

**Investigation of a developmentally linked transcriptional gene
silencing mechanism involving an antisense RNA**



Robert J S Murray

**A thesis submitted for the degree of Doctor of Philosophy
of the University of Leicester**

2009

Abstract

Investigation of a developmentally linked transcriptional gene silencing mechanism involving an antisense RNA

An increasing number of RNA targeted gene repression mechanisms are being discovered operating in mammalian cells. Antisense-RNA mediated silencing of CpG island-associated genes is not limited to imprinting and X inactivation, but can occur at autosomal non-imprinted loci leading to disease. An antisense silencing phenomenon observed in a patient with an inherited form of anaemia was recently recreated in differentiating mouse embryonic stem cells, demonstrating that the silencing and methylation of the oppositely transcribed tissue specific alpha globin CpG island exclusively occurs in the presence of antisense RNA.

The aim of this thesis is to use differentiating mouse embryonic stem cells to further characterise this silencing mechanism through analysis of the histone modifications associated with repression of the alpha globin gene upon differentiation, and to further define the developmental period in which repression is established. Also, the requirement for antisense RNA transcription through the alpha globin gene is examined through insertion of a transcriptional terminator to truncate antisense RNA transcription.

The gene silencing mechanism is tested with other genes in place of the alpha globin gene to investigate whether repression is limited to the alpha globin gene or is also applicable to other tissue specific or ubiquitously expressed genes. Analysis of the tissue specific gene *MYOD* and the ubiquitously expressed gene *UBC* demonstrated that the silencing mechanism could effect other genes in a similar manner, though not all genes were susceptible to repression as the ubiquitously expressed gene *ACTB* remained unsilenced by antisense RNA expression.

These observations establish that this silencing mechanism can play a more general role in repression, thereby suggesting its potential role as a constituent of the multivariate repressive pathways within the mammalian cell

Acknowledgement

There is a long list of people that should be thanked for their involvement, support and friendship during the course of my PhD. Without them it could not have been done.

First and foremost I must thank my supervisor Dr Cristina Tufarelli for allowing me to explore the fascinating world of RNA mediated repression. Without her unwavering enthusiasm and support over the last four years this somewhat lengthy tome could never have been written. I hope this is only the first thesis of many to grace her shelves. I feel I also owe a big thank you to Dr Hazel Cruickshanks, the post doc in our small group. The help and advice she provided throughout my PhD was utterly invaluable, and the tenacity she showed towards her work provided a wonderful inspiration.

My family have been unceasingly supportive, with my time spent at home allowing me to gain some perspective when I was getting too obsessive and stressed about work. Truly there is no greater catharsis than bashing seven shades of hell out of overgrown rhododendrons when you can't get your damned cells to grow properly.

I must also thank the many friends I have made during my PhD, mostly other PhD students undergoing the same experiences - we helped each other through the bad times.

Chapter 1 Introduction.....	11
1.1 DNA methylation	12
1.2 Mammalian DNA methyltransferases	16
1.3 CpG islands	20
1.4 Tissue specific methylated regions	23
1.5 Histone modifications.....	24
1.5.1 Bivalent Domains.....	28
1.5.2 H3K9 trimethylation	28
1.5.3 Acetylation	30
1.5.4 Ubiquitination.....	31
1.5.5 Histone modifications states define a genes transcriptional state	32
1.6 Mammalian H3K9me histone methyltransferases	33
1.7 Polycomb Repressor Complexes	38
1.7.1 PRC roles in repression beyond 'classical' bivalent domains	41
1.8 RNA and gene regulation.....	44
1.8.1 Transcription of non-coding RNA within the cell.....	44
1.8.2 Conservation and function of non-coding RNAs	45
1.8.3 Transcription of non-coding RNAs at transcriptional start sites	48
1.8.4 Small double stranded RNAs and argonaute family proteins.....	49
1.9 Non-coding RNAs and repressive nuclear domains	52
1.9.1 Imprinting.....	53
1.9.1.1 The Igf2r cluster	55
1.9.1.2 The Kcnq1 cluster	57
1.9.2 X inactivation	60
1.9.2.1 The X inactivation control centre.....	61
1.9.2.2 The RepA repeats	62
1.9.1.3 Initiation of X inactivation	64
1.9.1.4 Maintenance of X inactivation	65
1.10 Pluripotency factors in pre-implantation development.....	66
1.11 A novel case of gene repression by CpG island methylation	69
1.12 Aims of the thesis.....	73
Chapter 2 Materials and Methods	77
2.1 Chemicals and reagents	
2.2 Bacterial cultures	
2.2.1 Storage and revival of bacterial strains	//
2.2.2 Culturing Bacterial cells for miniprep	77
2.2.3 Extraction of DNA from Bacterial Strains.	78
2.2.3.1 Miniprep.....	78
2.2.3.2 Maxiprep.....	79

2.3 Phenol/chloroform extraction of DNA	79
2.4 Engineering of DNA constructs.....	80
2.4.1 Restriction digest of DNA	80
2.4.2 Amplification of large regions by PCR.....	80
2.4.2 Blunt-ending of a linearised plasmid	81
2.4.3 Dephosphorylation of plasmid DNA	81
2.4.4 Gel electrophoresis and purification	81
2.4.5 Ligation of insert into vector	82
2.4.6 Transformation of plasmid into bacterial cells	82
2.4.7 PCR screening of transformed bacterial cells	83
2.5 Growth and maintenance of mESC lines and EBs	83
2.5.1 Growth and maintenance of mESCs	83
2.5.2 Passaging of ESCs by trypsinisation and disaggregation.....	85
2.5.3 Revival of ESCs from frozen down aliquots	85
2.5.4 Long term storage of ESC lines	85
2.5.5 Differentiation of ES cells into Embryoid Bodies	86
2.5.6 Electroporation of DNA constructs into ESCs	86
2.5.6.1 Electroelution of DNA construct for electroporation	86
2.5.6.2 Electroporation by BioRad Gene Pulser Xcell (165-2660).....	87
2.5.6.3 Electroporation by Eppendorf Multiporator (4308 000.015)	88
2.5.6.4 Selection of ESC colonies for further culturing	88
2.5.7 Extraction of DNA from ESCs and EBs.....	89
2.5.7.1 Extraction of DNA using phase lock tubes.....	90
2.5.8 Extraction of RNA from ESCs and EBs.....	90
2.5.9 Media used in the propagation of ES cells and EBs	92
2.5.9.1 Complete ES Media	92
2.5.9.2 Freezing media	92
2.5.9.3 IMDM+P/S+MTG.....	92
2.5.9.4 Preparation Media.....	92
2.5.9.5 Differentiation Media 1	92
2.5.9.6 Differentiation media 2 (100ml).....	93
2.5.9.7 Gelatin 0.1% (100ml)	93
2.5.10 Cell Lysis buffers	93
2.6 Polymersase chain reaction (PCR).....	93
2.6.1 PCR Buffers.....	94
Table of standard PCR primers2.6.2 Electrophoresis of PCR products.....	96
2.6.2 Electrophoresis of PCR products.....	96
2.6.3 Methylation sensitive PCR (msPCR).....	96
2.6.4 Methylation sensitive multiplex PCR	97
2.6.5 Multiplex PCR for pluripotency factors	98

2.6.6 Realtime PCR	98
2.6.6.1 Creation of a Standard Curve	100
2.6.6.2 Taqman probe based detection	101
2.6.6.3 SYBR Green dye based detection	102
2.6.6.4 SYBR green analysis of ChIP samples	103
2.7 Reverse transcription	104
2.8 Crosslinking of cells for Chromatin Immunoprecipitation	104
2.8.1 Preparation of ES cells for crosslinking	104
2.8.2 Differentiation of ES cells to EBs and preparation for crosslinking	105
2.8.3 Crosslinking of cells for ChIP	105
2.9 Chromatin Immunoprecipitation using the EZ-Magna ChIP A Kit (Millipore catalog #17-408).	106
2.9.1 Cell preparation for sonication	106
2.9.2 Sonication of cells using Diagenode Biorupter™ 200	106
2.9.3 Immunoprecipitation (IP)	107
2.9.4 Clean-up of immunoprecipitated samples	107
2.9.5 Analysis of control samples	108
2.9.6 Antibodies used for Chromatin Immunoprecipitation	108
2.10 Southern Blotting	109
2.10.1 Preparation of nylon membrane	109
2.10.2 Hybridization of membrane	110
2.10.3 End labeling of DNA probe/ DNA Ladder	111
2.10.4 Random prime labeling of DNA probe	111
Chapter 3 Examination of the timing of repression of the alpha globin gene, and the requirement for antisense RNA in the silencing process	112
3.1 Introduction	112
3.2 Methods	115
3.2.1 Growth and analysis of 2A3 time course	115
3.2.2 Creation of the Zfx antisense Stop construct	115
3.2.3 Electroporation of Zfx antisense Stop construct into ESCs and screening for successful incorporation	116
3.2.4 Analysis of cell lines	119
3.3 Results	122
3.3.1 Analysis of epigenetic changes at the alpha globin gene during 2A3 differentiation	122
3.3.1.1 Expression profiling	122
3.3.1.2 DNA methylation profiling	126
3.3.1.3 Histone modification profiling	129
3.3.2 Investigating the requirement for AS-RNA	137
3.4 Discussion	150

3.4.1 Silencing at the alpha globin gene	150
3.4.2 Prevention of antisense RNA through the alpha globin gene	156
Chapter 4 Investigation of molecular mechanisms potentially involved in the antisense RNA initiated silencing effect	158
4.1 Introduction	158
4.1.1 Structure of AGO proteins	159
4.1.2 AGO1 and TGS in <i>S.pombe</i>	160
4.1.3 AGO1 and transcriptional gene silencing in mammals	162
4.2 Materials and Methods.....	167
4.2.1 Screening for small RNAs	167
4.2.1.1 RNA preparation	167
Table 4.14.2.2.2 Transfer by Electroblothing	169
4.2.2.2 Transfer by Electroblothing	170
4.2.2.3 Chemical Crosslinking.....	170
4.2.2.4 Hybridisation of membrane	171
4.2.2 Engineering of Ago1 Kd cell lines.....	171
4.2.2.1 Selection of siRNA target sequences	171
4.2.2.2 Cloning of target sequences into expression vector	172
4.2.4.3 Electroporation and selection of stable clones.....	176
4.2.2.4 Analysis of Ago1 shRNA stable clones	178
4.3 Results	179
4.3.1 Detection of small RNAs in differentiating 2A3 cells	179
4.3.2 Analysis of Ago1 knockdown cell lines in the 2A3 background.....	185
4.4 Discussion.....	199
Chapter 5 Role of antisense RNA in silencing of the tissue specific gene MYOD1	205
5.1 Introduction	205
5.2 Methods	207
5.2.1 Construction of the pZERO MYOD + LUC7L plasmid.....	207
5.2.2 Electroporation of MYOD construct into ESCs and screening for	211
5.2.3 Analysis of cell lines	211
5.3 Results	213
5.4 Discussion.....	223
Chapter 6 Role of antisense RNA in silencing ubiquitously expressed genes	228
6.1 Introduction	228
6.2 Methods	231
6.2.1 Engineering of constructs for electroporation into ESCs.....	231
6.2.1.1 Construction of the <i>ACTB</i> + <i>LUC7L</i> plasmid	231
6.2.1.2 Construction of the <i>UBC</i> + <i>LUC7L</i> plasmid	238
6.2.2 Electroporation of Constructs into ESCs and screening for successful incorporation	241

6.2.2.1 Electroporation of the <i>ACTB</i> construct into mESCs	241
6.2.2.2 Electroporation of the <i>UBC</i> construct into ESCs	242
6.2.3 Analysis of cell lines	242
6.3 Results	245
6.3.1 Analysis of ACTB cell lines.....	245
6.3.2 Analysis of <i>UBC</i> cell lines.....	256
6.4 Discussion.....	269
Chapter 7 Final Discussion	276
7.1 Antisense RNA mediated silencing.....	276
7.2 Involvement of H3K27me3 in gene silencing.....	277
7.3 Small RNAs as mediators of transcriptional gene silencing	283
7.4 An inducible silencing system	285
7.5 A useful silencing model	286
7.6 Concluding remarks	288
Appendix A: Testing of antibodies for specificity and enrichment in mESCs and EBs	290
Appendix B: Calibration of sonication for ChIP	295
Appendix C: comparing IgG enrichment to histone modification specific enrichment for the different primer pairs used during ChIP experiments.....	300
References	302

List of abbreviations

°C	degrees centigrade
µl	microlitre
AIDA	Advanced Image Data Analyzer
bHLH	basic helix loop helix
bp	base pair
BSA	bovine serum albumin
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
cpm	counts per minute
d	day
d.p.c	days post conception
Dcr	dicer
dCTP	2'-deoxycytosine 5'-triphosphate
DEPC	diethylpyrocarbonate
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTPs	2'-deoxyribose 5'-triphosphates

dsRNA	double stranded RNA
EBs	embryoid bodies
EDTA	ethylenediaminetetraacetic acid
ENCODE	ENCyclopedia Of DNA Elements
ENDO	endogenous control
ESCs	embryonic stem cells
EtBr	Ethidium Bromide
FANTOM	Functional Anotation of the Mammalian Genome
FCS	fetal calf serum
FISH	fluorescence in situ hybridization
G418	geneticin
GOI	gene of interest
HCP	high CpG content promoter
HDAC	histone deacetylase
HMTase	histone methyltransferase
Hrs	hours
ICM	inner cell mass
ICR	imprint control region
Kb	kilobases
Kd	knockdown
KO	knockout
LCP	low CpG content promoter
M	molar
MEF	mouse embryonic fibroblast
mESCs	mouse embryonic stem cells
mRNA	messenger RNA
ncRNA	non coding RNA
NEB	new england biolabs
ng	nanogram
NPC	neural progenitor cells
O.N	over night
P/C	phenol/chloroform
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
pMol	pico mol
Pol II	DNA dependent RNA polymerse II
PRC	polycomb repressor complex
PRE	polycomb repressor element
PTGS	postranscriptional gene silencing
RE/REs	restriction endonucleases
RISC	RNA induced silencing complex
RITS	RNA induced transcriptional silencing complex
RNA	ribose nucleic acid
RT-PCR	reverse transcription polymerase chain reaction
ssRNA	single stranded RNA
STDEV	standard deviation
T-DMR	tissue specific differentially methylated region
TE	Tris-EDTA
TGS	transcriptional gene silencing
tiRNAs	transcription associated RNAs

TRC	trithorax repressor complex
TRE	trithorax repressor element
TSS	transcriptional start site
V	volt
w/v	weight per volume
Wt	wildtype
X _A	active X
X _i	inactive X
βME	betamercaptoethanol
μM	micromolar

Chapter 1 Introduction

Following the success of the various genome sequencing projects, it is clear that knowledge of the genetic code is just the first of many steps towards a true appreciation of the complexities of the genome and its regulation, with epigenetics representing the next major challenge to our understanding of how the genome is regulated. Epigenetic modifications, which allow a large degree of control over a genes transcriptional state and are stably inherited through cell division without alteration of the DNA sequence, effect all aspects of gene expression and repression, and are central to genomic control during development and cellular differentiation. Genetic material in eukaryotes is packaged within the cell nucleus through the interaction of histone and non-histone proteins to form a dynamic structure known as chromatin. Epigenetic modifications to histone proteins alter chromatin structure, and can make the underlying DNA sequence easier or harder to access - thereby altering a genes transcriptional potential. The DNA base cytosine can become modified by the addition of a methyl group, and this form of DNA methylation is an important component of epigenetic regulation in mammalian cells.

Histone modifications and DNA methylation, along with the enzymes responsible for addition of these epigenetic marks, form a large and complex system of genomic control, supported by a web of interactions between the various components of the epigenome. In the following sections, the role and distribution of these epigenetic marks - and the enzymes responsible for their regulation - will be discussed, along with the emerging role played by non-coding RNA in genomic regulation.

1.1 DNA methylation

DNA methylation is essential for genome stability, as it plays a crucial role in maintaining the transcriptional silence of the many repeat sequences and parasitic elements that make up almost 40% of the mammalian genome (Yoder *et al.* 1997). DNA methylation has also been increasingly shown to play a central and vital role in the repression of developmental genes during early development (Epsztejn-Litman *et al.* 2008, Tachibana *et al.* 2008), as well as during imprinting, in both the germline and somatic cells, and in gene repression on the silenced X chromosome during dosage compensation (Lee & Jaenisch 1997).

DNA methylation is the transfer of a methyl group from the universal methyl donor, S-adenosyl-L-methionine (AdoMet), to the 5' carbon position of cytosine, creating 5-methylcytosine (Tost 2008). In mammals, methylation of cytosine mainly occurs within the dinucleotide sequence CpG, with cytosine immediately 5' to guanine. Methylation at CpG dinucleotides in mouse embryonic stem cells (ESCs) has been shown to be approximately 60%, whereas in adult tissues levels can be as high as 80% (Ramsahoye *et al.* 2000).

DNA methylation can act directly to block binding of transcription factors (Bell & Felsenfeld 2000), though it is thought that its main mode of action is through promoting the recruitment of a myriad of other repressive factors that mediate local chromatin changes (Miranda & Jones 2007). Methylation of DNA appears to be directly antagonistic to certain histone modifications that promote an open and accessible form of chromatin (Okitsu & Hsieh 2007), and has also been

shown to alter nucleosome occupancy thereby blocking transcription factor and Pol II binding (Patel *et al.* 1997).

Methylated CpG dinucleotides create binding sites for the Methyl Binding Domain (MBD) family of proteins, which act as platforms for the recruitment of chromatin modifiers and transcriptional repressors (Miranda & Jones 2007, Clouaire & Stancheva 2008). Six members of the MBD family have been identified so far; MeCP1, MeCP2, MBD1, MBD2, MBD3, and MBD4 (Meehan *et al.* 1989, Lewis *et al.* 1992, Hendrich & Bird 1998). The interaction between methylated DNA and the MBD domain was studied in MBD1, where it was shown that five core amino acids of the MBD domain form a hydrophobic patch that specifically recognises the methyl groups of the CpG sequence allowing interaction with DNA via the major groove where the two methyl groups of a reciprocally methylated CpG dinucleotide are exposed on the exterior of the DNA double helix (Ohki *et al.* 2001).

MeCP2 is the best characterised member of the MBD family, and is capable of binding to a single reciprocally methylated CpG dinucleotide in conjunction with an AT rich run of greater than four nucleotides (Klose *et al.* 2005, Nan *et al.* 1996). Upon binding, MeCP2 can recruit the HDAC-containing Sin3A complex (Nan *et al.* 1998, Jones *et al.* 1998) resulting in histone deacetylation, and also the SWI/SNF ATPase-dependent remodelling complex capable of inducing a condensed chromatin state (Harikrishnan *et al.* 2005). MeCP2 is also capable of directly affecting chromatin structure, with *in vitro* studies demonstrating that purified MeCP2 can induce a compact chromatin state (Nikitina *et al.* 2007).

MBD3 cannot directly bind to methylated DNA (Hendrich & Bird 1998), but interacts with HDACs and members of the SWI/SNF family of ATP dependent helicases to form the nucleosome remodelling and histone deacetylation (NuRD) complex (Zhang *et al.* 1997, Zhang *et al.* 1999, Wade *et al.* 1999) which is targeted to methylated CpG sites through interaction with MBD2 and causes both deacetylation and chromatin condensation (Clouaire & Stancheva 2008).

MBD1 and MeCP1 both occur in the same HDAC-containing complex (Cross *et al.* 1997, Ng *et al.* 1999) and have been shown to mediate DNA methylation dependent repression at euchromatic regions, though the complex also shows some localisation to regions of pericentric heterochromatin (Fujita *et al.* 1999). MeCP1 requires ten or more methylated CpG dinucleotides in order to mediate repression (Meehan *et al.* 1989).

Repressive factors are also recruited directly through interaction with the DNA methyltransferases responsible for the creation and maintenance of methylated DNA (Tachibana *et al.* 2008, Vire *et al.* 2006, Esteve *et al.* 2006).

In mammals, DNA methylation levels are high in both the sperm and the egg at fertilization, but global methylation levels decrease during the first few days of development, until they reach their lowest levels around blastocyst implantation, whereupon a wave of *de novo* methylation occurs within the inner cell mass, giving rise to lineage specific methylation patterns that are maintained in differentiated tissues (Fulka *et al.* 2008, Li 2002).

Demethylation of the paternal genome is an active process, occurring prior to replication, when the paternal genome is unpackaged and protamines are replaced with acetylated histones. Demethylation of the maternal genome occurs in a passive manner following Zygote formation. Germline imprints in both the paternal and maternal genomes are maintained during de-methylation (Fulka *et al.* 2008, Li 2002, Reik *et al.* 2001). In the mouse, methylation in the embryo begins upon implantation at 3.5dpc, with DNA methylation levels increasing rapidly in the primitive ectoderm that gives rise to all the tissues of the embryo. DNA methylation levels in the trophoblast and the primitive endoderm, which give rise to, respectively, the yolk sac and the placenta, show only a modest increase in methylation levels (Li 2002). The dynamic process of de-methylation and re-methylation in the early embryo is shown in Figure 1.1.

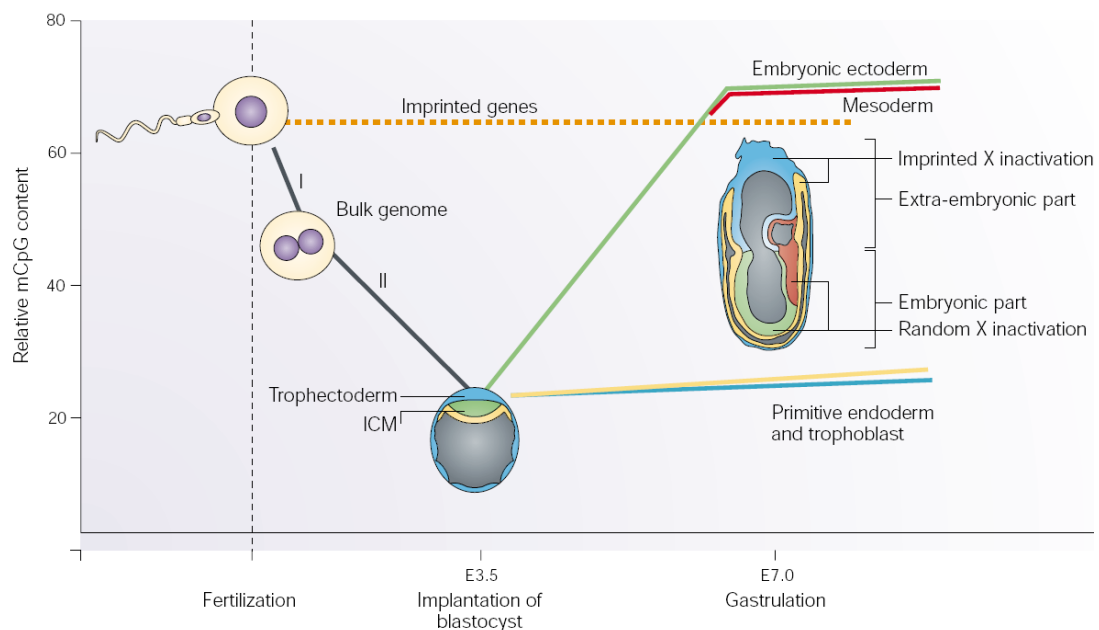


Figure 1.1 DNA methylation patterns in the developing mouse embryo. Upon fertilization, the paternal genome is actively demethylated prior to replication (I), followed by maternal genome demethylation in a passive manner (II). Following implantation a wave of *de novo* methylation within the primitive ectoderm re-establishes DNA methylation (green line and tissue). DNA methylation is not re-established in non-embryonic tissues (blue/yellow lines and tissues). Germ line imprints are maintained throughout demethylation (orange dashed line). Figure and information from Li 2002.

1.2 Mammalian DNA methyltransferases

Methylation at CpG dinucleotides is mediated by the DNA methyltransferase (Dnmt) family of proteins. Dnmt1 is the main maintenance methyltransferase, while *de novo* DNA methylation is carried out by Dnmt3a and Dnmt3b, aided by Dnmt3L (Tost 2008, Okano *et al.* 1999, Ooi *et al.* 2007). The DNA methylation machinery is a fully integrated component of the wider transcriptional repression machinery, and during both *de novo* and maintenance methylation the Dnmt family are capable of interaction with each other, and a wide variety of proteins involved in different aspects of repression, with interactions frequently enhancing methyltransferase ability or ensuring localised enrichment (Tost 2008, Vire *et al.* 2006, Ooi *et al.* 2007, Chedin *et al.* 2002, Fuks *et al.* 2003, Geiman *et al.* 2004, Li *et al.* 2007, Robertson *et al.* 2000a). This high level of cooperation serves to highlight both the interdependency of these different repressive mechanisms, and the redundancy that has evolved to ensure that gene repression is stable and enduring.

Dnmt1 is a ubiquitously expressed protein with elevated expression levels in S phase of the cell cycle during DNA replication (Robertson *et al.* 2000b), where it preferentially binds hemi-methylated DNA that results from the semi-conservative replication of methylated regions. Dnmt1 is essential for DNA methylation maintenance, with knockout of Dnmt1 resulting in the almost complete loss of DNA methylation following replication (Li *et al.* 1992). Dnmt1 is the predominant methyltransferase in somatic cells, where the levels of other methyltransferases are low (Schermele *et al.* 2007).

Dnmt1 plays an important role in maintenance of repression by interacting with a range of chromatin modifying enzymes ensuring that the full suite of repressive marks are present at a silenced loci. It directly interacts with G9a, Suv39h, HP1, HDAC1 (Esteve *et al.* 2006, Fuks *et al.* 2000, Smallwood *et al.* 2007), and multiple components of Polycomb Repressor Complex 2 (PRC2) (Vire *et al.* 2006), thereby enhancing, respectively, methylation of H3K9 at both euchromatic and heterochromatic regions, repressive chromatin remodeling, de-acetylation of histones, and enrichment for H3K27me3.

During DNA replication, Dnmt1 is recruited to replication foci through interaction with the protein Np95, which preferentially binds hemi-methylated DNA, and is essential for maintenance of DNA methylation levels (Sharif *et al.* 2007). Dnmt1 is also recruited through dynamic interaction with the homotrimeric ring protein PCNA, which encloses the DNA double helix and acts as the scaffold for tethering of the replication machinery (Schermelleh *et al.* 2007). PCNA interaction helps to ensure localized enrichment for Dnmt1, though interaction is dynamic to prevent the highly processive replication machinery from becoming stalled while Dnmt1 catalyses the comparatively slow addition of methyl groups to CpG dinucleotides. It is during replication that Dnmt1 has been shown to interact with G9a, causing its recruitment to replication foci. In contrast to this, Suv39h interaction occurs in a post-replicative manner, suggesting that maintenance of H3K9me levels at heterochromatic regions occurs after the replication machinery has passed through (Esteve *et al.* 2006).

DNA methylation patterns are established after implantation of the blastocyst during early embryogenesis in a lineage specific manner, beginning in the inner cell mass (ICM), with levels of methylation increasing rapidly in the primitive ectoderm that gives rise to all the tissues of the embryo (Li 2002). Dnmt3a and Dnmt3b are the principal *de novo* methyltransferases, and have both overlapping and non-overlapping roles in establishment of DNA methylation patterns. Both are essential for development (Okano *et al.* 1999).

Dnmt3b is expressed earlier than Dnmt3a in the developing embryo, with expression detectable in both the ICM and subsequently in both the epiblast and developing ectoderm. Expression levels reduce about the same time as Dnmt3a levels become detectable, and Dnmt3b levels generally remain lower than those of Dnmt3a. Both are expressed in adult tissues (Tost 2008, Li 2002, Okano *et al.* 1998). Dnmt3a has been shown to be required for methylation at both imprinting loci and during X inactivation, while Dnmt3b is responsible for methylation of minor satellite repeat methylation (Okano *et al.* 1999) and is generally associated with heterochromatic regions (Geiman *et al.* 2004). Dnmt3a and Dnmt3b also function together as components of the same complex in a synergistic manner, and are essential for repression of pluripotency factors during early development (Li *et al.* 2007).

It has been suggested that Dnmt3a and Dnmt3b have a role in maintenance of DNA methylation, as depletion of both in ESCs results in the eventual loss of DNA methylation patterns after prolonged culture, and it is also possible that Dnmt3a and Dnmt3b ensure fidelity in DNA methylation maintenance by filling

in any gaps left by the Dnmt1 enzyme (Chen *et al.* 2003). While this has been demonstrated in ESCs, where Dnmt3a and Dnmt3b expression levels are both high, it remains to be seen if such a mechanism exists in somatic tissues where expression levels are low.

Dnmt3L shares significant homology with both Dnmt3a and Dnmt3b but lacks a core catalytic domain and is therefore incapable of methyltransferase activity (Ooi *et al.* 2007). Although Dnmt3L cannot bind DNA directly (Suetake *et al.* 2004) it is essential for proper development, and knockout causes loss of imprinting that is lethal in female embryos and causes severe defects in male germline development (Bourc'his *et al.* 2001). Dnmt3L occurs in the same complex as Dnmt3a and Dnmt3b (Li *et al.* 2007), and has been shown to stimulate both DNA binding and the rate of AdoMet binding for Dnmt3a, both *in vitro* and *in vivo*. Dnmt3L is also capable of Dnmt3b stimulation *in vitro*, though this was not observed *in vivo* (Chedin *et al.* 2002, Gowher *et al.* 2005).

The Dnmt3 proteins have been shown to interact with each other, and also with other components of the various transcriptional silencing pathways. In a similar manner to Dnmt1, both Dnmt3a and Dnmt3b interact with EZH2, EED and SUZ12, all components of the PRC2 complex (Vire *et al.* 2006), and both also interact with Suv39h1, G9a, HDAC1 and Hp1 proteins, with Dnmt3a associating with Hp1 β , while Dnmt3b is associated with both Hp1 α and Hp1 γ (Tachibana *et al.* 2008, Fuks *et al.* 2003, Geiman *et al.* 2004). Dnmt3L, while incapable of directly binding DNA, acts as an intermediate capable of binding to all four core

histones (Ooi *et al.* 2007), as well as directly binding to HDAC1 (Aapola *et al.* 2002).

1.3 CpG islands

The distribution of CpG dinucleotides within the genome is non-linear, with the majority of the genome having a low density of mostly methylated CpG dinucleotides, but with regions of high CpG density, CpG islands, which occur at promoter regions for the majority of housekeeping genes, and almost half of genes expressed in a tissue specific manner (Jones 1999). Unlike regions of low CpG density, CpG islands are nearly always unmethylated irrespective of gene expression (Meissner *et al.* 2008, Weber *et al.* 2007, Irizarry *et al.* 2009, Shen *et al.* 2007). Methylation of CpG dinucleotides has a bimodal distribution, in that at any given location they will be either mostly all unmethylated or methylated. Within the human genome approximately 80% of CpG dinucleotides in adult tissues are methylated, with unmethylated CpGs normally occurring within CpG islands (Meissner *et al.* 2008).

