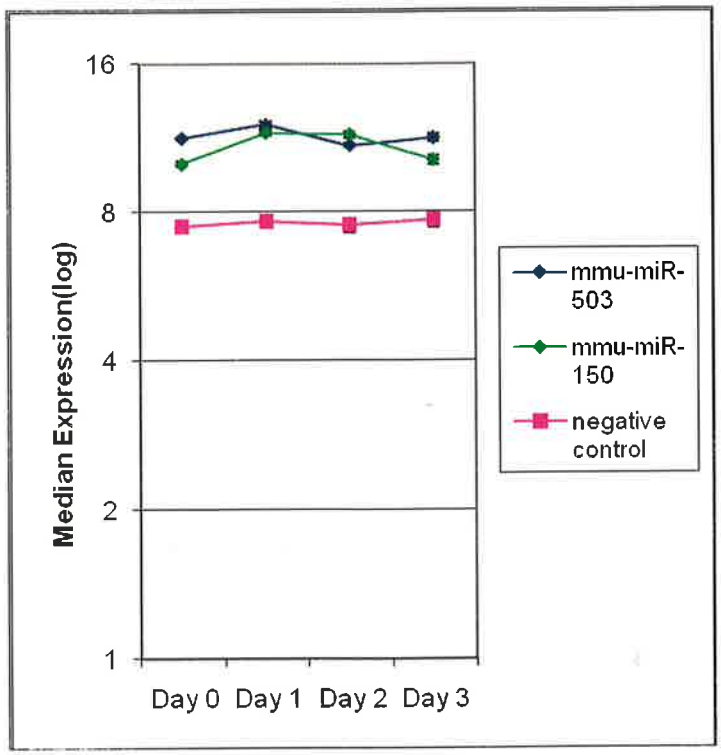


Figure 4.11: MicroRNAs showing increased transient expression during differentiation of stem cells into embryoid bodies. Negative control probe is also plotted. Also shown are error bars corresponding to standard deviation obtained for three independent replicates.

A few microRNAs like mmu-miR-503 and mmu-miR-150, show a transient increase in expression between days 0 and 3 (Figure 4.11), implying that they might be required during intermediate stages of differentiation of stem cells into endodermal and ectodermal layers. One also observes very small standard deviation observed between replicates at individual time points.

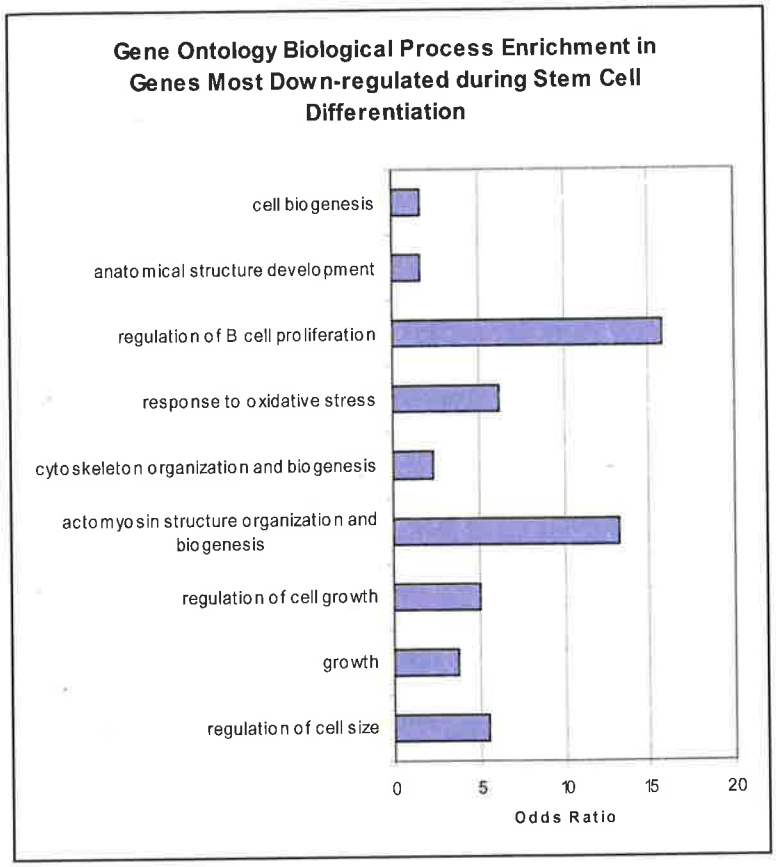
As one observes excellent correlation between replicates for both mRNA and microRNA expression measurements, I have relied on plotting median expression in subsequent plots without error bars.

4.4.2 Analysis of Gene Ontology Biological Process Enrichment in Differentially Expressed mRNAs

Differential expression of mRNA transcripts between days 1-0, 2-0, 3-0, 2-1 and 3-2 was carried out using eBayes method in limma (Smyth, 2004), a Bioconductor package in R. There were 11,798 and 11,733 probes detected on day 0 and 3 respectively (Appendix Table 1). There were 10,533 probes detected on both days. Figure 4.10B demonstrates changes in mRNA at adjusted p-value (Benjamini Hochberg correction) of 10^{-13}.

Functional analyses of genes most differentially expressed between days 0 and 3, based upon Gene Ontology Biological Process enrichments (Ashburner et al., 2000) was performed using GOstats (Carey et al., 2005).

This analysis revealed high over-representation of 'DNA-replication' and 'blood-vessel formation' biological process terms in genes that were up-regulated between day 0 and 3 (Figure 4.12). In contrast, there was a high enrichment observed for 'actomyosin structure organisation and biogenesis' and 'B-cell proliferation' biological process terms in genes which were most down-regulated during differentiation. Growth of embryoid bodies, is marked by increased cell proliferation which could explain the requirement of DNA replication products. Hence genes in this category would need to be highly expressed.



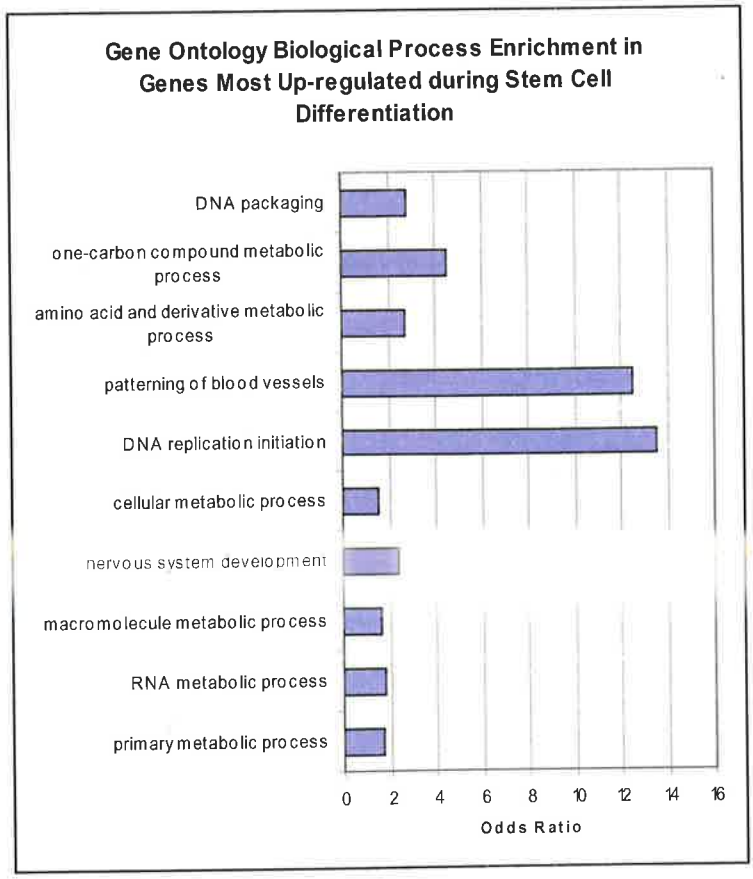


Figure 4.12: Odds Ratios of Gene Ontology Biological Process Term Enrichment in genes which are most up and down-regulated during differentiation of embryonic stem cells into embryoid bodies. (Acknowlegement: Siarhei Manakou)

ES cells have often served as model systems for studying vasculogenesis both *in vivo* and *in vitro* (Doetschman et al., 1993). Also, when undifferentiated ES cells are injected subcutaneously or intraperitoneally into syngeneic mice, they form solid and cystic tumours called teratocarcinomas which are not only angiogenic (in that they attract and induce growth of host endothelial cells) but also differentiate into vascular endothelium (Doetschman et al., 1993). Hence, our finding that genes corresponding to 'blood vessel formation' are highly overrepresented in embryoid bodies, could explain the tendency of these cells to undergo this developmental process in culture.

4.4.3 Sequence Biases and MicroRNA Seed Enrichments in 3'UTRs of Most Differentially Expressed Genes

In order to estimate whether there are any unique sequences enriched in the 3'UTRs of mRNA molecules which are most differentially expressed from days 0 to 3 of stem cell differentiation into embryoid bodies, 'Sylamer Plots' were generated for all possible 4-mer sequence motifs in these gene sets. As described previously, 'Sylamer' is a program developed by Stijn van Dongen (manuscript in preparation) which ranks genes in order of decreasing expression, comparing two different conditions, and calculates cumulative enrichment of sequence 'words' or motifs in the entire spectrum of differentially expressed genes. Sylamer plots can help detect specific enrichments of sequence motifs and microRNA seeds in the 3'UTRs of entire transcriptomes.

Genes expressed differentially between days 3 and 0 were ranked horizontally, with genes on the right expressed most highly in day 3 with respect to day 0. As can be observed in Figure

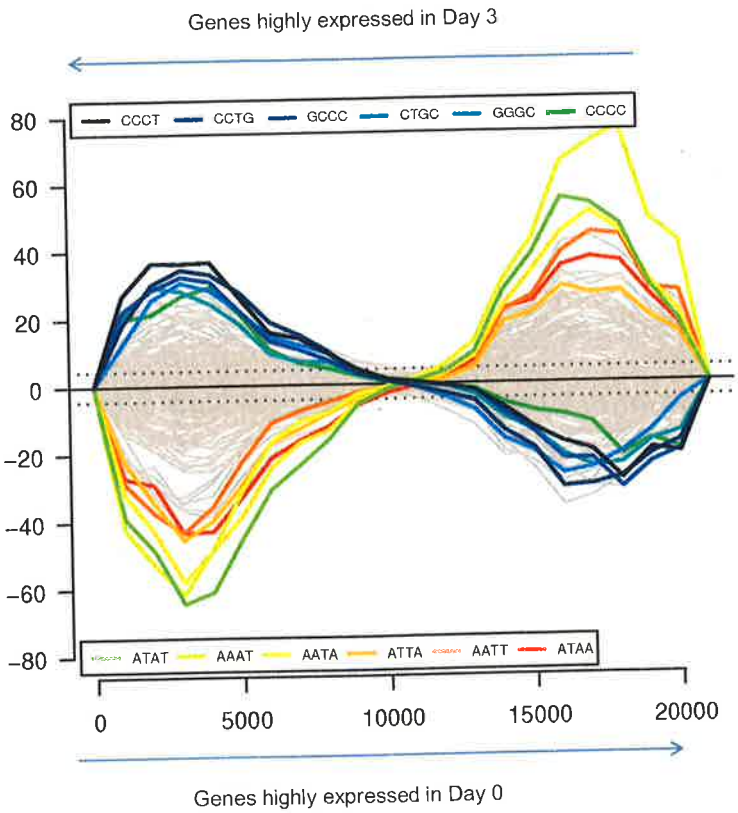


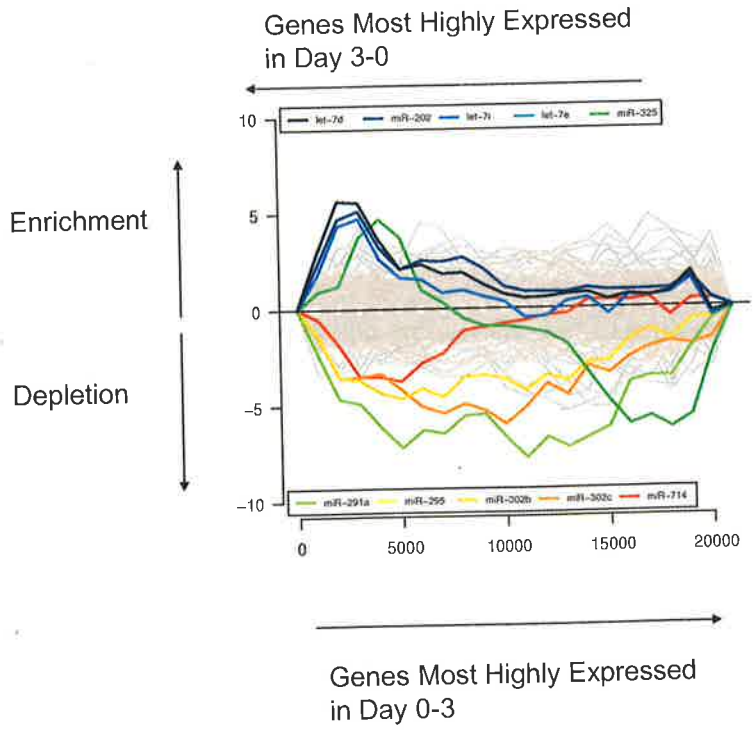
Figure 4.13: Sylamer Plot for sequence enrichment in 3'UTRs of genes most differentially expressed between day 0 and 3 of embryoid body formation. The x-axis represents individual genes which have been ranked horizontally in decreasing order of their expression from day 3 to 0. Genes on the left are most highly expressed in day 3 while those on the right are most highly expressed in day 0. These genes have been scanned in windows consisting of 1000 genes for enrichments or depletions of sequence motifs. The y-axis represents log of p-values of cumulative hyper-geometric test performed for detecting enrichments or depletions in these gene sets. Positive values indicate enrichments, while negative values depict depletions. Each grey line depicts enrichment for a particular sequence motif across ranked genes. Also shown are error lines for p-value:0.0001. As can be observed there is a considerable enrichment of GC-rich sequences and a depletion of AT rich sequences in 3'UTRs of genes most highly expressed on day 3 of stem cell differentiation. This plot has been generated very kindly by Dr. Cei Abreu-Goodger. In contrast, there is an enrichment of AT rich sequences and a depletion of GC rich sequences in genes most highly expressed in day 0 compared to day 3.

4.13, there is a considerable enrichment of GC-rich sequences (log p-value~ 40) and a depletion of AT-rich (log p-value~60) sequences in 3'UTRs of genes most highly expressed on day 3 of stem cell differentiation. In contrast, there is an enrichment of AT rich sequences and a depletion of GC rich sequences in genes most highly expressed in day 0 compared to day 3. It is also interesting to note the 'sine-wave' periodicity of both AT and GC enrichments across differentially expressed genes.

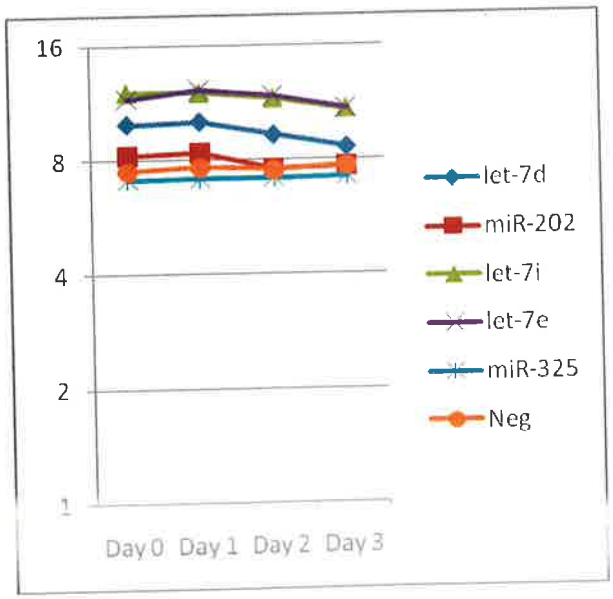
In order to investigate whether this bias in sequence composition of 3'UTRs of genes subject to differential expression could be due to microRNAs, I compared the frequencies of AT and GC bases in the seeds of microRNAs which are most differentially expressed at these time points to seeds of all the microRNAs known to be expressed in stem cells using a perl script.

A Chi-square test was performed to test the differences between these frequencies. As seen in the Table 4.2, there was a slight increase in the frequency of occurrence of GC bases (0.57) in seed sequences belonging to microRNAs which are most down-regulated between days 0 and 3, although the p-value is not significant (0.83). There is no change in AT composition of seeds belonging to microRNAs that are most down-regulated (p-value 0.99). These results suggest that microRNAs might be responsible for the substantial over-representation (p-value~10^{-40}) of GC rich 3'UTRs observed on day 3, based upon the observation that seeds corresponding to most down-regulated microRNAs are also GC rich. The causes of fluctuations with genes with high AT enrichment in their 3'UTRs are not known.

Table 4.2: Frequency of AT/GC base-pairs in Seeds of Differentially expressed MicroRNAs	**AT**	**GC**	**Chi-Square p-value**
Seeds of MicroRNAs most Up-regulated at Day 3	0.53	0.47	0.99
Seeds of MicroRNAs most Down-regulated at Day 3	0.43	0.57	0.83
Seeds of all MicroRNAs Expressed in ES cells	0.54	0.46	-



B.



C.

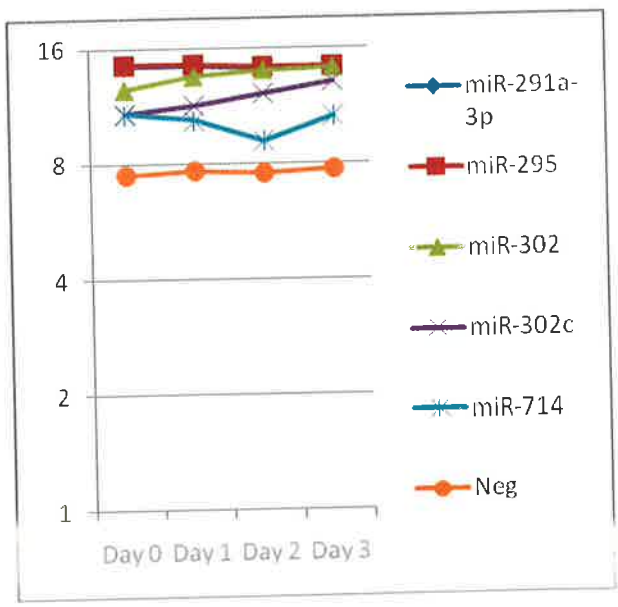


Figure 4.14: A. Sylamer Plots for maximum enrichments and depletions for microRNA seeds across all differentially expressed genes. The x-axis represents individual genes which have been ranked horizontally in decreasing order of their expression from day 3 to 0. Genes on the left are most highly expressed in day 3 while those on the right are most highly expressed in day 0. These genes have been scanned in windows consisting of 1000 genes for microRNA seed enrichments. The y-axis represents log of p-values of cumulative hyper-geometric test performed for detecting enrichments or depletions in these gene sets. Positive values indicate enrichments, while negative values depict depletions. Each grey line depicts enrichment of a microRNA seed across ranked genes, while seeds or words showing maximum enrichments and depletions have been coloured. This plot was generated very kindly by Dr. Cei Abreu-Goodger. B. and C. Expression changes for microRNAs with maximum seed enrichments in day 3 versus 0 (B) and maximum seed depletions (C). The y axis represents the median expression value across three replicates, while the x axis depicts individual time-points of differentiation. Negative controls 1 and 5 have also been plotted.

In order to assess whether microRNA seed enrichment and depletion patterns were due to changes in microRNA expression levels, median expression values of these microRNAs were plotted (Figure 4.14). One indeed observes a generalised reduction in expression levels of microRNAs let-7d, mmu-miR-202, let-7i and let-7e. In contrast, an increased expression is observed for microRNAs mmu-miR-302 and mmu-miR-302c. Little changes in expression are observed for mmu-miR-714 and mmu-miR-325. These results indicate opposing trends between expression of microRNAs let-7d, mmu-miR-202, let-7i, let-7e, mmu-miR-291a, mmu-miR-295, mmu-miR-302 and mmu-miR-302c and their seed enrichment and depletion patterns in differentially expressed mRNAs.

These observations for increased expression of microRNAs mmu-miR-302 and mmu-miR-302c, and a corresponding depletion of their seeds in mRNAs co-expressed with these microRNAs are in accordance with previous findings whereby expression of microRNAs and their targets is localised in a different space-time during formation of different cell types during differentiation (Farh et al., 2005). One also observes that a reduction in expression of microRNAs let-7d, mmu-miR-202, let-7i, let-7e and mmu-miR-325 during stem cell differentiation results in a corresponding increase in enrichment mRNAs containing their seeds on day 3 of embryoid body formation implying that these microRNAs might help regulate the expression of their target genes in embryonic stem cells.

4.4.4 mRNA Expression Analysis of Key ES Pluripotency Transcripts

As discussed in Chapter two, stem cell pluripotency is dependent upon certain transcription factors like Oct4, Sox2, Nanog, cMyc,Klf4, Esrrb,Stat3, Tcf7,Lrh1 and Sall4 (Zhou et al., 2007). The normalized median expression values of three replicates from the Illumina mRNA experiment were plotted across different time points of stem cell differentiation. There is a distinct up and down-regulation observed for key pluripotency transcripts during this early differentiation phase

(Figure 4.15). Certain transcripts like Oct4, Sall4, Tcf7 and cMyc are up-regulated. However, a marked reduction in the expression of many transcripts belonging to Nanog, Klf4, Esrrb, Stat-3, Sox2 and Lrh1 transcription factors is also observed. Comparison of these results and previously published findings involving expression analysis of some of these transcripts during stem cell differentiation, reveals considerable overlap in expression trends observed for Oct4, Sox2, Klf4, Esrrb, Nanog and Stat3 (Niwa et al., 2000, Glover et al., 2006, Palniqvist et al., 2005). For cMyc, one observes an increase in expression levels contrary to previous findings which have reported reduction in its expression during stem cell differentiation (Cartwright et al., 2005).

However, one must note that previous studies have carried out differentiation of embryonic stem cells after LIF withdrawal or after retinoic acid treatment. I have used the methodology for differentiating stem cells developed by Wobus and Guan (Guan et al., 1999). It is possible that different procedures applied for eliciting differentiation of stem cells might result in alteration of the external environment in ways which would affect expression signatures of key pluripotency transcripts. Moreover, embryoid bodies consist of a mixed population of cells which might exhibit differences in their gene expression patterns from one another. Hence the expression trends observed above indicate over-all changes in gene expression occurring in whole embryoid bodies and not individual cell-types.

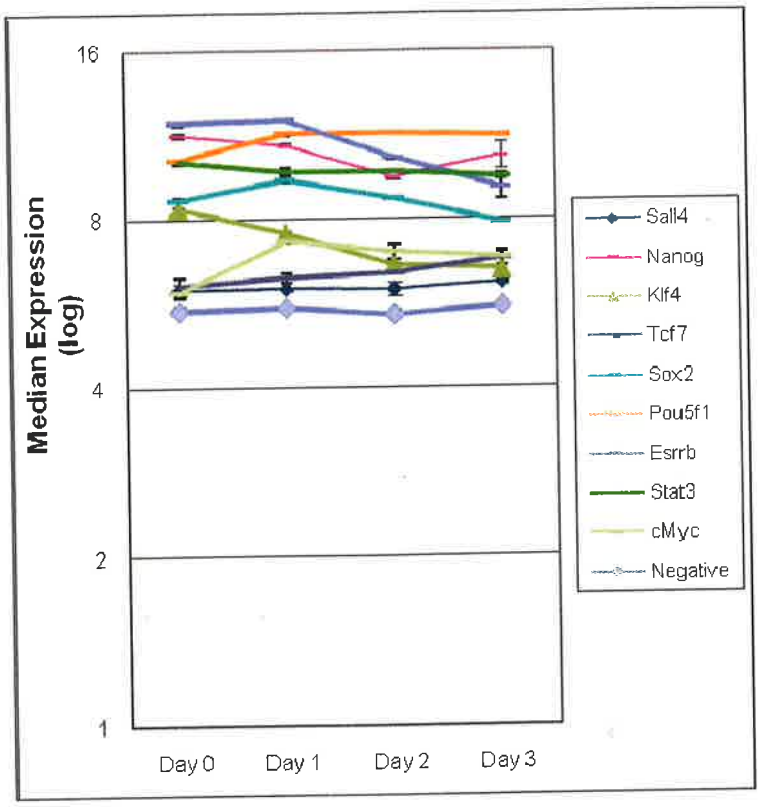


Figure 4.15: Expression profiling of key pluripotency transcripts (Nanog, Pou5f1(Oct4), Esrrb, Stat3, Tcf7, Sall4, Klf4, cMyc and Sox2) across three days of stem cell differentiation into embryoid bodies. (Acknowledegment: Dr. Cei Abreu-Goodger) The y-axis represents the median expression values. Also shown are error bars corresponding to standard deviation across three replicates.

4.4.5 Determining Correlations between Expression Values of Key Pluripotency Transcripts and their Predicted MicroRNA Regulators

As stated in Chapter two, microRNAs expressed in stem cells might function in negatively regulating stem cell pluripotency, by inhibiting expression of transcripts belonging to the core pluripotency network. It was decided to investigate, whether there were any positive or negative correlations between the expression values of these pluripotency transcripts and their predicted microRNA interacting partners using miRanda miRBase targets, TargetScan and Pictar. Negative correlations between microRNAs and their predicted targets would possibly hint at a regulatory effect on mRNA expression mediated by microRNAs. However, it is also known that microRNAs can activate translation of transcripts under conditions of cell cycle arrest (Vasudevan et al., 2007). Cell cycle arrest often occurs early in the differentiation program of certain cell types like the

myotubes (Walsh and Perlman, 1997) and dopamine expressing neurons (Castro et al., 2001). It is known that during the process of cavitation in embryoid bodies, many cells undergo programmed cell death (Coucouvanis and Martin, 1995). It is possible that this might be preceded by some kind of cell cycle-arrest. In this scenario one would expect increased expression of target messenger RNAs, positively correlating with increased microRNA expression. However correlation may not always imply causation, and it is possible that some of these positive or negative correlations might be observed by chance alone.

Figures 4.16 and 4.17 depict the trends in expression levels observed for selected pluripotency transcripts and their predicted microRNA interacting partners. Table 4.3 shows the calculated Pearson's correlation values to depict the linear relationship between expression values observed in Figures 4.16 and 4.17. For Oct4, Sox2, Klf4, Stat3, cMyc, Sall4, Tcf7 and Lrh1 one observes a negative correlation between these and their predicted microRNA interacting partners described in Chapter two.

Positive correlations are observed for Nanog. A mixture of positive and negative correlations are observed for Oct4, Sox2, Klf4, Sall4, Stat3 and Lrh1. Negative correlations are observed for cMyc and Tcf7. Carrying out real-time PCRs, microRNA over-expression and luciferase assays in stem cells would help validate these results further. Nevertheless, these results provide mechanistic insights into expression dynamics of key pluripotency transcripts and their predicted microRNA interacting partners. Loss of pluripotency in stem cells might not only depend on down-regulation of transcription factor transcripts, but also on down-regulation microRNAs which are predicted to interact with these transcripts. Further, increased expression of many pluripotency transcripts correlates with the simultaneous down-regulation of their predicted microRNA interacting partners, hinting at a regulatory relationship between the two.

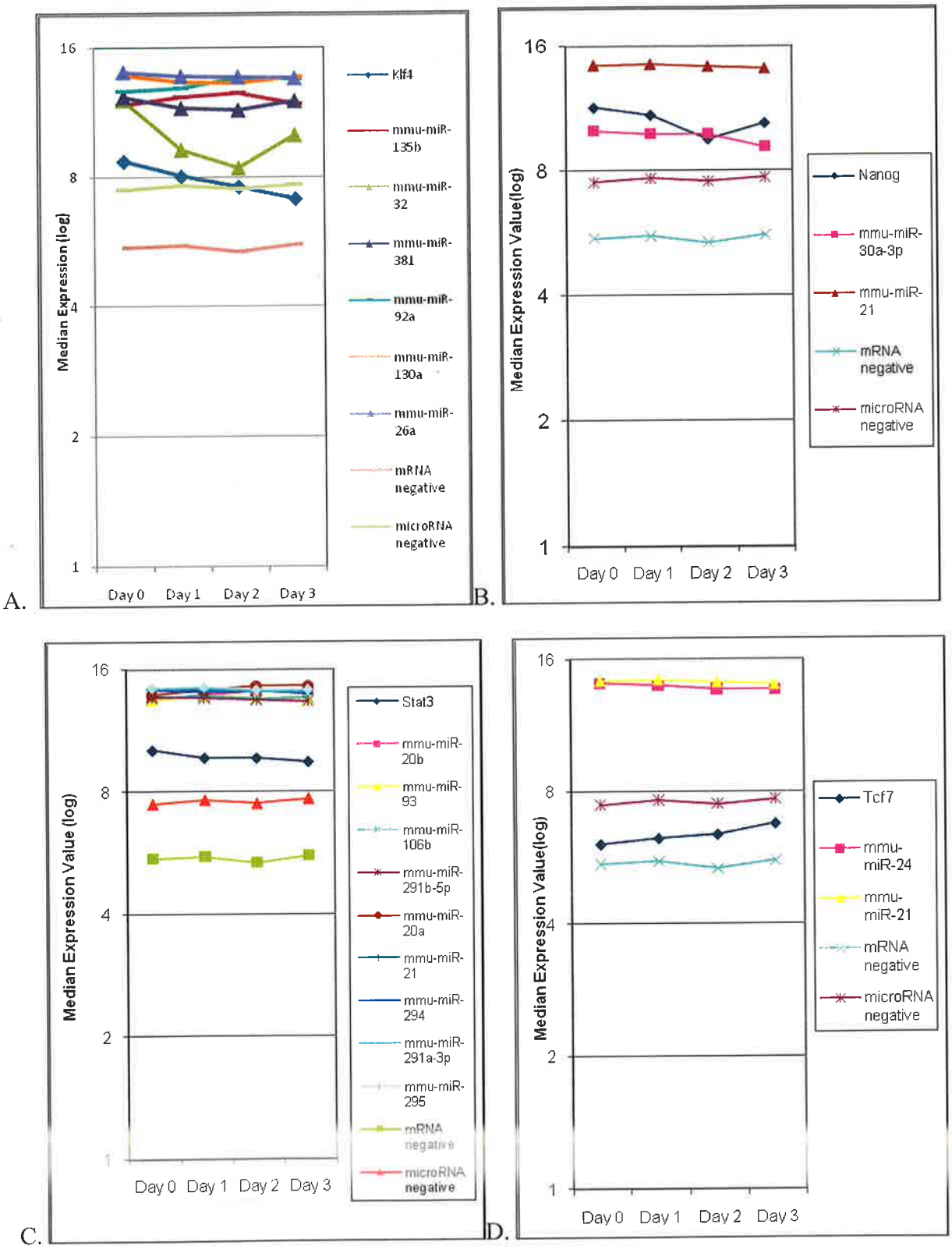
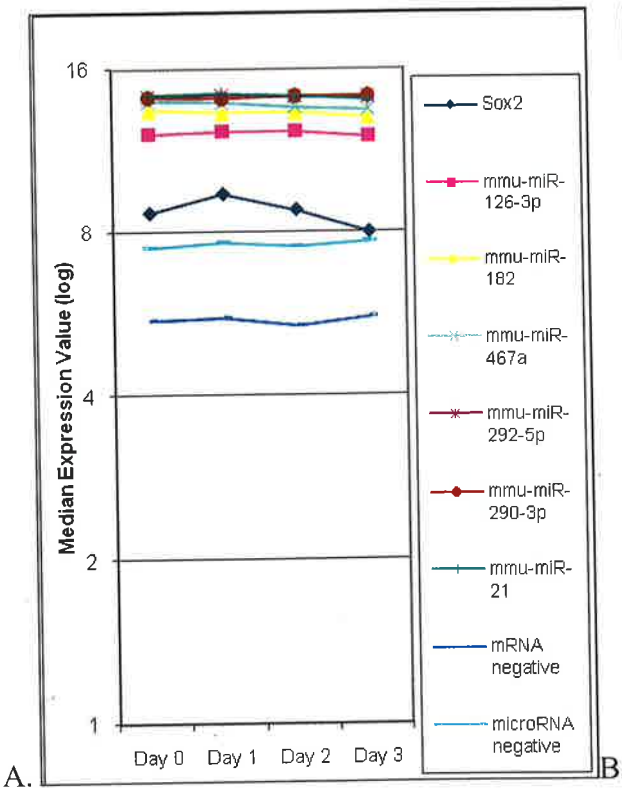
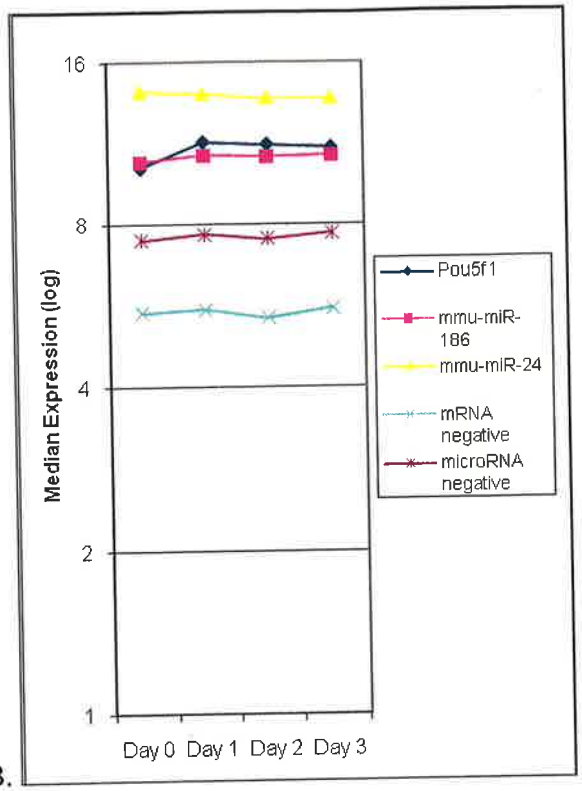


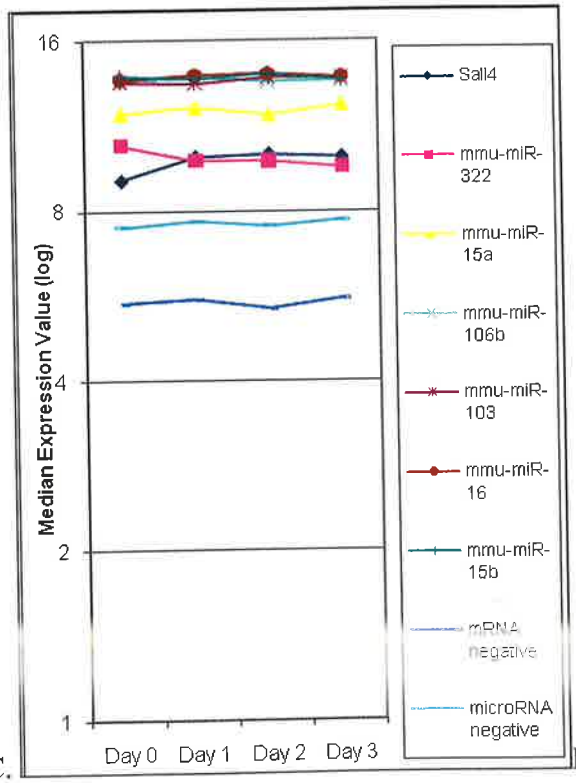
Figure 4.16: Expression changes for key pluripotency transcripts Klf4, Nanog, Stat3 and Tcf7 and their predicted microRNA interacting partners (identified in Chapter two) were plotted to determine correlations between them. Negative controls 1 and 5 (for microRNAs) and eGFP-S (for mRNA) and AmpR (for mRNA) are depicted in each plot. The y-axis represents the median expression values in log-scale.


16
8
4
2
1
Median Expression Value (log)
Day 0
Day 1
Day 2
Day 3
Sox2
mmu-miR-126-3p
mmu-miR-182
mmu-miR-467a
mmu-miR-292-5p
mmu-miR-290-3p
mmu-miR-21
mRNA negative
microRNA negative

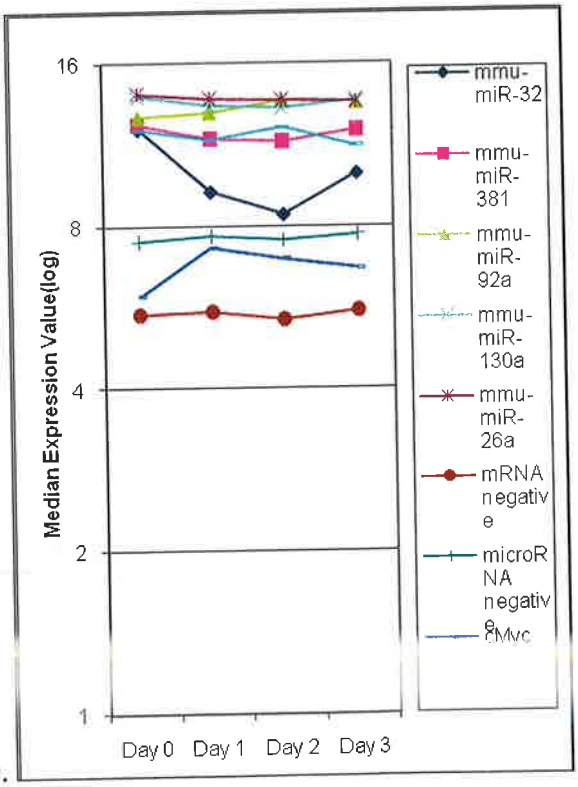
A.


16
8
4
2
1
Median Expression (log)
Day 0
Day 1
Day 2
Day 3
Pou5f1
mmu-miR-186
mmu-miR-24
mRNA negative
microRNA negative

B.


16
8
4
2
1
Median Expression Value (log)
Day 0
Day 1
Day 2
Day 3
Sall4
mmu-miR-322
mmu-miR-15a
mmu-miR-106b
mmu-miR-103
mmu-miR-16
mmu-miR-15b
mRNA negative
microRNA negative

C.


16
8
4
2
1
Median Expression Value(log)
Day 0
Day 1
Day 2
Day 3
mmu-miR-32
mmu-miR-381
mmu-miR-92a
mmu-miR-130a
mmu-miR-26a
mRNA negative
microRNA negative
cMyc

D.

E.