CpG dinucleotides are under-represented in the human genome, with only about a fifth of the expected numbers present (Brena *et al.* 2006, Zilberman 2007). The distinct pattern of CpG dinucleotide distribution in mammals is thought to have arisen from deamination of inherently unstable 5-methylcytosine, which breaks down to form thymine. This results in a DNA mismatch with thymine base-paired with guanine. Resolution of this error by the DNA repair pathway can result in removal of guanine instead of thymine, resulting in depopulation of the genome of methylated CpG dinucleotides over an evolutionary timescale (Weber *et al.* 2007, Antequera & Bird 1999). In

support of this, a recent comparison of gene promoters between humans and chimpanzees showed loss of CpG dinucleotides from among those constitutively methylated, while CpG regions with no methylation show high levels of conservation (Weber *et al.* 2007).

Gene promoters in humans can be divided into two groups, those with a high CpG content, high CpG promoters (HCPs), which includes over 70% of promoters, while the remaining 30% of promoters have a low CpG content, referred to as low CpG content promoters (LCPs) (Saxonov *et al.* 2006). These two classes of promoter display very different DNA methylation patterns and distinct histone modifications.

Regulation of gene expression from HCPs is generally via repressive or active histone modifications, with even inactive genes marked with H3K4me2 that provides protection against DNA methylation, and even repressed genes marked with H3K27me3 remain unmethylated (Meissner *et al.* 2008, Weber *et al.* 2007). DNA methylation of a CpG island is a rare event, but upon methylation the gene is deeply, and potentially permanently, repressed. LCPs generally contain methylated CpG dinucleotides, but demonstrate no link between methylation of CpG dinucleotides and gene expression (Meissner *et al.* 2008, Saxonov *et al.* 2006). A recent study that examined Pol II occupancy at LCPs also found no correlation between methylation and repression, further suggesting that DNA methylation is inconsequential at LCPs (Weber *et al.* 2007).

The differing impact of DNA methylation on HCPs and LCPs is thought to stem from the differences in CpG density, where methylation of three or four CpG dinucleotides within a ~500bp LCP is insufficient for repression, but methylation of the (on average) nineteen CpG dinucleotides within a HCP would have a profound effect upon chromatin structure (Meissner *et al.* 2008, Weber *et al.* 2007, Brena *et al.* 2006, Zilberman 2007). The subsequent multiplying force of repression due to the many interactions between Dnmts and chromatin modifiers such as Suv39h or G9a (Tachibana *et al.* 2008, Fuks *et al.* 2003, Geiman *et al.* 2004) being recruited to the multiple CpG dinucleotides methylated at HCPs would then ensure gene repression.

CpG islands avoid methylation during the wave of *de novo* DNA methylation that occurs in the early embryo shortly after implantation (Li 2002). This is thought to occur through the transcription of CpG island containing genes, including low level expression of tissue specific genes during early embryogenesis (Antequera & Bird 1999). Global studies of gene expression and histone modifications lend credence to this theory as they show low level transcription of HCPs in embryonic stem cells, and a bivalent marking of gene promoters with both active and repressive marks that allow continued low level expression of genes that have tissue specific expression patterns later in development (Mikkelsen *et al.* 2007). CpG islands have also been demonstrated to act as origins of replication, replicating early in S phase when Dnmt1 levels are low (Delgado *et al.* 1998). Although this may confer some protection from DNA methylation, CpG islands that are methylated tend to replicate late in S phase, therefore CpG islands can be susceptible to *de novo*

methylation under the right circumstances, which is then readily maintained by Dnmt1 (Caiafa & Zampieri 2005).

1.4 Tissue specific methylated regions

Recent global studies of the methylation status of CpG islands in both mouse and human has confirmed that CpG islands are nearly always unmethylated, but has also revealed the existence of a core group of HCPs that are methylated in a tissue dependent manner. Excluding regions methylated during X inactivation or imprinting, tissue-differentially methylated regions (T-DMRs) at CpG islands may represent about 5% of mouse and human CpG islands (Meissner *et al.* 2008, Weber *et al.* 2007, Irizarry *et al.* 2009, Shen *et al.* 2007, Song *et al.* 2005).

Gene repression strongly correlates with DNA methylation, as well as loss of Pol II occupancy. Many of the studies, upon further characterization of the methylation and expression profile of identified T-DMRs, showed that the majority were densely methylated in all adult tissues examined, but were unmethylated and highly expressed in sperm, and to a lesser extent, testis (Weber *et al.* 2007, Shen *et al.* 2007, Song *et al.* 2005) suggesting that this subset of CpG island-containing promoters form a specific group of genes required for germline propagation. A small subset of HCP T-DMRs were not germ-line specific, but were tissue specific transcription factors, probably becoming methylated to ensure that alternative differentiation pathways would be permanently silenced thereby safeguarding tissue-specific development (Weber *et al.* 2007).

While most of the studies were restricted to CpG islands within the immediate confines of a genes promoter, several studies examined methylation at CpG islands away from classical promoter regions (Song *et al.* 2005), and also within the proximity of HCPs, but at regions up to 2Kb from the TSS where the CpG density was lower than that of the CpG island proper, referred to as CpG island shores (Irizarry *et al.* 2009). This revealed a far greater number of T-DMR regions than had previously been shown, with 76% of T-DMRs occurring at CpG island shores, and approximately half occurring over 2Kb from the nearest annotated gene (Irizarry *et al.* 2009, Song *et al.* 2005). Gene expression studies showed that the majority of T-DMRs that were located some distance from a genes' TSS were infact controlling expression from alternative TSSs, thereby regulating alternative transcription. Methylation at all T-DMRs examined showed a strong inverse relationship with gene expression, with repression alleviated in Dmmt1/Dnmt3b KO cells, as well as cells treated with the DNA methylation inhibitor 5-azacytidine, highlighting the central role of DNA methylation in ensuring repression (Irizarry *et al.* 2009).

1.5 Histone modifications

Histone proteins have a globular structure, with the exception of their N-terminal regions, or tails, which are unstructured. Two of each of the four histone proteins H2A, H2B, H3 and H4, combine together with DNA to form a nucleosome, the most basic structure of chromatin, which is then packaged into higher order chromatin structures (Peterson & Laniel 2004, Kouzarides 2007). The unstructured tails of histone proteins provide a platform for modifying enzymes that catalyze the addition of different histone modifications to specific residues, which can directly affect chromatin structure, but also provide binding

sites for proteins involved in gene regulation. Together, histone modifications and the protein complexes they recruit, control the chromatin structure and therefore the biological role played by the underlying DNA sequence (Kouzarides 2007, Zhang & Reinberg 2001).

Acetylation, methylation, phosphorylation, and ubiquitination can occur at a large number of specific sites, with the H3 and H4 tails in particular being extensively modified (Zhang & Reinberg 2001, Lachner *et al.* 2003). See Figure 1.2 for a diagram of the histone tails and their potential modifications. Many different histone modifications combine together, producing a histone code to dictate the overall transcriptional state of the underlying DNA. Histone modification can be highly dynamic with modifications being added and removed rapidly (Kouzarides 2007, Strahl & Allis 2000). Different combinations of histone modifications at gene promoters, and throughout the body of the gene, controls the transcriptional state of that gene, with the correct combination dictating transient repression, high levels of transcription, or permanent silencing.

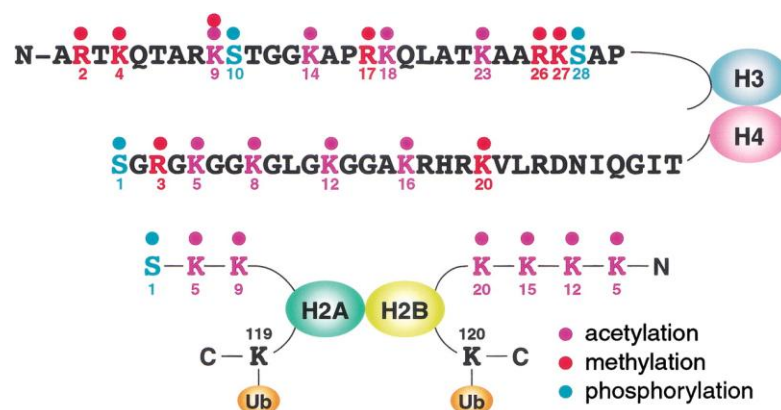


Figure 1.2 Histone tail modifications. The histone tails of H3, H4, H2A and H2B can all be covalently modified at specific residues. Histone modifications directly and indirectly affect transcription of the underlying DNA sequence. Figure adapted from Zhang and Reinburg (2001).

Methylation of lysine residues is one of the most intensively studied groups of histone modifications, and includes many modifications associated with both gene expression and gene repression. Lysine residues can be modified by up to three methyl groups, with monomethylated residues having clear and distinct roles compared to trimethylation (Kouzarides 2007). An overview of the main histone modifications is shown in Table 1.1

H3K9 methylation
H3K9me1 associates at promoters prior to gene activation. H3K9me2 and me3 at gene promoters both strongly associated with gene repression in euchromatin and heterochromatin respectively. H3K9me3 occurs at pericentric heterochromatin, satellite repeats, transposon sequences, imprinted genes and on the silenced X chromosome.
H3K27 methylation
H3K27me1 has a very similar localisation profile as H3K9me1, localising to gene promoters prior to their activation. H3K27me3, catalysed by the PRC2 component EZH2, is the major repressive mark at gene promoters in the mammalian genome. It frequently occurs with the activating mark H3K4me3 at bivalent domains.
H3K4 methylation
As with previously described modifications, monomethylation of H3K4 accumulates at gene promoters prior to their activation. H3K4me2 and me3 are both enriched at active genes, with H3K4me3 generally showing a greater level of enrichment that is more tightly focussed on the promoter. The level of enrichment is proportional to the transcriptional level of the gene. At genes poised for activation, H3K4me3 and H3K27me3 are both present.
H3K36 methylation
Monomethylation has a slight enrichment at active genes, but its role in gene regulation is unclear. H3K36me3 correlates with transcriptional elongation, and is highly enriched downstream of the TSS of active genes, with enrichment increasing further towards the 3' end of the gene.
H3K79 methylation
Very highly transcribed genes show a strong enrichment for H3K79me3 within the promoter. Most active genes show localised enrichment around the TSS, H3K79me3 is also present at many inactive genes but has a more diffuse pattern of enrichment.
H4K20 methylation
H4K20me1, as with many of the other monomethylations for other histone modifications, associates with promoters just prior to their activation. H4K20me3 has a near-identical distribution to H3K9me3 methylation and is associated with constitutive silencing at centromeric regions and repeat sequences.

Table 1.1 Histone lysine modifications. Information from Barski *et al.* 2007, Cui *et al.* 2009, Mikkelsen *et al.* 2007, Pan *et al.* 2007, Zhao *et al.* 2007

1.5.1 Bivalent Domains

H3K4me3 methylation occurs at actively expressed genes, with approximately half of all gene promoters, and practically all CpG island-containing promoters having this mark in ESCs (Mikkelsen *et al.* 2007, Pan *et al.* 2007). H3K27me3 is present at repressed genes, and is deeply involved in long-term silencing of the inactive X chromosome, with over half of identified H3K27me3 regions occurring on the inactive X (Zhao *et al.* 2007). Gene promoters purely associated with H3K27me3 are rare in ESCs, where most H3K27me3 regions coincide with H3K4me3. Bivalent domains, regions with both active and repressive marks, occur at approximately a quarter of CpG island containing promoters in mammalian ESCs, and include many developmental transcription factors (Mikkelsen *et al.* 2007, Pan *et al.* 2007, Ku *et al.* 2008). Bivalent domains allow a low level of gene expression, but essentially keep the gene in a repressed state, yet poised for activation during differentiation.

As cells differentiate, the number of promoters purely marked with H3K4me3 decreases as more genes become bivalently marked. With further differentiation, the pluripotency of the cell becomes increasingly reduced, with those genes required for a particular cellular lineage remaining bivalent or becoming active, and those that are no longer required becoming marked solely by H3K27me3 (Mikkelsen *et al.* 2007, Ku *et al.* 2008).

1.5.2 H3K9 trimethylation

The majority of H3K9 trimethylation is associated with silencing at regions of pericentric heterochromatin, or satellite and long terminal repeats. H3K9me3 is

also involved in repression during imprinting, with silenced genes often being marked with H3K9me3 (Mikkelsen *et al.* 2007, Lehnertz *et al.* 2003, Rice *et al.* 2003, Peters *et al.* 2001, Barski *et al.* 2007). H3K9me3 methylation often correlates with H4K20me3 at heterochromatic regions (Mikkelsen *et al.* 2007). *In vitro* studies show that trimethylation of H4K20 changes both the orientation of the residue and also alters the position of neighbouring sidechains, resulting in a condensed chromatin state without the need for additional chromatin modifiers recognising and binding at the modification (Lu *et al.* 2008). H3K9me3 also shows some correlation with H3K27me3 (Barski *et al.* 2007).

H3K9me3 is associated with long-term repression of sequences that must be repressed to preserve genomic stability, such as repetitive elements and parasitic DNA sequences like retrotransposons. H3K9me3 is one of the few histone modifications examined that showed little change in disposition as cell differentiate (Mikkelsen *et al.* 2007, Cui *et al.* 2009). Approximately three quarters of H3K9me3 occurs in intergenic regions, while less than 1% of gene promoter regions are H3K9me3 marked (Cui *et al.* 2009).

Regions of H3K9me3 do exist at unique sites within the mammalian genome, however many of these are within 2Kb of repetitive elements - suggesting that H3K9me3 enrichment may be the result of spreading (Mikkelsen *et al.* 2007). Despite the clear correlation of H3K9 trimethylation with repeat sequences and heterochromatin, H3K9me3 can occur in euchromatic regions at gene promoters. One study identified 126 genes involved in preimplantation and germline development that are repressed by H3K9me3 methylation upon

differentiation (Epsztejn-Litman *et al.* 2008), and global histone modification studies suggest that for those promoters that are methylated at H3K9me3, enrichment occurs over a large area, with methylation extending over a 10Kb region around the silenced promoter (Barski *et al.* 2007).

Interestingly, for a mark that has such strong associations with repression, modest levels of H3K9me3 have been observed starting approximately 5Kb downstream of the TSS of some genes upon transcriptional elongation by Pol II (Vakoc *et al.* 2006, Vakoc *et al.* 2005). Global studies of mammalian genomes have also observed a correlation between gene expression and H3K9me3 enrichment within the body of the gene, though the number of examples to date is small suggesting that this is a rare occurrence (Mikkelsen *et al.* 2007, Barski *et al.* 2007, Cui *et al.* 2009).

1.5.3 Acetylation

Lysine acetylation is a highly dynamic modification associated with decondensed and active chromatin, with many locations on all four histone tails that can potentially become modified (Kouzarides 2007, Zhang & Reinberg 2001, Mellor *et al.* 2008). See Figure 1.2. Acetylation at specific residues is essential for recruitment of transcription initiation factors, with recruitment of complexes leading to further acetylation. H3K9Ac and H3K14Ac in particular are required for transcriptional initiation as they recruit the transcription initiation factor TFIID (Agalioti *et al.* 2002). In ESCs these two acetyl marks occur at 70% of genes, with a near perfect correlation with both H3K4me3 enrichment and gene transcription (Guenther *et al.* 2007). Acetylation at specific residues can also directly affect chromatin structure, for example; H4K16Ac has been shown

to inhibit formation of a compact 30nm fibre, thereby preventing development of higher order chromatin, although the manner in which it does this has yet to be elucidated (Shogren-Knaak *et al.* 2006).

It has previously been suggested that acetylation of histones functioned via interruption of the electrostatic charge between the negatively charged DNA backbone and positively charged histone tails, but this is no longer the prevalent view as, although it is true that histone octomers with twelve acetylated residues disrupt higher order chromatin structure (Tse *et al.* 1998), histone tails are not released from DNA upon hyperacetylation (Peterson & Laniel 2004). Additionally, significant histone acetylation would be required to disrupt the ionic interaction between DNA and histones, but gene expression is capable of being initiated with acetylation of just a handful of lysine residues (Agalioti *et al.* 2002).

1.5.4 Ubiquitination

Ubiquitination occurs at two residues, H2AK119 and H2BK120, with diametrically opposite effect. Monoubiquitination at H2BK120 is common at promoters of genes undergoing transcriptional activation, where it enhances binding of an important chromatin remodelling factor essential for transcriptional elongation, whereas H2AK119 monoubiquitination inhibits binding of the same chromatin remodelling factor thereby stalling transcription at pre-elongation (Zhou *et al.* 2008). Ubiquitination at H2AK119 is catalysed by Rnf2, a component of the polycomb repressor complex 1 (Stock *et al.* 2007) discussed in detail in 1.7.

1.5.5 Histone modifications states define a genes transcriptional state

The transcriptional state of a gene is defined by the histone modifications associated with its promoter region, as well as throughout the gene body. Transcriptional control involves the interplay of many different histone modifications, though there are certain key modifications, such as H3K4me3, H3K27me3, or H3K9me3 whose presence at a domain clearly indicate a genes' transcriptional state. Recent studies examining global histone modification patterns for many different modifications, and in various different cell types has begun to reveal the complexity and extent of the histone modifications involved during gene expression and repression (Mikkelsen *et al.* 2007, Barski *et al.* 2007, Cui *et al.* 2009).

Prior to activation of gene expression Cui *et al.* demonstrated that bivalently modified genes are enriched with monomethylation of H3K4, H3K9, and H2K27, as well as Pol II, at gene promoter regions. This is associated with a basal level of gene expression, possibly to maintain the activation potential of the gene (Cui *et al.* 2009). Upon gene expression, the promoter loses any H3K27me3 enrichment, but retains the monomethylated modifications present prior to activation. It also becomes enriched for both dimethylation and trimethylation of H3K4. H3K36me3 methylation develops downstream of the TSS, and at highly expressed genes, H3K79me3 becomes enriched across the promoter. H2AK5me1 and H3R2me1, whose roles in regulation have not fully been explored, also become enriched (Mikkelsen *et al.* 2007, Barski *et al.* 2007). An example of the spatial enrichments of different histone modifications at an active promoter is shown in Figure 1.3A.

The modifications at inactive genes are less complex, with H3K27me3 normally predominant, with enrichment for H3K4me3 also present if the domain is bivalent. H3K4me3 enrichment is normally a small domain centred on the promoter, while H3K27me3 enrichment can cover an area up to 3Kb in size, which includes the promoter region. H3K79me3 enrichment may also be present, as it is observed at both the active and inactive promoters of most genes, with a significant correlation only existing for genes that exhibit high levels of expression (Ku *et al.* 2008, Barski *et al.* 2007). Typical histone modifications and their enrichment relative to promoter regions is shown in Figure 1.3B.

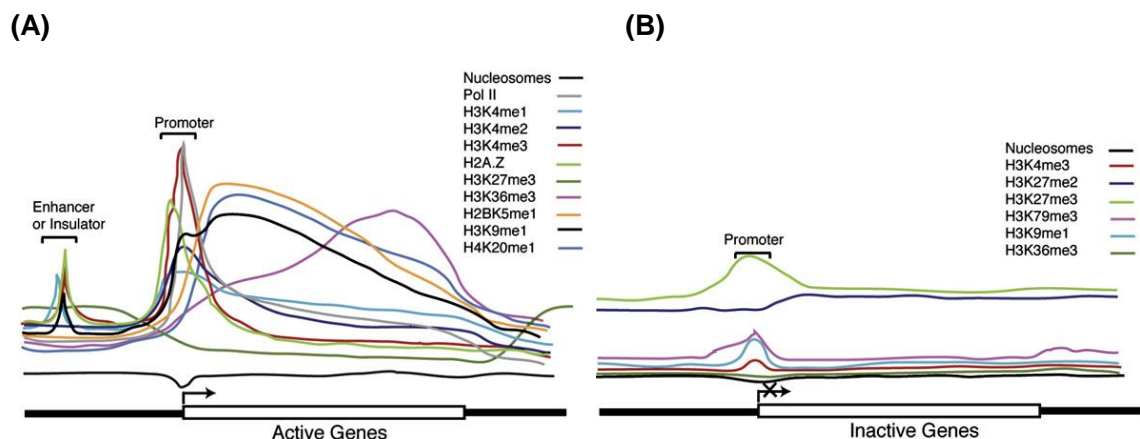


Figure 1.3 Histone modifications at active and inactive genes.

Representation of the histone modifications and their relative enrichment at a gene promoter and across the body of the gene when the gene is active (A) or silenced (B). Multiple histone modifications at an actively expressed gene form a complex code that define the gene's expression state. Figure adapted from Barski *et al.* (2007)

1.6 Mammalian H3K9me histone methyltransferases

There are six known histone methyltransferases (HMTs) that can methylate H3K9 in mammals. Suv39h1 and Suv39h2 are highly similar and are jointly

responsible for most H3K9 methylation at heterochromatic regions (Peters *et al.* 2001, O'Carroll *et al.* 2000), while G9a and GLP function together as a single unit to carry out HMT activity as the main mediators of euchromatic H3K9 methylation (Tachibana *et al.* 2002, Tachibana *et al.* 2005). Two additional methyltransferases are known, ESET – another euchromatic HMT involved in transcriptional repression (Schultz *et al.* 2002), and CLL8, which had yet to be characterised, though shares significant homology with ESET (Mabuchi *et al.* 2001) and therefore may occupy a similar role.

The HMTs SUV39H1 and SUV39H2 are the main proteins responsible for H3K9me3 at heterochromatic regions (Rice *et al.* 2003, Peters *et al.* 2001). The *suv39h* family is highly conserved in eukaryotes, with mammalian *suv39h* evolutionarily related to Su(Var)3-9 in *D.melanogaster* and Ctr4 in *S.pombe*. Mammalian *suv39h* proteins have near-identical expression patterns during embryogenesis, with *suv39h1* continuing to be widely expressed in adult tissues, while expression of *suv39h2* is limited to the testes where it is thought to play a role in maintaining sex chromosome integrity during pairing of meiotic chromosomes (Peters *et al.* 2001, O'Carroll *et al.* 2000)

Mammalian Suv39h HMTs are responsible for the silencing of satellite repeats at pericentric heterochromatin, and do so in a semi-redundant manner, whereby knockout of one *suv39h* results in only a minor reduction in methylation, whereas knockout of both abolishes silencing (Peters *et al.* 2001) and results in severe loss of H3K9me3 at heterochromatic loci, loss of DNA methylation at Major Satellite Repeats and an increase in the level of Major Satellite transcripts

(Lehnertz *et al.* 2003). Suv39h activity is essential for genome stability and knockout results in chromosomal instabilities, tumourogenesis, and severe impairment of viability in mice (Peters *et al.* 2001).

Suv39h-mediated H3K9me3 provides a binding site for other components of the transcriptional repression machinery such as HP1 α , HP1 β , DNMT1, DNMT3b, and meCP2, all of which have been observed co-localising with suv39h at heterochromatic loci. Additionally; Suv39h1 directly interacts with HDACs to cause further transcriptional repression (Lehnertz *et al.* 2003, Lachner *et al.* 2001, Vaute *et al.* 2002). Suv39h contains a chromodomain, whereby H3K9me3 will serve to recruit further Suv39h thereby propagating the silencing effect (Lachner *et al.* 2001).

G9a is the main HMT in euchromatin, and functions as a heterodimer with GLP, another SET domain containing protein that is essential for G9a HMTase activity. Both proteins are ubiquitously expressed, and localise to euchromatic regions within the nucleus, with no localisation at pericentric heterochromatin or nucleoli. Knockout of G9a or GLP is lethal in early embryogenesis, approximately 9.5 dpc in the mouse (Tachibana *et al.* 2002, Tachibana *et al.* 2005).

Several studies have demonstrated that knockout of G9a causes almost total loss of H3K9me2, and a severe reduction in H3K9me1 levels in cultured mammalian cells, but knockout has little or no impact on global H3K9me3 levels (Rice *et al.* 2003, Tachibana *et al.* 2005). The G9a/GLP heterodimer is essential

for silencing lineage-specific genes in ES cells (Dong *et al.* 2008), and It has also been implicated in the repression of genes involved in cell cycle progression and DNA replication, thereby maintaining a cell in G₀ (Ogawa *et al.* 2002).

Based upon immunofluorescence and western blot analysis of cells knocked out for G9a, there was no discernable impact upon H3K9me3 levels, therefore G9a was putatively labelled as capable of only monomethylation and dimethylation of H3K9me3. However, it was noted that G9a was capable of trimethylation *in vitro* (Rice *et al.* 2003, Tachibana *et al.* 2005), a discovery confirmed by another study demonstrating that *in vitro* purified G9a was efficient at trimethylation in a processive manner (Patnaik *et al.* 2004). Subsequently, G9a mediated H3K9me3 has been shown to be involved in repression of a group of over one hundred early developmental genes (Epsztejn-Litman *et al.* 2008), and G9a mediated H3K9me3 also has a role in the repression of placentally imprinted genes at the Kcnq1 locus, where it has been shown to colocalise with the ncRNA essential for implementing transcriptional repression at the imprinted loci (Wagschal *et al.* 2008, Pandey *et al.* 2008).

G9a mediates gene silencing through establishment of H3K9 methylation, but also directly recruits DNMT3a and DNMT3b via its ankyrin repeats to establish DNA methylation. It has been shown that both these repressive marks can be established independent of the other, and that the presence of either mark at a locus is sufficient for silencing (Epsztejn-Litman *et al.* 2008). G9a is also capable of reducing the levels of H3K4me2 and H3Ac, histone marks

associated with gene expression, in a SET domain independent manner (Epsztejn-Litman *et al.* 2008, Tachibana *et al.* 2008). The G9a/GLP heterodimer has been observed in a silencing complex that includes components of the PRC, transcriptional repressors, and also HP1 γ (Tachibana *et al.* 2008, Ogawa *et al.* 2002), and loss of G9a has specifically been shown to prevent HP1 γ localisation (Tachibana *et al.* 2002, Ikegami *et al.* 2007).

ESET is another HMT of H3K9 that localises to mainly euchromatic regions (Schultz *et al.* 2002, Yang *et al.* 2002), and is known to be expressed in the post-implantation embryo (Dodge *et al.* 2004), though its expression in adult tissues has yet to be fully characterised. Knockout results in early embryonic lethality within 5.5 dpc, though there are no observable changes in levels of H3K9 methylation (Dodge *et al.* 2004). ESET directly interacts with DNMT3A and DNMT3B (Li *et al.* 2006), with HDACs, and the transcriptional repressors mSin3A and mSin3B, with which it forms a transcriptional repressor complex capable of gene silencing (Yang *et al.* 2003).

Few precise roles for ESET have yet been elucidated, though it has been shown to bind with the KAP1 co-repressor that is involved in the repression of over 200 genes (Schultz *et al.* 2002), while in striatal neurons the over expression of ESET is a suggested effect of Huntington's disease and was shown to cause expansion in the size of pericentric heterochromatic regions (Lee *et al.* 2008), though when expressed at normal levels ESET is not observed at pericentric heterochromatin (Schultz *et al.* 2002).

1.7 Polycomb Repressor Complexes

Polycomb group (PcG) proteins are evolutionarily conserved transcriptional repressors responsible for controlling the expression of several thousand genes in mammals, many with important roles in development, such as homeodomain genes (Ku *et al.* 2008, Pasini *et al.* 2004). Complexes of PcG proteins bind within a genes' promoter region and mediate addition of the repressive histone modifications H3K27me3 and H2A119u1. The repressive action of PcG complexes is counterbalanced by the Trithorax group (TrxG) proteins that act as transcriptional activators and catalyse the addition of H3K4me3 (Schuettengruber *et al.* 2007). It is thought that the PcG/TrxG system acts as a buffer for cell pluripotency by reinforcing repression of differentiation factors (Ku *et al.* 2008). The majority of promoters targeted by these chromatin remodelling complexes contain CpG islands, with PcG complexes in mammals binding within 1Kb of the promoter in 90% of cases (Ku *et al.* 2008, Boyer *et al.* 2006).

In ESCs, many genes are bound by both PcG and TrxG proteins, with the result that approximately a quarter of CpG island-containing genes are bivalently marked by both H3K27me3 and H3K4me3 (Mikkelsen *et al.* 2007, Ku *et al.* 2008). Genes bivalently marked are essentially repressed, with only low levels of gene expression, but are poised to be either fully activated or repressed upon differentiation, upon loss of one of the diametrically opposed histone modifications (Bernstein *et al.* 2006a).

PcG proteins form two main complexes, polycomb repressor complex 1, PRC1, and PRC2. The core components of PRC2 are EZH2, EED, and SUS12. EZH2

is a SET domain containing HMTase responsible for H3K27me₃, while both EED, and SUZ12 are essential for EZH2 function (Pasini *et al.* 2004, Boyer *et al.* 2006, Cao & Zhang 2004). See Figure 1.4. Deletion of the main components of PRC2 causes embryonic lethality around day 7.5 of development in mice, resulting from incomplete gastrulation (Pasini *et al.* 2004, O'Carroll *et al.* 2001). PRC2 essentially occupies all bivalently marked domains, with over 3000 sites bound by PRC2 components in the human genome. Approximately half of these sites are also bound by PRC1 (Ku *et al.* 2008), which leaves an equal number of loci enriched with PRC2 alone.

The central components of PRC1 are Cbx, Rnf2, and Bmi1 (Schuettengruber *et al.* 2007). Cbx is a chromobox containing protein that binds to H3K27me₃ (Bernstein *et al.* 2006b), while Rnf2 is a RING motif containing ubiquitin ligase that adds the histone modification H2A119u1 at targeted sites (Stock *et al.* 2007). PRC1 function is summarised in Figure 1.4. Knockout of Rnf2 is lethal during early embryogenesis, with a very similar phenotype to loss of PRC2 components, with development halting during gastrulation - though embryos continue to develop until day 10.5 (mouse model) suggesting PRC1 is essential slightly later in development compared to PRC2 (Voncken *et al.* 2003).

Although the two PRC complexes do not physically interact, because the PRC1 complex is recruited through interaction H3K27me₃ mediated by PRC2, it has been suggested that the PRC1 complex is subservient to PRC2 (Hernandez-Munoz *et al.* 2005). Indeed; in the human genome, the majority of the 1500 observed PRC1 sites occur at regions where PRC2 is also present (Ku *et al.*

2008), and observations using immunofluorescence do show a significant loss in PRC1 localisation upon knockdown of Ezh2 (Hernandez-Munoz *et al.* 2005). However, global studies of PRC1 binding have identified a subset of around 130 sites in humans where it localises without the presence of PRC2 (Ku *et al.* 2008, Bracken *et al.* 2006), with a similar example seen in mouse (Boyer *et al.* 2006).

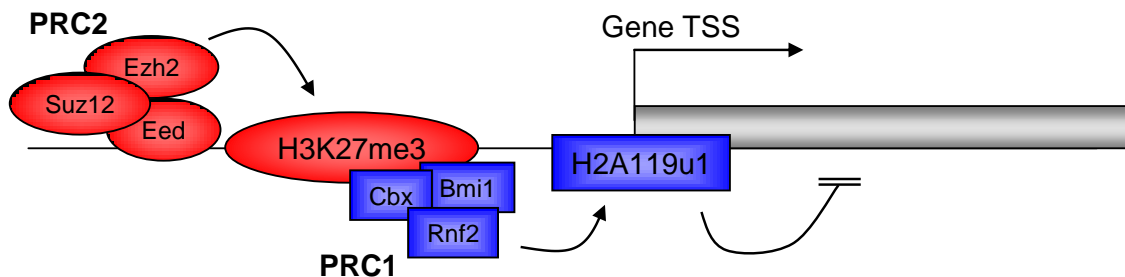


Figure 1.4 Classical repression by polycomb repressor complexes. Ezh2 (PRC2) catalyses H3K27me3 via its SET domain. H3K27me3 is then bound by the Cbx (PRC1) via its chromodomain. The PRC1 component Rnf2 catalyses H2A119u1, a histone mark that physically blocks binding of the transcriptional elongation protein FACT (1.5.4). The majority of PRC1 binding sites are at loci also marked by PRC2 mediated H3K27me3, though PRC1 is also capable of localisation independent of H3K27me3, with Cbx capable of binding to both H3K9me3 and RNA. Information derived from Bernstein *et al.* 2006; Boyer *et al.* 2006; Cao and Zhang 2004; Pasini *et al.* 2004; Schuettengruber *et al.* 2007; Stock *et al.* 2007; Zhou *et al.* 2008.