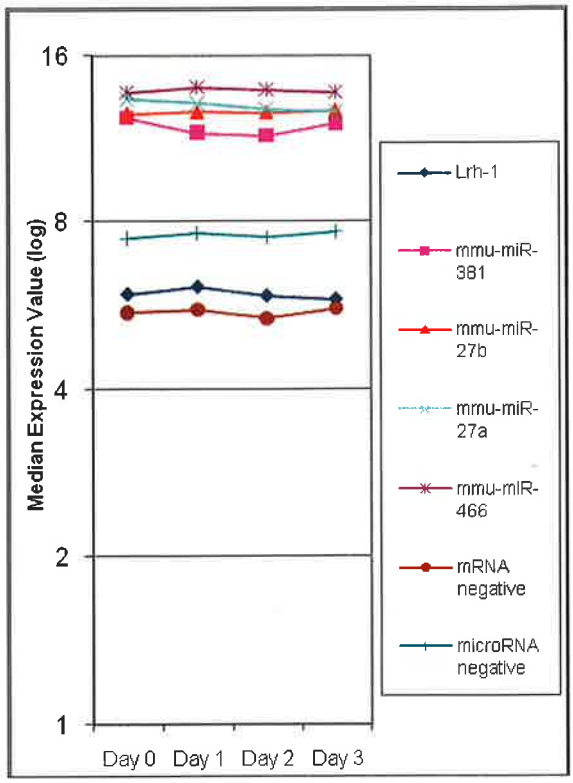


Figure 4.17: Expression changes for key pluripotency transcripts Sox2, Pou5f1 (Oct4), Sall4 ,cMyc and Lrh1 and their predicted microRNA interacting partners (identified in Chapter two) were plotted to determine correlations between them. Negative controls 1 and 5 (for microRNAs) and eGFP-S (for mRNA) and AmpR (for mRNA) are depicted in each plot. The y-axis represents the median expression values in log-scale.

Table 4.3: Pearson's Correlation Values for Expression Values of Pluripotency Transcripts and their Predicted Interacting Partners.

	Pou5f1	Nanog	Sox2	Klf4	Stat3	cMyc	Tcf7	Sall4	Lrh1
mmu-miR-186	0.97								
mmu-miR-24	-0.62						-0.88		
mmu-miR-30a-3p		0.26							
mmu-miR-21		0.45	0.93		0.62		-0.88		
mmu-miR-126-3p			0.87						
mmu-miR-182			0.57						
mmu-miR-467a			0.69						
mmu-miR-292-5p			0.77						
mmu-miR-290-3p			-0.83						
mmu-miR-135b				-0.23					
mmu-miR-32				0.65					
mmu-miR-381				0.38					
mmu-miR-92a				-0.90					
mmu-miR-130a				0.36					
mmu-miR-26a				0.96					
mmu-let-7c						-0.31			
mmu-miR-125b						-0.40			
mmu-let-7d						-0.40			
mmu-miR-132						0.40			
mmu-miR-20b					-0.93				
mmu-miR-93					-0.49				
mmu-miR-106b					-0.15				
mmu-miR-291b-5p					0.86				
mmu-miR-20a					-0.91				
mmu-miR-294					0.88				
mmu-miR-291a-3p					0.69				
mmu-miR-295					0.62				
mmu-miR-322								-0.93	
mmu-miR-15a								0.34	
mmu-miR-106b								0.19	
mmu-miR-103								0.48	
mmu-miR-16								0.88	
mmu-miR-15b								-0.15	
mmu-miR-381									-0.36
mmu-miR-27b									-0.32
mmu-miR-27a									0.61
mmu-miR-466									0.84

4.5 Summary

Expression analysis of microRNAs using Illumina bead arrays reveals the presence of around 322 microRNAs in stem cells. These numbers are around six to eight fold greater than previously reported estimates of microRNAs expressed in stem cells (Houbaviy et al., 2003,

Landgraf et al., 2007).

MicroRNA expression analysis reveals simultaneous down-regulation of multiple microRNAs expressed in stem cells. This down-regulation coincides with stem cell differentiation and loss of pluripotency implying that these might be important in maintenance of pluripotency and stem cell homeostasis in general. Similarly microRNAs that are up-regulated during differentiation might be essential in formation of new cellular types.

Analysis of over-representation of Gene Ontology Biological Process (GOBP) terms in mRNAs that are most differentially expressed between day 0 and 3, reveals the presence of 'DNA-replication' and 'blood vessel patterning' enriched in genes which are most up-regulated in day 3 embryoid bodies. Among genes which are most down-regulated, one finds an enrichment of terms: 'actomyosin structural organisation' and 'B-cell proliferation'.

Sylamer (unpublished) program has been applied to scrutinize 3'UTRs of entire transcriptional networks for specific sequence biases and microRNA seed over-representations. This analysis has resulted in revealing a fascinating AT-GC sequence bias in the 3'UTRs of genes most differentially expressed between days 0 and 3. It is possible that this bias might be caused by microRNAs with similar sequence biases. An increased GC frequency in seeds belonging to microRNAs which are most down-regulated on day 3 was observed, although the difference is not statistically significant (p:0.83). This result correlates well with the corresponding GC bias observed for genes most up-regulated on day 3. However, one cannot completely rule out other factors which might be involved in causing this phenomenon.

MicroRNAs and 3'UTRs containing their over-represented seeds are inversely related with respect to their expression patterns.

Correlation analysis of temporal expression values of key pluripotency transcripts and their predicted microRNA interacting partners reveals both negative and positive correlations between transcripts and their predicted microRNA interacting partners. These results would have to be further validated by experiments such as luciferase assays and microRNA over-expression in stem cells.

CHAPTER FIVE: MAKING TARGETING VECTORS FOR MICRORNA LOSS OF FUNCTION ANALYSIS IN ES CELLS

5.1 Introduction

As discussed in the previous chapters, microRNAs regulate gene expression at the post-transcriptional level by silencing hundreds of genes simultaneously. In order to decipher their functions, gene targeting vectors were constructed to make homozygous deletions of microRNAs in stem cells. The miR-192 poly-cistronic cluster (consisting of miR-192 and miR-194) and miR-345 were chosen for making homozygous deletions. The miR-192 cluster is expressed in embryonic stem cells (Thomson et al., 2004). miR-345 is expressed specifically during mammalian brain development (Miska et al., 2004).

Stem cells containing homozygous knock-outs of microRNA genes could be further studied in the following ways:

1. Messenger RNA expression profiling of miR-192 homozygous null embryonic stem cells could be used to identify genes which are up-regulated and down-regulated with respect to wild type stem cells. MicroRNA targets could be identified in these gene sets. It is known that most microRNAs silence genes by triggering the mRNA degradation pathway (Hutvagner and Zamore, 2002, Seggerson et al., 2002, Jing et al., 2005). Hence, one would expect most microRNA targets to localize within the down-regulated gene sets. However, as discussed in the Introduction section, there are some microRNAs which might affect gene expression by interfering with translational initiation and elongation steps and causing ribosome drop-off from mRNAs (Liu et al., 2005, Sen and Blau, 2005, Kedersha et al., 2005, Pillai et al., 2005). It is possible that the expression of a particular mRNA molecule is detected, even if it is in a state of inhibition. It has also been shown that microRNAs can

activate translation of certain genes under conditions of cell-cycle arrest (Vasudevan et al., 2007). Therefore, in this case, some target molecules might be detected in the up-regulated gene sets as well. Classification of targets based upon their over-representation in both up and down-regulated gene sets would help us gain more insights into their mode of action. It is might be possible to do proteomic studies in the future to identify changes in protein levels in a high-throughput manner, to answer questions regarding the nature of microRNA mediated repression or activation.

2. miR-345 is expressed exclusively during mammalian brain development (Miska et al., 2004). miR-345 knock-out cells could be used to analyze any defects in neurogenesis by carrying out differentiation of these neurons *in vitro*. Alternatively, homozygous null mouse embryos could also be studied to identify any defects in brain development.

5.2 Making Targeting Vectors by Recombineering

Gene targeting involves homologous recombination between a targeting vector and chromosomal DNA present inside embryonic stem cells (Capecchi, 1989). This technology has been widely used for creating and analyzing mice with specifically mutated genes, an approach that has been vital in our efforts to characterise the function of genes and their potential role in development and disease.

Originally, construction of targeting vectors involved screening of a genomic library with cDNA probes to isolate corresponding genomic fragments obtained by restriction enzyme digestion (Capecchi, 1989) and cloning them with ligases into appropriate vectors containing antibiotic resistance markers for selection of targeted ES cells. Although using restriction endonucleases and ligases has been very useful in the generation of recombinant DNA molecules, there are some practical limitations imposed by them (Copeland et al., 2001). First, DNA molecules can only be

combined if appropriate restriction sites are present. This problem has been overcome by performing PCR amplifications with restriction sites incorporated into primer sequences. However, cloning of large fragments, especially those required for construction of homology arms of a targeting vector, limits the choice of appropriate restriction sites available for this purpose. One is also limited by the presence or absence of restriction sites in choosing exact genomic coordinates for making a specific gene deletion. Moreover, this procedure has been quite laborious and time consuming.

These difficulties have now been overcome with the advent of recombineering technology which enables DNA modifications inside recombination defective RecA- *E.coli* strains expressing lambda phage proteins to mediate homologous recombination between two DNA molecules (Ellis et al., 2001, Copeland et al., 2001). The expression of the three main recombineering proteins- exo, beta and gam is regulated by a temperature dependent repressor and thus, a temporary shift of cultures to 42°C for 15 minutes leads to expression of these proteins and gain of recombination functions in *E.coli* (Copeland et al., 2001).

Exo is a 5'-3' exonuclease that acts on linear double stranded DNA to generate 3' single stranded DNA overhangs for recombination (Muniyappa et al., 1984, Takahashi and Kobayashi, 1990). Beta binds to the ssDNA overhangs that are no longer than 35nt long and mediates annealing of homologous complementary strands (Carter and Radding, 1971, Little, 1967). Gam inhibits RecBCD-dependent linear degradation activities to protect the DNA molecules undergoing recombination (Murphy, 1991). A number of approaches have been applied to engineer DNA inside *E.coli*. Recombination has been mediated between two circular DNA molecules (Poustka et al., 1984, Oconnor et al., 1989) and between linear PCR molecules and intact recipients such as a BAC (Bacterial Artificial Chromosome) (Muyrers et al., 2000) or the *E.coli* genome (Murphy et al., 2000, Zhang et al., 1998). In order to generate targeting vectors for making homozygous deletions of microRNAs miR-192 cluster and miR-345 in stem cells, I have used the recombineering system

devised by Pentao Liu's laboratory (Chan et al., 2007). This procedure is described in Figure 5.1 and in the Materials and Methods Section. This procedure has also being applied in the initial attempts for construction of microRNA targeting vectors in a high throughput manner at the Wellcome Trust Sanger Institute.

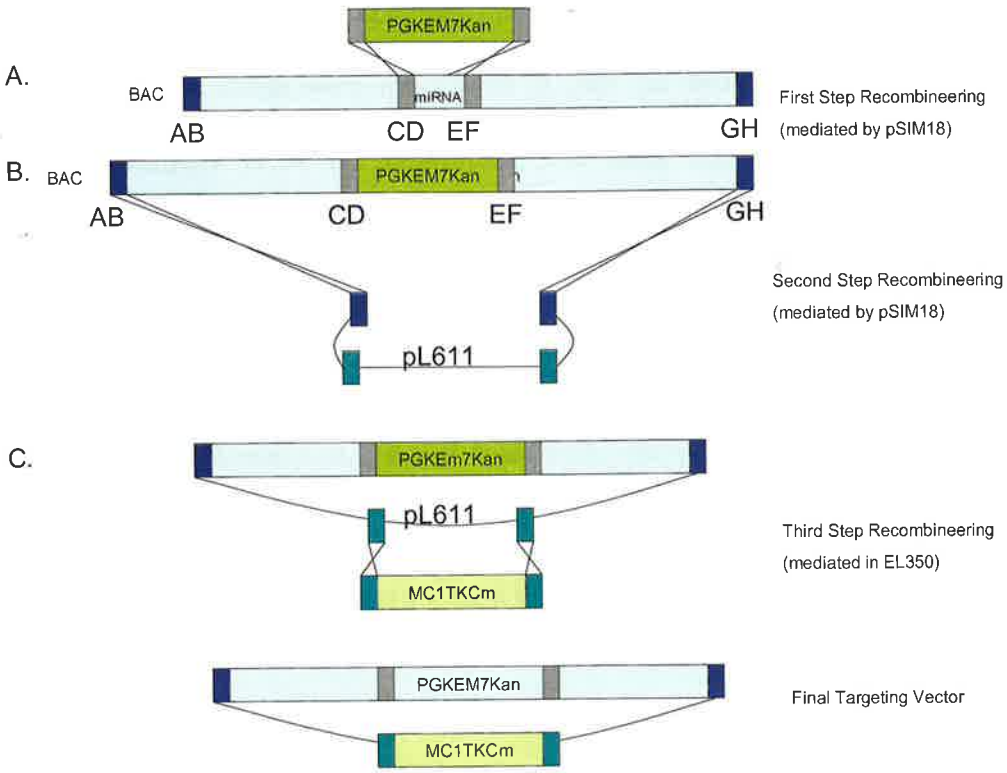


Figure 5.1: Recombineering procedure. *E.coli* cutures with BACs carrying microRNA genes were grown at 37°C and transformed with pSIM18 plasmid which carries the lambda phage proteins. Cultures grown at 32°C were made electro-competent. A. PCR products containing 90bp of homology with DNA flanking the microRNA were electroporated into *E.coli* containing pSIM18 plasmid, and recombineering was mediated by a shift of temperature transiently to 42°C. This replaced the microRNA gene with PGKEM7-Kanamycin cassette. B. Bacterial cultures with BACs containing PGKEM7-Kanamycin cassette were further electroporated with retrieval plasmid pL611 (with ampicillin resistance marker for selection in *E.coli*) containing short regions of homology to DNA lying 3.5 and 4kb upstream and downstream of the microRNA gene. Induction of recombineering proteins resulted in retrieval of homology arms of the vector. C. The pL611 vector back-bone was finally replaced with a negative selection marker containing cassette (with chloramphenicol resistance marker for selection in *E.coli*) in a third recombineering step. This resulted in formation of the final targeting vector. See Materials and Methods for further details.

5.3 Results

5.3.1 First Step Recombineering

The expression of recombineering proteins, exo, beta and gam was induced by a temperature shift to 42°C in *E.coli* cultures containing BACs with microRNA genes and pSIM18

recombineering plasmid. Un-induced *E.coli* were used as controls. These cells were made electrocompetent and were subsequently electroporated with PCR products for mediating recombination between BAC DNA and linear PCR DNA. Electroporated *E.coli* were recovered in plain LB medium, without any antibiotics at 32°C for one hour. Recovered cells were plated on LB agar plates containing kanamycin and hygromycin antibiotics and incubated overnight at 32°C. There were three colonies obtained for miR-192 and two for miR-345. No colonies were obtained for the un-induced cultures indicating that there was no leaky expression of the recombineering proteins at 32°C. These were inoculated in 2ml LB broth containing the same antibiotics and incubated overnight at 32°C. A miniprep was made with 1mL of each overnight culture. The remaining 1mL was used for making frozen glycerol stocks. PCR was performed using the miniprep DNA as template and primers flanking the upstream (U-PCR) and downstream (D-PCR) junctions of the BAC and PGKEm7-kanamycin cassette as demonstrated in Figure 5.2. As seen in the figure, two out of three colonies tested positive for both upstream and downstream junctions for miR-192. The third colony tested positive for the downstream junction but showed multiple PCR products for the upstream junction. For miR-345, both the colonies tested positive for upstream and downstream junctions.

5.3.2 Second Step Recombineering

After the first step of recombineering was confirmed, the glycerol stocks of positive *E.coli* clones were re-inoculated into LB broth containing kanamycin and hygromycin for the second recombineering phase. These cultures were incubated overnight at 32°C. On the next day, recombineering proteins were induced in the overnight cultures by a temperature shift to 42°C. Un-induced cultures were used as controls. Cells were made electro-competent and were subsequently electroporated with PCR products consisting of pL611 retrieval back bone and short regions of homology to the BAC DNA, ~4kilobases upstream and downstream of the targeted microRNA gene. After recovery in plain LB medium at 32°C for one hour, the E.coli cells were plated on LB agar containing ampicillin and kanamycin antibiotics. 21 and 45 colonies were obtained for miR-

192 and miR-345 respectively. No colonies were obtained for the un-induced cultures. 5 colonies were picked and inoculated into 2ml LB broth containing kanamycin and ampicillin. Cultures were incubated overnight at 32°C and minipreps were made the next day. Plasmid minipreps were carried out for retrieved cultures the next day. Minipreps were tested with restriction enzyme digestions (gel pictures not shown).

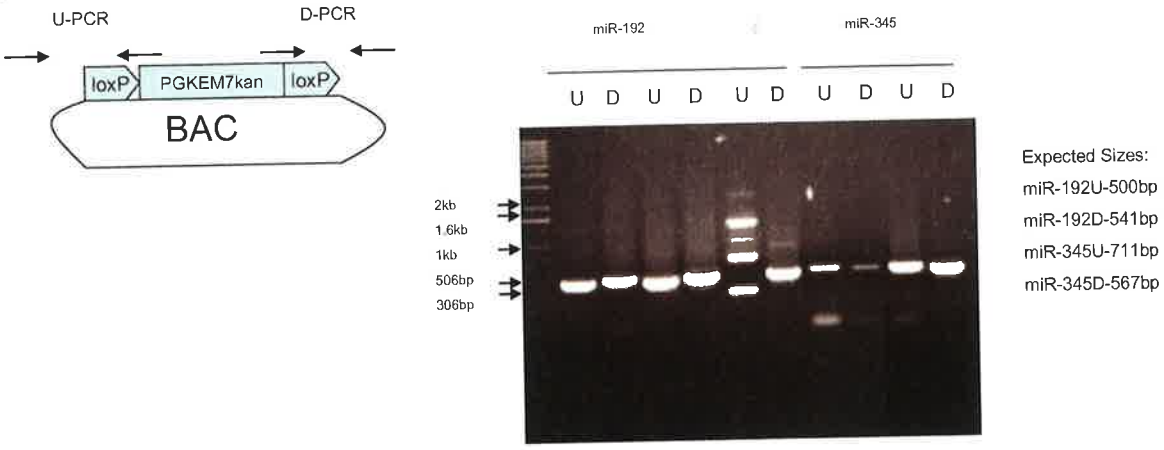
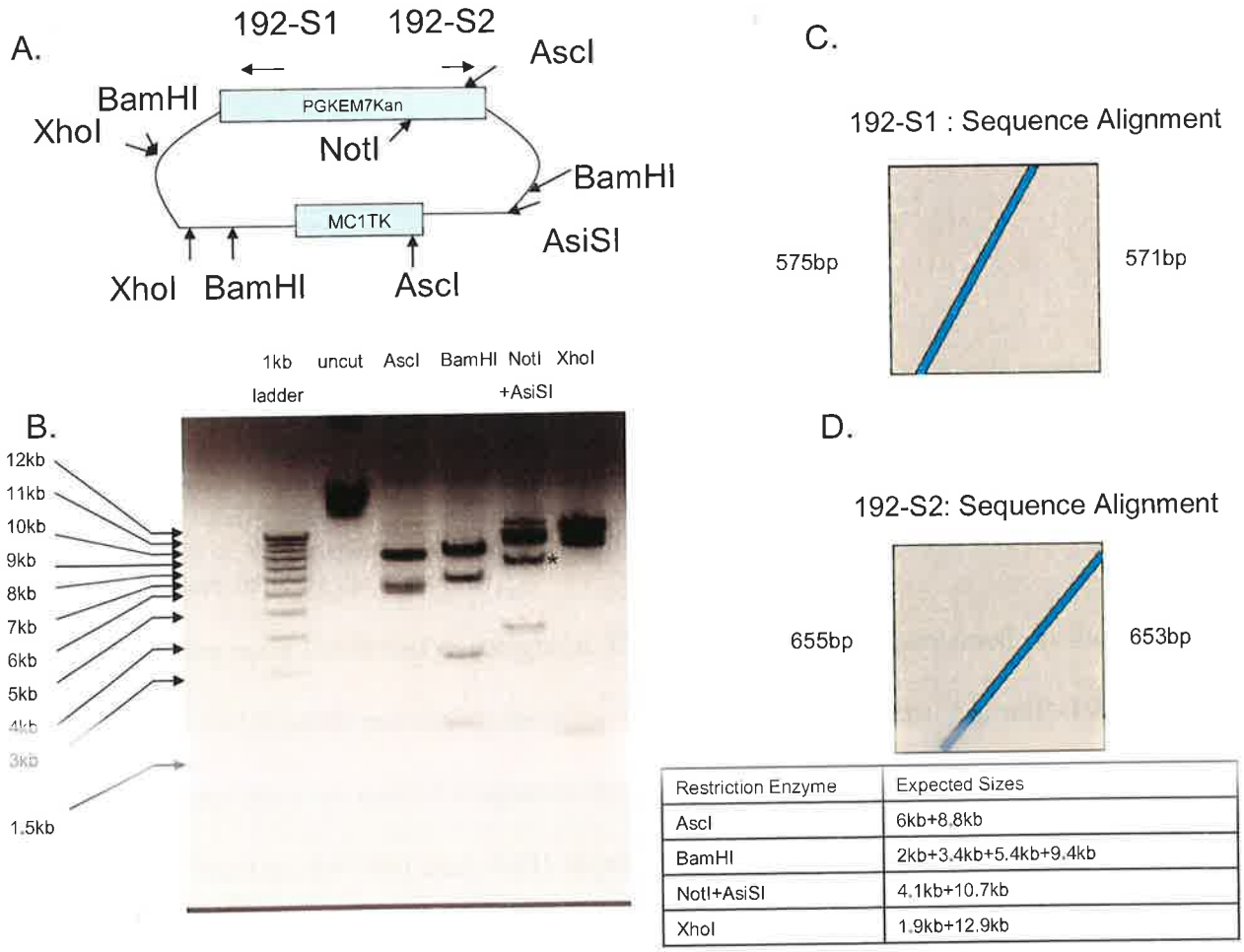


Figure 5.2: PCR diagnosis of successful recombineering events. Primers flanking the upstream (U-PCR) and downstream junction (D-PCR) of BAC and the PGKEM7-kanamycin cassette which replaces the microRNA gene on the BAC were used. Lanes labelled U and D represent U and D-PCRS respectively for miR-192 and miR-345 replaced BACs. Expected sizes of PCR products are given on the right of the gel image.

5.3.3 Third Step Recombineering

The retrieved vectors were re-transformed into EL350 cells (*E.coli* recombineering strain (Liu et al., 2003)) which were subsequently plated on LB agar plates containing kanamycin and ampicillin antibiotics. These plates were incubated overnight at 32°C. Single colonies obtained were re-inoculated into LB broth. The final targeting constructs were verified by restriction enzyme digestion (Figures 5.3 and 5.4) containing the same antibiotics. Overnight cultures were made electro-competent and electroporated with linearized MC1-TK fragments. After recovery at 32°C for one hour, cells were plated in LB agar plates containing chloramphenicol and kanamycin.



Restriction Enzyme	Expected Sizes
AscI	6kb+8.8kb
BamHI	2kb+3.4kb+5.4kb+9.4kb
NotI+AsiSI	4.1kb+10.7kb
XhoI	1.9kb+12.9kb

Figure 5.3: Final Targeting Vector for miR-192. A. *Asc*I, *Bam*HI, *Not*I and *Asi*SI and *Xho*I digestions were performed. Moreover, the prep was sequenced at the junctions of the PGKEM7-Kanamycin selection cassette and the homology arms using primers 192-S1 and 192-S2. B. 0.8% electrophoresis gel showing uncut vector (~14.6kb) and, *Asc*I, *Bam*HI, *Not*I and *Asi*SI and *Xho*I. 1kb ladder was used as a marker. C. and D. Sequence alignments for vectors using primers flanking the junctions.
*: Band of unexpected size is observed.

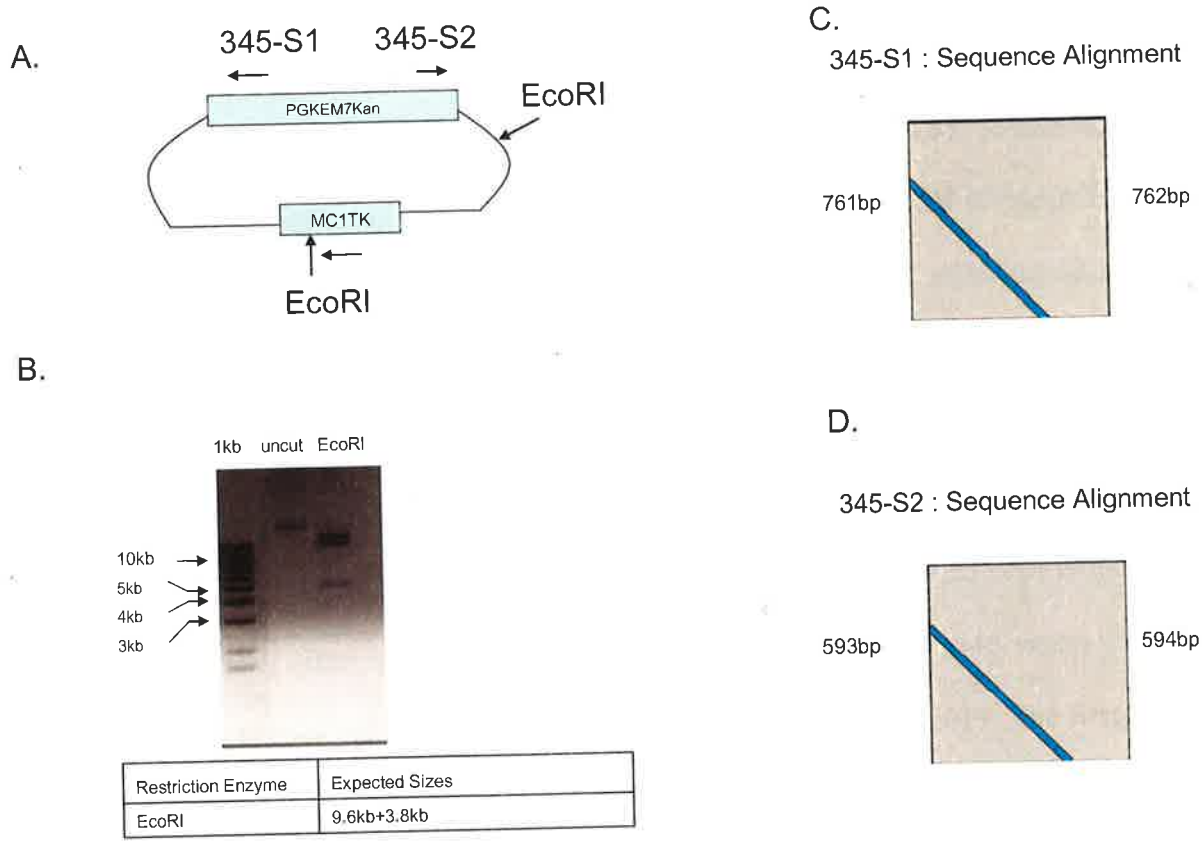


Figure 5.4: Final Targeting Vector for mir-345. A. The vector was digested with *Eco*RI restriction enzyme with arrows showing restriction sites. Moreover, the vector was sequenced at the junctions of the PGKEM7-Kanamycin selection cassette and the homology arms using primers 345-S1 and 345-S2. *Eco*RI digestion is expected to yield two fragments corresponding to 9.6kb and 3.8 kilo-bases B. 0.8% electrophoresis gel showing uncut vector (~13kb) and *Eco*RI digested fragments. An extra band for *Eco*RI digestion might indicate star activity of the enyme or partial digestion.1kb ladder was used as a marker. C. and D. Sequence alignments for vectors flanking the junctions.

These plates were incubated overnight at 32°C. Single colonies obtained on the next day, were inoculated into LB broth containing the same antibiotics for minipreps. For miR-192 targeting vector, although one observes correct fragments for *Asc*I, *Bam*HI and *Xho*I digestion, there seem to extra 9kb band observed for *Not*I and *Asi*SI digestion. For miR-345, after *Eco*RI digestion, one observes that both long and short bands are greater than the expected sizes. For both vectors, it is not known whether this is caused due to some star activity of enzymes or whether there has been some re-arrangement in the vector caused by recombineering. It is also possible that the original BACs used in these experiments could have certain mutations or rearrangements resulting in these discrepancies.

5.3.4 Sequence Verification of Targeting Vectors

miR-192 and miR-345 targeting vector samples were sent in for sequencing using primers across the junction of the left and right homology arms and PGKEM7-kanamycin cassette. Sequences were compared using 'BLAST 2 Sequences' algorithm at NCBI (Figures 5.3 and 5.4C and D). For miR-192 vector (Figure 5.5), there were a few nucleotide substitutions observed at positions 4619, 4984, 5034 and 5087. These substitutions lie in the non-coding vector sequences of the PGKEM7-Kanamycin cassette. There was a single base pair deletion detected at position 3010 lying in the vector arm exterior of the targeting cassette. As this lies in the intergenic region location of the mouse genome, it is not known, whether using this vector to target ES cells would have any significant cis or trans effect on the expression of other genes. For miR-345 vector (Figure 5.5) there were two single base pair deletions detected at positions 3723 and 6016. The first deletion at position 3723 lies in the non-coding vector region of the pGKEM7-kanamycin cassette, while the second deletion at 6016 lies on the right vector arm. Complete sequencing of vector should help determine whether there are any mutations or recombineering artefacts in the vector.

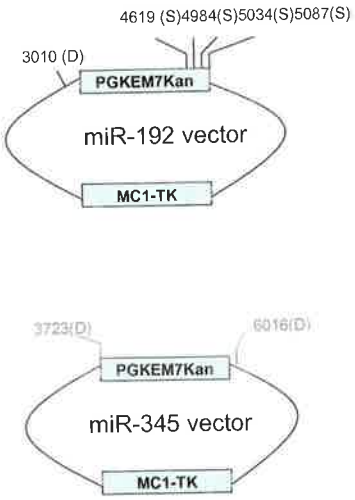


Figure 5.5: Graphical representations of nucleotide substitutions and deletions detected in mir-345 and mir-192 vectors. Deletion: D; Substitution :S

5.4 Summary

Targeting vectors for making homozygous microRNA knock-outs in mouse embryonic stem cells have also been constructed. These vectors should be sequenced completely to determine their fitness for use in targeting experiments. Loss of function analysis in embryonic stem cells is expected to reveal further information regarding the nature of microRNA –target interactions. Further, the role of microRNAs in stem cell differentiation could also be studied by making knock-outs using these vectors.

CHAPTER SIX: CONCLUSIONS AND DISCUSSION

In my thesis I have complemented computational approaches with experiments, which has led to the discovery of novel regulatory relationships between microRNAs and target transcripts in stem cells undergoing differentiation.

Pluripotency and self-renewal in stem cells is an outcome of interactions between numerous transcription factors and their downstream target genes. DGCR8 is one of the key microRNA processing enzymes whose deficiency in embryonic stem cells results in failure of stem cells to differentiate (Wang et al., 2007). These cells also lose the ability to down-regulate the expression of certain markers like Oct4, Nanog and Sox2 during stem cell differentiation. These previously reported findings led me to devise three hypotheses for the role of microRNAs in regulating expression of key stem cell specific transcription factor transcripts and transcriptional networks as described in Chapter two.

Through an elaborate computational analysis procedure I have identified a number of microRNAs interacting with key transcription factor transcripts. I have used three target prediction algorithms: miRanda (John et al., 2004), TargetScan (Grimson et al., 2007) and Pictar (Krek et al., 2005)for this analysis. This has resulted in the identification of numerous microRNA-transcript interactions which might be crucial in maintaining stem cell homeostasis. Combining results from miRanda, TargetScan and Pictar provides better coverage of targets than achievable by using any of these algorithms individually. However, many of these sites are poorly conserved across species indicating that they might be important in the mouse alone and redundant in most other mammalian species. Using these results, hypothetical regulatory networks were built to graphically depict interactions between microRNAs and these key stem cell specific transcripts in relation to stem cell pluripotency. Moreover, scanning the entire set of genes down-stream of these transcription factors also reveals significant enrichment of microRNAs interacting with these transcriptional networks in embryonic stem cells as well as differentiated tissues. These results provide important mechanistic

insights regarding the role of microRNAs in maintaining stem cell homeostasis and in regulating genes down-stream of these transcription factors.

Moreover, changes in microRNA and mRNA expression were measured using Illumina bead arrays during the process of stem cell differentiation into embryoid bodies. Illumina bead arrays represent a highly robust and sensitive expression measurement platform. Through this study, microRNAs found to be expressed in stem cells are in the range of 322. This is around six times greater than previously reported estimates of microRNAs in mouse embryonic stem cells. It is possible that some of these probes might show positive signals due to non-specific hybridisations on the array. It would be useful to confirm these results with real-time PCR experiments.

One also observes dramatic changes in expression of both microRNA and mRNA molecules during embryoid body formation. However, one must note that embryoid bodies represent a heterogenous population of cells and it is possible that individual cell types might exhibit differences in gene expression compared to one another. Transcripts belonging to key pluripotency transcription factors undergo both up and down-regulation during differentiation of stem cells into embryoid bodies. Performing *in situ* hybridisations for microRNAs and mRNAs found to be up-regulated during this process, should also help reveal the specific cell types these might be expressed in. There have been numerous examples described previously where increase in microRNA expression co-incides with appearance of new cell types, like the photoreceptor cells in *Drosophila* (Li and Carthew, 2005), Th2 cells in mice (Rodriguez et al., 2007a) and muscle cells in Xenopus laevis (Chen et al., 2006). Additionally, real-time PCRs could be performed to further validate these results.

There is an enrichment of Gene Ontology (GO) terms 'DNA-replication' and 'blood-vessel formation' in day 3 embryoid bodies. These results could explain the tendency of embryonic stem

cells to undergo cell proliferation and vasculogenesis (Doetschman et al., 1993) (Risau et al., 1988) in culture.

Sylamer (paper submitted) program has been applied to scrutinize 3'UTRs of entire transcriptional networks for specific sequence biases and microRNA seed over-representations. This analysis has resulted in revealing a fascinating AT-GC sequence bias in the 3'UTRs of genes most differentially expressed between days 0 and 3. It is possible that this bias might be caused by microRNAs with similar sequence biases. Indeed, one detects an increased GC frequency in seeds belonging to microRNAs which are most down-regulated on day 3. The exact causes for fluctuations in expression of genes with a high AT content are not known. This result correlates well with the corresponding GC bias observed for genes most up-regulated on day 3. However, one cannot completely rule out other factors which might be involved in causing this phenomenon. It is speculated that the post-transcriptional silencing machinery composed of Argonaute proteins and RISC might exploit this sequence composition bias to increase the efficiency of mRNA breakdown to facilitate rapid changes in gene expression which are characteristic of developmental processes.

One also observes, that microRNAs and 3'UTRs containing their over-represented seeds are inversely related with respect to their expression patterns.

Additionally, I have determined correlations between microRNAs and their computationally predicted interacting partners to discover whether there exists any kind of a regulatory relationship between them. Negative correlations might indicate that microRNAs might repress the expression of their predicted targets while positive correlations might imply that microRNAs might either activate transcription of some transcripts undergoing cell cycle arrest during the process of embryoid body formation. One cannot rule out the possibility that their might exist some kind of a regulatory feed-back loop between microRNAs and their predicted partners that maintains expression of these interacting molecules in a state of homeostasis. However, correlation does not

always imply causation and it is possible that some of these negative and positive correlations between microRNAs and their predicted interacting partners might be due to chance alone. Further confirmation of these results by real-time PCRs or luciferase assays would be useful.

Future work involving construction of homozygous microRNA knock-outs in embryonic stem cells should help reveal valuable target information for deleted microRNAs by monitoring changes in transcript and protein levels through expression and proteomic studies respectively. These cells could also be analyzed for their ability to undergo normal differentiation both *in vitro* and *in vivo*.

CHAPTER SEVEN: MATERIALS AND METHODS

7.1 Identification of MicroRNAs Targeting Key Pluripotency Transcripts

The ENSEMBL transcripts for Oct4, Nanog, Sox2, Sall4, Klf4,Tcf7, Sall4, Stat3, cMyc and Lrh1 were scanned for the presence of microRNA binding sites using miRanda (John et al., 2004) on miRBase (Griffiths-Jones et al., 2006), TargetScan (Grimson et al., 2007) and Pictar (Krek et al., 2005). MiRanda microRNA-targets were filtered according to *org p-value* (which denotes the probability of the same microRNA family binding to multiple transcripts within the same orthologous group) of 0.01 (Griffiths-Jones et al., 2006). The identity of microRNAs expressed in stem cells was determined from microRNA.org according to median expression cut-off value above zero.

7.2 Determining Significance of microRNA-mRNA Interactions

In order to determine whether the enrichment of binding sites for stem cell microRNAs on key stem cell transcripts is observed due to chance alone, ten thousand iterations on sets of ten transcripts with miRanda (John et al., 2004) was performed and number of observations corresponding to ES microRNA hits were counted for four out of ten transcripts. This was performed using a Perl script written by Dr. Harpreet Saini. P-values were calculated for 0, 1, 2, and 3 hits.

7.3 Identification of Up-regulated and Down-regulated Genes in the Oct4 and Nanog Transcriptional Network

The microarray data with gene expression values from mouse embryonic stem cells obtained after knock-down of Oct4 and Nanog was downloaded from GEO DataSets available from NCBI. These were subject to pre-processing and normalization using GEPAS (Herrero et al., 2003), a microarray analysis tool available online. The genes were sorted for differential expression using t-rex and classified into up-regulated and down-regulated sets according to FWER (Family wise error rate) <0.001 and <0.01.

7.4 Identification of MicroRNA Targets in Different Gene Sets

MicroRNA 6,7,8 nucleotide and double six-mer patterns were counted in each UTR of the whole mouse genome, across the Affymetrix Gene Probe sets and specific gene sets covering both up-regulated and down-regulated genes in the transcriptional network . This was preformed using PERL scripts written by Dr. Anton Enright. There were 11 and 30 genes in the down-regulated and up-regulated sets respectively which passed this filtering in the Nanog transcriptional network. Similarly, 76 and 123 of the down-regulated and up-regulated genes respectively, passed this filtering for the Oct4 transcriptional network. At a higher FWER cut-off threshold of 0.01, there were 40 genes which were down-regulated and 66 genes which were up-regulated in the Nanog transcriptional network. For the Oct4 transcriptional network, there were 209 and 328 genes in the down-regulated and up-regulated sets respectively.

MicroRNA targets spanning both the up-regulated and down-regulated genes in the Oct4 and Nanog transcriptional obtained were analyzed for tissue specificity using miRBase. The overrepresentation of microRNAs was calculated based on the formula:

OR = Sample counts per kb of 3'UTRs of the Sample Gene Set / Background Counts in 3'UTRs per kb of the entire Mouse Genome,

where OR: Overrepresentation.

7.5 ES Cell Culture

AB2.2 wild-type ES cells were maintained on SNL76/7 feeder cell layers mitotically inactivated treated by γ-irradiation (Ramirezsolis et al., 1993). ES cells were grown in M15 medium (See Table 7.1). Cells were cultured at 37°C with 5% CO_2. ES cell medium was changed daily. Upon reaching 80-85% confluence, ES cells were ready for passaging. The media was changed about two hours before passaging. After two hours, media was aspirated and the plate was washed once with PBS. 2ml of trypsin was added to each 90-mm plate. The plate was incubated in a tissue culture incubator at 37°C for 15 minutes. 8ml of fresh M15 media was added to each well. After dispersing the cells by pipetting them up and down vigorously, the cell suspension was distributed equally to three 90-

mm feeder plates. The plates were incubated in a tissue culture incubator at 37°C and 5% CO_2.