Mono ubiquitination of H2A119, mediated by Rnf2, is essential for PcG mediated gene repression (Stock *et al.* 2007, van der Stoep *et al.* 2008), with silencing caused by inhibition of RNAPII elongation. Bivalently marked genes are bound by high levels of Pol II at promoter regions, but despite this concentration of transcriptional machinery, transcription from bivalently marked promoters is low (Stock *et al.* 2007). The chromatin remodelling protein FACT, a dimeric protein that alters the position of the H2A/H2B dimer within the nucleosome, is essential to relieve chromatin mediated repression of

elongation. FACT recruitment is blocked by RNf2 mediated monoubiquitination, thereby stalling transcription at an early stage of elongation (Zhou *et al.* 2008).

There are two distinct classes of bivalent domain in mammals. Promoters that are bound by both PRC complexes account for the vast majority of developmental regulator genes, which are clearly controlled by the presence of the repressor complexes. These domains contain large H3K27me3 domains over 3Kb in length. Those regions bound only by PRC2, show poor maintenance of H3K27me3, containing domains generally less than 1Kb in size. CpG islands bound by both complexes are almost twice the size of those bound by PRC2 alone, and 'PRC2 only' domains show weak evolutionary conservation, suggesting that they could be the result of inadvertent PRC2 targeting as a side effect of the mechanisms that recruit and maintain PRC1/PRC2 domains (Ku *et al.* 2008).

1.7.1 PRC roles in repression beyond 'classical' bivalent domains

The PRC1 component Cbx belongs to a family of five chromobox containing proteins in mammals, and although most investigations have focussed upon its ability to bind to H3K27me3, members of the Cbx family have also been shown to have an affinity for H3K9me3 in *in vitro* binding assays (Bernstein *et al.* 2006b). Bernstein *et al.* showed that of the five chromobox family members, three bound to H3K9me3. Cbx2 and Cbx7 bound with similar affinity to both K9 and K27 trimethylated residues, while Cbx4 has a significant preference for trimethylated K9. Further analysis of Cbx7 demonstrated that it bound dimethylated and trimethylated H3K9, and H4K20me1 in an *in vivo* system, alongside its H3K27me3 binding ability - all characteristic markers of facultative

heterochromatin, and especially histone modifications on the silenced X chromosome. Both Cbx2 and Cbx7 localised to the X chromosome undergoing silencing in a differentiating ESC model, highlighting the role of PcG proteins in the silencing process (Bernstein *et al.* 2006b). Interestingly, it was observed that Cbx7 was capable of binding to RNA in a sequence independent manner, preferentially binding to single stranded RNA, and that this interaction was important for association with the inactive X chromosome, through interaction with Xist, but also for its general association throughout chromatin (Bernstein *et al.* 2006b).

H3K9me3 is not an observed feature of bivalent domains in mammals (Mikkelsen *et al.* 2007, Ku *et al.* 2008, Bernstein *et al.* 2006a), though; since a subgroup of PRC1 complexes are not bound at bivalent domains or PRC2 associated (Ku *et al.* 2008, Boyer *et al.* 2006, Bracken *et al.* 2006) this does not preclude its involvement at such regions. Taken together with the interactions between Cbx7 and RNA, these observations are suggestive of a repressive role aside from its demonstrated involvement in mainstream PcG/TrxG gene regulation system.

This is especially interesting given recent findings that demonstrate a role for the PRC2 complex in siRNA mediated gene repression involving the argonaute protein AGO1 (Hawkins *et al.* 2009, Kim *et al.* 2006), and also the interactions between the PRC2 component SUZ12 and a large ncRNA during gene repression at the HOXD locus (Rinn *et al.* 2007). Alongside the observations concerning PRC1, these interactions demonstrate that both PcG complexes can

interact with RNA in different ways and are capable of becoming targeted through that interaction - an ability not suggested by their role at bivalent domains where their recruitment is thought to occur via the DNA sequence (Ku *et al.* 2008).

PRC2 has been implicated in aberrant DNA methylation and suppression of gene expression in cancer. In colon cancer cell lines, genes subject to hypermethylation in cancer were enriched for PRC2 components, and were marked with H3K27me3 in both normal and cancer cells, as well as being enriched for both *de novo* methyltransferases and Dnmt1, suggesting that normal silencing via the PcG protein complexes could cause aberrant recruitment of Dnmts (Schlesinger *et al.* 2007).

Another study screened for genes that were normally enriched for PRC2 components and H3K27me3 in ESCs, and were hypermethylated in primary tissue from colorectal tumours. They showed that three-quarters of those genes hypermethylated in cancer exhibited PRC2 recruitment, H3K27me3, or both, in undifferentiated ES cells. This suggests that cancer progression can be traced back to aberrant events in ESCs or early precursor cells, where crosstalk between DNA methylation machinery and PcG repressor complexes results in aberrant DNA methylation causing permanent silencing of important oncogenes. These epigenetic changes would only become apparent after differentiation and could then facilitate cancer progression. Further work looking for low frequency dense promoter methylation in hemopoietic progenitor cells revealed that genes with bound components of PRC2 and/or H3K27me3 were

twice as likely to become aberrantly methylated compared to genes without marks (Widschwendter *et al.* 2007).

The link between polycomb repressors and aberrant DNA methylation in cancer is suggestive of a link between Polycomb proteins and DNA methyltransferases (DNMTs). However, over expression of proteins in cancerous cells and outside of their normal patterns of expression would allow protein: protein interactions to occur that would otherwise not take place under normal cellular conditions. Nevertheless; the observations in cancer are at least suggestive that an interaction between Dnmts and PRC components can theoretically occur.

1.8 RNA and gene regulation

The investigation into the role of non-protein coding RNAs (ncRNAs) in transcriptional regulation is in its nascency, though there are already many examples of RNAs involved in transcriptional repression, with recent studies finding a surprising levels of ncRNA within the nucleus, and also a degree of conservation of many long ncRNAs that suggests a functional role.

1.8.1 Transcription of non-coding RNA within the cell

Recent investigations into the levels of non protein-coding RNA within the cell has not only revealed a vast number of RNAs that have potential regulatory roles, but also suggests that the extent of the genome which is actively transcribed far exceeds previous estimates. Such discoveries have fundamentally altered perceptions on genome regulation and suggest a far greater role for ncRNAs in regulation of gene expression than previously thought. Various studies examining transcription in both mouse and human

genomes have found examples of ncRNA transcripts in their thousands, arranged in both sense and antisense orientation to known genes, joining areas of the genome many kilobases apart, and being transcribed from regions thought to be transcriptionally silent (Lavorgna *et al.* 2004, Guttman *et al.* 2009, Kapranov *et al.* 2007, ENCODE Project Consortium *et al.* 2007, Katayama *et al.* 2005).

The pilot phase of the ENCODE project, a collaboration of many groups that examined in high detail the transcription of approximately 30Mb of the human genome (equivalent to about one percent of the total) suggested that, potentially, 93% of the regions they surveyed could be transcribed (Kapranov *et al.* 2007, ENCODE Project Consortium *et al.* 2007). Another study that created transcriptional maps for ten human chromosomes showed that, on average, the majority of transcripts were non-polyadenylated, with more than half of mature mRNAs in the cytosol not corresponding to known protein-coding genes, while the proportion of transcripts in the nucleus that did not match any previously annotated region of the genome was almost 80% (Cheng *et al.* 2005). Together these two studies indicate both the potential scale of transcription, and the proportion of it that produces transcripts of unknown function - many of them un-polyadenylated and restricted to the nucleus.

1.8.2 Conservation and function of non-coding RNAs

Large numbers of ncRNAs have been identified (Katayama *et al.* 2005, Okazaki *et al.* 2002, Numata *et al.* 2007, Kiyosawa *et al.* 2003), and while studies have found specific examples of ncRNAs with biological roles, questions concerning the numbers of ncRNAs that are functional compared to those that are

biologically meaningless transcripts, possibly representing a background of 'noise' produced by the inherent leakiness of transcription, have yet to be satisfactorily answered (Okazaki *et al.* 2002, Mattick 2005, Wang *et al.* 2004). This is partially because ncRNAs may not necessarily require to have a high level of conservation of their primary sequence if it is not the transcript *per se* that is required but rather its transcription, which affects the chromatin structure of the regions they are transcribed though (Numata *et al.* 2007, Struhl 2007).

Several studies have examined conservation of long ncRNAs, or macroRNAs, in mammals (Guttman *et al.* 2009, Ponjavic *et al.* 2007). Ponjavic *et al.* looked at the evolutionary conservation of over three thousand long ncRNAs in mammals that had previously been discovered by the FANTOM consortium, and saw evidence of selective pressure, especially at promoter regions - clearly suggesting a biological role (Ponjavic *et al.* 2007). Guttman *et al.* identified over one and a half thousand evolutionarily conserved long ncRNAs using chromatin state maps to identify regions likely to code for previously unannotated ncRNAs that overlapped with known protein coding loci. Most interestingly; some subsets of the long ncRNAs discovered contained binding sites for known transcriptional regulators within their promoter regions, including the pluripotency factors Oct4 and Nanog, providing clues to their biological function (Guttman *et al.* 2009). Conservation of smaller ncRNAs may not be so prevalent; a study that examined sequence preservation of *cis* encoded RNA pairs, where both sense and antisense strands are coded for in the same location, saw extremely low conservation of just 6.6% between mouse and human of roughly two thousand pairs, with nearly all conserved pairs occurring

within protein coding regions (Numata *et al.* 2007). However; this study only examined a specific subset of ncRNAs, and conservation levels may not represent those of the population as a whole.

The roles that ncRNAs can play in gene regulation appears diverse and potentially complex. They have been demonstrated to be essential for cell viability, and capable of interacting with multiple proteins (Willingham *et al.* 2005), while sense/antisense pairs can regulate expression in a reciprocal manner in some examples examined, with ncRNAs antisense to a coding gene causing repression (Katayama *et al.* 2005).

Non-coding RNAs have been demonstrated to play a role in the regulation of the HOX gene transcription factors. HOX genes are organised into four clusters, which have very specific spatial and temporal expression patterns, and control anterior-posterior axis development in the early embryo. Rinn *et al.* identified 231 ncRNAs expressed from the different HOX loci in a distinct spatial manner, with almost three quarters expressed in an antisense orientation. Characterisation of one ncRNA, the highly conserved 2.1Kb spliced and polyadenylated transcript HOTAIR identified its role as an *in trans* transcriptional repressor. HOTAIR is transcribed from the HOXC locus on chromosome 12 and causes repression of a 40Kb region of the HOXD locus on chromosome 2, with repression occurring via Suz12-mediated H3K27me3 (Rinn *et al.* 2007).

1.8.3 Transcription of non-coding RNAs at transcriptional start sites

A collection of recent studies has characterised the surprising extent of ncRNA transcription that occurs around the TSS of many genes. Transcription of these small non-coding RNAs occurs in both sense and antisense orientations, and shows strong evolutionary conservation, indicating a biological role (Taft *et al.* 2009, Seila *et al.* 2008, Core *et al.* 2008, He *et al.* 2008). The transcription associated RNAs (tiRNAs) are small, on average 18-20nts long, with a peak of transcription at the TSS, with the number of sense and antisense transcripts rapidly tailing off, downstream and upstream respectively, within no more than 100bp of the TSS (Taft *et al.* 2009, Seila *et al.* 2008).

tiRNAs were observed at approximately 90% of CpG island containing promoters, and showed correlation with high levels of gene expression, as well as some correlation with genes in a state of pre-initiation (Taft *et al.* 2009, Seila *et al.* 2008, Core *et al.* 2008). Promoters enriched for tiRNAs also showed both TF and Pol II enrichment, as well as H3K4 methylation - though this may merely be due to their correlation with active genes. It is possible that tiRNAs may play a direct role in regulation of chromatin modifications (Taft *et al.* 2009), or ongoing transcription could act as a barrier against nucleosome occupancy, opening up the chromatin and enhancing TF binding, while the localisation of Pol II complexes could provide a reservoir for gene transcription. TiRNAs could also recruit protein complexes directly via RNA binding proteins (Taft *et al.* 2009, Seila *et al.* 2008, Core *et al.* 2008). Whatever their role, tiRNAs are evolutionary conserved in human, mouse, chicken and drosophila, suggesting

that they may be an ancient mechanism conserved in all metazoans (Taft *et al.* 2009).

1.8.4 Small double stranded RNAs and argonaute family proteins

RNA interference (RNAi) is the process by which double stranded RNAs (dsRNAs) are processed into short RNA duplexes 17 to 25 nucleotides long by the dsRNA-specific endonuclease Dicer. Short RNA duplexes are then included in multimeric protein complexes that are targeted by the incorporated RNA to cause either posttranscriptional gene silencing (PTGS) if the RNA is complementary to the mRNA sequence, or transcriptional gene silencing (TGS) if the RNA is complementary to a genes' promoter region. A central component of the complexes that mediate both PTGS and TGS is the argonaute family of proteins (Meister & Tuschl 2004, Matzke & Birchler 2005).

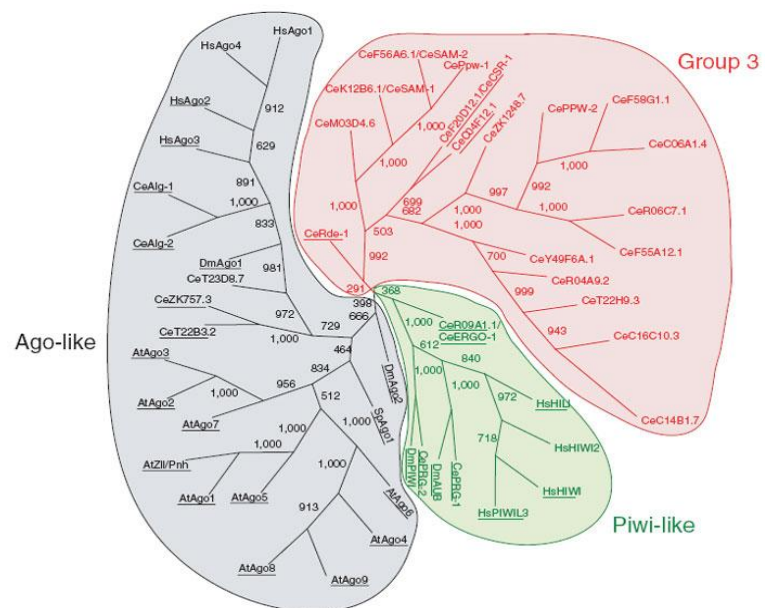
There are many sources for dsRNA substrates that are processed into short RNAs by dicer, including hybridisation of homologous sequences such as repeats, or sense/antisense pairs produced *in cis* (Numata *et al.* 2007, Meister & Tuschl 2004). Double stranded RNAs can also be produced by RNA strands that contain self complementarity; MicroRNAs (miRNAs) a species of conserved small RNAs with essential roles in maintenance of pluripotency and differentiation, are transcribed as long single stranded RNAs that self-hybridize to form a stem-loop structure, which are then cleaved to produce mature miRNAs.

The argonaute family is divided into three evolutionarily related groups (Figure 1.5). The AGO subfamily's role in gene silencing is conserved between species,

whereas the PIWI subfamily appears to have gained different roles when comparing mammals with *D. melanogaster* and *C. elegans* (Sasaki *et al.* 2003). In mammals, the AGO subfamily is ubiquitously expressed, while the PIWI subfamily is expressed mainly in the male germline - predominantly in spermatocytes, with some expression shown in non-differentiated cell populations, such as haemopoietic progenitors (Sasaki *et al.* 2003, Peters & Meister 2007). Mammalian PIWI proteins are important for spermatogenesis and appear to have a role in stem cell self renewal (Tolia & Joshua-Tor 2007). PIWI proteins in *Drosophila* are expressed in both male and female germ cells, and are thought to silence retroviral elements (Parker & Barford 2006). Group3 argonaute proteins are an evolutionarily distinct group arising in nematodes, and are thought to perform unique functions (Tolia & Joshua-Tor 2007).

Figure 1.5 Phylogenetic tree of Argonaute family.

The argonaute family is divided into three groups; AGO-like that are similar in structure to the protein Ago1 in *A. thaliana*; PIWI-like that are similar to the protein PIWI in *D. melanogaster*, and Group3, a nematode specific clade (Figure and information derived from Tolia & Joshua-Tor 2007).



The AGO subfamily of argonaute proteins is central to both transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) in *S.pombe*, *D.melanogaster*, and in mammals (Corey 2005, Grimaud *et al.* 2006, Sigova *et al.* 2004). *S.pombe* has a single AGO protein, Ago1, which mediates both TGS and PTGS (Verdel *et al.* 2004). The AGO family has expanded in higher

eukaryotes, with two AGO proteins in *D.melanogaster*, and four in mammals (Sasaki *et al.* 2003, Rehwinkel *et al.* 2006) where it is thought that the diversification of the AGO family is a result of specialization, with different AGO proteins responsible for TGS and PTGS (Sigova *et al.* 2004).

In mammals, AGO2 is the central component of the RNAi induced silencing complex (RISC), the complex responsible for PTGS. It binds to a small RNA duplex, and cleaves the sense strand while leaving the antisense strand intact to act as a guide. The antisense strand targets the RISC to complementary mRNA sequences, and AGO2 uses its catalytic 'slicer' activity to cleave the target mRNA. If there is not exact complementarity between the guide strand and the targeted mRNA, translational repression occurs instead; the RISC/mRNA relocates to a cytoplasmic body, a sight of mRNA degradation within the cytosol, which contains exonucleases and de-capping enzymes (Tolia & Joshua-Tor 2007, Liu *et al.* 2004).

TGS is well characterised in *S.pombe*, where RNA from heterochromatic repeats target an AGO protein-containing complex to the site of transcription where it mediates repression by promoting heterochromatin formation including methylation of H3K9 and recruitment of the *S.pombe* homolog of HP1 (Grewal & Elgin 2007, Hall *et al.* 2002, Volpe *et al.* 2002). Much less is known about TGS in mammals, where targeting of gene promoters with exogenously introduced short interfering RNAs (siRNAs), have been shown to cause a variety of changes to both histone modifications and DNA methylation status that results in differing degrees of repression.

An investigation into TGS in humans demonstrated siRNA dependent silencing of a reporter construct involving DNA methylation, with gene repression prevented by treatment with inhibitors of histone deacetylases and DNA methyltransferases (Morris *et al.* 2004). DNA methylation was again observed within the CpG island promoter of the gene *RASSF1A* where siRNAs caused partial DNA methylation, and reduced expression of the gene by half (Castanotto *et al.* 2005), and again in another study where siRNAs were targeted at promoter regions (suzuki *et al.* 2005). However, DNA methylation is not always observed; as siRNAs targeted at the CpG island containing promoter of the cancer suppressor gene *CDH1* did not cause DNA methylation but did result in gene repression via enrichment of the repressive histone modification H3K9me2 (Ting *et al.* 2005).

More recent work has demonstrated siRNA-mediated gene repression causing enrichment for both H3K9me2 and H3K27me3, as well as modest DNA methylation and demonstrated a requirement for the argonaute protein AGO1 in establishment of gene repression - suggesting that AGO1 is responsible for TGS in mammals (Hawkins *et al.* 2009, Kim *et al.* 2006). The involvement of AGO1 in TGS, and the similarities in silencing between *S.pombe* and mammals is discussed in detail in chapter 4.

1.9 Non-coding RNAs and repressive nuclear domains

The processes of X inactivation and Imprinting both utilize long ncRNAs to repress gene expression. The two processes have recently been shown to share many mechanistic features, with both X inactivation and silencing of

imprinted clusters involving ncRNA-mediated development of a repressive domain enriched for repressive histone modifications and devoid of transcriptional machinery. The main difference between X inactivation and imprinting is the scale of gene repression as, in X inactivation, the ncRNA mediated repressive domain grows to engulf the entire silenced X chromosome. Both mechanisms occur in a temporally similar manner; immediately after embryonic implantation, or within the first few days of mESC differentiation.

1.9.1 Imprinting

Imprinting is a form of epigenetic regulation where a gene is expressed from either the paternally or maternally inherited allele, with the other allele repressed early in postimplantation development. Around eighty imprinted genes have been identified with the majority occurring in clusters. Each cluster of imprinted genes is under the control of a *cis* acting regulatory element, or Imprint Control Region (ICR) which usually show differential methylation (Edwards & Ferguson-Smith 2007, Ideraabdullah *et al.* 2008). DNA methylation of the ICR occurs in a parentally specific manner during gametogenesis and is maintained during the wave of demethylation that occurs in the early pre-implantation embryo (Tilghman 1999, da Rocha & Ferguson-Smith 2004). Repression by DNA methylation controls the function of the ICR, and ultimately the transcriptional state of the genes in the imprinted cluster.

Two different mechanisms have been documented controlling gene expression at imprinted clusters. Most maternally methylated ICRs contain promoters for long and oppositely transcribed ncRNAs (Ideraabdullah *et al.* 2008). These paternally expressed ncRNAs cause imprinted gene repression via interactions

with histone methyltransferases and result in the repressive histone modification H3K9me3 and, in a few cases, DNA methylation (Pandey *et al.* 2008, Edwards & Ferguson-Smith 2007, Nagano *et al.* 2008, Lewis *et al.* 2004). Paternally methylated ICRs do not act as gene promoters, and cause imprinting in other ways (Edwards & Ferguson-Smith 2007). For example, at the Igf2 cluster the ICR lies between the two imprinted genes Igf2 and H19 and, when unmethylated, acts as an insulator to prevent enhancer elements from interacting with the Igf2 promoter, allowing the enhancer to promote H19 expression, but when the ICR is methylated it becomes incapable of acting as an insulator, resulting in expression of Igf2 and repression of H19 (Edwards & Ferguson-Smith 2007, Ideraabdullah *et al.* 2008, da Rocha & Ferguson-Smith 2004).

Imprinting clusters are a mix of genes imprinted in either embryonic tissues or extra-embryonic tissues, such as the placenta, and also genes imprinted in both tissue types. Repression of placentally imprinted genes occurs in a different but related manner compared to embryonic tissues, despite repression being regulated in both cases by the same ncRNA. This may be due to the differences in epigenetic regulation in these distinct tissues, as DNA methylation levels remain low in both trophoblast and the primitive endoderm while the ectoderm becomes progressively methylated post-implantation (Li 2002). It has been suggested that imprinting first evolved in extra-embryonic tissues, and was subsequently adapted in embryonic lineages, with DNA methylation occurring to stabilize repression (Lewis *et al.* 2004). The Igf2r and Kcnq1 imprinted clusters both have ICRs that contain a paternally expressed ncRNA that mediates

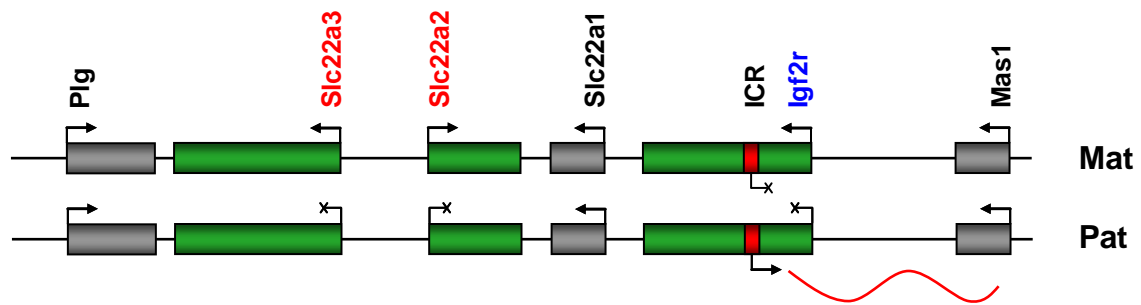
imprinting at the loci. In both cases, the ncRNA mediates repression through establishment of a repressive nuclear domain (Wu & Bernstein 2008).

1.9.1.1 The Igf2r cluster

The Igf2r cluster contains three maternally expressed genes, along with several non-imprinted genes, and the paternally expressed ncRNA Airn. The Airn promoter is within intron2 of the Igf2r gene, with Airn expressed in the antisense orientation to the protein coding gene. The genes Slc22a2 and Slc22a3 are both imprinted on the paternal allele in placental tissues, while the Igf2r gene is paternally imprinted in both the placenta and the embryo (Ideraabdullah *et al.* 2008, da Rocha & Ferguson-Smith 2004, Lewis *et al.* 2004). A map of the Igf2r cluster is shown in Figure 1.6A.

Airn is transcribed by pol II, and is both capped and polyadenylated, with only a small proportion of transcripts spliced. It is localised to the nucleus and is unstable compared to protein coding genes (Seidl *et al.* 2006). Airn is essential for the imprinting process, but its full length may not be required, as Igf2r can still become silenced even when the Airn transcript is truncated to only 3Kb in length (Sleutels *et al.* 2003). Nagano *et al.* examined Airn localisation in mouse placental tissue by joint DNA/RNA FISH, and observed Airn forming a 'cloud' like structure surrounding the imprinted gene Slc22a3. This was accompanied by enrichment for H3K9me3, which was also observed at Slc22a2, the other gene repressed only in placenta, but not at the ubiquitously imprinted Igf2r gene. Knockout of G9a abolished H3K9 enrichment, demonstrating that it was responsible for the repressive histone mark. Additionally; interaction between G9a and Airn was indicated by RNA ChIP at the Slc22a3 gene, suggesting that

(A) The Igf2r imprinted cluster



(B) The Kcnq1 imprinted cluster

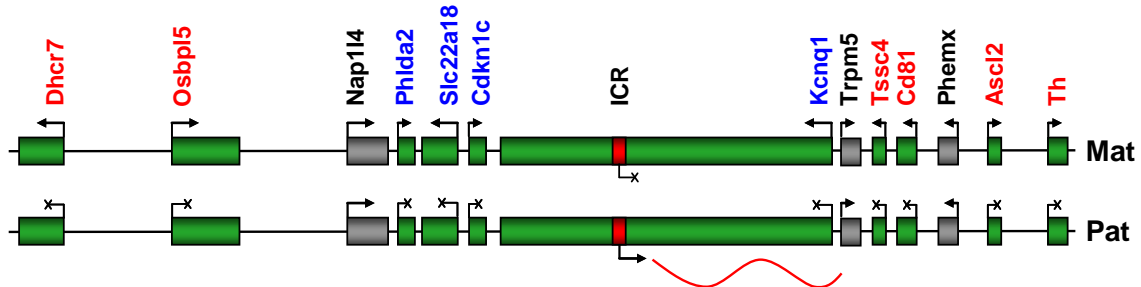


Figure 1.6 Paternally imprinted Loci. Schematic representation of the Igf2R (A) and Kcnq1 (B) imprinted clusters. Imprinted genes are green boxes, nonimprinted genes are grey boxes. The ICR is a red box. Genes in blue are imprinted in both the embryo and the placenta, genes in red are imprinted in the placenta only. Pat; paternal copy, Mat; Maternal copy. The ncRNA expressed from the paternal ICR is shown in red. Expressed genes are marked with an arrow, repressed genes are marked with a cross. Information derived from Terranova *et al.* 2008.

Airn may recruit the HMTase (Nagano *et al.* 2008). Further RNA FISH experiments carried out on trophoctoderm stem cells by Terranova *et al.* observed Airn co-localisation with the PRC1 component Rnf2, as well as the H2A119u1 mark catalysed by PRC1 (Terranova *et al.* 2008). Terranova *et al.* also observed some co localisation of Airn with H3K27me3 in both placental and embryonic tissues, contradictory to chromatin immunoprecipitation experiments carried out by other studies which saw no enrichment for this mark at imprinted genes in placental tissues (Latos *et al.* 2009, Regha *et al.* 2007), though H3K27me3 co localisation in the embryo was most likely a result of the

non-imprinted repression of *Slc22a2* or *Slc22a3*; neither are expressed on either allele in the embryo where both genes are contained within large H3K27me3 domains (Regha *et al.* 2007) in an analogous manner to silencing at HCP genes during development (Mikkelsen *et al.* 2007).

It is as yet unclear if *Airn* mediates imprinting within the embryo in the same manner as observed for *Slc22a3*. Paternally repressed *Igf2r* is enriched for H3K9me3 and the Hp1 isoform Hp1 β , and shows no enrichment for either H3K4 methylation or H3K9 acetylation (Regha *et al.* 2007), while the promoter region shows methylation at 70% of CpG sites examined (Stoger *et al.* 1993).

1.9.1.2 The *Kcnq1* cluster

The *Kcnq1* imprinted cluster is larger and more complex than the *Igf2r* cluster, though it shares many similar features of imprinted repression. The cluster contains a 91.5Kb ncRNA, *Kcnq1ot1*, which is paternally expressed in the antisense orientation to the protein coding *Kcnq1* gene. The cluster contains thirteen genes, with the four 'inner' genes closest to the site of ncRNA transcription paternally imprinted in both the embryo and the placenta, and six 'outer' genes located in the peripheral regions of the cluster that are paternally imprinted in the placenta only. There are also three non-imprinted genes scattered throughout the cluster that evade the varying forms of repression that silence the imprinted genes (Pandey *et al.* 2008, Wu & Bernstein 2008, Terranova *et al.* 2008). The layout of the *Kcnq1* cluster is shown in Figure 1.6B.

The ncRNA *Kcnqot1* is expressed by Pol II, is unspliced, and nuclear localised. In a similar manner to *Airn*, it also has a low level of stability in comparison to

mRNAs from protein coding genes (Redrup *et al.* 2009). *Kcnqot1* is essential for imprinting at the *Kcnq1* cluster, responsible for transcriptional repression of a 400Kb domain in embryonic tissues that encompasses the four ubiquitously imprinted genes, and a larger 780Kb domain in placental tissues that includes all ten placentally imprinted genes. *Kcnq1* ncRNA co-localises with imprinted genes right across the cluster (Pandey *et al.* 2008) and causes the formation of a repressive chromatin domain (Terranova *et al.* 2008, Redrup *et al.* 2009) involving both H3K9 and H3K27 trimethylation, though - as was observed at the *Igf2r* cluster - imprinting occurs in a different yet related manner in placental and embryonic tissues (Pandey *et al.* 2008, Lewis *et al.* 2004, Terranova *et al.* 2008, Umlauf *et al.* 2004).