Table 7.1: Media Composition (Wang, 2005)	
Medium Name	**Composition**
M15	Knockout Dulbecco's Modified Eagle's Medium (DMEM, Gibco/Invitrogen), supplemented with 15% foetal bovine serum (FBS, Gibco/Invitrogen),2mM L-glutamine, 50 units/ml penicillin,40µg/ml streptomycin and 100µM β-Mecaptoethanol (BME)
Differentiation Medium	For 100 ml Knockout Dulbecco's Modified Eagle's Medium (DMEM, Gibco/Invitrogen), 25ml FBS (Gibco/Invitrogen), 1.25ml 200 mM 100X l-glutamine stock (Gibco/Invitrogen), 1.25ml 10mM BME stock (10mM) and 1.25ml 100X nonessential amino acids (NEAA) stock (Gibco/Invitrogen)were added)
Trypsin	For 5 L, add 35 g NaCl, 5g D-Glucose, 0.9 g $Na_2HPO_4.7H20$,1.85g KCl, 1.2g KH_2PO_4, 2g EDTA, 12.5g Trypsin (1:250), 15g Tris base. Adjust the pH to 7.61 with HCl,add phenol to get pink colour. Filter sterilize and aliquot into 50ml falcon tubes and store at -20°C.

7.6 Embryoid Body Formation

ES cells grown on 90-mm plates upto 80% confluence were fed 2-3 hours before trypsinization. The plates were washed in PBS and trypsinized for 15 minutes. The cells were re-suspended in 10ml M15 medium and counted using a hemocytometer. The cells were diluted in Differentiation Medium (See Table 7.1) to a final concentration of 1000 cells per 20 µl (Kurosawa, 2007). 20 µl drops were plated on the lids of 140-mm plates, which were carefully inverted over the Petri-dishes filled with approximately 15ml PBS. The dishes were carefully placed in the incubator at 37°C and 5% CO_2. Embryoid body formation was closely monitored daily for three days.

7.7 Total RNA Isolation from ES cells and Embryoid Bodies

ES cells and differentiating embryoid bodies at day 1, 2 and 3 were pelleted and total RNA was isolated using mirVana miRNA isolation kit. Cells and embryoid bodies were lysed by adding 600µl

of lysis solution. Pellets were vortexed vigorously for a minute. 1/10 volume of miRNA Homogenate Additive was added to the lysate and the mixture was mixed by vortexing and incubated on ice for 10 minutes. Acid-Phenol:Chloroform in volume equal to the lysate was added to the original lysate. The mixture was vortexed for 30-60 seconds and centrifuged for 5 minutes at 10000g to separate the aqueous and organic phases. The upper aqueous phase was carefully removed without disturbing the lower phase and transferred to a fresh tube. The volume removed was noted and 1.25 volumes of room temperature 100% ethanol was added to the aqueous phase. The filter cartridge provided with kit, was placed into one of the collection tubes supplied. 700 µl of the lysate mixture was applied to the cartridge at a time, and the mixture was centrifuged through the filter at 10000g for 15 seconds. After discarding the flow-through, centrifugation was repeated with the remaining sample. The filter cartridge was washed with 700 µl of Wash Solution 1 and centrifuged for 5-10 seconds. The flow-through was discarded and the washing step was repeated with 500 µl of Wash Solution 2/3 two times for 5-10 seconds. After discarding the flow through from the last wash, the filter cartridge was centrifuged for an additional 1 minute to remove any residual fluid. After this, the cartridge was transferred into a new collection tube. 20 µl of pre-heated (95°C) Elution solution was added to the centre of the cartridge and the tube was spun at 10000g for 20-30 seconds to elute the RNA. Isolated RNA was stored at -80°C and the quality was assessed with the help of Bioanalyzer (Agilent).

7.8 MicroRNA Expression Analysis Using miRCURY™ Exiqon Microarrays

In brief, RNA isolated was treated with Calf Intestinal Phosphatase (CIP), labelled with Hy3 and Hy5, with the two samples being mixed, denatured and hybridized onto the arrays overnight. Hybridized arrays were washed in buffer A (56°C), B (23°C) and C (23°C). After washing and drying, the arrays were scanned and the images were downloaded using the relevant GAL files. Individual steps are described in detail below:

Fluorescent dyes were dissolved by adding 29 µl of nuclease free water followed by brief vortexing and centrifugation to collect the contents of the tube. All kit components were thawed on ice for 15-

20 minutes followed by brief centrifugation. 12 samples corresponding 0, 1, 2 and 3 days of ES cell differentiation were diluted to a concentration of 75ng/µl. 5 µl of each sample was taken to form the reference sample pool. Reactions were set up for labelling the samples with Hy3 and the reference with Hy5 and their corresponding dye-swaps, where the samples were labelled with Hy5 and the reference was labelled with Hy3 respectively. Reagents were combined in RNAse free tubes as shown in the Table 7.2. Spike-in microRNAs consist of 10 different synthetic unlabelled microRNAs in a range of concentrations. The miRNAs can be spiked into an RNA sample prior to labelling and the synthetic spike-in kit will hybridize to corresponding capture probes on the miRCURY™ LNA microRNA Array. These were dissolved in 30 µl nuclease free water prior to use.

Table 7.2

	Volume µl
Total RNA	2
Spike-in miRNA	1
CIP buffer	0.5
CIP enzyme	0.5

Samples were incubated at 37°C for 30 minutes in a PCR cycler with a heated lid. The enzyme reaction was stopped by incubation at 95 µl for 5 minutes followed by snap cooling on ice. After incubating the samples on ice for 2-15 minutes, 4 µl of CIP reaction from previous step was taken and the reagents were added as shown in Table 7.3.

Table 7.3

	Volume µl
CIP reaction	4
Loading buffer	3
Fluorescent label	1.5
DMSO	2
Labelling enzyme	2

The reagents were mixed and centrifuged briefly and incubated at 16°C for one hour. The reactions were protected from light and were incubated at 4°C until hybridisation.

Prior to hybridisation, Hy3 and Hy5 samples were mixed and 20 µl of hybridisation buffer was added to the mixture. The samples were heated at 95°C for 2 minutes and then incubated on ice.

The slides were loaded onto the TECAN chamber and flushed with ~50-100 µl of hybridisation

buffer. The dye mixture was loaded onto the slides and incubated overnight at 56°C while protecting the samples from light. On the next day, the slides were washed with buffers A, B and C which were prepared by mixing components as provided in the kit.

7.9 Exiqon Microarray Data Analysis

Data was analysed in R/Bioconductor (Team, 2007) using limma package . Between arrays and within arrays normalisation was performed using print-tip loess and aquantile methods respectively. Differential expression analysis was performed using ebayes method (Smyth, 2004). Relative expression changes were calculated for most up and down-regulated genes.

7.10 Illumina Bead Arrays

RNA samples isolated from three days of stem cell differentiation into embryoid bodies were sent to the Sanger Microarray facility for microRNA and mRNA expression profiling using Illumina Bead Arrays (Sentrix array Matrix Universal Probe Set 5A Probe Set for microRNA expression profiling and mouse chip version 1.1 for mRNA expression profiling).

7.10.1 Illumina MicroRNA Expression Profiling

RNA samples were normalised to a concentration of 50ng/µl with DEPC treated H_2O. 5 µl of PAS was added to each well of the PAP plate. The plate was sealed with microplate heat seal and vortexed at 2300 rpm for 20 seconds. The PAP plate was placed on a preheated 37°C heat block and incubated for 60 minutes. The PAP plate was transferred to a preheated 70°C heat block and incubated for 10 minutes to inactivate the PAP enzyme. The PAP plate was pulse centrifuged to 250g for 1 minute.

8 µl of polyadenylated RNA sample from each well of the PAP plate to the corresponding well of the CSP plate which was then sealed with a microplate heat seal. The sealed plate was vortexed at 2300 rpm for 20 seconds and pulse centrifuged for 1 minute. The CSP plate was placed on the preheated heat block and the lid was closed. The plate was incubated at 42°C for 60 minutes.

The heat block was left at 70°C. The MAP reagent tube and OB1 tube was thawed to room temperature and mixed by vortexing to re-suspend the solution. The entire contents of the OB1 tube were poured into a sterile reservoir.

5 µl of MAP was added to each well of the ASE plate. 30 µl of OB1 was added to each well of the ASE plate. After carefully removing the heat seal from the CSP plate, 15 µl of biotinylated cDNA was transferred the corresponding well of the ASE plate. The ASE plate was heat-sealed with a microplate heat sealer. The plate was pulse centrifuged at 250g for 1 minute. The ASE plate was vortexed at 1600 rpm for 1 minute or until all beads are completely re-suspended. The ASE plate was placed on the preheated 70°C heat block and the lid was closed. The temperature of the heat block was immediately changed to 40 °C and the ASE plate was left in the heat block for exactly 2 hours while it cooled to 40°C.

The ASE plate was removed from the heat block which was reset to 45°C. The ASE plate was placed on the raised-bar magnetic plate for 2 minutes or until the beads are completely captured. The heat seal was removed from the ASE plate. Using an 8-channel pipette with new tips, all the liquid from the wells was removed leaving the beads in the wells. 50 µl of AM1 was added to each well of the ASE plate placed on the raised-bar magnetic plate. The ASE plate was sealed with clear adhesive film. The ASE plate was vortexed at 1600 rpm for 20 seconds or until all beads are re-suspended. The ASE plate was placed on the raised-bar magnetic plate for approximately 2 minutes or until the beads were completely captured. Using the same 8-channel pipette with the same tips, all the AM1 reagent was removed from the well leaving the beads in the wells. Repeat addition of AM1 followed by all the steps up to removal of AM1 reagent leaving the beads in the wells once. The ASE plate was removed from the raised-bar magnetic plate for approximately 2 minutes or until the beads were completely captured. The entire UB1 reagent was removed from each well, leaving the beads in the wells. Steps involving addition of UB1 up to their removal were repeated once more. 37 µl of MEL was added to each well of the ASE plate, which was sealed with clear adhesive film and vortexed for 1 minute to remove re-suspend the beads. The ASE plate was incubated on the preheated 45°C heat block for 15 minutes. In the meanwhile the Make PCR

process was performed.

64 µl of Illumina-recommended DNA polymerase was applied to the SCM tube. 50 µl of Uralic DNA glycosidase was added to SCM tube. The tube contents were mixed several times by inverting the tube. 30 µl of SCM mixture was added to each well of the PCR plate. The plate was sealed with an adhesive film. After incubation of the ASE plate was complete, it was placed on the raised-bar magnetic pole for 2 minutes or until the beads were completely captured. The supernatant was removed and from all the wells. The plate was left on the raised-bar magnetic plate for 2 minutes or until the beads were completely captured. The supernatant was discarded from all the wells of the ASE plate. The plate was sealed with clear adhesive film and vortexed at 1800rpm for 1 minute. The ASE plate was placed on the raised-bar magnetic plate for 2 minutes or until the beads were completely captured. The new tips were placed on an 8-channel pipette and the plate wells were pipette up and down 3-4 times to mix. 30 µl of supernatant was transferred from each well of the ASE plate to the corresponding well of the PCR plate. The PCR plates were sealed and pulse centrifuged to 250 g for 1 minute. The PCR plate was immediately transferred to the thermal cycler and cycled at:

37°C- 10 minutes

95°C- 3 minutes

95°C- 35 seconds

56°C- 35 seconds

72°C- 2 minutes

72°C- 10 minutes

4°C- 5 minutes

After completion, the PCR plate was pulse centrifuged for 1 minute at 250g. 20 µl of MBP was added to each well of the PCR plate. Mix the solution with the beads and transfer the mixed solution from each well of the PCR plate into the corresponding well of the filter plate. The filter plate was covered with the filter plate lid and stored at room temperature, protected from light for

60 minutes. The filter plate was placed on an empty unlabelled 96-well V-bottom plate. The filter plate containing the bound PCR products were centrifuged at 1000g for 5 minutes at 25°C. 30 µl of MH1 was added to each well of the INT plate. The waste plate was replaced with the INT plate, which was oriented such that A1 of the INT plate coincided with the A1 of the filter plate. After discarding the waste plate, 30 µl of 0.1N NaOH was added to all wells of the filter plate. After replacing the filter plate lid, the plate was centrifuged at 1000g for 5 minutes at 25°C. The filter plate was discarded and the Illumina-supplied Humidity Control Wells Template was placed beneath the 384-well HYB. 30 µl of UB2 was transferred to each orange well of the hyb plate.

The Illumina supplied Sample Wells Template was placed underneath the HYB plate exactly matching up with the edges of the plate. The contents of the plate were mixed by pipeting the contents up and down slowly several times. 50 µl of neutralised hyb solution was transferred from the first column of the INT plate into each blue well of the HYB plate. The HYB plate was sealed with clear adhesive film and centrifuged at 3000g for 4 minutes at 25°C.

The Sentrix Array Matrix (SAM) was unpackaged carefully and placed with the barcode facing up and the fiber bundles facing down into the UB2 containing tray. The SAM was agitated gently for 10 seconds and left in the tray for 3 minutes. The SAM was moved into the NaOH conditioning tray which was left for 30 seconds. The SAM was moved back into the UB2 conditioning tray and left for at least 30 seconds to neutralize NaOH.

The SAM Hyb Cartridge was loaded into the SAM Hyb Cartridge frame. After all HYB plate/SAM pairs were assembled and stacked, they were placed into the 60°C oven. After 30 minutes, the oven was reset to 45°C and incubated for at least 14 hours.

The BeadArray Reader was turned on at least 1-2 hours before imaging. A 50/50 mixture of 95% EtOH and 2-butanol was prepared and added to IS1 bottle and mixed well. 70ml of UB2 was dispensed into two trays. 70ml of IS1 was dispensed into a third tray. All trays were covered with plastic covers until further use.

After incubating the SAM in the oven, it was carefully separated from the HYB plate and was placed immediately in the first tray with UB2. Gentle agitation was carried out was carried out by

shaking it side by side ten times. It was allowed to stand at room temperature for 1 minute at room temperature. The array was then transferred into a second tray with UB2 and gently agitated as described above and then left in the tray for 1 minute at room temperature. The array was then transferred into the IS1 tray and left for 5 minutes and lifted out of the IS1 several times to ensure complete buffer exchange. The arrays were then placed on an empty array and air dried for 20 minutes in a closed drawer at room temperature. The arrays were scanned in the BeadArray Reader. Raw data files were obtained and subject to analysis.

7.10.2 Illumina Whole Genome Expression Profiling

32ml of 100% ethanol was added to the cDNA Wash Buffer Concentrate and mixed well. 20ml of 100% ethanol was added to cRNA Wash Solution and mixed well. The cRNA binding mix was prepared by combining 10µl of RNA binding beads, 4 µl of bead resuspension solution and 6 µl of 100% ethanol. 50 µl of cRNA Binding Buffer concentrate was added to 20 µl of bead mixture.

50ng of RNA was placed into wells of a PCR plate and nuclease free water was added to a final volume of 11 µl. Reverse Transcription Master Mix was prepared by adding 2 µl of 10X first strand buffer, 4 µl of dNTP mix, 1 µl of RNase inhibitor and 1 µl of ArrayScript reverse transcriptase. After mixing the contents well by vortexing, they were centrifuged for 5 seconds to collect the master mix and placed on ice. 9 µl of the master mix was transferred to each RNA sample and mixed gently by pipeting up and down 3-4 times. The reactions were incubated with aluminium sealing foil. The reaction was incubated for 2 hours at 42°C and centrifuged briefly to collect the samples afterwards. The reactions were placed on ice.

Second Strand Master Mix was prepared by adding reagents (63 µl nuclease free water, 10 µl Second Strand Buffer, 4 µl dNTP mix, 2 µl DNA Polymerase and 1 µl RNase H) and mixed gently by rocking the reservoir back and forth. 80 µl of Second Strand Master Mix was transferred to each sample in the PCR Plate. The samples were mixed gently by pipetting up and down 3-4 times. The reactions were covered in aluminium foil and incubated in a thermal cycler at 16°C for 2 hours and then placed on ice covered with aluminium foil.

180 µl of re-suspended cDNApure was added to each second strand cDNA synthesis reaction. The samples were transferred from PCR plate into a U-bottom Plate and gently shaken for at least 2 minutes to thoroughly mix the sample with cDNAPure. The U-bottom Plate was moved to a magnetic stand and the magnetic beads were captured for 5 minutes. The supernatant was carefully aspirated and discarded. The plate was removed from the magnetic stand. 150 µl of cDNA wash buffer was added to each sample and shaken at moderate speed for 1 minute. The plates were moved to a magnetic stand and the beads were captured. The supernatant was carefully aspirated and discarded without disturbing the magnetic beads. The plate was moved to a shaker and shaken vigorously without disturbing the beads. This washing step with cDNA Wash Buffer was repeated once more. The plate was then shaken vigorously for 2 minutes to evaporate all the ethanol. The cDNA was eluted from the beads by adding 20 µl of preheated (50-60 °C) nuclease free water to each sample. The plate was vigorously shaken for 3 minutes and then checked to make sure the magnetic beads were dispersed. 17.5 µl of eluted cDNA was transferred to wells of a new PCR plate. 7.5 µl of IVT master mix was added to each sample and the reaction was incubated for 4-14 hours at 37°C.

70 µl of cRNA and 95 µl ethanol was transferred to U-Bottom Plate and shaken gently for 2 minutes. The RNA binding beads were captured on a magnetic stand and the supernatant was carefully discarded. 100 µl of cRNA wash solution was added to each sample and shaken at moderate speed for 1 minute. The plate was moved to a magnetic stand and the RNA binding beads were captured as in the previous step. This washing step was repeated a second time and the plate was moved to a shaker and the plate was shaken vigorously for 3 minutes to completely disperse the beads. The RNA binding beads were captured and the supernatant containing the eluted cRNA was transferred a nuclease free PCR plate for storage. The quality of cRNA was assessed by measuring absorbance at 260nm. A positive control reaction with 2 µl of HeLa cells RNA is expected to yield >= 5 µg of cRNA.

7.10.3 Hybridisation, Washing and Signal Detection

The Hyb Chamber Gasket was placed into the Hyb chamber and 200 µl of HCB was added into 8 humidifying buffer reservoirs in the Hyb chamber. The chamber lid was closed and locked and it was left at room temperature until DNA samples were ready for loading. The BeadChip was placed in a Hyb Chamber Insert, orienting the barcode end so that it matched the barcode symbol on the Hyb Chamber Insert. The tube containing the assay sample was preheated at 65°C for 5 minutes. The tube was vortexed and pulse centrifuged at 250g. The sample was cooled to room temperature. Using a single channel precision pipette, the assay sample was loaded onto the center of each inlet port. Four Hyb Chamber Inserts containing samples were loaded into each BeadChip Hyb Chamber. The Illumina Hybridisation oven was set to 58°C and the Hyb Chamber was placed into the oven after ensuring it was locked properly. The chamber was incubated for 16 hours at 58°C. In preparation for the next day's washes, 1X High-Temp Wash Buffer was prepared by adding 50ml of 10X buffer to 450ml RNAase-free water. 500ml of prepared buffer was added to the Hybex waterbath. The Hybex heating base was set to 55°C and the buffer was warmed overnight.

6ml of E1BC buffer was added to 2L RNAse-free water to make the Wash E1BC solution. Block E1 buffer was pre-warmed to room temperature for 30 minutes. Block E1 buffer was prepared by mixing 2ml of buffer with 2 µl of 1mg/ml streptavidin-Cy3 stock.