Within the repressive chromatin compartment in placental tissues, Pol II and H3K4me3 are excluded and the repressive marks H3K9me3, H3K27me3 and H2A119u1 are enriched along with multiple components of both PRC1 and PRC2 (Terranova *et al.* 2008) but, unimprinted genes as well as the promoter for *Kcnq1ot1*, remain free of repression with the ncRNA promoter enriched for both Pol II and H3K4me3 (Pandey *et al.* 2008, Terranova *et al.* 2008). RNA chromatin immunoprecipitation showed co localisation of *Kcnq1ot1*, G9a, and PRC2 components Suz12 and Ezh2, suggesting that they may interact (Pandey *et al.* 2008). The roles played by PRC-mediated repression were investigated by depletion of Rnf2 and Ezh2, the catalytic components of, respectively, PRC1 and PRC2. In both cases there was no loss of silencing, despite loss of H3K27me3 or H2A119u1, but the silenced domain remained uncondensed compared to the more concentrated repressive foci observed in wild type cells,

suggesting that these modifications contribute to higher order restructuring of chromatin perhaps enhancing repression (Terranova *et al.* 2008).

Kcnqot1 marked out a similar but smaller domain in embryos, in keeping with repression of only the four 'inner' genes of the cluster (Redrup *et al.* 2009). While there was similar levels of H3K9me3 enrichment compared to the placenta, H3K27me3 levels, with the exception of the gene Cdkn1c, were greatly reduced suggesting that this mark plays much less of a role in imprinting of the embryonic Kcnq1 cluster (Pandey *et al.* 2008). Two of the four 'inner' genes have been shown to have paternal DNA methylation in both placental and embryonic tissues (Lewis *et al.* 2004), and interestingly Dnmt1 has been shown to play a role in repression of all four of the embryonic imprinted genes. Lewis *et al.* knocked out the catalytic domain of Dnmt1 and observed loss of imprinting of all four 'inner' genes, while genes imprinted only in the placenta were unaffected (Lewis *et al.* 2004).

The level of enrichment for repressive epigenetic modifications correlates with the duration for which the genes are imprinted. Of the four genes imprinted in the embryo only Cdkn1c remains imprinted throughout adult tissues, while the other three genes lose their imprinted status during differentiation (Umlauf *et al.* 2004), and Cdkn1c is the only gene to be strongly marked by DNA methylation, H3K9me3 and H3K27me3 in the embryo (Pandey *et al.* 2008, Lewis *et al.* 2004).

At the two clusters, sense and antisense RNAs overlap, Airn with Igf2r and Kcnq1 with Kcnq1ot1, which has led to the suggestion that dsRNA could form, which could be processed by Dicer into siRNAs that could target complexes capable of transcriptional gene silencing. However deletion of the Igf2r promoter, which caused complete loss of Igf2r expression and therefore any potential dsRNA formation, had no impact upon the imprinting within the cluster (Sleutels *et al.* 2003), while imprinting at the Kcnq1 cluster occurred as normal when a cell line with a conditional Dicer knockout was used (Redrup *et al.* 2009). Both of these observations make it clear that it is the long ncRNA that mediates repression, and that it does so without help from the RNAi system.

Repression at both the Igf2r and Kcnq1 clusters is reliant upon a ncRNA that is capable of coating the region and establishing a repressive domain that, in embryos, involves both DNA methylation and H3K9 trimethylation. These discrete foci do not spread to the surrounding region and are highly selective, as they do not silence non-imprinted genes within the same cluster, nor do the surrounding repressive modifications block transcriptional elongation of these genes (Nagano *et al.* 2008, Terranova *et al.* 2008, Regha *et al.* 2007).

1.9.2 X inactivation

X inactivation, the method by which mammals ensure dosage compensation between the sex chromosomes, involves silencing of one X chromosome during early female development (Reik & Lewis 2005). Repression of the silenced X chromosome (Xi) is caused by expression of a long ncRNA, the X inactive specific transcript (Xist) (Nesterova *et al.* 2003), which promotes the rapid loss of transcription machinery and enrichment for the repressive histone

modifications H3K27me3, H2A119u1, as well as dimethyl and trimethylation of H3K9 (Ideraabdullah *et al.* 2008, Chow & Heard 2009). Extensive methylation of CpG islands within gene promoter regions also occurs on the Xi, though slightly later in development suggesting that it is required more for long term maintenance (Lee & Jaenisch 1997).

1.9.2.1 The X inactivation control centre

Silencing of the X chromosome involves the interplay of multiple different ncRNAs, all expressed from the X inactivation control centre (Xic) (Latham 2005). A schematic of the Xic is shown in Figure 1.7.

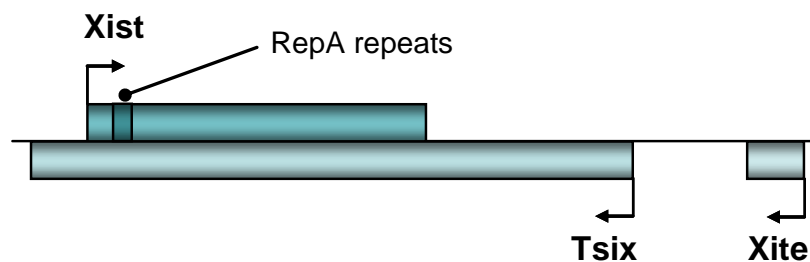


Figure 1.7 The X inactivation control centre. Xist produces a mature transcript 17Kb long that is both spliced and polyadeylated. The RepA repeat is located within exon one of the Xist gene. Tsix regulates Xist by transcription of a 40Kb antisense ncRNA though the Xist gene. Tsix expression is regulated by the 1.2Kb enhancer Xite in a developmental manner. Information derived from Ohata *et al.* 2008; Stavropoulos *et al.* 2005; Wutz *et al.* 2007; Zhao *et al.* 2008.

The ncRNA Tsix is 40Kb in size, and transcribed in the antisense orientation through the entirety of the Xist gene, including the promoter. Tsix transcription through Xist is essential in order to control Xist expression, with truncation of Tsix resulting in aberrant Xist upregulation (Ohhata *et al.* 2008). Tsix is regulated in turn by a region upstream of the Tsix promoter, Xite, which transcribes many ncRNAs at low levels in the same orientation as the Tsix gene, some of which overlap the Tsix promoter (Ogawa & Lee 2003). Xite, X-

inactivation intergenic transcription elements, acts as an enhancer and promotes expression of Tsix in mESCs but not in differentiated cell lines (Stavropoulos *et al.* 2005). Transcription of ncRNAs from Xite through the Tsix promoter region is not required for Tsix regulation (Ogawa & Lee 2003). The Xist transcript is 17Kb long, and is both spliced and polyadenylated. Another ncRNA is produced from the Xist region, from within the first exon. This region contains the RepA repeats that produce a 1.6Kb transcript transcribed in the sense orientation to Xist, and is thought to hold the key to the initiation of X inactivation as well as the means of control that Tsix maintains over Xist in undifferentiated cells (Zhao *et al.* 2008).

1.9.2.2 The RepA repeats

Immediately prior to upregulation of Xist, the promoter and gene body of Xist becomes enriched for H3K27me3 (Sun *et al.* 2006). This exceptional requirement for enrichment of a histone modification that is unambiguously associated with transcriptional repression has yet to be satisfactorily explained, and suggests that Xist could belong to a unique class of genes that require a heterochromatic environment for full expression (Sun *et al.* 2006). Alternatively, enrichment for uncharacterised histone modifications that perhaps modify the effect of this normally repressive mark may occur, or further regulatory steps, as yet unidentified, are involved in Xist upregulation that require the presence of this repressive mark.

H3K27me3 is catalysed by the PRC2, which is recruited through direct interaction with the RepA repeats. RepA consists of a series of tandem repeats that form a conserved dual stem-loop structure (Figure 1.8). This structure is

directly bound by the catalytic subunit of PRC2, Ezh2 (Zhao *et al.* 2008). The RepA stem-loop structure is also present in both Xist and Tsix ncRNAs, and it is the competition between the Xist and 1.6Kb RepA transcripts on one side, verses the Xist transcripts on the other that is thought to control X inactivation (Zhao *et al.* 2008). In mESCs, Xist transcript levels are low, around ten copies per cell, while Tsix levels are in excess, and it has been suggested that this disparity allows the Tsix transcript to out-compete the others in the recruitment of PRC2 complexes (Zhao *et al.* 2008). Alternatively, the Tsix transcript may act directly to block transfer of PRC2 onto chromatin thereby preventing catalysis of H3K27me3 within the Xist gene. In support of this second theory, RepA/PRC2 complexes were observed in ESCs but PRC2 was not enriched on Xist DNA until differentiation occurs - suggesting that binding was blocked (Zhao *et al.* 2008).

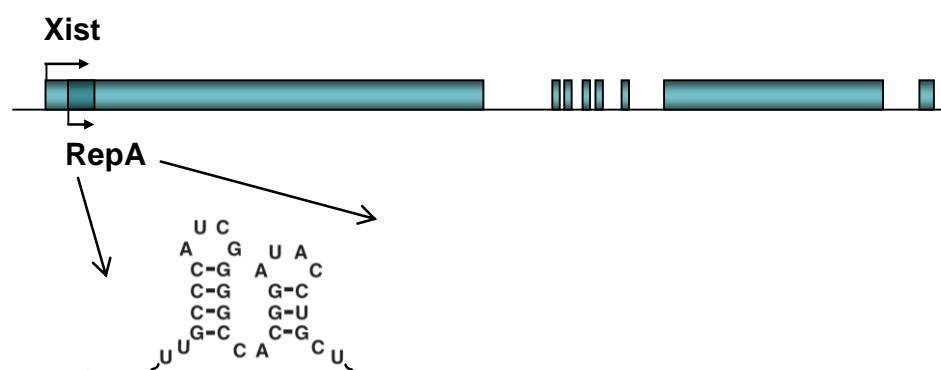


Figure 1.8 Xist and the RepA repeats. The Xist gene consists of seven exons, which contain a number of repeat sequences. The RepA repeats within exon1 produce a transcript ~1.6Kb long, and contains a series of 28nt repeats that form a dual stem-loop structure that physically interacts with the PRC2 component Ezh2. The RepA repeats are the epicentre of PRC2 binding during X inactivation. Information from Chow and Heard 2009; Zhao *et al.* 2008.

1.9.1.3 Initiation of X inactivation

X inactivation occurs immediately after post-implantation prior to gastrulation, or within the first few days of *in vitro* differentiation of ESCs (Wutz 2007, Chaumeil *et al.* 2006). The Xic regions of the two X chromosomes physically interact, which is thought to allow a form of 'counting' to occur (Wutz 2007). Immediately after interaction, initiation of X inactivation occurs, with down-regulation of Tsix and upregulation of Xist by up to 100 fold on the Xi. RepA transcript levels also increase, though only by a modest 1.8 fold (Zhao *et al.* 2008). Xist recruits PRC1, independent of PRC2 recruitment by RepA, and both H3K27me3 and H2A119u1 are rapidly enriched on the Xi following induction of Xist (Zhao *et al.* 2008, Schoeftner *et al.* 2006). Enrichment for H3K9me2 occurs at the same time and with similar kinetics, but which HMTase is responsible or how it is recruited has not yet been shown, though G9a has been implicated and is essential for maintenance of H3K9 methylation (Rougeulle *et al.* 2004). Xist coats the X chromosome creating a silencing domain that then expands, with genes relocating inside and becoming repressed. The domain rapidly excludes Pol II, and other transcription factors, becoming enriched for repressive histone modifications at the same time as histone acetylation and H3K4me3 is lost (Chaumeil *et al.* 2006). In *in vitro* differentiation, silencing is virtually complete by d4, in both embryo body and Retinoic acid differentiation (Chaumeil *et al.* 2006, Keohane *et al.* 1996).

Prior to X inactivation, the promoter region of Xist displays partial methylation on both X chromosomes, with DNA methylation of Xist becoming lost on the Xi

and complete on the active X chromosome (Xa) through recruitment of Dnmt3a by Tsix, which has been shown together in the same complex (Sun *et al.* 2006). DNA methylation on the wider X chromosome is present before silencing occurs, but shows a differential pattern between the Xi and the Xa upon differentiation. DNA methylation on the Xi occurs within CpG islands at gene promoters, enhancing gene repression, while DNA methylation on the Xa occurs within the bodies of genes or within intergenic regions, and is inconsequential for gene expression (Hellman & Chess 2007, Weber *et al.* 2005, Norris *et al.* 1991). DNA methylation occurs later in X inactivation than enrichment for repressive histone modifications, and is essential for maintenance rather than initiation of X inactivation, with DNA methylation thought to stabilize repression and help maintain silencing throughout the cell cycle (Lee & Jaenisch 1997, Heard *et al.* 1997).

1.9.1.4 Maintenance of X inactivation

The inactive X chromosome is enriched for H3K27me3, H3K9me2 and me3, H2A119u1, H4K20me3 and the histone variant macroH2A (Ideraabdullah *et al.* 2008, Chow & Heard 2009). The Xi displays perinucleolar localization soon after silencing, and replicates late in S phase. While lack of expression of Xist does not appear to effect repression in the short term (Ogawa & Lee 2003), deletion of the portion of the Xic containing the Xist promoter and the RepA repeats results in 'erosion' of heterochromatin, causing complete loss of H3K27me3 and the re-expression of repressed genes indicating that continued expression of Xist, the RepA transcript, or both, are essential for maintenance of X inactivation (Zhang *et al.* 2007).

A role for the siRNA pathway in X inactivation has long been suggested, with a recent study demonstrating the presence of small RNAs within the Xic centre, though these were in the size range of 25 to 27nts (Ogawa *et al.* 2008), which is greater in size than the short RNAs normally associated with the RNAi machinery. Interestingly these small RNA species, found in both sense and antisense orientations, were located at the RepA repeat region (Ogawa *et al.* 2008). Using an Rnase A protection assay Ogawa *et al.* also detected a small number of double stranded RNAs corresponding to the RepA region, however these were most likely the stem-loop structures later characterised by Zhao *et al.* 2008. A central role for the RNAi machinery appears now to have been ruled out by Kanellopoulou *et al.*, who deleted Dicer in mESCs and demonstrated that it had no effect upon either establishment or maintenance of silencing on the X chromosome, in both an inducible system in mESCs and in mESCs partially differentiated using retinoic acid, though curiously the extent of coating of the inactive X by Xist was reduced in a significant manner in Dicer KO cells - indicating that Dicer may be involved in some subsidiary manner (Kanellopoulou *et al.* 2009).

1.10 Pluripotency factors in pre-implantation development

The transformation of the zygote from a mass of cells into an early embryo with a distinct internal structure occurs in a series of well defined stages. First the zygote undergoes cell division resulting in the morula, a ball of cells, which then takes on a defined structure referred to as the blastocyst: The outer morula cells compact to form the trophectoderm, an outer layer of extra-embryonic epithelium that will develop into the yolk sac and provide the embryo with nutrients during the early stages of growth, while the inner cells of the morula

develop into the blastocyst cavity that contains the Inner Cell Mass (ICM). The ICM subsequently divides into two lineages, primitive ectoderm that will go on to form all the different tissues of the embryo, and the primitive endoderm that will give rise to the placenta (Li 2002, Rossant & Tam 2009).

Early development is tightly controlled by the transcription factors which are expressed in a specific temporal manner, with Oct4, Nanog, and Sox2 all playing central roles in the development of the ICM (Cavaleri & Scholer 2003, Nichols *et al.* 1998, Chambers *et al.* 2003, Mitsui *et al.* 2003, Avilion *et al.* 2003). Oct4 is essential for the development of the first cells of the embryonic lineage, with maternally inherited protein present in the zygote, and embryonically transcribed Oct4 detectable during the eight cell stage of the morula, just prior to blastocyst formation (Rossant & Tam 2009, Nichols *et al.* 1998). Oct4 expression is maintained in pluripotent early embryo cells in the ICM until implantation (Nichols *et al.* 1998). Nanog is expressed at a slightly later stage, and is detectable in the compacted morula, in cells that will become the primitive ectoderm of the ICM. Nanog expression continues throughout pre-implantation, restricted to the ectoderm and its subsidiary tissues, before being down-regulated at the implantation stage in a similar manner to Oct4 (Chambers *et al.* 2003, Mitsui *et al.* 2003). Sox2 is also expressed in the late morula and then within the ICM, though it also plays a role in extra-embryonic development, with expression detectable in the primitive endoderm as well as the ectoderm (Avilion *et al.* 2003).

The three transcription factors show a large degree of overlap in the genes they regulate. A global screening approach by Boyer *et al.* using microarrays and gene expression data revealed that together, Oct4, Nanog, and Sox2 regulate the transcriptional state of over 2000 genes (Boyer *et al.* 2005). Approximately half of these genes are actively expressed and include many transcription factors involved in pluripotency and components of cell signalling pathways such as TGF β or Wnt, while the other approximately one thousand genes are repressed, with many of these genes encoding transcription factors involved in lineage commitment and tissue specific development (Boyer *et al.* 2005). Interestingly; both Oct4 and Nanog have been shown to associate with components of the NuRD or Sin3A repressor complexes, and are therefore capable of directly targeting repressive complexes to mediate gene repression (Liang *et al.* 2008).

Embryonic Stem Cells, ESCs, are derived from ectodermal cells of the ICM of the pre-implantation embryo, and are considered to be the *ex vivo* counterpart of early stem cells in the embryo. ESCs are effectively immortal, having the ability to undergo symmetrical cell division and therefore indefinitely self-renew. They are also pluripotent; they have the ability to differentiate into all foetal and adult lineages (Smith 2001). ESCs when maintained in an undifferentiated state represent the pluripotent cells of an embryo just prior to implantation, but with any further development suspended. As such, ESCs express the same pluripotency markers as observed in the ICM, with high levels of Oct4, Nanog and Sox2 expression detectable (Nichols *et al.* 1998, Mitsui *et al.* 2003, Boyer *et al.* 2005).

1.11 A novel case of gene repression by CpG island methylation

The human alpha globin gene cluster is located on chromosome 16 (16p13.3), and spans approximately 30 kb. It includes five protein coding genes and two pseudogenes (Flint *et al.* 1997). A schematic of the alpha globin cluster is shown in Figure 1.9A. The cluster contains two copies of the alpha globin gene, HBA1 and HBA2, therefore in a diploid genome there are four copies of the gene. The HBA genes code for the alpha globin chain of the oxygen carrier hemoglobin.

Hemoglobin is composed of two alpha globin and two beta globin chains, with any imbalance in the production of alpha or beta globin chains potentially resulting in assembly of hemoglobin incorporating an incorrect ratio of alpha or beta globin chains. Under production of the alpha globin chain results in alpha thalassaemia, caused by the relative overabundance of beta globin chains resulting in hemoglobin that is inefficient at delivering oxygen to tissues. Alpha thalassaemia is normally the result of genomic deletion within the alpha globin locus leading to the loss of one or more HBA genes. Loss of a single alpha globin copy does not result in an obvious phenotype with the individual normally unaware of the deletion, whereas loss of two copies results in alpha thalassaemia trait that causes mild anemia (Nussbaum *et al.* 2007).

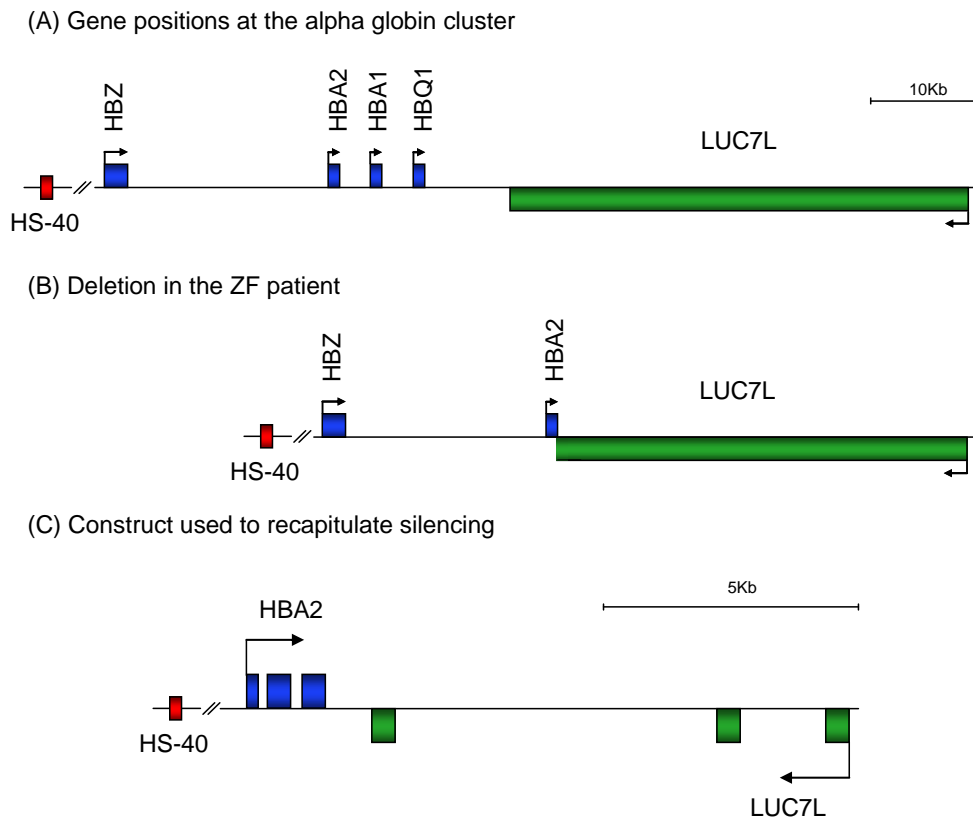


Figure 1.9 The alpha globin locus. (A) Layout of the alpha globin cluster showing the four main genes, the HS-40 alpha globin regulatory element, and the *LUC7L* gene transcribed antisense to the other genes on the cluster. The HS-40 regulator is situated 40Kb upstream of HBZ (B) Shows the alpha globin cluster in the patient after deletion of a 18.3Kb section of the cluster that has removed *HBA1*, *HBQ1*, and the last three exons of the *LUC7L* gene including the transcriptional terminator. (C) The construct (ZF alpha antisense) used to recreate the silencing effect *in vivo* in a mouse model and *in vitro* in a differentiating mESC system, containing the HS-40 2Kb upstream of the HBA2 gene and the promoter and first three exons of the *LUC7L* gene. The distance between the HS-40 and the HBA2 gene is approximately 60Kb in its genomic context. Information from Tufarelli *et al.* 2003 and Higgs *et al.* 1990.

An individual was discovered with only three extant copies of the alpha globin gene due to an aberrant recombination event within their maternally inherited alpha globin cluster, which had resulted in the loss of an ~18Kb section, including loss of HBA1, and causing juxtaposition of a truncated *LUC7L* gene next to the remaining alpha globin gene, *HBA2*. This is shown in Figure 1.9B (Barbour *et al.* 2000). The individual was suffering from alpha thalassaemia trait, normally only associated with the loss of function of two alpha globin

genes. Following investigation it was revealed that the remaining copy of the alpha globin gene, *HBA2*, on the altered copy of chromosome 16 was repressed, with very little transcription detectable. Gene repression was occurring *in cis*, with the alpha genes on the normal copy of chromosome 16 unaffected (Barbour *et al.* 2000). Initial analysis by Barbour *et al.* revealed dense CpG island methylation at the alpha globin promoter within peripheral blood samples, as well as in cell lines created through fusion with the aberrant chromosome 16, with the only exception being spermatocytes where no DNA methylation was detected. Despite dense DNA methylation there was no observed alteration in the timing of replication of the repressed alpha globin loci. Normal DNA methylation patterns were re-established 2Kb upstream of the *HBA2* TSS, demarcating the extent of the silencing effect (Barbour *et al.* 2000).

The deletion removed the terminator and last three exons of the widely expressed *LUC7L* gene, suggesting that transcription from *LUC7L* could occur through the alpha globin gene in the antisense orientation. The *LUC7L* gene normally produces a 41Kb pre-mRNA transcript that is spliced to produce a 1.9Kb mature transcript. An antisense RNA transcribed through the alpha globin gene from *LUC7L* was detected in both EBV transformed lymphocytes from the patient, as well as in mouse erythroleukemia (MEL) cells incorporating the aberrant chromosome 16 (Tufarelli *et al.* 2003). A construct, ZF α Antisense, containing the alpha globin gene, with the *LUC7L* gene included in the antisense orientation (Figure 1.9C), was then used to recapitulate gene repression in an *in vivo* mouse model, where gene silencing was shown to occur in the early embryo, and was accompanied by expression from *LUC7L*,

with the antisense transcript being spliced in a normal manner. In contrast; no repression occurred in a construct where the alpha globin and *LUC7L* genes were both in the sense orientation and antisense transcription though the alpha globin promoter did not occur (Tufarelli *et al.* 2003). Tufarelli *et al.* further characterized CpG island methylation at the alpha globin promoter detecting methylation as early as day 7.5 in the yolk sac as well as in all adult tissues examined in the founder population. Additionally in a mouse model, gene expression occurred in an unrepressed manner from an incorporated BAC transgene that contained the entire alpha globin cluster, demonstrating that silencing of the alpha globin gene was occurring *in cis* and was not affecting other copies of the alpha globin that were present (Tufarelli *et al.* 2003).

The constructs used in the creation of transgenic mice were used to create stable mESC lines that were then differentiated into embryoid bodies (EBs) in order to examine gene repression during the early stages of differentiation. Both alpha globin and *LUC7L* were expressed in mESCs without repression, but after differentiation for 7 days, in the absence of LIF and in the presence of factors that promote hematopoietic development, the alpha globin gene became repressed, with DNA methylation of its CpG island (Tufarelli *et al.* 2003), and enrichment for the repressive histone modification H3K9me3, along with loss of both histone acetylation and H3K4me3 (Tufarelli *et al.*, unpublished data). Additionally; expression of antisense RNA was shown to halt approximately 2Kb upstream of the alpha globin gene in the same region where DNA methylation patterns returned to their normal distribution (Barbour *et al.* 2000, Tufarelli *et al.*, unpublished data).

The LUC7L gene contains many repeat sequences, with the gene body region insensitive to Dnase1 digestion suggesting it may have a closed chromatin environment (Flint *et al.* 1997, Tufarelli *et al.* 2001). To investigate if repression of the alpha globin gene was related to the juxtaposition of this potentially heterochromatic environment next to the gene, the entire LUC7L gene was replaced with the 1.2Kb promoter region of the ubiquitously expressed gene UBC, as defined by Schorpp *et al.* (Schorpp *et al.* 1996). This did not effect repression, with gene silencing occurring upon differentiation when antisense RNA was transcribed, indicating that repression was not a result of the apparent heterochromatic environment of the *LUC7L* gene body (Tufarelli *et al.* 2003).

1.12 Aims of the thesis

Recent discoveries concerning the extent of genomic transcription, coupled with the increasingly vast number of annotated ncRNAs have transformed the manner in which the genome, and more importantly its regulation, is viewed (Kapranov *et al.* 2007, Mattick 2005). This reconsideration of the mechanisms by which genes are expressed or repressed has occurred apace with the global analysis of histone modification patterns within the many different tissues and developmental stages of mammalian organisms (Mikkelsen *et al.* 2007, Barski *et al.* 2007). Taken together; recent advances reveal an epigenetic landscape that utilizes multiple combinations of epigenetic modifications to fine-tune control over individual genes, with an expanding number of ncRNAs being identified that are central to the targeting of these modifications in various examples of repression (Rinn *et al.* 2007, Nagano *et al.* 2008, Terranova *et al.* 2008, Zhao *et al.* 2008, Yu *et al.* 2008).

The complexes that mediate gene repression have been revealed to be both multimeric and multifunctional, with histone methyltransferases, DNA methyltransferases and histone deacetylases frequently found within the same regulatory complexes, with the different components often capable of direct interaction with each other (Epsztejn-Litman *et al.* 2008, Esteve *et al.* 2006, Li *et al.* 2007, Lehnertz *et al.* 2003), thus ensuring a multifaceted approach to gene repression. Additionally; components of these various repressive complexes are now being demonstrated to interact directly with, or to occur in the same complex as, long ncRNAs that target silencing at their respective loci (Rinn *et al.* 2007, Nagano *et al.* 2008, Terranova *et al.* 2008, Zhao *et al.* 2008). Given the increasing rate at which both ncRNAs and the epigenetic landscape are being annotated, many more examples of RNA mediated repression are likely to be found.

The work described in this thesis has used a differentiating mouse embryonic stem cell model system to further explore the unique silencing effect observed at the alpha globin locus.

Previous work has shown that repression of the alpha globin gene occurs following at least seven days of *in vitro* differentiation, and that repression is accompanied by DNA methylation and H3K9me3 enrichment (Tufarelli *et al.* 2003). A time course study was conducted to further characterize the timing of gene repression, and allow the relationship between loss of gene expression and the alteration in epigenetic modifications to be studied in detail to compare

how repression at the alpha globin locus compares to other characterized silencing events that occur within the same time period. Additionally; the requirement for antisense transcription through the alpha globin gene for gene repression was examined (Chapter 3).

There is an increasing number of long ncRNAs now being identified that are directly involved in gene repression, but small RNA duplexes, siRNAs, are also capable of mediating gene repression when targeted at gene promoter regions. In mammals, the main component of the silencing complex responsible is Ago1 (Hawkins *et al.* 2009, Kim *et al.* 2006), with various incidences of promoter targeted siRNAs causing DNA methylation and H3K9 methylation (Kim *et al.* 2006, Castanotto *et al.* 2005, Morris *et al.* 2004). With expression of both alpha globin and LUC7L in undifferentiated cells there is potential for the formation of dsRNA that could then be processed into siRNAs by Dicer1. To address the potential role of this mechanism in silencing of the alpha globin gene, mESC cells carrying the ZFaAS transgene were stably transfected with constructs expressing hairpin RNAs directed to Ago1 and changes in the degree of silencing at the alpha globin assessed and correlated to the reduction in Ago1 RNA levels. In addition, the presence of small RNA species corresponding to the alpha globin gene were for tested by northern blot (Chapter 4).

Silencing of the alpha globin gene may be a unique event as a result of specific regulatory sequences, or other features particular to the alpha globin gene, its sequence, or its erytheroid specific regulation. To investigate whether a tissue specific gene from a different developmental lineage is equally susceptible to

this form of silencing, the skeletal muscle specific transcription factor *MYOD* has been incorporated into a construct that allows *LUC7L* transcription through the *MYOD* gene's promoter region in a manner analogous to the alpha globin silencing construct (Chapter 5).

Both the alpha globin and *MYOD* genes are normally subject to repression in tissues in which they are not expressed, and are therefore already capable of the recruitment of, or being targeted by, silencing factors when they are to be repressed. This ability may be a prerequisite for repression of the kind observed by Tufarelli *et al.* To examine this possibility and further characterize the abilities of this form of gene repression the ubiquitously expressed genes *UBC* and *ACTB* were incorporated into constructs in the same manner as *MYOD* to investigate whether such widely expressed genes that are not normally subject to repression can be silenced in the same manner as observed for the alpha globin gene (Chapter 6).

Chapter 2 Materials and Methods

2.1 Chemicals and reagents

All commonly used stocks, solutions and buffers were prepared as outlined in Current Protocols in Molecular Biology, Ausubel *et al.* 1994 – 1998; or Sambrook *et al.* 1989. Unless otherwise stated, all chemicals were supplied by Sigma Aldrich, analytical grade or higher.

2.2 Bacterial cultures

The ultracompetent bacterial strain XL10 GOLD produced by Stratagene was used as it was suitable for the propagation of plasmids with large inserts rich in CpG dinucleotides.