The Hyb Chamber was removed from the oven and placed on the lab-bench and all the BeadChips were removed from the Hyb chamber and submerged at the bottom of the beaker containing 1l of E1BC buffer. The coverseal was removed from the BeadChip while it was suspended in the beaker. The BeadChip was then transferred to the slide rack submerged in the staining dish containing 250ml Wash E1BC solution. The rack containing all the slides after washing was transferred into the Hybex Waterbath containing High-Temp Wash buffer. And incubated static for 10 minutes with the lid closed. During the 10-minute incubation, 250ml of the diluted E1BC was added into a clean staining dish, and the slide rack was transferred into the staining dish containing the buffer. The rack was plunged in and out of the solution 5 to 10 times. The staining dish was placed on a orbital shaker for 5 minutes. The rack was transferred to a clean staining dish containing 250ml of fresh

100% ethanol. The slide rack was plunged in and out of the solution 5-10 times. The rack was then transferred to a clean staining dish containing 250ml wash E1BC buffer and placed in an orbital shaker at room temperature for 2 minutes.

The BeadChip was tray was placed on a rocker mixer and 4 ml of block E1 buffer was added to the wash tray. The BeadChip was transferred into the BeadChip wash tray and placed on a rocking platform for 10 minutes.

2ml of Block E1 buffer with a 1:1000 dilution of streptavidin-Cy3 was prepared and 2ml of Block E1 buffer+ streptavidin was added into a new wash tray. Using tweezers, the BeadChip was grasped at the barcode end via the well in the blocker-wash tray and transferred into a tray containing streptavidin-Cy3. The tray was covered and placed on the rocker-mixer and shaken for 10 minutes. The third washing was carried out in a staining dish containing 250ml E1BC buffer shaken in an orbital shaker for 5 minutes. The slides were centrifuged for 4 minutes at room temperature and once they were completely dry they were scanned on the Illumina BeadArray Reader.

7.11 Data Analysis of Illumina Bead Arrays

All data analysis was performed on R/Bioconductor (Team, 2007) using lumi (Du et al., 2008) and limma packages. Quality control for both raw and normalised data was performed by plotting box-plots, density plots, sample-relations and MA plots. mRNA and microRNA raw data was transformed using log2 conversion and normalised using quantile method. Differential expression analysis was carried out using ebayes method (Smyth, 2004) and contrasts were determined for pair-wise comparisons of differential expression between day 0-1, 0-2, 0-3, 1-2 and 2-3. Median expression of three replicates was determined for plotting expression changes. The heatmap was plotted using gplots on R. Gene Ontology enrichment was determined by GOstats package (Falcon and Gentleman, 2007) on R. The Gene Universe consisted of all probes from the arrays, which mapped uniquely to Entrez Ids. It was ensured that there was no redundancy in Entrez gene lists. Sylamer was used to scan genes in sets of 1000 to calculate cumulative enrichment of words or microRNA seeds. I wrote a perl script to calculate AT-GC enrichments in microRNA seeds.

7.12 Using Sylamer

Sylamer is a tool for efficient computation of p-values associated with small sequence word occurrences in sequences. The p-values in all the plots generated have been cumulative hypergeomentric. For Figure 3.9, word size taken was 7. For, Figure 4.13 word size was 4. All plots have been generated by Dr. Cei-Abreu Goodger.

7.13 Making Vectors by Recombineering

Primers for amplifying linear recombineering products were designed manually. The sequences for amplifying the PGKEM7-Kanamycin cassette are as follows:

The sequences for the retrieval step are as follows:

1ng of template (ScaI digested pL452 (Liu et al., 2003) plasmid vector for amplifying PGKEm7-Kanamycin cassette and *Eco*RI and *Bam*HI digested pL611 (Chan et al., 2007) for amplifying retrieval cassette) and 10 μM of primers were used for all PCR reactions. The conditions were set as follows:

Step 1: 94°C-4 minutes

Step 2: 94°C-30 seconds

Step 3: 55°C-30 seconds

Step 4: 68°C-2 minutes (for pL452)/3 minutes(for pL611)

Step 5: Go to Step 2 (35 cycles)

Step 6: 68°C -4 minutes

Step 7: 4°C-for ever

Five PCR reactions were set up for each recombineering experiment. PCR products were gel purified using QIAGEN gel purification kit. Approximately 3 μg of PCR product for each reaction was pooled in a volume of 50μl of double distilled water.

The following steps were subsequently followed:

BACs were inoculated into 1ml LB with chloramphenicol (12.5 μg/ml), for overnight growth at 37°C with shaking. Cells were pelleted and washed with cold water thrice. 1ng of pSIM18 (Chan et al., 2007) plasmid was electroporated in 50 μl water at the following conditions:

200 Ohms (capacitance, 25 microFD, 0.1 cm gap cuvette, 1.8 kV)

1 ml LB was added to the transformation mixture and was incubated for 1 h at 32°C. The cells were plated onto a LB-Hygro (75 μg/ml) plates and incubated at 32°C for overnight. On the next day, one Hygromycin resistant colony was picked and inoculated into 1 ml LB with chloramphenicol and hygromycin (75 μg/ml), for overnight growth at 32°C with shaking.

On the next day ~30μl of the overnight culture was inoculated into four 15-ml tubes (or into four wells of a 96-well deep plate) with 1 ml fresh LB in each tube. Cultures were shaken at 32°C for 2 h. Cells were transferred to a 42°C water bath and incubated for 15 min without shaking. The cells were subsequently transferred into an ice bucket and chilled on ice for 5 min. Cells were then transferred into 1.5-ml eppendorf tubes and centrifuged at maximum speed for 25 s and washed with ice-cold distilled water thrice and electroporated with PGKEm7-kanamycin PCR product (3 μg) in 50 μl water. 1 ml LB was added to the transformation mixture, which was incubated for 1 h at 32°C before plated onto a LB-Chloramphenicol, hygromycin and kanamycin (20 μg /ml) plates.

Chloramphenicol and kanamycin resistant colonies were picked to test for successful recombineering events. These were inoculated into 2ml of LB containing chloramphenicol and kanamycin. On the next day 1ml of culture for all inculations was frozen at -80°C in glycerol and miniprep was made of the remaining culture. 1ng of minipreped BAC was used as a template to carry out PCRs using primers flanking the up-stream and down-stream junctions of the PGKEM7-Kanamycin cassette using the following primers:

Table 7.4 Junction PCR Primers (5'-3')	
Name	**Sequence**
Mir-192-upstreamjunctionpL452F	GGTGTGGGCCTTGGATGA
Mir-192-upstreamjunctionpL452R	CCTACCGGTGGATGTGGAAT
mir-192-downstreamjunctionpL452R	GGGAGGATTGGGAAGACAAT
mir-192-downstreamjunctionpL452R	TCAAAGCCCGAAGACCTATG
mir-345upstreamjunctionpL452F	GCTAGGTGCAGGACAAGGAG
mir-345upstreamjunctionpL452R	GGGGAACTTCCTGACTAGGG
mir-345downstreamjunctionpL452F	GGGAGGATTGGGAAGACAAT
mir-345downstreamjunctionpL452R	CAAGGGCTGAAGAGATGAGG

PCR Conditions were as follows:

Step 1: 94°C-4 minutes

Step 2: 94°C-30 seconds

Step 3: 55°C-30 seconds

Step 4: 68°C-1 minute

Step 5: Go to Step 2 (35 cycles)

Step 6: 68°C -4 minutes

Step 7: 4°C-for ever

Glycerol stocks of colonies that tested positive for both the upstream and down-stream junctions were used for subsequently inoculating cultures for the next recombineering steps. Glycerol stocks were also inoculated in LB containing ampicllin (50 μg/ml) to confirm that they were ampicillin sensitive.

Chloramphenicol^{R} / Hygromycin^{R} / Kanamycin ^{R} colonies were inoculated into 1 ml LB with the antibiotics for overnight growth at 32°C. Recombineering was performed as described above

using electroporation with PCR-amplified pL611 (~3 μg). Transformants were plated on an Amp plate (50 μg/ml) and grown at 32°C for overnight. Individual Ampicllin^R colonies were inoculated into LB containing ampicllin and kanamycin. Minipreps were made the following day. 1.0 ng plasmid was transformed into EL350 *E.coli* strain (Liu et al., 2003). Transformed cells were plated on LB agar containing kanamycin and ampicillin and incubated overnight at 32°C. Single colonies obtained were re-inoculated into LB containing ampicllin and kanamycin. Overnight cultures (~30μl) were re-inoculated the next day into LB and incubated at 32°C for 2 hours. Cells were electroporated (as described above) with 3 μg of MC1-TK-Cm 4.2kb fragment obtained after digestion of plasmid with NotI and SalI. Electroporated cells were recovered in LB for one hour at 32°C and plated on LB agar containing chloramphenicol and kanamycin. Cells were incubated overnight at 32°C and single colonies were inoculated into LB medium containing the same antibiotics. Vectors were isolated from overnight cultures grown at 32°C on the next day by miniprep. Restriction enzyme digestion was performed to detect the presence of final vectors. miR-192 vector was digested with NdeI and XhoI and mir-345 vector was digested with EcoRI.

Table 7.5 Recombineering Primers (5'-3')

Name	Sequence
mir-192-pL452-F-23-09	GCCGAGAGGCCTGCGGGCATGAGTGGGCAGCTG TGGGGCTCTTCCCAGGATCCTGCAGAGAGGAGGT CCCTGCACCTGTGCTGACCCCCCACCATGATTACG CCAAGCTC
mir-192-pL452-R-23-09	GTGTGAGGAAGGAGGAGGGACCTTGACTTTGATA CCCTCCCCCACCCTGACTCCCACCCTGCTCCTTCG AGTCGTTATGGCTGGCATTGAGTTTTCCCAGTCAC GACGTT
mir-192Retrieval 1pL611	ACATCTGCTTCCCCTCTCCCTCACGCACACTTTTC CCGAACTACTAGAGACCAGCATACAGGTCATGAC CCTTCACCATGAAAATTTTCATTAGTGAGGGTTAA TTATCG
mir-192Retrieval 2pL611	TGCTACGATTATTGGTTTTACCACGGTTGGGGGGA GCTGTCGGATCACACCCCCCACGGC ACCTGTCAGTCCCTTCTGCTCATGGCCCCTATACG ACTCACTATAGGGAG
mir-345-pL452-forward-29-09	TTAAAGTTGAGGTGGGCTAAGAACCCGACGCGCC GCCCCCCACCCCCCCCAGTTGCTGGTAATATCTACA ACAGCAAGGAGTCTGGAAGCAACCATGATTACGC CAAGCTC
mir-345-pL452-reverse-29-09	AAAGAGCCGTCTTGTCTCATCGGATGCCCTCCCG TTTCTTAGAACCAGAAACCAGGGTGGAGGCGTG AAGACCCGAGCTTTACCACCGGAGTTTTCCCAGT CACGACGTT
mir-345Retrieval 1pL611	TTGGCTGGTGTGCCTTGATGTGAGTACCCACTTCA GGCTTCCAAGAGGGAAGTGAAGCATAGAGTACA GCCAAATCAGATTGCCAGCCCCTTAGTGAGGGTT AATTATCG
Mir-345Retrieval 2pL611	TTGCTTTAAAGACACAGGCACATCCTTGGCACCG TCTGGTGACATTTCTAAAGTCCGCACATAAAATCT CAGAGGCGTCCTGAGAGAACAATACGACTCACTA TAGGGAG

Table 7.6 Sequencing Primers (5'-3')

Name	Sequence
mir-192-upstreamjunctionpL452R	CCTACCGGTGGATGTGGAAT
mir-192-downstreamjunctionpL452F	GGGAGGATTGGGAAGACAAT
mir-345upstreamjunctionpL452R	GGGGAACTTCCTGACTAGGG
mir-345downstreamjunctionpL452F	GGGAGGATTGGGAAGACAAT

BIBLIOGRAPHY

AMBROS, V. & HORVITZ, H. R. (1984) Heterochronic Mutants of the Nematode Caenorhabditis-Elegans. *Science,* 226, 409-416.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M., SHERLOCK, G. & GENE ONTOLOGY, C. (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics,* 25, 25-29.

BARTEL, D. P. (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell,* 116, 281-297.

BAULCOMBE, D. C. (1996) RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Molecular Biology,* 32, 79-88.

BENJAMINI, Y. & HOCHBERG, Y. (1995) CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. *Journal of the Royal Statistical Society Series B-Methodological,* 57, 289-300.

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. & HANNON, G. J. (2003) Dicer is essential for mouse development. *Nature Genetics,* 35, 215-217.

BOHNSACK, M. T., CZAPLINSKI, K. & GORLICH, D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna-a Publication of the Rna Society,* 10, 185-191.

BORCHERT, G. M., LANIER, W. & DAVIDSON, B. L. (2006) RNA polymerase III transcribes human microRNAs. *Nature Structural & Molecular Biology,* 13, 1097-1101.

BRADLEY, A., EVANS, M., KAUFMAN, M. H. & ROBERTSON, E. (1984) Formation of Germ-Line Chimeras from Embryo-Derived Teratocarcinoma Cell-Lines. *Nature,* 309, 255-256.

BRENNECKE, J., HIPFNER, D. R., STARK, A., RUSSELL, R. B. & COHEN, S. M. (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell,* 113, 25-36.

CALABRESE, J. M., SEILA, A. C., YEO, G. W. & SHARP, P. A. (2007) RNA sequence analysis defines Dicer's role in mouse embryonic stem cells (vol 104, pg 18097, 2007). *Proceedings of the National Academy of Sciences of the United States of America,* 104, 21021-21021.

CAPECCHI, M. R. (1989) Altering the Genome by Homologous Recombination. *Science,* 244, 1288-1292.

CAREY, V. J., GENTLEMAN, R., HUBER, W. & GENTRY, J. (2005) Bioconductor software for graphs. *Bioinformatics and Computational Biology Solution Using R and Bioconductor.*

CARTER, D. M. & RADDING, C. M. (1971) Role of Exonuclease and Beta Protein of Phage Lambda in Genetic Recombination .2. Substrate Specificity and Mode of Action of Lambda Exonuclease. *Journal of Biological Chemistry,* 246, 2502-&.

CARTWRIGHT, P., MCLEAN, C., SHEPPARD, A., RIVETT, D., JONES, K. & DALTON, S. (2005) LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development,* 132, 885-896.

CASTRO, D. S., HERMANSON, E., JOSEPH, B., WALLEN, A., AARNISALO, P., HELLER, A. & PERLMANN, T. (2001) Induction of cell cycle arrest and morphological differentiation by Nurr1 and retinoids in dopamine MN9D cells. *Journal of Biological Chemistry,* 276, 43277-43284.

CHAMBERS, I., COLBY, D., ROBERTSON, M., NICHOLS, J., LEE, S., TWEEDIE, S. & SMITH, A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell,* 113, 643-655.

CHAN, W., COSTANTINO, N., LI, R. X., LEE, S. C., SU, Q., MELVIN, D., COURT, D. L. & LIU, P. T. (2007) A recombineering based approach for high-throughput conditional

knockout targeting vector construction. *Nucleic Acids Research,* 35.

CHEN, J. F., MANDEL, E. M., THOMSON, J. M., WU, Q. L., CALLIS, T. E., HAMMOND, S. M., CONLON, F. L. & WANG, D. Z. (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genetics,* 38, 228-233.

CHIU, Y. L. & RANA, T. M. (2003) siRNA function in RNAi: A chemical modification analysis. *Rna-a Publication of the Rna Society,* 9, 1034-1048.

COPELAND, N. G., JENKINS, N. A. & COURT, D. L. (2001) Recombineering: A powerful new tool for mouse functional genomics. *Nature Reviews Genetics,* 2, 769-779.

COUCOUVANIS, E. & MARTIN, G. R. (1995) Signals for Death and Survival - a 2-Step Mechanism for Cavitation in the Vertebrate Embryo. *Cell,* 83, 279-287.

DENLI, A. M., TOPS, B. B. J., PLASTERK, R. H. A., KETTING, R. F. & HANNON, G. J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature,* 432, 231-235.

DOENCH, J. G. & SHARP, P. A. (2004) Specificity of microRNA target selection in translational repression. *Genes & Development,* 18, 504-511.

DOETSCHMAN, T., SHULL, M., KIER, A. & COFFIN, J. D. (1993) Embryonic Stem-Cell Model Systems for Vascular Morphogenesis and Cardiac Disorders. *Hypertension,* 22, 618-629.

DU, P., KIBBE, W. A. & LIN, S. M. (2008) lumi: a pipeline for processing Illumina microarray. *Bioconductor package version 1.5.18.*

ELLIS, H. M., YU, D. G., DITIZIO, T. & COURT, D. L. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America,* 98, 6742-6746.

EVANS, M. J. & KAUFMAN, M. H. (1981) Establishment in Culture of Pluripotential Cells from Mouse Embryos. *Nature,* 292, 154-156.

FALCON, S. & GENTLEMAN, R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics,* 23, 257-258.

FARH, K. K. H., GRIMSON, A., JAN, C., LEWIS, B. P., JOHNSTON, W. K., LIM, L. P., BURGE, C. B. & BARTEL, D. P. (2005) The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science,* 310, 1817-1821.

FORSTEMANN, K., HORWICH, M. D., WEE, L., TOMARI, Y. & ZAMORE, P. D. (2007) Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by Dicer-1. *Cell,* 130, 287-297.

GIRALDEZ, A. J., MISHIMA, Y., RIHEL, J., GROCOCK, R. J., VAN DONGEN, S., INOUE, K., ENRIGHT, A. J. & SCHIER, A. F. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science,* 312, 75-79.

GLOVER, C. H., MARIN, M., EAVES, C. J., HELGASON, C. D., PIRET, J. M. & BRYAN, J. (2006) Meta-analysis of differentiating mouse embryonic stem cell gene expression kinetics reveals early change of a small gene set. *Plos Computational Biology,* 2, 1463-1474.

GREGORY, R. I., CHENDRIMADA, T. P., COOCH, N. & SHIEKHATTAR, R. (2007) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell,* 131, 63-73.

GREGORY, R. I. & SHIEKHATTAR, R. (2005) MicroRNA biogenesis and cancer. *Cancer Research,* 65, 3509-3512.

GRIFFITHS-JONES, S., GROCOCK, R. J., VAN DONGEN, S., BATEMAN, A. & ENRIGHT, A. J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research,* 34, D140-D144.

GRIFFITHS-JONES, S., SAINI, H. K., VAN DONGEN, S. & ENRIGHT, A. J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Research,* 36, D154-D158.

GRIMSON, A., FARH, K. K. H., JOHNSTON, W. K., GARRETT-ENGELE, P., LIM, L. P. & BARTEL, D. P. (2007) MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Molecular Cell,* 27, 91-105.

GUAN, K., ROHWEDEL, J. & WOBUS, A. M. (1999) Embryonic stem cell differentiation models: cardiogenesis, myogenesis, neurogenesis, epithelial and vascular smooth muscle cell differentiation in vitro. *Cytotechnology,* 30, 211-226.

HALEY, B. & ZAMORE, P. D. (2004) Kinetic analysis of the RNAi enzyme complex. *Nature Structural & Molecular Biology,* 11, 599-606.

HAN, J. J., LEE, Y., YEOM, K. H., NAM, J. W., HEO, I., RHEE, J. K., SOHN, S. Y., CHO, Y. J., ZHANG, B. T. & KIM, V. N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell,* 125, 887-901.

HAY, D. C., SUTHERLAND, L., CLARK, J. & BURDON, T. (2004) Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells,* 22, 225-235.

HERRERO, J., AL-SHAHROUR, F., DIAZ-URIARTE, R., MATEOS, A., VAQUERIZAS, J. M., SANTOYO, J. & DOPAZO, J. (2003) GEPAS: a web-based resource for microarray gene expression data analysis. *Nucleic Acids Research,* 31, 3461-3467.

HOUBAVIY, H. B., MURRAY, M. F. & SHARP, P. A. (2003) Embryonic stem cell-specific MicroRNAs. *Developmental Cell,* 5, 351-358.