2.2.1 Storage and revival of bacterial strains

To create a glycerol stock; 350µl of bacteria growing in LB media with the appropriate antibiotic was added to 350µl sterile 30% glycerol solution in a screw top 1.5ml Eppendorf tube, vortexed to ensure mixing, then stored at -80°C. To revive the bacterial strain, a sterile needle was used to remove approximately 50µl of the glycerol stock and used to inoculate an LB agar plate containing the appropriate antibiotic. The plate was then incubated overnight at 37°C.

2.2.2 Culturing Bacterial cells for miniprep

A colony was removed from the plate with a sterile toothpick and used to inoculate a 12ml sterile tube containing 5ml LB media containing the

appropriate antibiotic. The tube was then incubated overnight at 37°C in a shaking incubator.

2.2.3 Extraction of DNA from Bacterial Strains.

All extraction methods are variations of the alkaline lysis purification method as described in Sambrook *et al.* 1989.

2.2.3.1 Miniprep

The Qiagen Plasmid Miniprep kit was used, as per the manufacturer's protocols. A more rapid protocol was also carried out when screening large numbers of minipreps and when purifying BACs and PACs. Briefly; a 1.5ml tube was filled with prepared culture, then centrifuged for 5min at 7000rpm. The supernatant was discarded and 100µl of solution P1 (50mM Glucose, 25mM Tris, pH8, 10mM EDTA pH8 +1µl of 10mg/ml Rnase A per ml) was added, the cell pellet re-suspended, and left for 5min at room temp. 200µl of solution P2 (200mM NaOH, 1% SDS) was added, the tube inverted several times and incubated on ice for 5min. 150µl of solution P3 (3M potassium acetate, pH 5.5) was added, the tube inverted several times, and then centrifuged at 13000rpm for 5min. The supernatant was moved to a new tube containing 350µl isopropanol, vortexed, then centrifuged for 15min at 13000rpm. The supernatant was discarded, and the pellet washed with 500µl 70% ETOH, then centrifuged at 13000rpm for 5min. The supernatant was then discarded and the pellet allowed to dry. The pellet was then re-suspended in 50µl TE.

2.2.3.2 Maxiprep

Large scale plasmid DNA extractions were performed using the Qiagen Plasmid Maxiprep kit according to the manufacturer's protocols. These protocols are based on a modified alkaline lysis purification method, followed by binding of plasmid DNA to an anion-exchange resin under appropriate salt and pH conditions. RNA, proteins, and low molecular weight impurities are removed by a medium salt wash. Plasmid DNA is then eluted in a high salt buffer, and concentrated/de-salted by isopropanol precipitation.

2.3 Phenol/chloroform extraction of DNA

50% of the samples volume in phenol, and 50% of the samples volume in chloroform was added, and shaken vigorously. The sample was centrifuged for 5min at 13000rpm. The aqueous phase was moved to a new tube. 100% of the samples volume in chloroform was added, the tube shaken vigorously, then centrifuged for 5min at 13000rpm. The aqueous phase was moved to a new tube and 10% 3M NaOAc pH5.5 and 200% volume of 100% ETOH added. The tube was inverted 10-15 times. If no pellet formed, it was placed at -20C for 30mins. The sample was centrifuged for 30mins at 13000rpm and the supernatant discarded. The pellet was washed with 200µl of 70% ETOH, and the sample was centrifuged for 15min at 13000rpm. The supernatant was then discarded and the pellet allowed to dry. The pellet was then re-suspended in 50µl TE.

2.4 Engineering of DNA constructs

2.4.1 Restriction digest of DNA

DNA was digested with Restriction Endonucleases (REs) from New England Biolabs (NEB) unless otherwise stated. Manufactures guidelines were followed when setting up restriction digests. Plasmid DNA was digested for 1 to 5 hours. Genomic DNA was digested for 3 hours to O.N. PCR products were digested for between 1 to 2 hours. When required, either P/C or QIA PCR Purification Kits were used to clean up samples following restriction digest.

2.4.2 Amplification of large regions by PCR

Long range PCR was carried out using 1µl of a 1/100 dilution of BAC/PAC DNA prepared by the miniprep method described in 2.2.3.1. A 10µl PCR reaction contained; 0.9µl 11.1xAJJ buffer (0.5M TRIS pH8.8, 122mM Ammonium sulphate, 50mM MgCl₂, 75mM Beta-mecaptoethanol, 0.05mM EDTA, 11mM dNTPs, 1.2mM BSA), 0.06µl Taq (Abgene 5U/µl), 0.02µl Pfu (NEB 2U/µl), 10µM forward and reverse primers, 0.7M Betaine, and 1ul of BAC/PAC dilution DNA template. A PCR cycle with a extended elongation time was used as shown below:

94°C	2min	
94°C	10s	} 30 cycles
annealing temp	30s	
68°C	8min	
68°C	20min	
15°C	1min	

After optimisation of primer annealing temperatures, large scale PCR reactions were used to produce sufficient quantities of PCR product. 5x100µl reactions were set up and run as stated. 2µl of each reaction was analysed by gel electrophoresis then reactions were combined together and cleaned up by P/C.

2.4.2 Blunt-ending of a linearised plasmid

Linearised plasmid was incubated with 1µl Klenow (NEB) per 5µg DNA, 1x final conc. Klenow buffer, and 33µM final conc. dNTPs for 20min at 37°C, then heat inactivated for 10min at 65°C.

2.4.3 Dephosphorylation of plasmid DNA

Plasmid DNA was incubated with Shrimp Alkaline phosphatase (SAP) and 1x final conc. 10xSAP buffer for 1hour at 37°C. The number of units of SAP added was calculated using the formula; (µg DNA/Kb DNA) x3.04. SAP was then inactivated by 15min at 65°C.

2.4.4 Gel electrophoresis and purification

DNA was combined with 1xfinal concentration bromophenol blue loading dye (6x concentration; 3.5g sucrose, 4ml 10xTBE, 0.025g bromophenol blue in a final volume of 10ml dH₂O) and loaded onto a 1xTBE agarose gel. Percentage concentration of agarose was dependent upon size of DNA fragment to be observed. Generally, for fragments above 1kb a 0.8% agarose gel was used, between 1Kb and 500bp a 1-1.5% gel was used, while DNA fragments smaller than 500bp were resolved on a 2% agarose gel. Voltage applied varied depending upon size of gel tank used.

Following electrophoresis, gels were stained by emersion in a 2.5µg per ml solution of ethidium bromide for 10min with gentle agitation before being imaged on a UV transiluminator. Bands of interest were excised using a scalpel blade and purified using the QIAquick gel extraction kit and following the provided protocol

2.4.5 Ligation of insert into vector

Concentration of vector and insert were determined by gel electrophoresis. For blunt ended ligations a ratio of 3:1 of insert to construct were prepared. For sticky ended ligations a ratio of 1:1 was used. The formula; $\text{ng of insert} = ((\text{ng of vector} \times \text{size of insert in Kb}) / \text{size of vector in Kb})$ was used to correct for the different sizes of the insert and vector. Ligations were carried out in a total volume of 10µl, with 1µl T4 ligase (NEB), 1µl 10xT4 ligase buffer. Ligation mixes were incubated at 16°C overnight. Negative control ligations were set up in an identical manner but without the addition of the DNA insert.

2.4.6 Transformation of plasmid into bacterial cells

100µl of XL10 Gold cells were thawed on ice, then 4µl of β mercaptoethanol (βME) was added. The cells were incubated for 10 minutes on ice. 52µl of cells were transferred into two Eppendorf tubes and 6µl of ligation mix, and the negative control, was added to the tubes, and mixed. The cells were then incubated on ice for 30min. After incubation, the cells were heat-shocked for 30s at 42°C then immediately placed on ice for 2min. 250µl of LB media, warmed at 37°C, was added and the solution incubated in a shaking incubator for 1hour at 37°C. The solution was then plated out onto LB agar plates

containing an antibiotic to allow selective growth. Platings were at 50µl, 100µl, and 150µl.

2.4.7 PCR screening of transformed bacterial cells

A 96well PCR plate was loaded with 50µl of dH₂O per well. A toothpick was used to transfer some of the colony to a well on the PCR plate, then to inoculate a fresh LB agar reference plate with a numbered grid. All colonies were transferred in this manner. The PCR plate was then sealed and heated at 100°C for 5min. It was then cooled on ice and pulse spun. 5µl was then used as the template for a PCR reaction screening for incorporation of the construct into the plasmid. The reference plate was then used to retrieve and grow colonies positive by PCR.

2.5 Growth and maintenance of mESC lines and EBs

2.5.1 Growth and maintenance of mESCs

Mouse embryonic stem cells are derived from the inner cell mass of early mouse pre-implantation embryos (*Evans & Kaufman 1981, Martin 1981*). The work in this thesis was carried out using the mouse embryonic stem cell line E14Tg2a (sub-line E14, Thioguanine resistant, well 2a) derived from day 3.5 blastocysts from the mouse strain 129/Ola (*Hooper et al. 1987*). In this study, ESCs were differentiated down the hematopoietic lineage through the addition of specific factors that enhance hematopoiesis; A final concentration of 4.5×10^{-4} M MTG was added to differentiation media as it has been demonstrated to enhance hematopoietic activity, as well as increase the plating efficiency of EBs whereby less ESCs need to be seeded as EBs appear less dependent upon

surrounding cell density to develop. (Wiles & Keller 1991). The same study demonstrated a similar but lesser effect with Ascorbic acid. Iron saturated transferrin is added for its normal physiological role as an iron transporter essential for erythrocyte development.

The differentiation method described (2.5.5 and 2.5.9.6) has been shown to be highly reproducible, with the timings associated with hematopoietic development shown to remain constant across cell lines and experiments. Additionally, hematopoietic development mirrors hematopoiesis *in vivo*, with the development of mesoderm-like tissue which then produces hematopoietic cells. Approximately 85% of EBs show differentiation down a hematopoietic cell lineage, though in any culture around 10% of EBs fail to develop properly and remain small (Keller et al 1993). EBs contain a mix of different cell types, including endothelial and muscle lineages besides the hematopoietic lineage (erythroid, myeloid and lymphoid lineages) (Keller 1995).

In this study, ESCs were grown on tissue culture grade plastics from NUNC or Greiner Bioline. A 0.1% Gelatin solution in PBS (2.5.9.7) was used to produce a layer to which the ESCs adhered. In order to maintain ESCs in an undifferentiated state, cells grown in Complete ES Media (2.5.9.1) were passaged every second day, and were not allowed to reach a density of greater than 70% confluence. ESCs were maintained in an incubator at 37°C, with a CO₂ level of 5%.

2.5.2 Passaging of ESCs by trypsinisation and disaggregation

Media was aspirated from the ESC monolayer, and the cell layer briefly washed with PBS. Trypsin (Trypsin-EDTA 1X, liquid - 0.25% Trypsin, 1mM EDTA 4Na Invitrogen 25200) was then added and the cells incubated at 37°C for 5min. Complete ES media was added and the cell suspension passed three times through a 18G needle to disaggregate cell clumps. The ESC suspension was then centrifuged for 5min at 1000rpm, and the supernatant removed. The cell pellet was re-suspended in complete ES media (37°C) and transferred to a gelatinised flask.

2.5.3 Revival of ESCs from frozen down aliquots

Aliquots of ESCs were stored in liquid nitrogen. An aliquot was removed and warmed in a 37°C water bath until thawed. The sample was then centrifuged for 5min at 1000rpm, then the supernatant was removed. The pellet was re-suspended in complete ES media (37°C) and transferred to a gelatinised flask containing complete ES media (37°C). The flask was gently shaken to ensure an even layer of cells.

2.5.4 Long term storage of ESC lines

ESCs were prepared as described in 2.5.2. Following centrifugation and removal of supernatant, cells were re-suspended in Freezing Media (2.5.9.2) at a concentration of no more than 1×10^6 cells per ml. Aliquots of 1.5ml were transferred to 2ml cryovials (Sarstedt). Vials were then placed in cotton-wool padded polystyrene containers and transferred to -80°C. After 48hrs, cryovials were transferred to liquid nitrogen storage tanks.

2.5.5 Differentiation of ES cells into Embryoid Bodies

ESCs, grown as described above, were cultured for 24hrs in Preparation Media (2.5.9.4). Cells were then passaged (2.5.2), with the exception that Differentiation Media 1 (2.5.9.5) was used to neutralise the trypsin in place of Complete ES media. 10cm Petri dishes were prepared with 10ml of Differentiation media 2 (2.5.9.6) and 1.5×10^5 cells were added to each dish and distributed evenly by a gently swirling action. Petri dishes were then incubated at 37°C, with a CO₂ level of 5%. To prevent adherence of developing EBs to the surface of the Petri dish, dishes were gently but thoroughly shaken every 24hrs. After 7 to 8 days of differentiation EBs were harvested. The contents of the Petri dish was transferred to a collection tube and centrifuged at 1000rpm for 5min. The supernatant was removed and the pellet was either stored at -80°C or used immediately in the preparation of DNA or RNA. Level of differentiation within the culture was assessed in two ways; firstly by the number of EBs that had developed blood islands (it was expected that 80%+ would show blood island development for differentiation to be termed successful), and secondly by assessing the level of the pluripotency transcription factors Oct4 and Nanog (2.6.5).

2.5.6 Electroporation of DNA constructs into ESCs

2.5.6.1 Electroelution of DNA construct for electroporation

Constructs were excised from plasmids by restriction digest, then separated from other fragments by gel electrophoresis using a 0.8% 300ml 1xTAE agarose gel. Electrophoresis was carried out at 40v O.N in a 3.5L gel tank containing 1xTAE buffer. A portion of the gel containing a DNA ladder and a

fraction of the lane containing the construct was separated from the main gel and stained (2.4.4). This was used as a guide to excise the fragment of the gel containing the construct. Following excision, the gel slice was sealed inside a section of dialysis tubing (MEDICELL International Ltd. Tube preparation: Heated at 100°C for 5min in 2% Sodium Bicarbonate, 1mM EDTA, then washed for 5min at 100°C in dH₂O, and stored at 4°C). The Gel fragment then underwent electrophoresis at 140V for 4hrs. The current was reversed for 30s to detach DNA that had become adhered to the dialysis tubing, then the liquid contents was transferred to a 1.5ml Eppendorf and cleaned up by phenol/chloroform extraction (2.3)

2.5.6.2 Electroporation by BioRad Gene Pulser Xcell (165-2660)

A minimum of 1.2×10^7 ESCs were cultured and prepared as a single cell suspension (2.5.2). The cells were centrifuged for 5min at 1000rpm, and re-suspended in PBS. The cells were then pelleted and re-suspended in PBS a second time to remove all traces of Complete ES Media. The cells were re-suspended at a concentration of 2×10^7 cells per ml. An electroporation curvette (4mm gap, Biorad) was then prepared with 600µl of cell suspension in PBS, ~25µg of DNA construct (2.5.4.1), and 2µg of a 1859bp fragment of the pPNT plasmid containing the geneticin (G418) resistance gene. Electroporation was then carried out at 450V with a capacitance of 25µF, and resistance (Ω) set to infinity. This produced a time constant of 0.5. The contents of the electroporation curvette was made up to 6ml and 1ml was pipetted into six 10cm NUNC tissue culture dishes prepared with 0.1% gelatin layer containing 10ml of Complete ES media. After 24hrs, the media was removed and replaced

with complete ES media + G418 at a final concentration of 0.3mg/ml. Subsequently the media was changed every second day for the next 6 days.

2.5.6.3 Electroporation by Eppendorf Multiporator (4308 000.015)

7.5×10^6 ESCs were washed in PBS twice, and re-suspended in 750 μ l of PBS, as described for the BioRad Gene Pulser 2.5.4.2. The ESCs were then loaded into a Biorad electroporation curvette (4mm gap), ~50 μ g of DNA construct was added, and the solution gently mixed. Electroporation was carried out at 450v, and with a time constant of 100 μ s. The contents of the electroporation curvette was mixed with 50ml of Complete ES media, and 10ml was distributed into five, gelatinised, 10cm NUNC tissue culture dishes. They were incubated for 24hrs, then the media was replaced with fresh media containing Puromycin at a final concentration of 2 μ g/ml. Media was subsequently changed every second day for a total of 8 days.

2.5.6.4 Selection of ESC colonies for further culturing

After 8 to 10 days growth, ESC colonies were visible to the naked eye. Colonies were selected based on their undifferentiated morphology as determined by monophase microscopy. Complete ES media was removed from the plates, and the colonies were washed with PBS twice, leaving ~5ml of PBS behind. Colonies were then removed from the plates by a pipette set to 15 μ l. Each colony was placed in 50 μ l of 0.25% trypsin for 5min. The cell suspension was then transferred to a gelatinised well of a 24 well plate containing 1ml of complete ES media + antibiotic and cell lines cultured (2.5). When passaging cell lines with a low number of cells, cells were not pelleted following

trypsinisation, but had 1ml Complete ES media added and were directly transferred to prepared wells of a 24well plate.

2.5.7 Extraction of DNA from ESCs and EBs

ESCs were collected following trypsinisation and stored as a cell pellet at -80°C before proceeding with DNA extraction using Lairds Lysis buffer with 50µg/ml Proteinase K (described in 2.5.10). Alternatively, media was removed from ESC monolayers, the cell layer was washed with PBS, then Lairds Lysis buffer with 50µg/ml Proteinase K was added directly onto the cell layer. DNA extraction of EBs was carried out using Tail Lysis Buffer with 100µg/ml Proteinase K (described in 2.5.10). After addition of lysis buffer, ESC and EB samples were incubated at 37°C O.N. Depending upon application for extracted DNA, an optional treatment of 25µg/ml Rnase A was added and the sample incubated at 37°C for 2hrs.

Samples were then transferred to 15ml polypropylene tubes for phenol/chloroform extraction. 50% of the samples volume in phenol, and 50% of the samples volume in chloroform was added, and shaken vigorously. The sample was centrifuged for 15min at 4000rpm. The aqueous phase was moved to a new tube. The P/C extraction was then repeated. After the aqueous phase was moved to a new tube, 100% of the samples volume in chloroform was added, the tube shaken vigorously, then centrifuged for 15min at 4000rpm. The upper phase was transferred to a fresh tube and 10% volume of 4M NaCl, and 2x volume of 100% ETOH was added. For EB samples, a volume of dH₂O equal to a quarter of the starting volume of lysis buffer was added. The tube was inverted until a cloud of DNA became visible, the precipitated DNA was

then transferred to a 1.5ml tube using a Pasteur transfer pipette, briefly centrifuged at 13000rpm, then the supernatant discarded and the pellet washed with 70% ETOH. If DNA did not precipitate upon inversion due to low concentration, the sample was stored at -20°C for 1hr then centrifuged for 30min at 4000rpm, the supernatant discarded, and the pellet washed with 70% ETOH. After washing with ethanol the DNA pellet was allowed to dry for ~ 20min at RT, then a suitable volume of TE was added and the sample left at RT overnight to re-suspend.

2.5.7.1 Extraction of DNA using phase lock tubes

Qiagen 15ml maXtract low density tubes (129025) were used during the extraction process of both ESC and EB DNA where cell pellets were $\sim 5 \times 10^6$ cells. Following cell lysis in 500 μ l of lysis buffer, and optional Rnase A treatment (2.5.7), samples were transferred to phase lock tubes. 500 μ l of phenol/chloroform solution (50% Phenol, 49% Chloroform, 1% Isomyalcohol) was added and the tube vigorously shaken. The sample was then centrifuged for 5min at 4000rpm. An additional 500 μ l of P/C solution was added, the sample mixed, and then centrifuged for 5min at 4000rpm. 500 μ l chloroform was added, the sample mixed, then centrifuged for 10min at 2000rpm. The upper phase was then poured off into a 15ml tube, and the DNA precipitated as detailed in 2.5.4.5.

2.5.8 Extraction of RNA from ESCs and EBs

All chemicals/equipment used was RNA Grade or RNA sterile and samples were kept on ice unless otherwise stated. ESCs and EBs were harvested and stored as pellets in 15ml polypropylene tubes. RNA extractions were carried out

on freshly collected material, and on material stored at -80°C using TRI reagent according to manufacturers instructions. Briefly; TRI Reagent was added, 1ml for every $\sim 1 \times 10^7$ cells, the tube was then vortexed to dissolve the cell pellet, and the sample left at RT for 5min. A volume of chloroform equal to a fifth of the starting volume of TRI reagent was added, the tube vortexed, and left at RT for 10min. The sample was then centrifuged for 30min at 1000rpm at 4°C . The aqueous phase was transferred to a 1.5ml Eppendorf containing a volume of isopropanol equal to half the starting volume of TRI reagent and vortexed. It was then left for an hour at -20°C , then spun for 30min at 4000rpm at 4°C . The supernatant was removed, the RNA pellet washed in 70% ETOH, and then centrifuged for 10min at 4000rpm at 4°C . The supernatant was discarded and the pellet left to dry. 100 μl dH_2O was added and the sample incubated for 10min at 65°C . At this stage the sample could be stored at -80°C .

If the sample had been frozen, it was defrosted on ice. The 100 μl of sample was transferred to a PCR 0.2ml PCR tube and 12 μl Dnase1 10xbuffer, and 6 μl Dnase 1 (Roche) was added, along with 2 μl Rnase inhibitor plus (Promega N2611). The sample was then incubated at 15°C for 45min. The sample was then transferred into a 1.5ml tube and the volume increased to a total of 400 μl by addition of dH_2O . The sample was cleaned up by P/C (2.3), with the exception that all centrifuge steps were at 4°C , and the P/C step was carried out twice to ensure removal of trace amount of phenol. Following washing of the RNA pellet with 70% ETOH, the sample was left to dry for $\sim 20\text{min}$, then re-suspended in an appropriate volume of dH_2O and incubated for 10min at 65°C . Samples were stored at -80°C .

2.5.9 Media used in the propagation of ES cells and EBs

2.5.9.1 Complete ES Media

DMEM + glutaMAX +4.5g L-Glucose -Pyruvate (Invitrogen 61965)	394ml
MEM NEAA x100 (Invitrogen 11140)	5ml
Sodium Pyruvate 100mM (Invitrogen 11360)	5m
Penicillin/Streptomycin 100x (Invitrogen 11140)	5ml
Fetal calf serum	90ml
Beta-mecaptoethanol, (100mM liquid)	500µl
LIF (Millipore ESG1107)	50µl
<i>(Beta-mecaptoethanol; final conc. of 2x10⁻⁴M)</i>	

Fetal calf serum (FCS) used during this investigation was purchased from Sera Laboratories International Ltd. Two batches were used; 304005 and 509005. The majority of work was carried out using Batch No. 509005.

2.5.9.2 Freezing media

9ml of FCS
1ml of DMSO
Filter sterilize using 0.4µm filter

2.5.9.3 IMDM+P/S+MTG

To a 500ml bottle of IMDM + L-Glu +25mM HEPES (Invitrogen 21980) the following was added;
5ml P/S
62µl of a 1:10 dilution of 11.2M MTG stock

2.5.9.4 Preparation Media

To 85ml of IMDM+P/S+MTG
add:
15ml FCS
10µl LIF

2.5.9.5 Differentiation Media 1

To 95ml of IMDM+P/S+MTG add:
5ml FCS

2.5.9.6 Differentiation media (100ml)

15ml FCS

1ml L-Glutamine (stock 100mM, final conc. 100µM)

1ml Transferrin (Roche 10625202001, 30mg/ml)

200ul of MTG (26ul of 11.2M stock diluted in 2ml IMDM)

1ml Ascorbic acid (Stock at 5mg/ml. final conc. 5µg/ml)

5ml Protein free hybridoma medium II (Invitrogen 12040-051)

76.8ml IMDM+P/S+MTG

(Final conc. of MTG: $3.97184 \times 10^{-4}M$)

2.5.9.7 Gelatin 0.1% (100ml)

5ml Gelatin, 2% in water, TC grade (Sigma G1393)

95ml PBS -CaCl₂ -MgCl₂ pH7.4 (Invitrogen 10010)

2.5.10 Cell Lysis buffers

Lairds Lysis Buffer:

100mM Tris

200mM NaCl

5mM EDTA

0.2% SDS

Tail Lysis buffer:

50mM Tris

100mM NaCl

100mM EDTA

1% SDS

2.6 Polymerase chain reaction (PCR)

The PCR method was used to amplify DNA fragments for a variety of purposes, including; cloning, generation of DNA probes, screening of bacterial clones and ESC lines, to detect DNA methylation by methylation sensitive PCR, and in conjunction with reverse transcription to screen for gene expression and for expression of antisense RNA. PCR was also used in conjunction with immunoprecipitation to determine enrichment levels for different histone modifications. Reaction conditions were optimised for each primer pair, and their specific conditions, including buffer used, annealing temperature and elongation time are displayed in the tables in the relevant chapters' Materials and Methods section. A standard PCR program is shown below:

94°C	2min	
94°C	30s	} 30 cycles
annealing temp	30s	
72°C	30s	
72°C	5min	
15°C	1min	

2.6.1 PCR Buffers

A range of different PCR buffers were used, depending upon the GC content of the template to be amplified. PCR amplification of templates with a GC content of less than ~60% were carried out in a volume of 25µl with 2.5U of Taq (Abgene or NEB), 0.2µl of 25mM dNTP mix (Aversham), 25pMol Primers, 50ng to 100ng template DNA and NEB standard 10xPCR buffer, or Abgene buffer IV (10x, with supplemented MgCl₂ to a final concentration of 1.5mM). Templates with GC content greater than ~60% were amplified using a 25µl PCR mix composed of 50% 2xDMSO buffer, 2.5U of Taq (NEB), 0.3µl of 25mM dNTP mix, 25pMol Primers, 50ng to 100ng template DNA, and 1.75µl of 25mM MgCl₂. These mixes are described below.

NEB 10x PCR buffer	Abgene buffer IV (10x)	2xDMSO buffer
100mM Tris-HCl 500mM KCl 15mM MgCl ₂	750mM Tris-HCL pH8.8 200mM (NH ₄) ₂ SO ₄ 0.1% w/v Tween 20	32mM ammonium sulphate 134mM TRIS pH8 20% w/v DMSO 20mM Beta-mecarptoethanol <i>Stored at 4°C in the dark</i>

Primer	Sequence, 5' to 3'	PCR buffer	Annealling temp °C	Notes
mAPRT2	GGAAATCCAGAAAGATGCCT	2xDMSO NEB buffer	60	Amplify a region of the mouse <i>Aprt</i> gene, 225bp (cDNA) or 335bp (genomic DNA). Also used for realtime PCR with the settings 95°C 10min, (95°C 15s, 60°C 15s, 72°C 15s) x40 cycles, expert mode 1+4. 25ng input.
mAPRT3	TCTAGCCAGCTCCTCAGTCA			
CT53	TCAGTTTACGGTTCCCCTTG	2xDMSO	60	Primers amplify 248bp region of <i>LUC7L</i> that contains 3 MspI/HpaII sites. Used as a control for complete digestion in msPCR
CT54	GCTTCAGCCTCCTTCCTATG			
Luc7LSYBR F	TCAGGGTTGGGATTTTGGTTCC	2xDMSO	60	Primers amplify a 144bp region of the <i>LUC7L</i> CpG island that does not contain any MspI/HpaII sites. Used in multiplex PCR in conjunction with the PCR primer pair AlphaF/ 493 amplifying a 295bp region of the alpha globin gene with 3 MspI/HpaII sites
Luc7LSYBR R	TAGTTTCTGGGGTGGGTGCTTAGG			
M114	ATTCTGGTTGTGCCGAGTTGCGAG	2xDMSO	60	Primers amplify a 489bp region of the Airn promoter which contains 3 MspI/HpaII that is maternally methylated in oocytes and all tissues of the embryo. Used as a control for detectable DNA methylation in msPCR
M115	CGTGGAACCTTTTGAAGTTCATCTCTC			
mACTB F	TCAAGATGGACCTAATACGGCTTT	ABI SYBR green	60	Primers amplify genomic DNA of the mouse Beta actin gene. Used as control for ChIP analysis. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
mACTB R	CGGTGTGGGCATTTGATGA			
436	CACCAAGGGTGACTTATAGAGCTGG	2xDMSO	60	Primers amplify 208bp region of <i>LUC7L</i> that contains no MspI/HpaII restriction sites. Used as positive control for msPCR
437	GCAACTCTCTTTTTTAAGGGTGGG			
mAPRT1	GGTAGCTCACAAAGGTCAC	-	-	Primer used to prime reverse transcription reaction

Table 2.1. general primer list

2.6.2 Electrophoresis of PCR products

For 25µl PCR reactions; following PCR, 5µl of bromophenol blue loading dye was added to each sample, and 20µl of sample loaded. Unless otherwise stated, electrophoresis was carried out using TBE agarose gels. PCR products less than ~600bp were run on 2% agarose gels, while PCR products in the range of approximately 600bp to 1kb were run on 1.5% agarose gels. EtBr staining was carried after electrophoresis by immersion in a 2.5µg/ml EtBr solution for 10min with gentle agitation and rinsed in water for 10min before being imaged on a UV transilluminator. For electrophoresis of PCR products for msPCR and multiplex PCR, where further analysis was to be carried out on gel images, a standardized protocol was followed. Samples were loaded onto 300ml 2% TBE gels. Electrophoresis was carried out in a 3.5L gel tank containing 3.5L of 1xTBE at 65v for 1hr, then 85v for 3hrs. Gels were stained with EtBr as described, then run at 85v for an additional 20min. Gels were photographed on a U.V transilluminator.

2.6.3 Methylation sensitive PCR (msPCR)

DNA was digested by HpaII, a methylation sensitive RE, then PCR was carried out using primers that flanked restriction sites. Only in the presence of DNA methylation, preventing restriction, would the template be intact and capable of priming a PCR reaction. As a control, an identical digestion was carried out using MspI, an isoschizomer of HpaII, that is not sensitive to DNA methylation and should always restrict the DNA template and prevent a PCR product from being produced.

500ng DNA was digested with 6ul of HpaII (10U/μl), and 20ul NEB buffer 1 in a total volume of 200μl for 2hrs at 37°C, then an additional 4μl of HpaII was added and the digest left O.N at 37°C. An identical digest was carried out using MspI, with the initial digest set up using 3μl MspI (20U/μl) and NEB buffer 2, with an additional 2μl of enzyme added after 2hrs at 37°C. Samples were then left O.N at 37°C. The REs were inactivated by heating at 65°C for 20min. DNA was then precipitated by addition of 400μl 100% ETOH, 20μl 3M sodium acetate, and 1μl glycogen. Samples were placed at -20°C for 1hr then centrifuged at 13000rpm for 30mins, the supernatant removed and the DNA pellet dried for ~20min at RT. Samples were then re-suspended in 50ul, and 3μl used for subsequent PCR reactions.

2.6.4 Methylation sensitive multiplex PCR

A multiplex PCR approach was developed to assess DNA methylation at the alpha globin locus. The PCR reaction amplified two regions of the alpha globin construct using two sets of primer pairs that amplified regions that had similar GC contents. A region of the alpha globin CpG island that contained three MspI/HpaII cut sites to assess methylation levels, and a region of the LUC7L gene that contained no restriction sites for MspI/HpaII. Following DNA Digestion and preparation (2.6.2), PCR was carried out using the standard PCR program (2.6), with an annealing temperature of 60°C in 2xDMSO buffer (2.6.1). and 25pMol of each primer. For each sample, the multiplex PCR was run in triplicate. Following PCR, electrophoresis was carried out (2.6.2) and gel images were analyzed using AIDA gel analysis software, with a local background set. The ratio of intensity for the band denoting DNA methylation was normalized to the intensity of the control band. This was done for all three

triplicate repeats. The mean value and standard deviation was then calculated, and normalized to 2A3 Wt levels.