HUMPHREYS, D. T., WESTMAN, B. J., MARTIN, D. I. K. & PREISS, T. (2005) MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proceedings of the National Academy of Sciences of the United States of America,* 102, 16961-16966.

HUTVAGNER, G., MCLACHLAN, J., PASQUINELLI, A. E., BALINT, E., TUSCHL, T. & 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.

HUTVAGNER, G. & ZAMORE, P. D. (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science,* 297, 2056-2060.

JING, Q., HUANG, S., GUTH, S., ZARUBIN, T., MOTOYAMA, A., CHEN, J. M., DI PADOVA, F., LIN, S. C., GRAM, H. & HAN, J. H. (2005) Involvement of MicroRNA in AU-rich element-mediated mRNA instability. *Cell,* 120, 623-634.

JOHN, B., ENRIGHT, A. J., ARAVIN, A., TUSCHL, T., SANDER, C. & MARKS, D. S. (2004) Human MicroRNA targets. *Plos Biology,* 2, 1862-1879.

JOHN, B., ENRIGHT, A. J., ARAVIN, A., TUSCHL, T., SANDER, C. & MARKS, D. S. (2005) Human microRNA targets (vol 2, pg 1862, 2005). *Plos Biology,* 3, 1328-1328.

JOHNSTON, R. J. & HOBERT, O. (2003) A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. *Nature,* 426, 845-849.

KEDERSHA, N., STOECKLIN, G., AYODELE, M., YACONO, P., LYKKE-ANDERSEN, J., FRITZLER, M. J., SCHEUNER, D., KAUFMAN, R. J., GOLAN, D. E. & ANDERSON, P. (2005) Stress granules and processing bodies are dynamically linked sites of mRNP remodeling (vol 169, pg 871, 2005). *Journal of Cell Biology,* 170, 847-847.

KIRIAKIDOU, M., NELSON, P. T., KOURANOV, A., FITZIEV, P., BOUYIOUKOS, C., MOURELATOS, Z. & HATZIGEORGIOU, A. (2004) A combined computational-experimental approach predicts human microRNA targets. *Genes & Development,* 18, 1165-1178.

KOSCIANSKA, E., BAEV, V., SKREKA, K., OIKONOMAKI, K., RUSINOV, V., TABLER, M. & KALANTIDIS, K. (2007) Prediction and preliminary validation of oncogene regulation by miRNAs. *BMC Molecular Biology,* 8, Article No.: 79.

KREK, A., GRUN, D., POY, M. N., WOLF, R., ROSENBERG, L., EPSTEIN, E. J., MACMENAMIN, P., DA PIEDADE, I., GUNSALUS, K. C., STOFFEL, M. & RAJEWSKY, N. (2005) Combinatorial microRNA target predictions. *Nature Genetics,* 37, 495-500.

KUHN, D. E., MARTIN, M. M., FELDMAN, D. S., TERRY JR, A. V., NUOVA, G. J. & ELTON, T. S. (2008) Experimental Validation of miRNA Targets. *Methods,* 44, 47-54.

KUROSAWA, H. (2007) Methods for inducing embryoid body formation: In vitro differentiation system of embryonic stem cells. *Journal of Bioscience and Bioengineering,* 103, 389-398.

LANDGRAF, P., RUSU, M., SHERIDAN, R., SEWER, A., IOVINO, N., ARAVIN, A., PFEFFER, S., RICE, A., KAMPHORST, A. O., LANDTHALER, M., LIN, C., SOCCI, N. D., HERMIDA, L., FULCI, V., CHIARETTI, S., FOA, R., SCHLIWKA, J., FUCHS, U., NOVOSEL, A., MULLER, R. U., SCHERMER, B., BISSELS, U., INMAN, J., PHAN, Q., CHIEN, M. C., WEIR, D. B., CHOKSI, R., DE VITA, G., FREZZETTI, D., TROMPETER, H. I., HORNUNG, V., TENG, G., HARTMANN, G., PALKOVITS, M., DI LAURO, R., WERNET, P., MACINO, G., ROGLER, C. E., NAGLE, J. W., JU, J. Y., PAPAVASILIOU, F. N., BENZING, T., LICHTER, P., TAM, W., BROWNSTEIN, M. J., BOSIO, A., BORKHARDT, A., RUSSO, J. J., SANDER, C., ZAVOLAN, M. & TUSCHL, T. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell,* 129, 1401-1414.

LEE, R. C., FEINBAUM, R. L. & 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. J., CHOI, H., KIM, J., YIM, J., LEE, J., PROVOST, P., RADMARK, O., KIM, S. & KIM, V. N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature,* 425, 415-419.

LEE, Y., KIM, M., HAN, J. J., YEOM, K. H., LEE, S., BAEK, S. H. & KIM, V. N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *Embo Journal,* 23, 4051-4060.

LEUSCHNER, P. J. F., AMERES, S. L., KUENG, S. & MARTINEZ, J. (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *Embo Reports,* 7, 314-320.

LEWIS, B. P., BURGE, C. B. & BARTEL, D. P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell,* 120, 15-20.

LEWIS, B. P., SHIH, I. H., JONES-RHOADES, M. W., BARTEL, D. P. & BURGE, C. B. (2003) Prediction of mammalian microRNA targets. *Cell,* 115, 787-798.

LI, X. & CARTHEW, R. W. (2005) A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the Drosophila eye. *Cell,* 123, 1267-1277.

LI, Y., MCCLINTICK, J., ZHONG, L., EDENBERG, H. J., YODER, M. C. & CHAN, R. J. (2005) Murine embryonic stem cell differentiation is promoted by SOCS-3 and inhibited by the zinc finger transcription factor Klf4. *Blood,* 105, 635-637.

LIN, S. Y., JOHNSON, S. M., ABRAHAM, M., VELLA, M. C., PASQUINELLI, A., GAMBERI, C., GOTTLIEB, E. & SLACK, F. J. (2003) The C-elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Developmental Cell,* 4, 639-650.

LITTLE, J. W. (1967) An Exonuclease Induced by Bacteriophage . II. NATURE OF THE ENZYMATIC REACTION *Journal of Biological Chemistry,* 242, 679-686.

LIU, J. D., VALENCIA-SANCHEZ, M. A., HANNON, G. J. & PARKER, R. (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nature Cell Biology,* 7, 719-U118.

LIU, P. T., JENKINS, N. A. & COPELAND, N. G. (2003) A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Research,* 13, 476-484.

LOH, Y. H., WU, Q., CHEW, J. L., VEGA, V. B., ZHANG, W. W., CHEN, X., BOURQUE, G., GEORGE, J., LEONG, B., LIU, J., WONG, K. Y., SUNG, K. W., LEE, C. W. H., ZHAO, X. D., CHIU, K. P., LIPOVICH, L., KUZNETSOV, V. A., ROBSON, P., STANTON, L. W., WEI, C. L., RUAN, Y. J., LIM, B. & NG, H. H. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genetics,* 38, 431-440.

LUND, E., GUTTINGER, S., CALADO, A., DAHLBERG, J. E. & KUTAY, U. (2004) Nuclear export of microRNA precursors. *Science,* 303, 95-98.

LYKKE-ANDERSEN, J. & WAGNER, E. (2005) Recruitment and activation of mRNA decay enzymes by two ARE-mediated decay activation domains in the proteins TTP and BRF-1. *Genes & Development,* 19, 351-361.

MA, J. B., YUAN, Y. R., MEISTER, G., PEI, Y., TUSCHL, T. & PATEL, D. J. (2005) Structural basis for 5 '-end-specific recognition of guide RNA by the A-fulgidus Piwi protein. *Nature,* 434, 666-670.

MARTIN, G. R. (1981) Isolation of a Pluripotent Cell-Line from Early Mouse Embryos Cultured in Medium Conditioned by Teratocarcinoma Stem-Cells. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 78, 7634-7638.

MARTIN, G. R., WILEY, L. M. & DAMJANOV, I. (1977) Development of Cystic Embryoid Bodies Invitro from Clonal Teratocarcinoma Stem-Cells. *Developmental Biology,* 61, 230-244.

MATRANGA, C., TOMARI, Y., SHIN, C., BARTEL, D. P. & ZAMORE, P. D. (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell,* 123, 607-620.

MINENO, J., OKAMOTO, S., ANDO, T., SATO, M., CHONO, H., IZU, H., TAKAYAMA, M., ASADA, K., MIROCHNITCHENKO, O., INOUYE, M. & KATO, I. (2006) The expression profile of microRNAs in mouse embryos. *Nucleic Acids Research,* 34, 1765-1771.

MISKA, E. A., ALVAREZ-SAAVEDRA, E., TOWNSEND, M., YOSHII, A., SESTAN, N., RAKIC, P., CONSTANTINE-PATON, M. & HORVITZ, H. R. (2004) Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biology,* 5.

MITCHELL, P., PETFALSKI, E., SHEVCHENKO, A., MANN, M. & TOLLERVEY, D. (1997) The exosome: A conserved eukaryotic RNA processing complex containing multiple 3'->5' exoribonucleases. *Cell,* 91, 457-466.

MITSUI, K., TOKUZAWA, Y., ITOH, H., SEGAWA, K., MURAKAMI, M., TAKAHASHI, K., MARUYAMA, M., MAEDA, M. & YAMANAKA, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell,* 113, 631-642.

MUNIYAPPA, K., SHANER, S. L., TSANG, S. S. & RADDING, C. M. (1984) Mechanism of the Concerted Action of Reca Protein and Helix-Destabilizing Proteins in Homologous Recombination. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 81, 2757-2761.

MURCHISON, E. P., PARTRIDGE, J. F., TAM, O. H., CHELOUFI, S. & HANNON, G. J. (2005) Characterization of Dicer-deficient murine embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America,* 102, 12135-12140.

MURPHY, K. C. (1991) Lambda-Gam Protein Inhibits the Helicase and Chi-Stimulated Recombination Activities of Escherichia-Coli Recbcd Enzyme. *Journal of Bacteriology,* 173, 5808-5821.

MURPHY, K. C., CAMPELLONE, K. G. & POTEETE, A. R. (2000) PCR-mediated gene replacement in Escherichia coli. *Gene,* 246, 321-330.

MUYRERS, J. P. P., ZHANG, Y. M., BENES, V., TESTA, G., ANSORGE, W. & STEWART, A. F. (2000) Point mutation of bacterial artificial chromosomes by ET recombination. *Embo Reports,* 1, 239-243.

NAKATAKE, Y., FUKUI, N., IWAMATSU, Y., MASUI, S., TAKAHASHI, K., YAGI, R., YAGI, K., MIYAZAKI, J., MATOBA, R., KO, M. S. H. & NIWA, H. (2006) Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells. *Molecular and Cellular Biology,* 26, 7772-7782.

NELSON, P. T., HATZIGEORGIOU, A. G. & MOURELATOS, Z. (2004) miRNP : mRNA association in polyribosomes in a human neuronal cell line. *Rna-a Publication of the Rna Society,* 10, 387-394.

NIWA, H., MIYAZAKI, J. & SMITH, A. G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics,* 24, 372-376.

NYKANEN, A., HALEY, B. & ZAMORE, P. D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell,* 107, 309-321.

OCONNOR, M., PEIFER, M. & BENDER, W. (1989) Construction of Large DNA Segments in Escherichia-Coli. *Science,* 244, 1307-1312.

OKAMURA, K., HAGEN, J. W., DUAN, H., TYLER, D. M. & LAI, E. C. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell,* 130, 89-100.

OKAMURA, K., PHILLIPS, M. D., TYLER, D. M., DUAN, H., CHOU, Y. T. & LAI, E. C. (2008) The regulatory activity of microRNA star species has substantial influence on microRNA and 3 ' UTR evolution. *Nature Structural & Molecular Biology,* 15, 354-363.

OLSEN, P. H. & AMBROS, V. (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biology,* 216, 671-680.

ORBAN, T. I. & IZAURRALDE, E. (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *Rna-a Publication of the Rna Society,* 11, 459-469.

PADDISON, P. J., CAUDY, A. A., BERNSTEIN, E., HANNON, G. J. & CONKLIN, D. S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes & Development,* 16, 948-958.

PALNIQVIST, L., GLOVER, C. H., HSU, L., LU, M., BOSSEN, B., PIRET, J. M., HUMPHRIES, R. K. & HELGASON, C. D. (2005) Correlation of murine embryonic stem cell gene expression profiles with functional measures of pluripotency. *Stem Cells (Miamisburg),* 23, 663-680.

PESCE, M. & SCHOLER, H. R. (2001) Oct-4: Gatekeeper in the beginnings of mammalian development. *Stem Cells,* 19, 271-278.

PETERSEN, C. P., BORDELEAU, M. E., PELLETIER, J. & SHARP, P. A. (2006) Short RNAs repress translation after initiation in mammalian cells. *Molecular Cell,* 21, 533-542.

PILLAI, R. S., BHATTACHARYYA, S. N., ARTUS, C. G., ZOLLER, T., COUGOT, N., BASYUK, E., BERTRAND, E. & FILIPOWICZ, W. (2005) Inhibition of translational initiation by Let-7 microRNA in human cells. *Science,* 309, 1573-1576.

POUSTKA, A., RACKWITZ, H. R., FRISCHAUF, A. M., HOHN, B. & LEHRACH, H. (1984) Selective Isolation of Cosmid Clones by Homologous Recombination in Escherichia-Coli. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 81, 4129-4133.

RAMIREZSOLIS, R., DAVIS, A. C. & BRADLEY, A. (1993) Gene Targeting in Embryonic Stem-Cells. *Guide to Techniques in Mouse Development,* 225, 855-878.

RAND, T. A., PETERSEN, S., DU, F. H. & WANG, X. D. (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell,* 123, 621-629.

REID, L. (1990) From Gradients to Axes, from Morphogenesis to Differentiation. *Cell,* 63, 875-882.

RISAU, W., SARIOLA, H., ZERWES, H. G., SASSE, J., EKBLOM, P., KEMLER, R. & DOETSCHMAN, T. (1988) Vasculogenesis and Angiogenesis in Embryonic-Stem-Cell-Derived Embryoid Bodies. *Development,* 102, 471-478.

ROBB, G. B. & RANA, T. M. (2007) RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Molecular Cell,* 26, 523-537.

RODRIGUEZ, A., VIGORITO, E., CLARE, S., WARREN, M. V., COUTTET, P., SOOND, D. R., VAN DONGEN, S., GROCOCK, R. J., DAS, P. P., MISKA, E. A., VETRIE, D., OKKENHAUG, K., ENRIGHT, A. J., DOUGAN, G., TURNER, M. & BRADLEY, A. (2007a) Requirement of bic/microRNA-155 for normal immune function. *Science,* 316, 608-611.

RODRIGUEZ, A., VIGORITO, E., CLARE, S., WARREN, M. V., COUTTET, P., VAN DONGEN, S., SOOND, D., ENRIGHT, A., OKKENHAUG, K., DOUGAN, G., TURNER, M. & BRADLEY, A. (2007b) miR-155/bic is essential for normal immune function in mice. *Immunology,* 120, 16-17.

RUBY, J. G., JAN, C. H. & BARTEL, D. P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature,* 448, 83-U7.

SAETROM, P., SNOVE, O., NEDLAND, M., GRUNFELD, T. B., LIN, Y., BASS, M. B. & CANON, J. R. (2006) Conserved MicroRNA characteristics in mammals.

Oligonucleotides, 16, 115-144.

SEARS, R., NUCKOLLS, F., HAURA, E., TAYA, Y., TAMAI, K. & NEVINS, J. R. (2000) Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes & Development,* 14, 2501-2514.

SEGGERSON, K., TANG, L. J. & MOSS, E. G. (2002) Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. *Developmental Biology,* 243, 215-225.

SEN, G. L. & BLAU, H. M. (2005) Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nature Cell Biology,* 7, 633-U28.

SETHUPATHY, P., MEGRAW, M. & HATZIGEORGIOU, A. G. (2006) A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature Methods,* 3, 881-886.

SINGH, S. K., KAGALWALA, M. N., PARKER-THORNBURG, J., ADAMS, H. & MAJUMDER, S. (2008) REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature,* 453, 223-U11.

SINKKONEN, L., HUGENSCHMIDT, T., BERNINGER, P., GAIDATZIS, D., MOHN, F., ARTUS-REVEL, C. G., ZAVOLAN, M., SVOBODA, P. & FILIPOWICZ, W. (2008) MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nature Structural & Molecular Biology,* 15, 259-267.

SKALSKY, R. L., SAMOLS, M. A., PLAISANCE, K. B., BOSS, I. W., RIVA, A., LOPEZ, M. C., BAKER, H. V. & RENNE, R. (2007) Kaposi's sarcoma-associated herpesvirus encodes an ortholog of miR-155. *Journal of Virology,* 81, 12836-12845.

SMYTH, G. K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol,* 3, Article3.

SONG, J. J., SMITH, S. K., HANNON, G. J. & JOSHUA-TOR, L. (2004) Crystal structure of argonaute and its implications for RISC slicer activity. *Science,* 305, 1434-1437.

STARK, A., BRENNECKE, J., RUSSELL, R. B. & COHEN, S. M. (2003) Identification of Drosophila MicroRNA targets. *Plos Biology,* 1, 397-409.

TAKAHASHI, K. & YAMANAKA, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126, 663-676.

TAKAHASHI, K. T., K. OHNUKI, M. NARITA, M. ICHISAKA, T. TOMODA, K. YAMANAKA, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131, 861-872.

TAKAHASHI, N. & KOBAYASHI, I. (1990) Evidence for the Double-Strand Break Repair Model of Bacteriophage Lambda Recombination. *Proceedings of the National Academy of Sciences of the United States of America,* 87, 2790-2794.

TAY, Y. M. T., W.L.; ANG, Y.S.; GAUGHWIN, P.M.; YANG, H.H.; WANG, W.; LIU, R.; GEORGE, J.; NG, H.H.; PERERA, R.J.; LUFKIN, T.; RIGOUTSOS, I.; THOMSON A.M.; LIM, B. (2007) MicroRNA-134 Modulates the Differentiation of Mouse Embryonic Stem Cells where it Causes Post-transcriptional Attenuation of Nanog and LRH1. *Stem Cells,* 26, 17-29.

TEAM, R. D. C. (2007) R: A language and environment for statistical computing. *R Foundation for Statistical Computing.*

THOMSON, J. M., PARKER, J., PEROU, C. M. & HAMMOND, S. M. (2004) A custom microarray platform for analysis of microRNA gene expression. *Nature Methods,* 1, 47-53.

TOMARI, Y., DU, T. & ZAMORE, P. D. (2007) Sorting of Drosophila small silencing RNAs. *Cell,* 130, 299-308.

VALENCIA-SANCHEZ, M. A., LIU, J. D., HANNON, G. J. & PARKER, R. (2006) Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development,* 20, 515-524.