2.6.5 Multiplex PCR for pluripotency factors

To assess the level of differentiation within populations of ESCs and EBs, a multiplex PCR combining primers for expression of the Oct4 and Nanog pluripotency factors, and for the ubiquitously expressed gene Aprt was developed. A standard PCR reaction (2.6) was carried out in NEB standard PCR buffer (2.6.1) with 25pMol of each primer added to the reaction mix. cDNA produced by a random primed reverse transcriptase reaction (2.7) was used as a template. 1µl of freshly prepared cDNA was used, or 1.5 to 2µl of previously prepared and frozen cDNA. The multiplex PCR was carried out in triplicate for each sample. Following PCR, electrophoresis was carried out (2.6.2) and gel images were analyzed using AIDA gel analysis software, with a local background set. The ratio of the intensity of the bands for Oct4 and Nanog were then normalized to the intensity of the band for the Aprt PCR product. For each sample, the mean and standard deviation of the ratios in triplicate for Oct4 and Nanog were calculated.

2.6.6 Realtime PCR

Realtime PCR was carried out to assess gene expression levels, and to quantify enrichment levels for different histone modifications using DNA enriched by chromatin immunoprecipitation. For all realtime experiments, an Applied Biosystems 7500 fast machine was used on standard setting, running SDS software version v2.0 or v2.0.1. Plastic consumables were purchased from Applied Biosystems. Standard curve data was produced for all realtime primer

pairs, threshold values were determined manually and normalized between plates, all realtime PCRs were run in triplicate, with the mean value from all three replicates (unless otherwise stated) being used for analysis by the standard curve method in Microsoft Excel.

For relative gene expression studies using both Taqman and SYBR green based chemistries, the mean Ct value for the gene of interest (GOI) was converted to a value in nanograms using data from the Standard Curve for that primer pair. The mean value of the GOI was normalized to the mean value of the endogenous control (ENDO), derived in the same manner as for the GOI. The Standard Deviation (STDEV) of the triplicate values (in nanograms) for the GOI and the ENDO was also calculated and combined together to provide a complete expression of error for the combined data of the GOI and the ENDO. These formulae are shown below.

Raw Ct data was converted to an amount in nanograms:

$$10^{\left[\frac{\text{mean CT value} - b}{-X} \right]}$$

Where “^” denotes “to the power of”, b is equal to the value at which the Standard Curve line intercepts the Y axis, and X is equal to the slope of that line (see section 2.6.6.1). The raw value for the GOI in ng is then normalized to the ENDO:

$$\text{Normalise} = \frac{\text{Value (ng)}_{\text{GOI}}}{\text{Value (ng)}_{\text{ENDO}}}$$

The STDEV of the replicate group for the GOI was calculated as described, then combined together with the STDEV of the ENDO:

$$Cv = \sqrt{Cv_{GOI}^2 + Cv_{ENDO}^2}$$

Where Cv_{GOI} and Cv_{ENDO} can be calculated by:

$$Cv = \frac{STDEV}{\text{mean value (ng)}}$$

This produces the combined formula:

$$Cv = \sqrt{\left[\frac{STDEV_{GOI}}{\text{Mean value (ng)}_{GOI}} \right]^2 + \left[\frac{STDEV_{ENDO}}{\text{Mean value (ng)}_{ENDO}} \right]^2}$$

In Excel, individual Cv values were calculated, then combined using the formula “=SQRT(SUMSQ(Cv_{GOI}, Cv_{ENDO}))” Once calculated, the combined Cv value was converted back to a STDEV:

$$\text{Combined STDEV} = Cv \times \text{normalised value}$$

2.6.6.1 Creation of a Standard Curve

Standard Curves were created by serial dilution of realtime PCR template, either genomic DNA, or cDNA produced by a random primed reverse transcriptase reaction. Serial dilutions generally covered a range of input values from ~50ng to ~0.01ng, though the effective range at which primers functioned was specific to each primer pair. Generally five different points within the this dilution range were assayed. Mean Ct values (Y axis) were plotted against the log of the template input amount in ng (X axis). An example of a Standard Curve is shown in Figure 2.2 below. In this example the slope of the line (x) is -

3.4388, and the point at which the line cuts the Y axis (b), representing the average CT value with 1ng of input material, is 29.128.

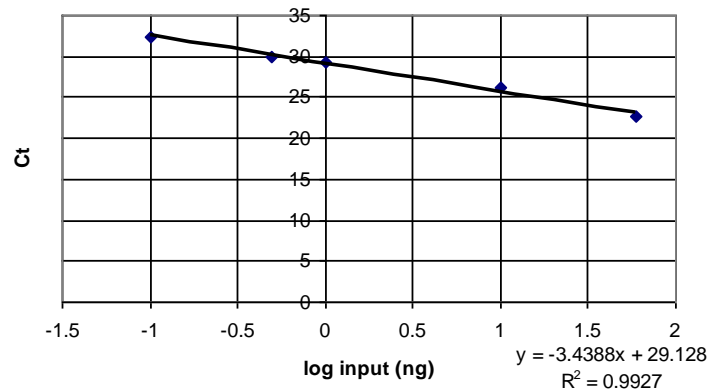


Figure 2.2 example of a Standard Curve

The value of the slope of the line (x) defines the efficiency of the reaction, with a value of -3.32 representing a theoretically 100% efficient PCR reaction as this is the number of cycles required for a 10 fold increase in product ($10 \text{ to } \log(2) = 3.321928$), so on a semi-log graph as shown in Figure 2.2, with exponential values converted to linear values using base 2, a 100% efficient reaction would double every 3.32 cycles. A larger negative value than -3.32 represents a PCR reaction that is not 100% efficient, which is normal for most primer pairs, while a value less negative than -3.32 indicates a reaction operating a greater than the theoretical maximum, suggesting the presence of more DNA in the standard curve serial dilution than thought.

2.6.6.2 Taqman probe based detection

For the quantification of gene expression of Ago1, two inventoried taqman realtime PCR assays were purchased from Applied Biosystems; Ago1 (Mm00462977_m1) and an assay for the endogenous control gene Gapdh (Mm99999915_g1). Reactions were set up in a total volume of 20µl with 10µl of

TaqMan[®] Universal PCR Master Mix No AmpErase[®] UNG (Applied Biosystems), 1µl of Assay, 1µl of dH₂O, and the equivalent cDNA of 10ng of the input RNA used for the reverse transcriptase reaction in a volume of 8µl. A PCR program of 95°C for 10min, and 40 cycles of: 95°C for 15s, 60°C 1min was used.

Conversion of RNA to cDNA was carried out using the Roche Expand reverse transcriptase kit (11785826001). 2µg RNA and 1µl of random primers at 65ng/µl (Promega C118A) was made up to a total volume of 24.4µl, and equally divided between two 0.2ml PCR tubes. One tube was used for the reverse transcription reaction (+), the other tube as the negative control (-). Tubes were incubated on a PCR block at 70°C for 10min, then left at RT for 10min. To the + tube; 4µl 5xbuffer, 2µl DTT, 0.4µl 25mM dNTP mix, 0.4µl Rnase Inhibitor Plus (Promega N2611), and 1µl (50U) Expand RT was added. The same reagents were added to the - tube with the exception of Rnase inhibitor and Expand RT, which were replaced with 1.4µl dH₂O. After mixing, the samples were incubated at 37°C for 1hr. cDNA was used immediately for realtime PCR.

2.6.6.3 SYBR Green dye based detection

SYBR Green dye based detection was used for both quantification of gene expression and for quantification of enrichment for histone modifications following chromatin immunoprecipitation. For gene expression studies, cDNA was prepared by random primed reverse transcription (2.7). Realtime PCR reactions were carried out in a total volume of 20µl, with 10µl Power SYBR[®] Green PCR Master Mix (Applied Biosystems), 0.8µl of a 5µM primer mix of forward and reverse primers, 1.2µl dH₂O, and 8µl of template dilution in the

range of 10ng to 50ng cDNA. Primers were designed using the recommended parameters for Primers 3 software (Applied Biosystems). Following primer design, primers were tested using the NCBI primer blast online database for non-specific binding, and tested using the apple based application MacVector for predicted primer dimer formation and annealing temperature. Primers were experimentally tested for their species specificity, whereby primers designed for human genes were tested for ability to amplify from a mouse template, and vice versa. Primer pairs were also tested for the ability to form primer dimers by melt curve analysis following realtime PCR amplification. Initial primer testing was carried out using 50ng of template, and with a standard realtime PCR program of 95°C for 10min, and 40 cycles of: 95°C for 15s, 60°C 1min. For primer pairs that displayed non-specificity under these conditions, alterations to annealing temperature and elongation time were made to ensure specificity. If non-specificity issues persisted, new primers were designed. The PCR program used for each primer pair is noted in the appropriate chapter.

2.6.6.4 SYBR green analysis of ChIP samples

DNA (2.9.4) was diluted so that 0.25µl of column elute was used in each real time PCR reaction. Real time PCR analysis was carried out as described (2.6.6, 2.6.6.1, 2.6.6.3). For analysis; calculated amounts (ng) were normalized to Input for each sample (2.9.3) in order to cancel out variation between cell lines due to copy number variation, then to the endogenous control ACTB to normalize for varying amounts of material within each immunoprecipitation reaction.

2.7 Reverse transcription

RNA was extracted (2.5.8) and the Roche Expand reverse transcriptase kit (11785826001) used for both random primed and strand-specific reverse transcription reactions. For strand-specific reverse transcription; 4µg of RNA was combined with 0.4µl of specific primer (25µM) and 0.4µl of internal control primer (25µM) and made up to a total volume of 25.6µl using dH₂O. The mix was divided equally between two 0.2ml PCR tubes. One tube was used for the reverse transcription reaction (+), the other tube as the negative control (-). Tubes were incubated on a PCR block at 70°C for 10min, then placed on ice. 4µl 5xbuffer, 2µl DTT, 0.4µl 25mM dNTP mix, 0.4µl Rnase Inhibitor Plus (Promega N2611), and 0.4µl (50U) Expand RT was added to the + tube. The same reagents were added to the - tube with the exception of Rnase inhibitor and Expand RT, which were replaced with 0.8µl dH₂O. Tubes were mixed and incubated at 42°C for 1hr. Prepared cDNA was then used immediately for subsequent PCR reactions, or frozen at -20°C. Random primed cDNA reactions were carried out in the same manner, with several exceptions; 2µl of random primers at 65ng/µl (Promega C118A) were added in place of strand-specific primers; following incubation at 70°C, samples were left at room temperature for 10min; and after addition of + and - reverse transcription mixes, samples were incubated at 37°C for 1hr.

2.8 Crosslinking of cells for Chromatin Immunoprecipitation

2.8.1 Preparation of ES cells for crosslinking

ES cells were cultured and collected following passaging (2.5.1 and 2.5.2 respectively) and transferred to a 50ml falcon tube, centrifuged at 1000rpm for

5min, and re-suspended in Complete ES Media at a concentration of 1×10^6 cells per ml.

2.8.2 Differentiation of ES cells to EBs and preparation for crosslinking

4.5×10^5 ES cells were differentiated in 30ml of Differentiation Media (2.5.5). EBs were harvested after 7 days, centrifuged at 1000rpm for 5min, and the supernatant removed. EBs were re-suspended in 5ml of PBS, and centrifuged at 1000rpm for 5min. The supernatant was removed and the EBs were re-suspended in 2-3ml of trypsin EDTA and incubated at 37°C for 5min. 2-3ml of FCS was added and the cells passed through a 19G needle to ensure single cell suspension. The cells were centrifuged at 1000rpm for 5min, re-suspended in Differentiation Media 2 (2.5.9.4), counted using a haemocytometer, and re-suspended to 1×10^6 cells per ml.

2.8.3 Crosslinking of cells for ChIP

Formaldehyde solution (38%) was added to a final concentration of 0.4%. and the sample was placed on a shaking platform for 10min at room temperature. 1M Glycin, chilled to 4°C, was added to a final concentration of 125mM, the sample was then inverted to mix. The sample was kept at 4°C from this point on. The sample was centrifuged at 1000rpm 5min 4°C and the supernatant removed. The crosslinked cells were re-suspended in 10ml of a chilled PBS/protease inhibitor solution (9.6ml PBS +0.4ml 50x proteinase inhibitor Roche complete EDTA-free Cat. No.11 873 580 001). This wash was then repeated once. The sample was centrifuged at 1000rpm 5min 4°C and the supernatant from the second wash removed, and the sample snap frozen in LN2 and Stored -80°C.

2.9 Chromatin Immunoprecipitation using the EZ-Magna ChiP A Kit (Millipore catalog #17-408).

All reagents were provided with the kit unless otherwise stated.

2.9.1 Cell preparation for sonication

The cell pellet was re-suspended in 500µl Cell Lysis Buffer (5mM PIPES (pH 8), 85mM KCl, 0.5% NP-40) per 1×10^7 cells. Protease Inhibitor cocktail II (200x) was added to Lysis buffer prior to use to a 1x final concentration. The sample was incubated on ice for 15min, with vortexing every 5min. The sample was centrifuged at 800xg for 5min 4°C and the supernatant removed. The pellet was re-suspended in 500µl Nuclear Lysis Buffer (50mM Tris-Cl (pH 8), 10mM EDTA, 0.8% SDS) per 1×10^7 cells. Protease Inhibitor cocktail II (200x) was added Nuclear Lysis buffer prior to use to a 1x final concentration.

2.9.2 Sonication of cells using Diaganode Biorupter™ 200

The sample in Nuclear Lysis Buffer was transferred to polystyrene 15ml tubes (BD Biosciences catalog No. 352095) as 500µl aliquots and sonicated at high power for 14 cycles of sonication (30s on/30s off). Ice was changed every 4 to 6 cycles. Following sonication the sample was centrifuged at 12000xg for 10min at 4°C. 5µl of the supernatant was removed to assess sonication efficiency, and the remainder of the supernatant was aliquotted as 50µl aliquots, each containing 1×10^6 cell equivalents. Aliquots were either used immediately for Immunoprecipitation or were snap frozen in liquid nitrogen and stored at -80°C.

2.9.3 Immunoprecipitation (IP)

For each immunoprecipitation reaction; 50µl of sonicated material (defrosted on ice if frozen) was combined with 447.75µl Dilution Buffer (kept at 4°C) and 2.25µl Protease Inhibitor Cocktail II. 5µl was removed for Input Control. 20µl of fully re-suspended protein A magnetic beads and antibody (description of antibodies used in 2.9.6) were added and the sample was incubated at 4°C overnight with constant rotation.

Protein A magnetic beads were pelleted using a magnetic separator rack and the supernatant discarded. The beads were then washed with Low Salt Immune Complex Wash Buffer, High Salt Immune Complex Wash Buffer, LiCl Immune Complex Wash Buffer, and TE Buffer (all chilled to 4°C). 500µl of each wash buffer was used, with the sample being incubated at 4°C on rotating rack for 3.5min, and the magnetic beads being precipitated between washes using the magnetic separator rack.

2.9.4 Clean-up of immunoprecipitated samples

100µl ChIP Elution Buffer with 1µl proteinase K (10mg/ml) was added to each immunoprecipitated sample and incubated at 62°C for 2hours in a rotating incubator. The ChIP Elution Buffer was heated prior to use to ensure that SDS was re-suspended. The sample was then incubated at 95°C for 10min, cooled to room temperature and briefly centrifuged. The magnetic separator rack was used to pellet the magnetic beads and the supernatant was moved to a new tube. 500µl of Bind Reagent A was added to the 100µl of sample, the sample was then transferred to a spin filter tube and centrifuged at 14000xg for 30s. The flow-through in the spin collection tube was discarded and 500µl of Wash

reagent B was added to the column. The spin filter tube was centrifuged at 14000xg for 30s and the flow-through was discarded. The spin filter tube was centrifuged at 14000xg for an additional 30s. The spin filter tube was placed in a clean microcentrifuge tube and 50µl of elution buffer C added to the centre of the spin filter membrane. The column was centrifuged at 14000xg for 30s. The purified DNA was then used for real time analysis.

2.9.5 Analysis of control samples

45µl of ChIP Elution Buffer, and 1µl of RnaseA (10mg/ml) was added to each 5µl sample and incubated at 37°C for 30mins. 1µl Proteinase K (10mg/ml) was then added and samples were incubated at 62°C for 2 hours. 10µl was loaded onto a 200ml 2% 1xTBE gel and a 100bp DNA ladder (NEB N3231) was used as a size marker. Electrophoresis was carried out (2.6.2) and gel lanes were analyzed using the AIDA software packages' lane and peak determination software. Lanes were divided up into sections based upon the 100bp DNA ladder, and the intensity for each size range determined as a proportion of total intensity for that lane. Samples from ESC and EB sonications were run on the same gel to avoid bias due to deviations in electrophoresis, EtBr staining and image capture. The percentage of total intensities pertaining to DNA fragments representing mono and di-nucleosomes (150bp to 450bp), tri-nucleosomes (450bp to 650bp) or tetra-nucleosomes (640bp to 850bp) were calculated to assess the efficiency of sonication, and the proportion of the sonicated material that was in the desired size range of mono or di nucleosomes.

2.9.6 Antibodies used for Chromatin Immunoprecipitation

Antibodies were purchased from Millipore. Antibodies used were;

- H3K9me3 (Catalog # 17-625, Lot # DAM1437101) 4µl used per IP
- H3K27me3 (Catalog # 17-622, Lot # DAM1478803) 4µl used per IP
- H3K4me2 (Catalog # 07-030, Lot # DAM1474881) 4µl used per IP
- H3K4me3 (Catalog # 05-745, Lot # DAM1460170) 8µl used per IP
- H3Ac (Catalog # 06-599B) 5µl used per IP
- Normal Rabbit IgG (Catalog # PP64B), 5µl used per IP

2.10 Southern Blotting

2.10.1 Preparation of nylon membrane

DNA was prepared (as described in 2.5.7) and 20µg of DNA was digested with an appropriate amount of RE (NEB) with the correct NEB buffer for 2hrs in a total volume of 200µl at the correct temperature for the RE. Additional RE was added, and the digest was incubated O.N. 400µl 100% ETOH, 20µl 3M NaOAc pH5.5, and 1µl glycogen was added and the sample placed at -20°C for 1hr. The sample was then centrifuged at 13000rpm for 30min. The supernatant was discarded, the pellet washed with 200µl 70% ETOH, and left to dry for ~20min. The pellet was then re-suspended in 20µl dH₂O. The sample was loaded onto a 300ml 1xTBE agarose gel, with an agarose content of between 0.8% and 1.5%. Radiolabelled ladders (2.10.3) were loaded in the far left and far right wells. Electrophoresis was then carried out in a 3.5L gel tank containing 1xTBE O.N (~16hrs) at 40v. The gel was stained by emersion in a 2.5µg/ml EtBr solution for 10min with gentle agitation, then washed by emersion in dH₂O for 10min with agitation. The gel was then imaged using a UV transilluminator and excess gel regions including the wells were removed. The gel was then placed in a 0.25M HCL solution using a slanting Perspex support so that the portion of the gel

containing DNA fragments greater than 4Kb were immersed. After 7.5min, the gel was transferred to a sheet of 3MM Whatmann paper that had been pre-soaked in 20xSSC and was suspended above a reservoir of 20xSSC with the edges of the whatmann paper submerged to act as wicks. A fitted section of nylon membrane (Zeta Probe GT+ membrane Biorad) was wetted in dH₂O and placed on top of the gel, with air bubbles being smoothed out. 3 layers of Whatmann paper were wetted in 20xSSC and placed on top of the membrane. Absorbent paper towels were then added along with a suitable weight, and the blot left O.N. The blot was dismantled and the nylon membrane rinsed in 2xSSC for 2min. The membrane was then place between two layers of Whatmann paper and left to dry for 20min, then crosslinked in a U.V transiluminator. The gel was stained with EtBr as previously described to check for efficiency of transfer.

2.10.2 Hybridization of membrane

The membrane was pre-hybridized with 50ml of Hyb solution (7.5ml 20xSSC, 1ml 0.5M EDTA, 1ml 10% PVP, 1ml 10% Ficoll, 1ml 10% BSA, 0.5ml 10% SDS, 1ml 2mg/ml salmon sperm DNA, 1ml 1mg/ml Heparin) for 3hrs at 65°C with constant lateral rotation. After 2hrs, 25ml of Hyb solution was removed and supplemented with 2.25g dextran sulphate. The Hyb solution was then heated to 65°C. The remaining 25ml of hyb solution with the membrane was discarded and the Hyb solution with dextran sulphate was added. A DNA probe was prepared as described in 2.10.4 (approximately 2×10^6 cpm per ml of buffer), and incubated with the membrane overnight at 65°C with constant lateral rotation. The membrane was washed twice for 5min at room temperature in 50ml 2xSSC, 0.1% SDS, then washed once for 15min at 65°C in 50ml

0.1xSSC, 0.1% SDS. The membrane was then examined using a Geiger counter to assess level of radioactivity. Subsequent washes of 0.1xSSC, 0.1% SDS at 65°C were applied as needed. After washing, the membrane was wrapped in clingfilm and exposed to x-ray film in an exposing cassette for 24 to 72 hours.

2.10.3 End labeling of DNA probe/ DNA Ladder

300ng of DNA ladder (1Kb ladder, Promega, G571A) was combined with 2.5µl 10PNK buffer, 25µCi γ-32P-ATP, 1µl T4polynucleotide kinase, made up to a final volume of 25µl, and incubated at 37°C for 1hr. The sample was then incubated at 68°C for 20min, then purified using a sephadex G50 column. A Geiger counter was used to assess the level of radiation, an a volume equivalent to 20 counts was loaded per ladder, in a mix with 1µl unlabelled ladder.

2.10.4 Random prime labeling of DNA probe

100ng of probe DNA was combined with 3µl of 1.5mg/ml Random Hexamers (Amersham 272166) and made up to a total volume of 14µl using dH₂O. The sample was incubated at 100°C for 5min, then placed on ice. 2.5µl 10xKlenow buffer, 1µl klenow, 2.5µl 0.5mM dNTP mix (without dCTP), and 5µl of α-32P-dCTP was added to the sample and incubated at RT for 4hrs. 100µl TE buffer was added, and 62.5µl phenol, 62.5µl chloroform was added. The sample was mixed then centrifuged at 13000rpm for 10min. The upper phase was then purified using a sephadex G50 column. The prepared probe was then incubated at 100°C for 5min and placed on ice.

Chapter 3 Examination of the timing of repression of the alpha globin gene, and the requirement for antisense RNA in the silencing process

3.1 Introduction

As described in the introduction, a unique case of alpha thalassemia in a patient (ZF) was recently reported that was due to silencing and transcriptional repression of the *HBA2* gene rather than mutation or deletion of the gene itself. Silencing was associated with methylation of the alpha globin CpG island, which is highly unusual as DNA methylation is not associated with transcriptional repression of the alpha globin gene even when the gene is not expressed. In the ZF patient, DNA methylation of the alpha globin gene on the altered copy of chromosome 16 was shown to be present in peripheral blood but not in the germline (Barbour *et al.* 2000). In a transgenic mouse model DNA methylation was detectable in the yolk sac at 7.5d.p.c, as well as in all adult tissues suggesting that DNA methylation of the alpha globin CpG island occurs early in embryonic development, probably around implantation. Consistent with this, introduction of the ZF α Antisense construct into mouse ESCs revealed that human alpha globin was expressed in undifferentiated cells and only became repressed upon differentiation, with DNA methylation and loss of alpha globin expression detectable in day 7 EBs (Tufarelli *et al.* 2003).

In addition, using the mESC system it was shown that replacement of the *LUC7L* promoter with promoters from other ubiquitously expressed genes resulted in repression of the alpha globin gene in a manner indistinguishable from when antisense RNA was driven by the *LUC7L* promoter, indicating that the ability to drive alpha globin gene silencing is not restricted to a single

specific promoter. These data suggest that either the antisense RNA or the process of transcription itself is necessary for this form of silencing (Tufarelli *et al.* 2003; Tufarelli unpublished data).

The alpha human globin gene is contained within a ~1.8Kb CpG island with a GC content of 67.5%. The gene consists of three exons, which are present in the overwhelming majority of alternative transcripts. The human alpha globin gene varies in structure compared to the mouse alpha globin gene, which does not contain a CpG island. Like most CpG island containing genes, the human alpha globin is bivalently marked in hESCs with both H3K27me3 and H3K4me3, with low levels of expression detectable (Pan *et al.* 2007, Ku *et al.* 2008, Guenther *et al.* 2007). Upon differentiation, the alpha globin gene is only expressed in erythroid cells, and is repressed in all other tissues with H3K27me3 enriched across the whole alpha globin locus, with localised binding of the PRC2 complex also detectable (Garrick *et al.* 2008). In contrast to human alpha globin, the mouse alpha globin gene lacks a CpG island within its promoter region and is repressed in ESCs by the histone modification H3K27me3.

The mouse ES model system provided an ideal opportunity to examine the hierarchy and dynamics involved in the establishment of repressive DNA and histone modifications during early stem cell differentiation. Indeed, the early stages of ESC differentiation are the subject of increasingly intense study, as various differentiating ESC model systems have shown that it is during these early stages of differentiation when X inactivation and somatic imprinting occurs

(Latos *et al.* 2009, Chaumeil *et al.* 2006) and when many genes with roles in pluripotency, germline maintenance and pre-implantation embryos become repressed (Epsztejn-Litman *et al.* 2008). Using the mESC model system, Gene expression, DNA methylation, and enrichment for selected histone modifications was examined during differentiation of the cell line 2A3, containing the construct ZF α Antisense, with special emphasises on the time period between day 2.5 and day 4.5.

Additionally, to further clarify the role of antisense RNA in the silencing process, the terminator from the beta globin gene was ligated into the ZF α Antisense construct to terminate transcription of the LUC7L derived antisense RNA before it reached the alpha globin gene. The state of the alpha globin gene was then examined upon differentiation to investigate the effect on gene repression.

3.2 Methods

A list of primers used is shown in Table 3.1

In general, expression of antisense RNA and the presence of DNA methylation was examined twice for each cell line. Gene expression analysis and Chromatin IP experiments were only carried out once for each cell line, unless otherwise stated.

3.2.1 Growth and analysis of 2A3 time course

The cell line 2A3, containing a single copy of the ZF α Antisense construct, was cultured and differentiated as described in 2.5 by Dr Cristina Tufarelli. Samples for analysis of DNA, RNA and Chromatin were taken at days: 0, 1, 2, 2.5, 2.75, 3, 3.25, 3.5, 3.75, 4, 4.25, 4.5, 5, and 6. Analysis of DNA methylation by Southern Blotting was carried out by Dr Cristina Tufarelli, and analysis of histone modifications was carried out by Dr Marco DeGobbi. Analysis of gene expression was carried out by SYBR green real-time PCR as described in 2.6.6.3, and analysis of expression of pluripotency factors was carried out as described in 2.6.5.

3.2.2 Creation of the Zf α antisense Stop construct

The plasmid containing the ZF α Antisense Stop construct was created by Dr Cristina Tufarelli by the addition of the terminator from the beta globin gene (Kind gift of Dr Nick Proudfoot) in the antisense orientation to the alpha globin gene so as to terminate transcription of antisense RNA from LUC7L prior to it reaching the alpha globin gene. The beta globin fragments consists of exon2,

intron2, exon3, the poly(A) site and the 1.6Kb region immediately downstream containing pre-determination cleavage sites essential for the initiation of transcriptional termination (Dye & Proudfoot 2001).

3.2.3 Electroporation of Zf α antisense Stop construct into ESCs and screening for successful incorporation

Purification of the ZF α Antisense Stop construct from its plasmid backbone, and its integration into mESCs by electroporation, growth of cells following electroporation, and extraction of DNA is described in section 2.5.6.

Primer	Sequence, 5' to 3'	PCR buffer	Annealling temp °C	Notes
493	GAAGGACAGGAACATCCTGCG	2xDMSO	60	Primer pair used to amplify 306bp region of alpha globin CpG island.
498	CCGCGCCCCAAGCATAAACCC			
201	TGGAGGGTGGAGACGTCCTG	-	-	Primer used to prime reverse transcription reaction
alphaSYBR F	TCTCCCCGAGGATGTTC	ABI SYBR green	60	Referred to as 'Alpha 3' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
alphaSYBR R	GGTCGAAGTGCAGGAAGTAG			
alphaSYBRF2	TCTAGGCCAGTGGGAGAGTCA	ABI SYBR green	60	Referred to as 'Alpha 1' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
alphaSYBRR2	ACTGGCTAAGCCCCAAGTCA			
alphaSYBRF3	GGGTGCGGGCTGACTTT	ABI SYBR green	65	Referred to as 'Alpha 2' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 65°C 1min) x40 cycles.
alphaSYBRR3	CAGCGCCACCCTTTCCT			
445	GAGAGAACCCACCATGGTGCTG	2xDMSO	60	Primer pair used to amplify human alpha globin cDNA. Specific in 2xDMSO buffer but amplifies mouse alpha globin in ABI SYBR green so unsuitable for realtime PCR. Product 426bp.
446	GTCAGCACGGTGCTCACAGAAG			
0163508-F	ACCTCCCCGCCGAGTTC	ABI SYBR green	60	Referred to as 'Alpha 4' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
0163508-R	AGGCTCCAGCTTAACGGTATTTG			
HBA2 F	GGCGAGTATGGTGCGGAGG	ABI SYBR green	63	Primer pair used for real-time PCR amplification of human alpha globin cDNA. program used; 95°C 10min, (95°C 15s, 63°C 1min) x40 cycles. 25ng input.
HBA2 R	CCAGCAGGCAGTGGCTTAGG			

Primer	Sequence, 5' to 3'	PCR buffer	Annealling temp °C	Notes
hc009a	CGACCTTACCCAGGCGGCCTTGA	2xDMSO	60	Primer pair amplify 423bp region of alpha globin CpG island.
hc009b	CGAAAGGAAAGGGTGGCGCTG			
alpha F	GGTCCCCACAGACTCAGAGAGAAC	2xDMSO	60	Primer pair amplify 456bp region of alpha globin CpG island.
alpha R	CGCCGCTCACCTTGAAGTTG			
mHba-a1 F	TGCTCTCTGGGGAAGACAAAAG	ABI SYBR green	60	Primer pair used for real-time PCR amplification of mouse alpha globin cDNA. program used; 95°C 10min, (95°C 15s, 60°C 15s, 72°C 15s) x40 cycles. 25ng input.
mHba-a1 R	CGTGGCTTACATCAAAGTGAGGG			

Table 3.1. Primer used for the HBA2 gene

The ZF α Antisense Stop construct was excised from the ZF α AS Stop plasmid by restriction digest with EcoRV (Figure 3.1A). Gel electrophoresis was used to separate the 12 845bp band containing the ZF α Antisense Stop construct from the remaining 3205bp band, then the construct was purified by band excision and electroelution. The geneticin (G418) resistance fragment was excised from the pPNT plasmid by restriction digest with EcoR1 and the 1859bp resistance fragment was purified from the remaining 5489bp band by gel electrophoresis followed by band excision and electroelution. Electroporation into ESCs was carried out using the BioRad Gene Pulser Xcell (2.5.6.2). After 8 days, 37 colonies were selected based upon their size and morphology, and cultured in 24 well plates until there were sufficient cells for DNA extraction, and for cryogenic storage. Extracted DNA was screened by PCR for the presence of the ZF α Antisense Stop construct using the primer pair 436/437 (Figure 3.1B) and 16 cell lines selected and labeled Y1 to Y16.

3.2.4 Analysis of cell lines

Five cell lines that had tested positive were cultured and expanded in tissue culture. ESC material was collected from all cell lines for RNA and DNA analysis, ESCs were differentiated into EBs over a period of seven to eight days, and more material prepared for RNA, DNA, and chromatin analysis. Preparation of DNA and RNA and ChIP pellets is described in 2.5.7 and 2.5.8 respectively. Differentiation of ESCs into EBs is described in 2.5.5. Subsequently, two of the cell lines were cultured and differentiated for analysis of histone

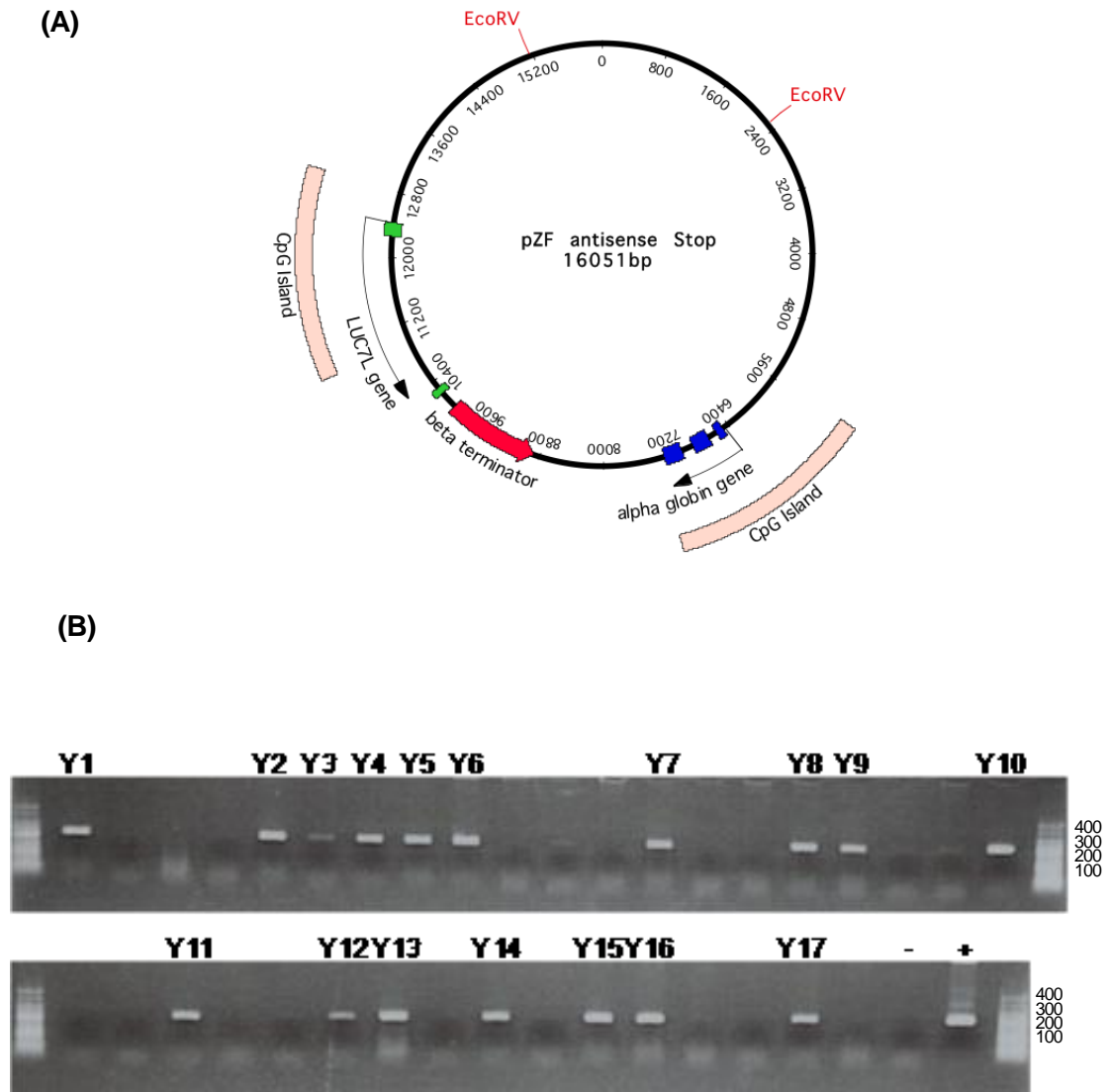


Figure 3.1 Design of Antisense stop construct. (A) the plasmid pZFaASstop, containing the alpha globin gene in the sense orientation, the LUC7L gene in the antisense orientation, and the beta terminator positioned in-between the two genes to cause termination of antisense transcripts was digested with EcoRV and electroporated along with the 1859bp pPNT fragment into mouse ES cells. PCR screening was carried out using the primer pair 436/437 amplifying a 208bp fragment of the alpha globin gene to detect successful incorporation of the construct (B).

modifications, with crosslinked pellets prepared for ChIP as detailed in 2.8. DNA and RNA was also extracted from these cell lines as detailed above to allow confirmation of previous observations in these cell lines and assess differentiation efficiency.

Expression of antisense RNA was tested for by strand specific RT-PCR as described in 2.7. The levels of the pluripotency factors *Oct4* and *Nanog* was also examined in ESCs and EBs as markers for successful differentiation of cells, this was carried out by multiplex PCR with primer pairs for *Oct4*, *Nanog* and *Aprt*. The PCR products were subject to gel electrophoresis and quantified using gel analysis software (2.6.5). DNA methylation was assessed by Southern blotting, and msPCR using primers that spanned various regions of the gene's promoter containing methylation sensitive HpaII sites (2.6.3). Gene expression levels were assessed by SYBR green real-time PCR (2.6.6.3). Histone modification changes between ESCs and EBs were examined using the protocol described in 2.9.

3.3 Results

3.3.1 Analysis of epigenetic changes at the alpha globin gene during 2A3 differentiation

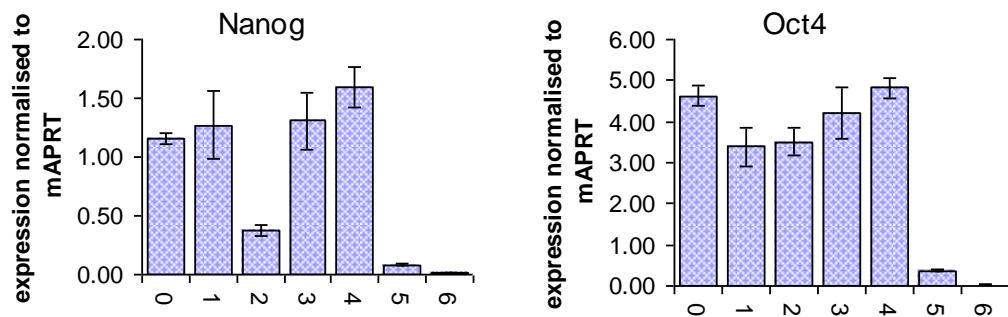
The mouse ESC line 2A3, containing a single copy of the ZF α antisense construct, was differentiated over a 6 day period. Samples were prepared for analysis of gene expression, DNA methylation, and histone modifications each day, or every 6 hours between day 2.5 and day 4.5 during which gene repression was thought to occur.

3.3.1.1 Expression profiling

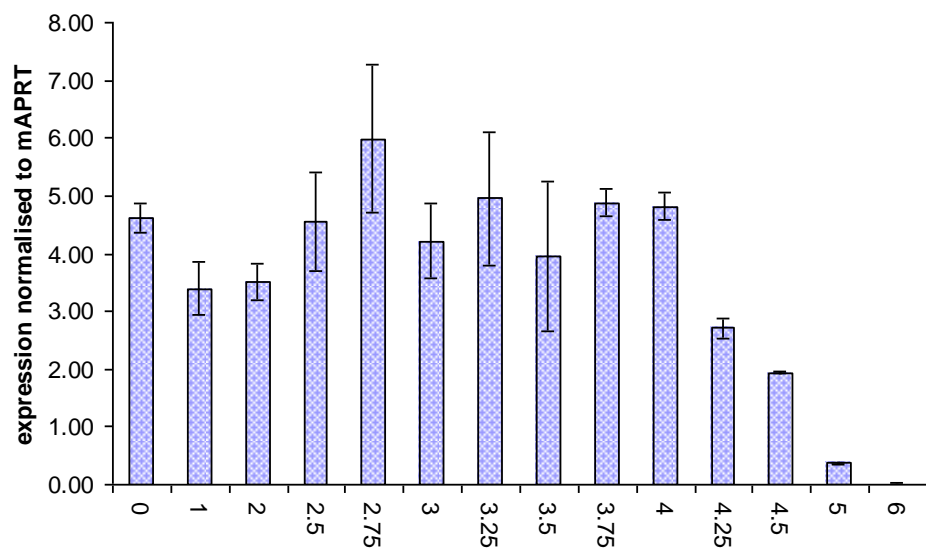
As ESCs differentiate, they form EBs that become visible to the naked eye and act as a clear indicator of differentiation, however EB populations are heterogeneous and a proportion of ESCs do remain. Therefore the level of differentiation within the EB population as a whole was assessed by examining the transcriptional expression levels of the pluripotency factors *Oct4* and *Nanog* at each time point, compared to expression of the ubiquitously expressed gene *Aprt* (Figure 3.2). Expression of *Oct4* remained fairly constant until day 5, when expression levels were severely inhibited (Figure 3.2A). Reduction in levels of detectable mRNA occurred at a steady linear rate throughout the 24 hour period between day 4 and day 5 (Figure 3.2B), with levels having reduced further by day 6. Expression of *Nanog* appeared more complex, with expression levels decreasing sharply at day 2, but with levels returning to pre-differentiated levels by day 2.5. Expression after day 2.75 increased steadily up to day 3.75 with levels greater than twice those observed in undifferentiated cells, before sharply

reducing over a 30hr period to very low levels of expression that had further reduced by day 6.

(A) Expression per day of differentiation



(B) Oct4 expression



(C) Nanog expression

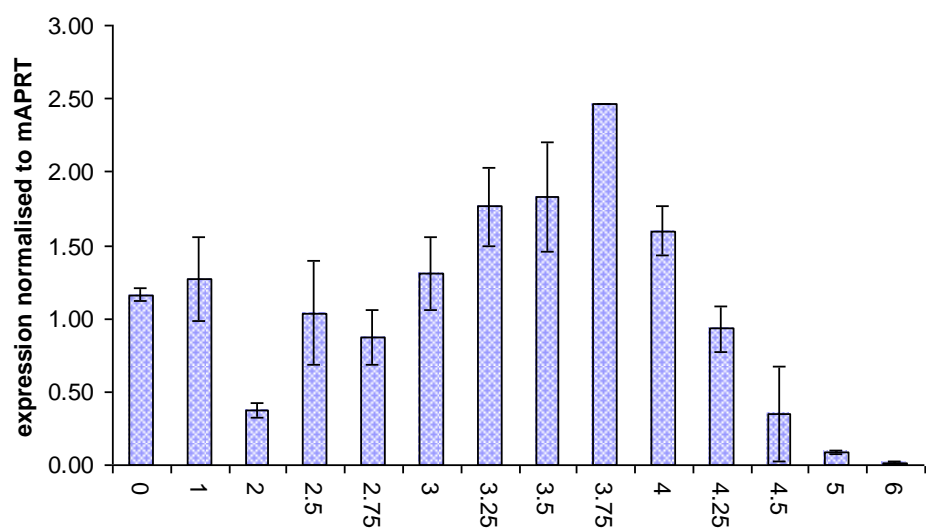


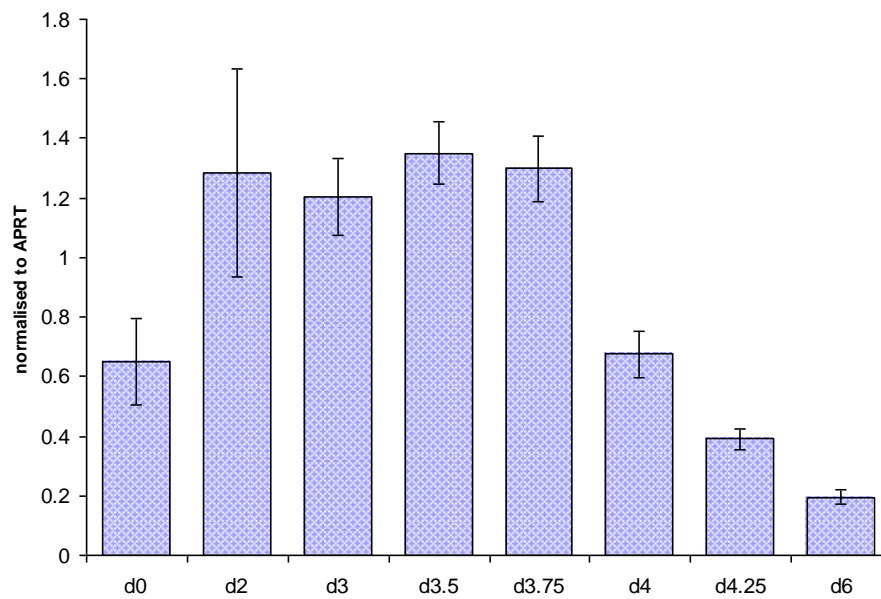
Figure 3.2 Expression of pluripotency factors during timecourse differentiation. (A) Expression of *Nanog* and *Oct4* normalised to levels of *Aprt* during differentiation over a six day period. (B) *Oct4* expression in detail during timecourse. (C) *Nanog* expression in detail during timecourse.

Expression of both human and mouse alpha globin was examined by SYBR green real-time PCR during the 6 day time course, concentrating on the period between day 3 and day 4 (Figure 3.3). In contrast to human alpha globin which is bivalently marked, the mouse alpha globin gene is monovalently enriched with H3K27me3 in ESCs, with expression levels correspondingly low. Data in Figure 3.3B is displayed using a logarithmic scale to allow mouse alpha globin expression levels in both ESCs and EBs to be displayed on the same graph. The human alpha globin construct includes the haematopoietic regulator HS-40 that is involved in alpha globin gene regulation, but expression levels from the construct are not as great as endogenous human alpha globin expression due to the absence of suspected long range enhancers. Therefore, while human alpha globin levels will not match mouse alpha globin expression levels in their intensity, they do correctly reflect the transcriptional state of the gene. Expression of human alpha globin increased within the first 24 hours then remained constant until day 3.75, when the level of detectable transcripts sharply decreased to less than a third in the 12 hour period to day 4.25 (Figure 3.3A) with levels having decreased still further by day 6. Expression of mouse alpha globin (shown as a logarithmic scale in Figure 3.3B) remained barely detectable until day 6 when expression became up regulated, indicating that the EBs were differentiating down a haematopoietic lineage

3.3.1.2 DNA methylation profiling

DNA methylation of the alpha globin CpG island was examined by southern blotting for each day of the time course (Figure 3.4). DNA was digested with PstI alone, or in a double digest with the methylation sensitive enzyme EagI. PstI digestion produced a 1489bp fragment of the alpha globin CpG island containing a single EagI restriction site and can be digested into 975bp and 515bp fragments in the absence of DNA methylation (Figure 3.4A). The Southern blot was hybridized with a probe covering the 1489bp region (Figure 3.4C). Levels of DNA methylation were quantified, with intensity of the 1489bp band in the double digest expressed as a fraction of the total band intensity. DNA methylation prevents EagI from cutting and results in a band 1489bp in size, while complete DNA digestion in the absence of DNA methylation results in two bands of sizes 975bp and 514bp. Quantified levels of DNA methylation are shown in Figure 3.4D. DNA methylation was low in the first two days of the timecourse, with levels increasing marginally at day 3. DNA methylation levels then increased dramatically between d3 and d4, so that over half the cell population exhibited alpha globin CpG island methylation by d4. Methylation then increased incrementally for the next 48hrs, and by day 6 DNA methylation in the cell population was almost 90%. This clearly indicated that the main stimulus in DNA methylation was occurring between days 3 and 4. The blot was stripped and hybridized with a probe for a region of *LUC7L* that contains EagI restriction sites and does not become methylated. The probe used binds to both mouse and human Luc7L (Figure 3.4B). Complete digestion had occurred for each of the different timepoints.

(A) Expression of human alpha globin during differentiation timecourse



(B) Expression of mouse alpha globin during differentiation timecourse

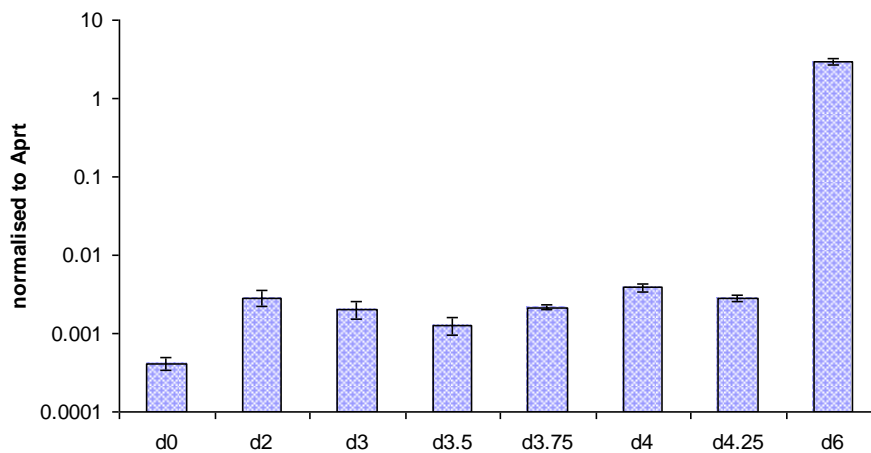
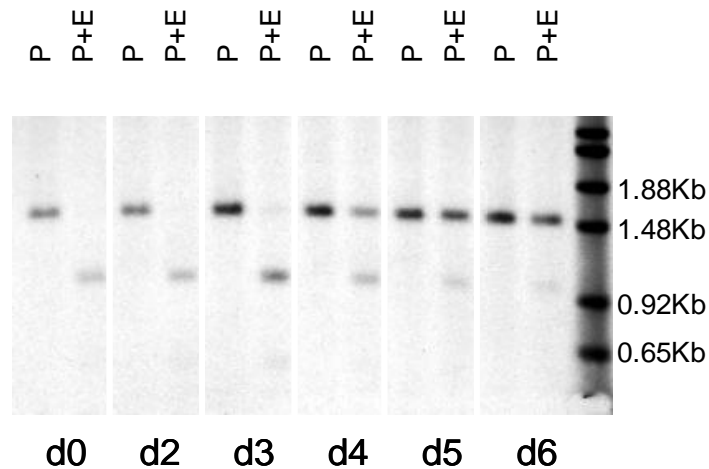
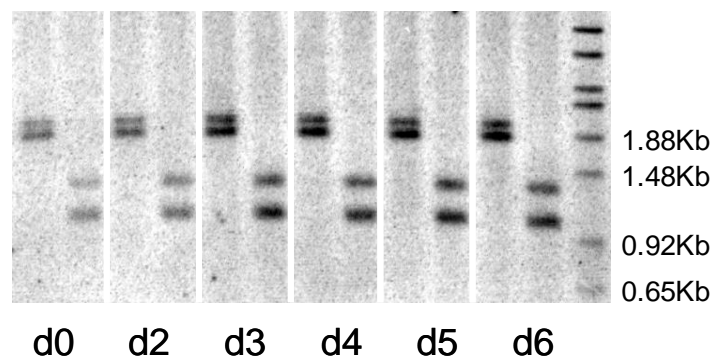


Figure 3.3 Expression of human and mouse alpha globin during timecourse differentiation. (A) Expression of human alpha globin normalised to expression of Aprt at selected time points during the differentiation timecourse. The level of expression of alpha globin in day 8 2A3 embryo bodies is shown in red. (B) Expression of mouse alpha globin normalised to Aprt at selected time points during the differentiation timecourse, shown with a logarithmic scale. The alpha globin construct contains the HS-40 hematopoietic regulator but lacks other long range regulatory elements resulting in a reduced intensity of human alpha globin expression. Mouse alpha globin is not expressed in ESCs as its promoter lacks a CpG island.

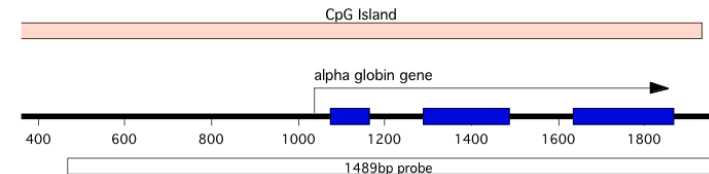
(A) Probing SB for methylation of alpha globin CpG island



(B) Probing SB to show complete digestion of DNA



(C) Schematic diagram showing location of CpG island probe



(D) Percentage of DNA methylation of the alpha globin CpG island

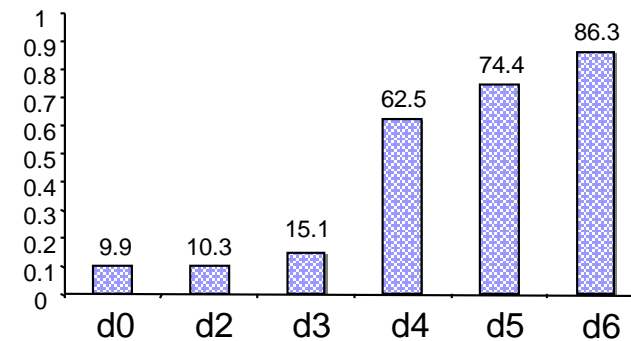


Figure 3.4 DNA methylation of the alpha globin CpG island. (A) timecourse SB screened for DNA methylation. PstI digest produces band of 1489bp, methylation sensitive PstI, EagI double digest produces 975bp and 515bp band when unmethylated, and 1489bp when methylated. (B) Control for complete digestion of DNA using CT20/CT29 probe to detect both mouse and human LUC7L. PstI digest produces a 1.84Kb band for human LUC7L and a 1.88Kb band for mouse Luc7L . PstI, EagI double digest produces bands of 1.02Kb, 0.27Kb, and 0.55Kb for human LUC7L and 1.43Kb and 0.45Kb for mouse Luc7L. (C) Position of alpha globin CpG island probe. (D) DNA methylation levels by image quantification, and normalization to total DNA levels shown in PstI digest. Southern Blot experiments were carried out by Dr Cristina Tufarelli.

3.3.1.3 Histone modification profiling

Histone modifications at the alpha globin gene were examined for each of the six days of the differentiation time course by taqman real-time PCR. Levels of H3K9me3, associated with gene repression, and levels of H3K4me2 and H3Ac, associated with gene expression, were examined. Comparison of day0 and day6 enrichments for the different histone modifications show that upon differentiation, H3K9me3 becomes enriched across the alpha globin gene, while H3K4me2 and H3Ac enrichment becomes greatly reduced (Figure 3.5). Examination of these three marks over the six day time period revealed that for the histone modifications examined, the major shift in enrichment levels occurred between the day 3 and day 4 time points (Figure 3.6 A-C). Direct comparison of changes in H3K9me3 and H3K4me2 methylation clearly showed that enrichment of the repressive mark H3K9me3 occurred concurrently with the loss of enrichment for the active histone modification H3K4me2 (Figure 3.6D).

At all four of the examined points within the alpha globin gene, H3K9me3 levels show little change prior to day3, then following the large increase in enrichment during the 24hr period between day 3 and 4, methylation levels continued to increase in small increments on each of the days assayed. H3K4me2 enrichment followed a very similar pattern, where the major alteration in dimethylation levels occurred in the 24hr period between days 3 and 4, with very little alteration in levels occurring outside of this time period. Enrichment for H3Ac was fairly low in ESCs and showed an initial reduction between day 1 and day 2, though this decrease also occurred at the LUCL promoter region. In line

with the other histone modifications examined, the main alteration in H3Ac levels occurred between day 3 and day 4.

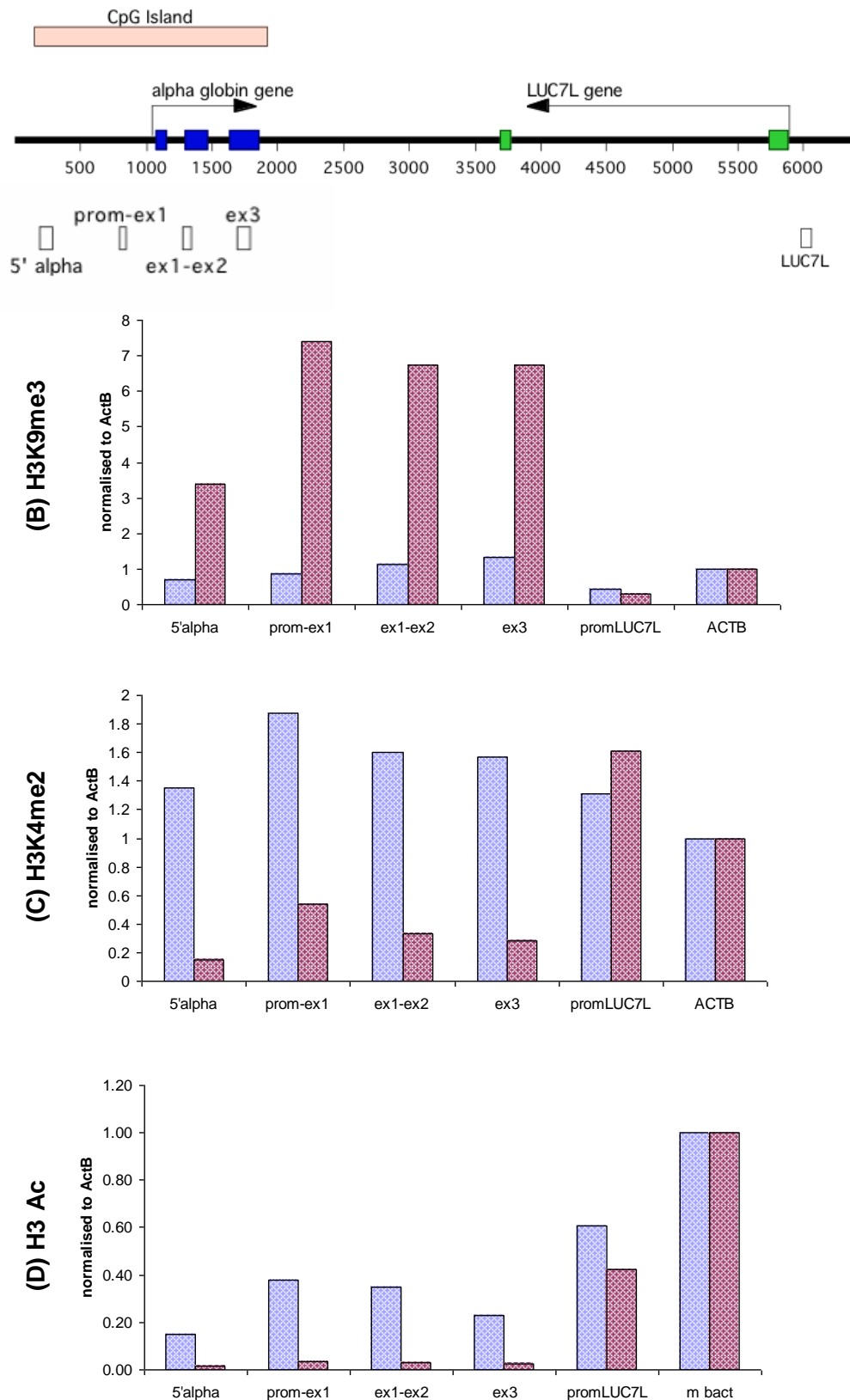
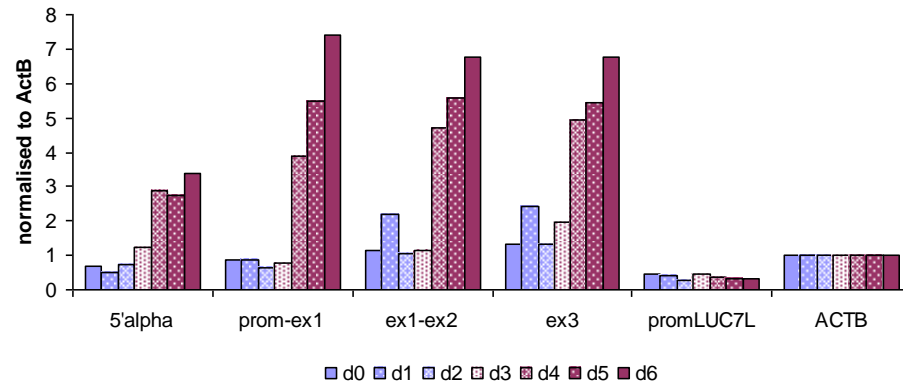
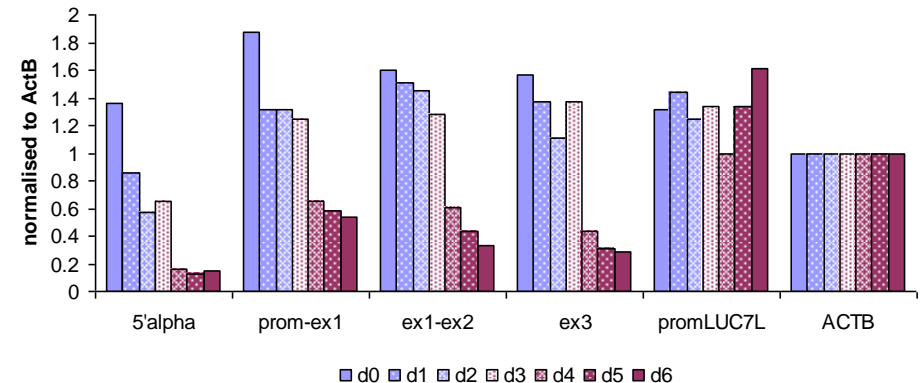


Figure 3.5 Histone modifications within the alpha globin gene at day 0 and day 6 of Differentiation timecourse. (A) Schematic diagram showing the position of the probes used for the taqman PCR. (B-D) Enrichment for different histone modifications in day 0 (blue) and day 6 (red). Chromatin IP experiments and realtime PCR carried out by Dr Marco DeGobbi

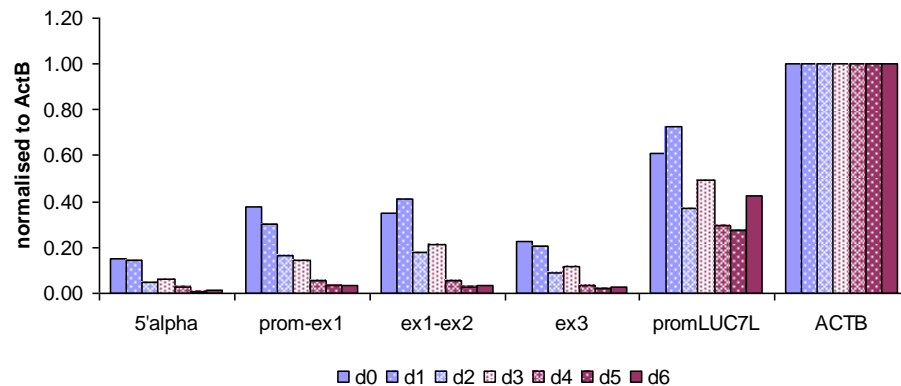
(A) H3K9me3



(B) H3K4me2



(C) H3 Ac



(D) H3K4me2 and H3K9me3 changes during timecourse

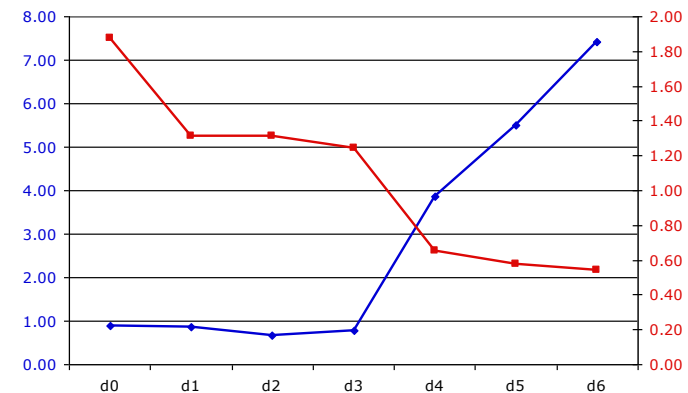


Figure 3.6 Histone modifications within the alpha globin gene during differentiation. Histone modification changes during the six days of the timecourse. H3K9me3, H3K4me3, and H3Ac are shown in A, B and C respectively. (D) compares H3K4me2 changes (red, right hand scale) and H3K9me3 (Blue, left hand scale) at primer position prom-ex1. Chromatin IP experiments and realtime PCR carried out by Dr Marco DeGobbi.