VANDERKROL, A. R., MUR, L. A., BELD, M., MOL, J. N. M. & STUITJE, A. R. (1990) Flavonoid Genes in Petunia - Addition of a Limited Number of Gene Copies May Lead

to a Suppression of Gene-Expression. *Plant Cell,* 2, 291-299.
VASUDEVAN, S. & STEITZ, J. A. (2007) AU-rich-element-mediated upregulation of translation by FXR1 and argonaute 2. *Cell,* 128, 1105-1118.
VASUDEVAN, S., TONG, Y. C. & STEITZ, J. A. (2007) Switching from repression to activation: MicroRNAs can up-regulate translation. *Science,* 318, 1931-1934.
VELLA, M. C., CHOI, E. Y., LIN, S. Y., REINERT, K. & SLACK, F. J. (2004) The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3 ' UTR. *Genes & Development,* 18, 132-137.
VISWANATHAN, S. R., DALEY, G. Q. & GREGORY, R. I. (2008) Selective blockade of MicroRNA processing by Lin28. *Science,* 320, 97-100.
WALSH, K. & PERLMAN, H. (1997) Cell cycle exit upon myogenic differentiation. *Current Opinion in Genetics & Development,* 7, 597-602.
WANG, W. (2005) Chromosome-specific Recessive Genetic Screen for Genes Involved in In Vitro Differentiation in Mouse Embyonic Stem Cells. *PhD Thesis.*
WANG, Y. M., MEDVID, R., MELTON, C., JAENISCH, R. & BLELLOCH, R. (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nature Genetics,* 39, 380-385.
WIGHTMAN, B., HA, I. & RUVKUN, G. (1993) Posttranscriptional Regulation of the Heterochronic Gene Lin-14 by Lin-4 Mediates Temporal Pattern-Formation in C-Elegans. *Cell,* 75, 855-862.
WU, L. G., FAN, J. H. & BELASCO, J. G. (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America,* 103, 4034-4039.
YU, J. Y., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, II & THOMSON, J. A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science,* 318, 1917-1920.
YUAN, H. B., CORBI, N., BASILICO, C. & DAILEY, L. (1995) Developmental-Specific Activity of the Fgf-4 Enhancer Requires the Synergistic Action of Sox2 and Oct-3. *Genes & Development,* 9, 2635-2645.
ZHANG, Y. M., BUCHHOLZ, F., MUYRERS, J. P. P. & STEWART, A. F. (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nature Genetics,* 20, 123-128.
ZHOU, Q., CHIPPERFIELD, H., MELTON, D. A. & WONG, W. H. (2007) A gene regulatory network in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America,* 104, 16438-16443.

Computational Resources Used:

1. http://microrna.sanger.ac.uk/sequences/
2. http://microrna.sanger.ac.uk/targets/
3. http://www.targetscan.org/
4. http://pictar.mdc-berlin.de/
5. http://www.ncbi.nlm.nih.gov/sites/entrez
6. http://gepas.bioinfo.cipf.es/
7. http://www.ebi.ac.uk/enright/software.html

Wahl et al.

Supplementary Table 1 :Differentially Expressed MicroRNAs

Day 1-0
Most up-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-491	1.29074	10.04269	12.10628	5.99E-08	1.03E-05	8.883188
mmu-miR-33	3.555742	9.991045	9.419214	8.64E-07	3.71E-05	6.210777
mmu-miR-136	1.771926	11.689124	9.256918	1.03E-06	3.99E-05	6.02782074
mmu-miR-210	1.117723	11.995751	8.948126	1.47E-06	5.15E-05	5.672013
mmu-miR-376a	2.049913	11.577381	8.175355	3.67E-06	1.09E-04	4.73498154
mmu-miR-345	1.282552	8.943806	7.734672	6.38E-06	1.54E-04	4.16895151
mmu-miR-188	1.760979	10.212154	6.787039	2.26E-05	3.80E-04	2.86742869
mmu-miR-32	2.636795	10.066506	6.425744	3.77E-05	5.39E-04	2.33920354
mmu-miR-500	1.136128	9.588544	6.060263	6.45E-05	7.78E-04	1.78617318
mmu-miR-193 (S)	1.403467	10.430948	5.725259	1.07E-04	1.15E-03	1.26254624
mmu-miR-741	1.533902	10.184236	5.101354	2.87E-04	2.26E-03	0.24498165
mmu-miR-367 (S)	1.324287	11.129411	4.941991	3.73E-04	2.82E-03	-0.02353257
mmu-miR-465-5p	1.118311	10.60768	4.86293	4.25E-04	3.07E-03	-0.15798903
mmu-miR-301b	1.10646	10.402373	4.78205	4.86E-04	3.26E-03	-0.29637332

Day 1-0
Most down-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-363	-1.771173	11.725734	-16.699125	1.72E-09	6.63E-07	12.28102568
mmu-miR-150	-1.495056	10.799975	-11.786489	8.01E-08	1.03E-05	8.59688459
mmu-miR-106a (S)	-1.123588	13.568531	-11.352586	1.20E-07	1.16E-05	8.19578637
mmu-miR-701	-1.163663	12.074118	-10.097112	4.18E-07	2.62E-05	6.94620393
mmu-miR-677	-1.248061	9.14659	-9.973384	4.75E-07	2.62E-05	6.81533709
mmu-miR-302b*	-1.714042	10.577843	-7.840927	5.57E-06	1.43E-04	4.30762482
mmu-miR-483	-1.001987	9.374949	-6.1596	5.56E-05	7.16E-04	1.93836318

Day 2-0
Most Down-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-34b	-2.177834	10.933298	16.862938	1.54E-09	1.28E-07	12.480455
mmu-miR-34c	-2.18344	10.422105	16.323653	2.21E-09	1.28E-07	12.108152
mmu-miR-210	-2.030344	11.995751	16.25428	2.32E-09	1.28E-07	12.059349
mmu-miR-491	-1.701273	10.04269	15.956797	2.85E-09	1.38E-07	11.84764
mmu-miR-199b	-1.352752	11.512946	14.996341	5.70E-09	2.37E-07	11.135946
mmu-miR-29b (S)	-1.463323	12.749397	13.19002	2.35E-08	6.67E-07	9.667103
mmu-miR-124a (S)	-1.844127	12.23974	10.874932	1.90E-07	3.29E-06	7.481269
mmu-miR-33	-3.87729	9.991045	10.271001	3.49E-07	4.99E-06	6.843326
mmu-miR-376a	-2.393702	11.577381	9.546437	7.51E-07	1.00E-05	6.034921
mmu-miR-345	-1.57444	8.943806	9.494961	7.95E-07	1.02E-05	5.975594
mmu-miR-32	-3.752903	10.066506	9.145647	1.17E-06	1.29E-05	5.566042
mmu-miR-425	-1.060969	9.618331	8.930875	1.50E-06	1.44E-05	5.308045
mmu-miR-499	-1.02883	12.614535	8.133917	3.87E-06	3.16E-05	4.307093
mmu-miR-188	-2.002417	10.212154	7.717572	6.52E-06	4.67E-05	3.755231
mmu-miR-136	-1.477156	11.689124	7.716976	6.53E-06	4.67E-05	3.754425
mmu-miR-193 (S)	-1.861691	10.430948	7.594524	7.54E-06	5.36E-05	3.588127
mmu-miR-339	-1.127981	11.08395	7.573552	7.85E-06	5.41E-05	3.559459
mmu-miR-708	-1.09331	10.225151	7.188314	1.31E-05	8.26E-05	3.023059
mmu-miR-714	-1.571761	10.212095	6.638239	2.78E-05	1.51E-04	2.223874
mmu-miR-99a	-1.036636	9.111045	6.486185	3.46E-05	1.80E-04	1.995845
mmu-miR-500	-1.168647	9.588544	6.233718	4.99E-05	2.47E-04	1.610283
mmu-miR-301b	-1.377028	10.402373	5.951429	7.59E-05	3.49E-04	1.168805
mmu-miR-367 (S)	-1.224049	11.129411	4.567921	6.98E-04	2.23E-03	-1.151043

Day 2-0
Most Up-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-363	3.156661	11.725734	-29.761907	2.39E-12	9.22E-10	18.79577
mmu-miR-302c*	1.811548	11.301051	-20.157974	2.05E-10	3.96E-08	14.51431
mmu-miR-670	1.671303	8.650449	-17.279108	1.17E-09	1.28E-07	12.759472
mmu-miR-302b*	3.573091	10.577843	-16.345191	2.18E-09	1.28E-07	12.123261
mmu-miR-370	1.606187	9.643685	-14.864768	6.29E-09	2.37E-07	11.034931
mmu-miR-323	1.270278	11.839625	-14.766776	6.77E-09	2.37E-07	10.959124
mmu-miR-106a	1.317096	13.568531	-13.307763	2.13E-08	6.67E-07	9.768568

mmu-miR-679	1.280563	10.90504	-9.322732	9.61E-07	1.12E-05	5.775197
mmu-miR-369-5p	1.123089	11.014395	-9.255076	1.04E-06	1.18E-05	5.695663
mmu-miR-376c	1.009868	12.300264	-9.011939	1.36E-06	1.35E-05	5.40599
mmu-miR-433-3p	1.100236	9.96204	-8.367469	2.91E-06	2.44E-05	4.607773
mmu-miR-302d	1.053993	13.748999	-8.098722	4.04E-06	3.18E-05	4.261237
mmu-miR-135a	1.757421	10.153996	-7.374722	1.02E-05	6.66E-05	3.284954
mmu-miR-128b	1.192515	10.778695	-6.113209	5.96E-05	2.80E-04	1.423162

Day 2-1
Most Up-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-503	1.157922	11.444441	10.118937	4.08E-07	1.79E-05	6.9410274
mmu-miR-34b	1.303917	10.933298	10.096213	4.18E-07	1.79E-05	6.9169035
mmu-miR-34c	1.19294	10.422105	8.918557	1.52E-06	4.89E-05	5.5969319

Day 2-1
Most Down-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-670	-1.789824	8.650449	-18.504462	5.41E-10	2.09E-07	13.4150001
mmu-miR-363	-1.385488	11.725734	-13.062782	2.62E-08	5.05E-06	9.7025821
mmu-miR-302c*	-1.100609	11.301051	-12.247007	5.28E-08	6.80E-06	9.0034734
mmu-miR-370	-1.14036	9.643685	-10.55368	2.62E-07	1.68E-05	7.3934036
mmu-miR-380-5p	-1.277346	9.198204	-9.764836	5.93E-07	2.29E-05	6.5595225
mmu-miR-369-5p	-1.102525	11.014395	-9.085614	1.25E-06	4.40E-05	5.7928742
mmu-miR-7b	-1.17486	10.1412	-8.738854	1.87E-06	5.55E-05	5.382763
mmu-miR-302b*	-1.859049	10.577843	-8.504264	2.47E-06	6.35E-05	5.0977582
mmu-miR-433-3p	-1.083994	9.96204	-8.243946	3.38E-06	8.05E-05	4.7741113
mmu-miR-135a	-1.151817	10.153996	-4.833407	4.46E-04	2.82E-03	-0.2805627

Day 3-2
Most Up-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-184	1.081874	10.078041	9.904063	5.12E-07	7.23E-05	6.7678257
mmu-miR-7b	1.283734	10.1412	9.54868	7.50E-07	7.23E-05	6.3961849
mmu-miR-150	1.085678	10.799975	8.559097	2.31E-06	1.78E-04	5.2897616

Day 3-2
Most Down-regulated

ID	logFC	AveExpr	t	P.Value	adj.P.Val	B
mmu-miR-302c*	-1.043925	11.301051	-11.616253	9.37E-08	3.62E-05	8.3905014
mmu-miR-367	-2.574884	11.129411	-9.608983	7.02E-07	7.23E-05	6.4601629
mmu-miR-705	-1.070114	8.284472	-7.034928	1.61E-05	5.43E-04	3.3543973
mmu-miR-712*	-1.218976	12.15739	-6.556359	3.13E-05	6.88E-04	2.682368
mmu-miR-714	-1.501039	10.212095	-6.33955	4.27E-05	8.25E-04	2.3671378
mmu-miR-376a	-1.284036	11.577381	-5.120924	2.78E-04	2.79E-03	0.4675334
mmu-miR-32	-1.830012	10.066506	-4.459651	8.39E-04	5.45E-03	-0.6509114

(S): expressed in stem cells (microRNA.org)
*: MicroRNA star sequence
logFC: log fold change in gene expression
AveExprs: Average Expression across replicates
t: t-statistic value
P-value: probability value of t-statistic
Adj.P.Val: adjusted p-value after correction for multiple hypothesis testing
B: B statistics(the log odds ratio that the gene is differentially expressed)

APPENDIX

miRanda microRNA-mRNA alignments

Klf4

mmu-miR-29a:

```
AUUGGCUAAAGU-CUACCACGAu
|||| |   |   | |||||||||
TAACAGCCTAAATGATGGTGCTt
```

cMyc

mmu-miR-196a:

```
GGGUUGUUGU--ACUUUGAUGGAu
::|||:| |:|   |   ||||:|||
TTCAGCCATAATTTTAACTGCCTc
```

mmu-miR-125b:

```
ucgaggguucucGGAUUGGGCa
             ||||||:||
-------actgaCCTAACTCGa
```

mmu-let-7d:

```
UUGAUACGUUGGAUGAUGGAGa
|:| ||     :: |||:||||
AGCCATAATTTTAACTGCCTCa
```

mmu-miR-132:

```
GCUGGUACCGACA-UCUGACAau
:|:|:: |||| | :||||||
TGGCTGGGGCTTTGGGACTGTaa
```

Nanog

mmu-miR-30e:

```
gaAGGUCAGUUCCUACAAAUGU
  ||::   || |||||||||:|
taTCTGTAGAAAGATGTTTATA
```

Sox2

mmu-miR-21:

```
gUUGUA-GUCA-GACUAUUCGAU
  ||::| || | :| |||||||:
```

```
caAATGTCCATTGTTTATAAGCTG
```

mmu-miR-126-3p:

```
gcguaaUAAUGAGUGCCAUGCu
      | | || ::||||||
nnnnaaAGTTCTAGTGGTACGt
```

Sall4

mmu-miR-15:

```
GUGUUUGGUAAUACACGACGAu
|||:|:   |  | |||||||
CACGAGAGTTGCTATGCTGCTt
```

mmu-miR-103:

```
aguaucgggacAUGUUACGACGA
          |:| ||||||||
atcacgagagtTGCTATGCTGCT
```

mmu-mmiR-107:

```
acuaucgggacAUGUUACGACGA
          |:| ||||||||
atcacgagagtTGCTATGCTGCT
```

mmu-miR-16:

```
GCGGUUAUAAAUGCACGACGAu
| | |: :||  : |||||||
CACGAGAGTTGCTATGCTGCTt
```

mmu-miR-322:

```
agGUUUUGUACUUAACGACGAc
  |:|:|  ||   |||||||
caCGAGAGTTGCTATGCTGCTt
```

Tcf7

mmu-miR-21:

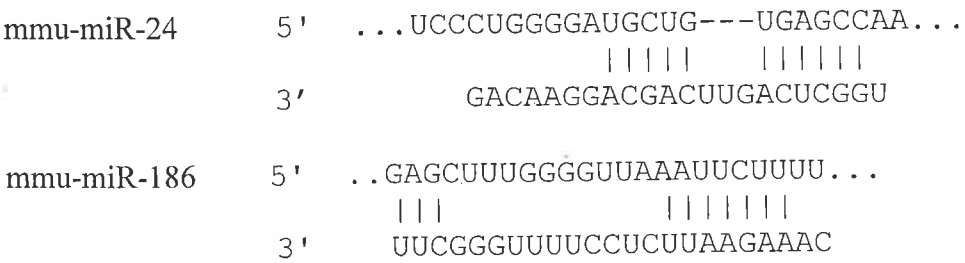
```
aGUUGUAGUCAGACUAUUCGAU
 ||:|   | || | |||||||:
cCAGCTGCTGT-TCATAAGCTG
```

mmu-miR-24:

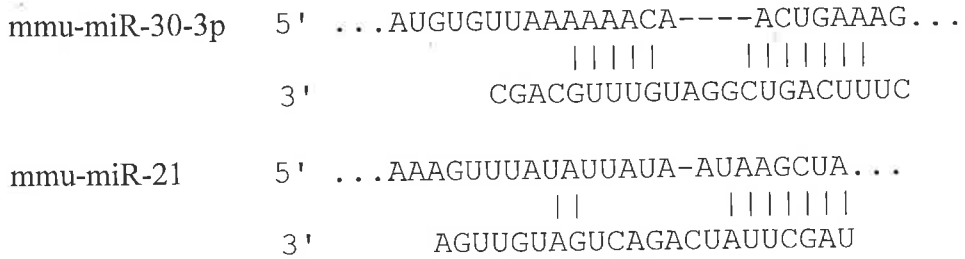
```
gacAAGGACGACUUGACUCGGU
   |||| ||| | |||||:||
catTTCCAGCTCATCTGAGTCA
```

TargetScan microRNA-mRNA Alignements

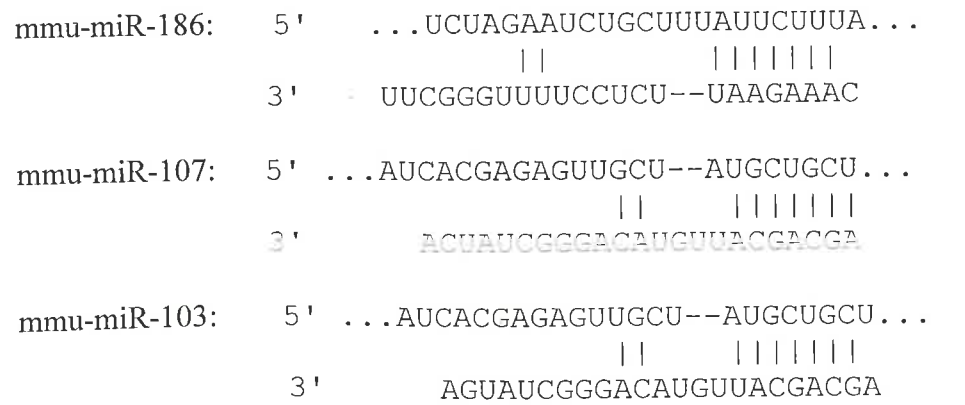
Pou5f1



Nanog



Sall4



```
mmu-miR-15b:     5' ...UCACGAGAGUUGCUAUGCUGCUU...
                                ||   |||||||
                 3'      ACAUUUGGUACUACACGACGAU

mmu-miR-136:     5' ...AAAAAAAAGAUGAAA--AAUGGAGG...
                               |||  |||||||
                 3'       AGGUAGUAGUUUUGUUUACCUCA
```

Stat3

Conserved

```
mmu-miR-106b: 5' ...GGAACUCCUGGCUCUGCACUUUC...
                           |||     |||||||
              3'     UAGACGUGACAGU—CGUGAAAU

mmu-miR-20b:  5'  ...GGAACUCCUGGCUCUGCACUUUC...
                           ||      |||||||
              3'     AUGGACGUGAUACU—CGUGAAAC

mmu-miR-20a:  5'    ...GGAACUCCUGGCUCUGCACUUUC...
                           ||      |||||||
              3'     GAUGGACGUGAUAUU—CGUGAAAU

mmu-miR-93:   5' ...GGAACUCCUGGCUCU---GCACUUUC...
                          ||||        |||||||
              3'        GAUGGACGUGCUUGUCGUGAAAC
```

Poorly Conserved

```
mmu-miR-106b: 5' ...CUUUGGGCAAUCUGGGCACUUUU...
                           |||     |||||||
              3'       UAGACGUGACAGUCGUGAAAU

mmu-miR-20b:  5' ...CUUUGGGCAAUCUGGGCACUUUU...
                           |||     |||||||
              3'      AUGGACGUGAUACUCGUGAAAC

mmu-miR-20a   5' ...CUUUGGGCAAUCUGGGCACUUUU...
                           |||     |||||||
              3'     GAUGGACGUGAUAUUCGUGAAAU

mmu-miR-93:   5' ...CUUUGGGCAAUCUGGGCACUUUU...
                           |||     |||||||
              3'     GAUGGACGUGCUUGUCGUGAAAC
```

Appendix Table 1:Probes detected on Illumina microRNA and mRNA Bead Arrays				
	at p<0.01		at p< 10-18	
	Probes detected	Probes absent	Probes Detected	Probes Absent
MicroRNA				
Day 0	322	58	302	78
Day 3	312	68	309	71
Expressed at Day 0 and Day 3	305	75	293	87
Messenger RNA				
Day 0	11798	34830	7712	38916
Day 3	11733	34895	7792	38836
Expressed at Day 0 and Day 3	10533	36095	7396	39252