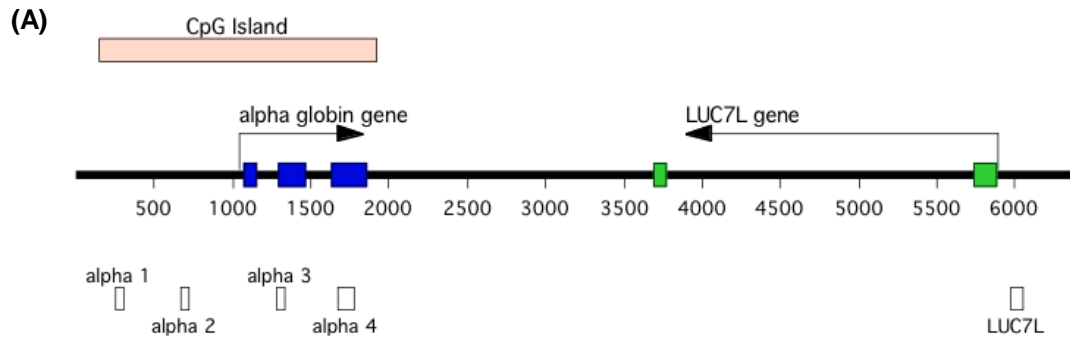
Antibodies used during previous experiments to examine histone modifications were no longer available, therefore new antibodies from Millipore were tested to see if similar patterns of H3K9 and H3K4 methylation enrichment at the alpha globin gene were obtained. Antibodies for H3K9me3, H3K4me2, H3K27me3, H3K4me3, and H3Ac were tested (Figures 3.7 and 3.8). New antibodies were tested for enrichment of known positive and negative controls to both check antibody specificity and establish enrichment levels. Antibody testing is detailed in Appendix A.

The H3K4me3 antibody showed a similar pattern of ES and EB enrichment across the alpha globin gene compared to H3K4me2 antibody previously used in that enrichment levels were higher in ESCs than in EBs (Figure 3.7B), however this decrease was not as substantial as the 3 to 4 fold decrease previously observed. For primer pairs alpha 2, 3, and 4 the detectable levels of enrichment halved, while the alpha1 pair showed a clear but modest decrease. The Millipore H3K4me2 antibody showed the same trend for a decrease in enrichment upon differentiation, however this decrease was minor for most of the primer sets tested (Figure 3.7C). Gapdh was tested as an endogenous control, but showed a decrease in H3K9me3 levels upon differentiation, suggesting that it is unsuitable as a control for this mark.

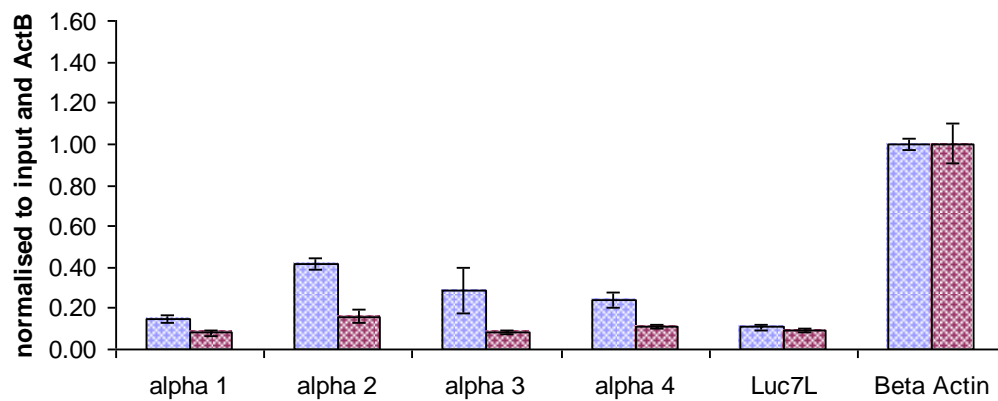
H3K9me3 enrichment at the alpha globin gene increased upon differentiation, though levels of enrichment in EBs were far less than had been observed with the previous antibody (Figure 3.8B). H3K27me3 levels were high in ESCs and became further enriched in EBs. H3K27me3 enrichment levels were not as high

as levels at the *Pcdh8* gene, a gene selected from the analysis of global histone modifications in ESCs (Mikkelsen et al 2007) that identified it as being highly enriched for this mark in ESCs and a range of *in vitro* differentiated tissues. H3K27me3 levels increased by approximately one third at positions alpha1, 3, and 4, though variation within the data suggested that the difference between ESCs and EBs may not be as great as this (Figure 3.8C). H3K27me3 levels at the *Gapdh* gene decreased upon differentiation. H3Ac levels decreased at all four regions examined (Figure 3.8D), in a very close approximation of previously observed enrichment, in terms of both the amplitude of enrichment and the primer specific levels of enrichment observed. The exception to this was position alpha 4 which showed a smaller decrease in H3Ac enrichment upon differentiation than had been previously shown.

Comparison of these results with those from antibodies previously used in the characterization of regions of the alpha globin gene show that, although levels of enrichment were somewhat reduced, the trends of enrichment upon differentiation were maintained, despite the use of different antibodies, and procedural differences in the preparation of the chromatin and data analysis.



(B) H3K4me3 (05-745 DAM 1460170)



(C) H3K4me2 (Millipore 07-030 DAM1474881)

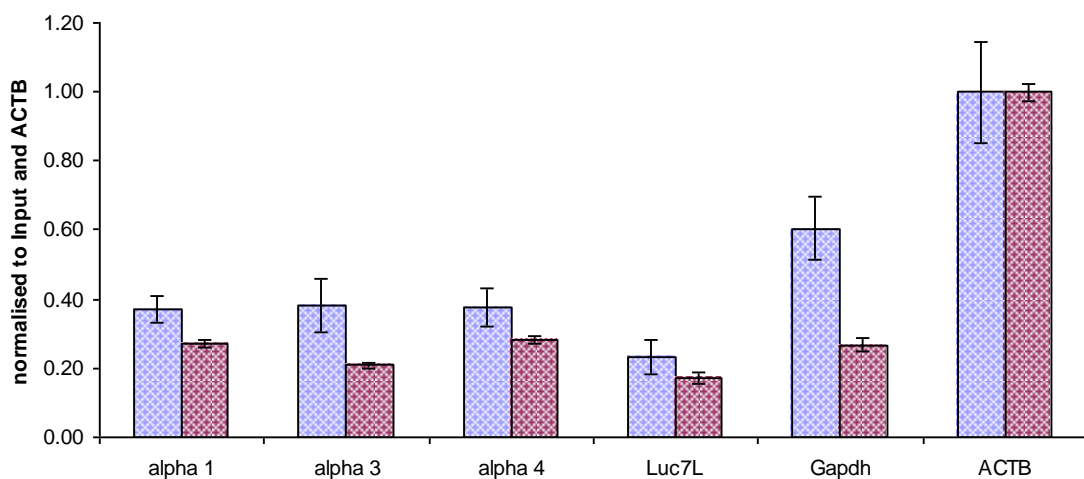


Figure 3.7 H3K4me methylation at the alpha globin locus. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the alpha globin gene. (B) enrichment for H3K4me3 within the alpha globin gene. (C) enrichment for H3K4me2 within the alpha globin gene. ES cells are blue, EBs are red.

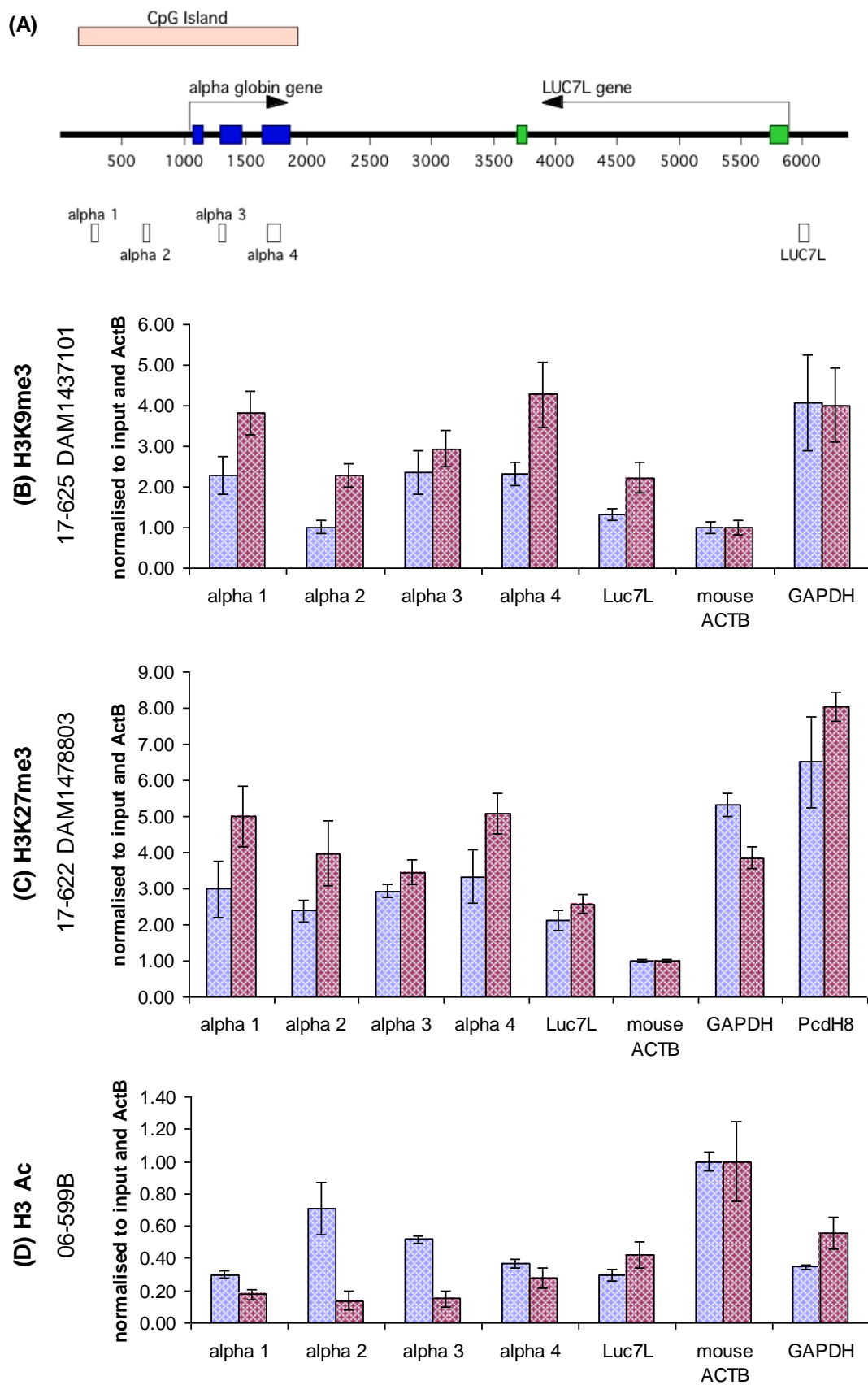


Figure 3.8 Histone modifications at the alpha globin gene. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the alpha globin gene. (B) to (D) shows enrichment levels for different histone modifications. Blue indicates the level in ESCs, red indicates the level in EBs. Enrichment is normalised to Input and to mouse ACTB.

3.3.2 Investigating the requirement for AS-RNA

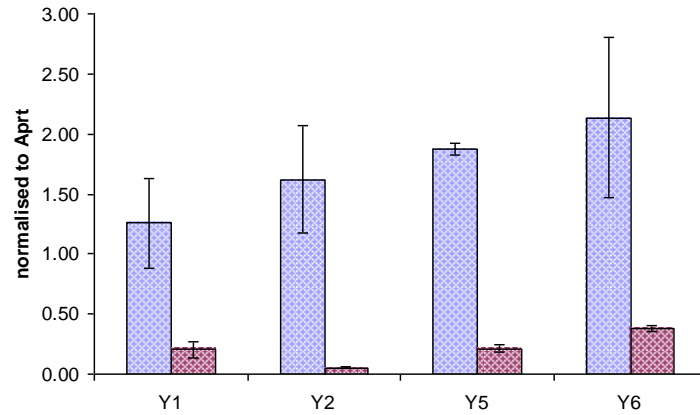
As described in 3.1, the data so far suggest that antisense RNA transcription is required for silencing to occur, similarly to what is observed at imprinted loci. However, all the promoters used to drive antisense transcription in the mESC system are CpG island containing housekeeping genes. To determine whether the close proximity of a CpG island belonging to a highly expressed gene could influence alpha globin expression, and to determine whether antisense transcription through alpha globin gene was required for gene repression, a transcriptional termination element from the β globin gene was cloned between the *LUC7L* promoter and the alpha globin gene to prevent antisense transcription through the alpha globin gene.

Mouse ESC cell lines containing the new construct, ZF α antisense Stop, were differentiated into EBs, and the level of differentiation within the cell populations were assessed by screening for expression of the pluripotency factors *Oct4* and *Nanog* using a multiplex PCR approach. Amplification of *Oct4*, *Nanog*, and *Aprt* cDNA was carried out in the same PCR reaction, in triplicate, and expression levels of *Oct4* and *Nanog* compared to *Aprt* levels by band intensity analysis using AIDA software. Expression levels of these two genes are shown in Figure 3.9.

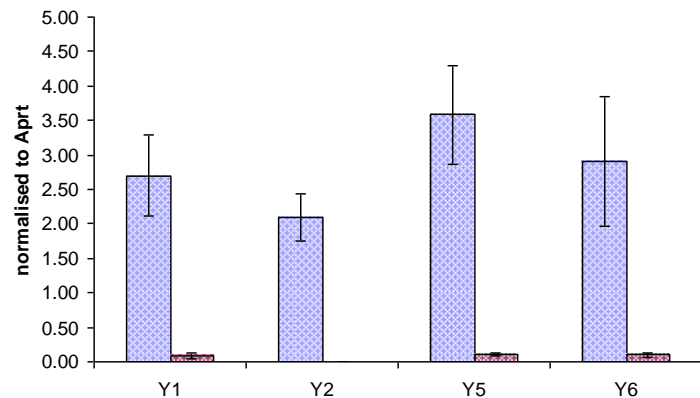
The four cell lines analyzed showed great reductions in *Oct4* and *Nanog* expression levels indicating a high level of differentiation (Figure 3.9A and B). A fifth cell line, Y4, was analyzed for antisense expression and methylation status but there was insufficient RNA for real-time analysis of gene expression. The

cell lines Y5 and Y6 were subsequently grown and differentiated for analysis of histone modifications. Expression of *Oct4* and *Nanog* in these cell lines is shown in Figure 3.9C. Cell line Y5 shows a high level of differentiation within the cell population, with strong repression of both *Oct4* and *Nanog*. Cell line Y6 shows incomplete repression of *Oct4* and continued expression of *Nanog*, indicating that a proportion of the cell population had not fully differentiated, despite the clear presence of embryoid body formation, and visible coloration characteristic of mouse alpha globin expression - though this was not as marked in Y6 as in the cell line Y5 (Figure 3.9D).

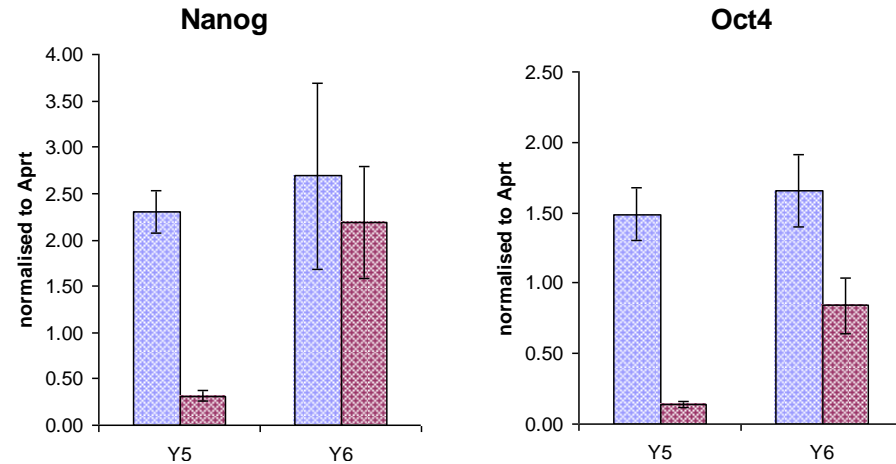
(A) Expression of Nanog in antisense stop cell lines



(B) Expression of Oct4 in antisense stop cell lines



(C) Expression of pluripotency factors in cell lines grown for Chromatin IP experiments



(D) Embryo body morphology on day of harvesting

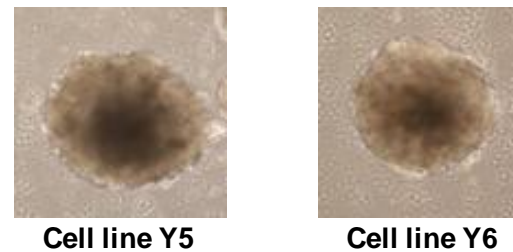


Figure 3.9 Expression of pluripotency factors for Antisense stop cell lines. Expression of Nanog (A) and Oct4 (B) in ESCs (blue bars) and in EBs (red bars) for antisense stop cell lines. (C) shows expression of Oct4 and Nanog in cell lines Y1 and Y6 grown for Chromatin IP experiments. (D) Shows EB morphology of cell lines grown for ChIP

The efficiency of the β globin terminator in terminating the antisense transcript from the LUC7L gene in Y cell lines containing the ZF α antisense Stop construct was determined by strand specific RT-PCR using an antisense specific primer, along with a primer for the endogenous control *Aprt* (Figure 3.10). Antisense RNA within the alpha globin gene was virtually undetectable, though there was some low levels detectable in some samples (Figure 3.10B). The presence of antisense RNA within the alpha globin gene was further characterized by semi-quantitative PCR comparing antisense RNA levels in ES and EBs for two additional Y cell lines, Y7 and Y8, at 24, 26, 28, and 30 cycles of PCR and contrasting this expression with that of the 2A3 cell line containing the ZF α antisense construct. This clearly demonstrated that, while there were minor levels of antisense RNA detectable in Y cell lines (Figure 3.10C and D), this was at a far lower level than antisense RNA expression levels in the 2A3 cell line, whose construct does not contain the β terminator (Figure 3.10E).

DNA methylation of the alpha globin CpG island was initially examined by southern blotting using a probe that covers a region of the alpha globin CpG island previously shown to become silenced (Figure 3.11). DNA from ESC and EBs for the cell lines Y1, Y2, Y4, Y5 and Y6 was digested with the restriction enzymes PstI and EagI in a double digest then transferred to a nylon membrane and hybridized with the alpha globin CpG island probe (2.10). All five cell lines showed two bands demonstrating a lack of DNA methylation (Figure 3.11B).

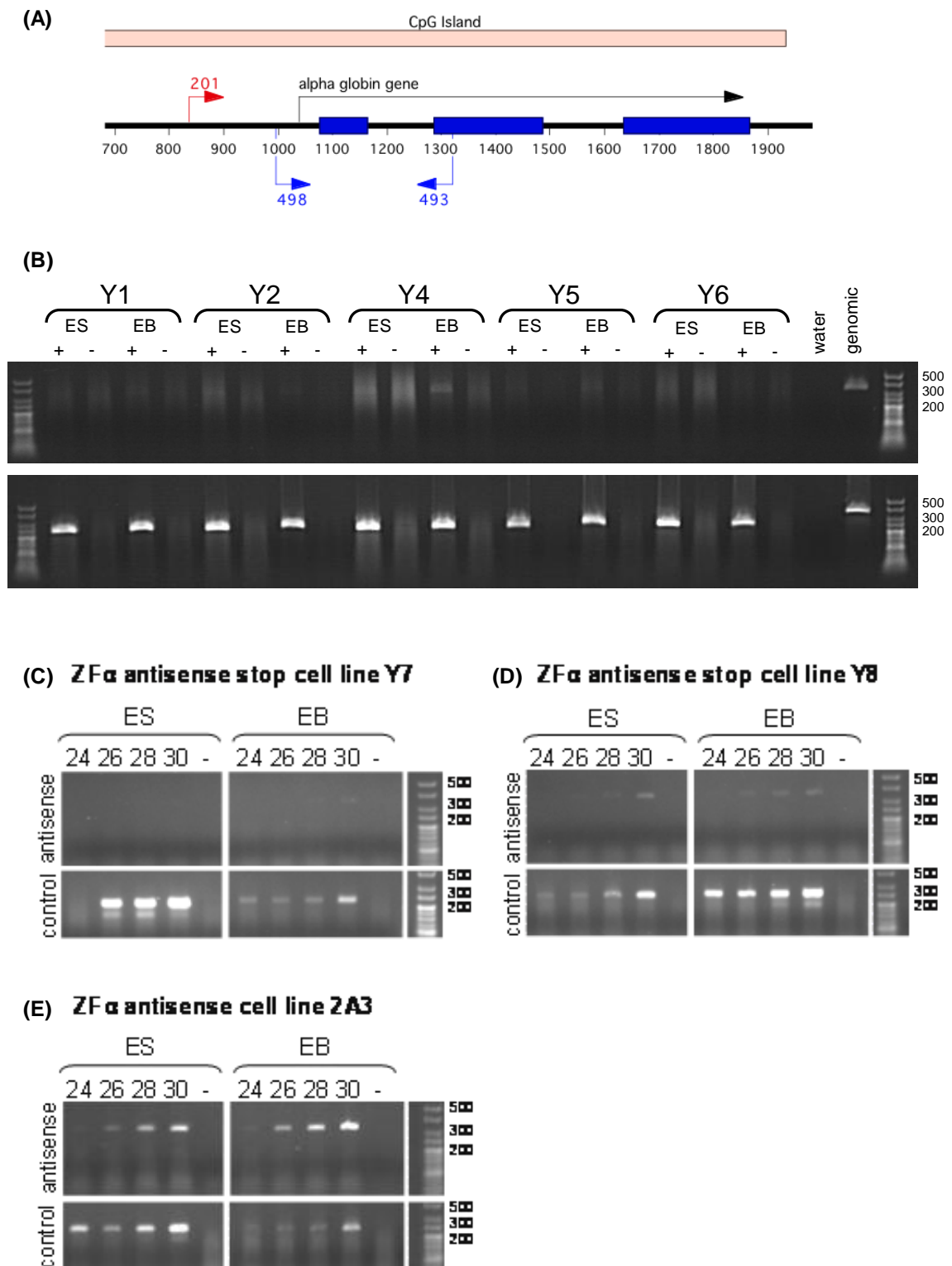


Figure 3.10 Expression of antisense RNA in antisense Stop cell lines. (A) schematic representation of the alpha globin construct with primer locations marked. (B) Strand-specific RT-PCR reaction screening for antisense RNA using the primers 493/498 (306bp), and amplification of control using mAPRT2/mAPRT3 (225bp cDNA, 335bp genomic DNA). (C) to (E) Semi-quantitative PCR with reactions of 24,26,28, and 30 cycles to compare antisense RNA levels in ZFα antisense Stop cell lines and the ZFα antisense cell line 2A3.

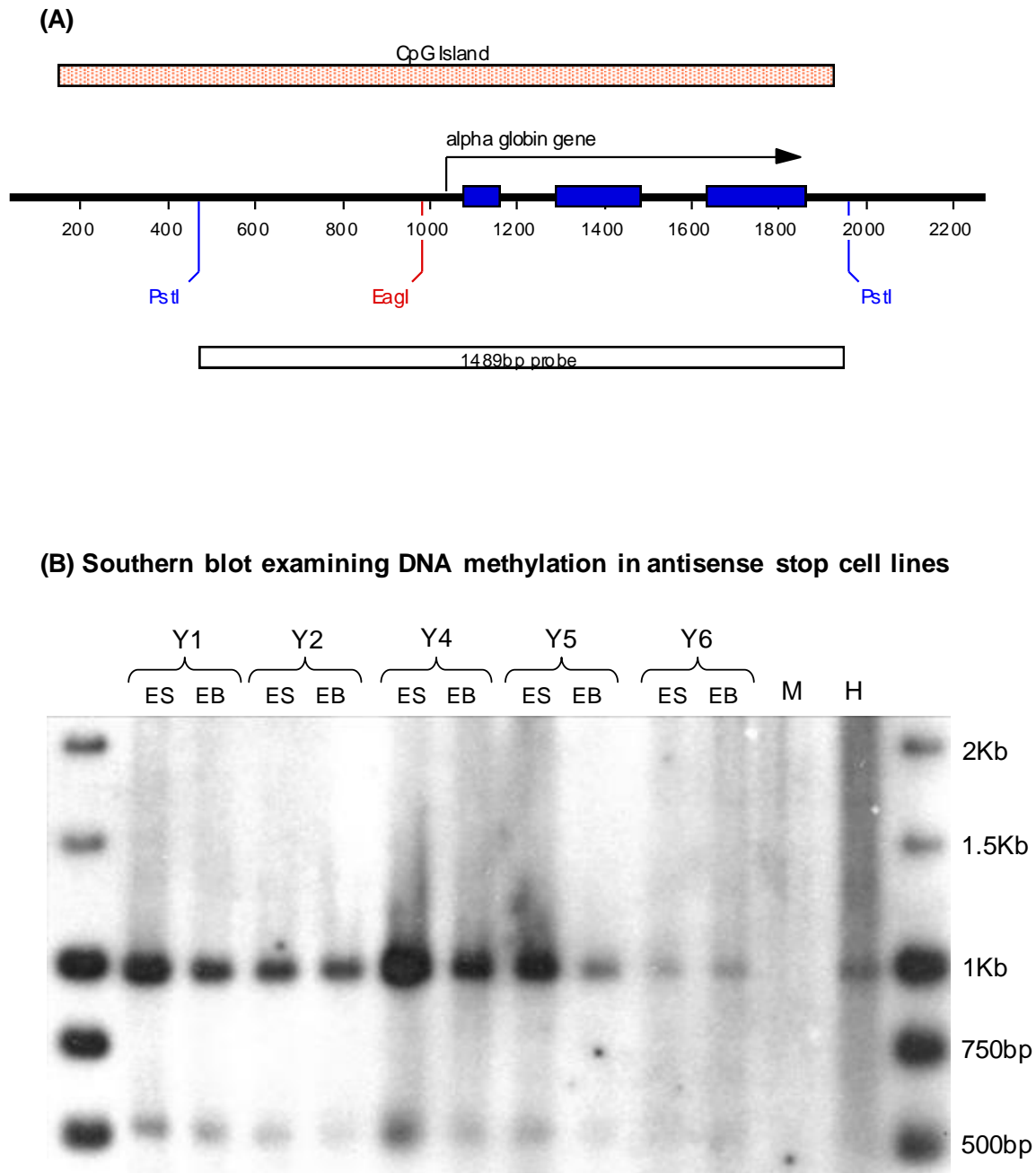


Figure 3.11 DNA methylation of antisense stop cell lines. (A) Location of 1489bp probe used to screen for DNA methylation of antisense stop cell lines by southern blotting. (B) Southern blot of antisense stop cell line DNA digested with PstI and the methylation sensitive enzyme EagI. M; mouse DNA, H; human DNA. Methylation of the alpha globin CpG island produces a 1489bp product from the restriction digest, if unmethylated, two fragments of 975bp and 514bp are detectable.

DNA methylation was also assessed by msPCR, using two sets of primer pairs that amplified across 3 and 4 HpaII sites within the alpha globin CpG island (Figure 3.12.A). The cell lines Y1, Y2, and Y4 from the initial examination of silencing in ZF α antisense Stop cell lines, and cell lines Y5 and Y6 grown for analysis of histone modifications, were analyzed by msPCR. None of the cell lines showed DNA methylation for either of the primer pairs tested, in agreement with results from DNA methylation analysis by southern blot. Controls showed that DNA was capable of acting as a template for PCR, that the digest had occurred in a methylation sensitive manner, and that the digestion was complete for each sample (Figure 3.12B).

Alpha globin gene expression was examined by SYBR green real-time PCR, with alpha globin levels being normalized to the ubiquitously expressed gene *Aprt*. Four out of five of the initially examined cell lines, Y1, Y2, Y5 and Y6, and both of the cell lines that were subsequently grown for examination of histone modifications were examined for expression of human alpha globin (Figure 3.13). Levels of expression were nearly identical for the cell lines grown in the initial study and for those grown later, showing a high level of reproducibility. For the cell lines Y1, Y2, Y4, and Y5 expression was low in ESCs, then showed a large increase in expression levels upon differentiation. This is in-line with upregulation of alpha globin expression in developing haematopoietic lineages, and in stark contrast to the silencing effect observed in ZF α antisense cell lines. The cell line Y6 showed reduced expression in ESCs compared to other cell lines and did not become highly up regulated upon differentiation, with levels remaining steady or showing only a minor increase.

Expression of mouse alpha globin was also examined by SYBR green real-time PCR for the same cell lines (Figure 3.13C). mouse alpha globin expression is shown with a logarithmic scale to allow visualization of expression in both ESCs and EBs on the same graph. For cell lines grown for in the initial study (Figure 3.13E), mouse alpha globin becomes highly up regulated upon differentiation, clearly indicating differentiation towards a haematopoietic lineage. The cell lines Y5 and Y6, grown later for analysis of histone modifications, also demonstrated expression of mouse alpha globin upon differentiation (Figure 3.13F). While expression in Y5 was high, expression in Y6 was lower than observed in the other cell lines, indicating that a proportion of cells were not differentiating down haematopoietic lineage.

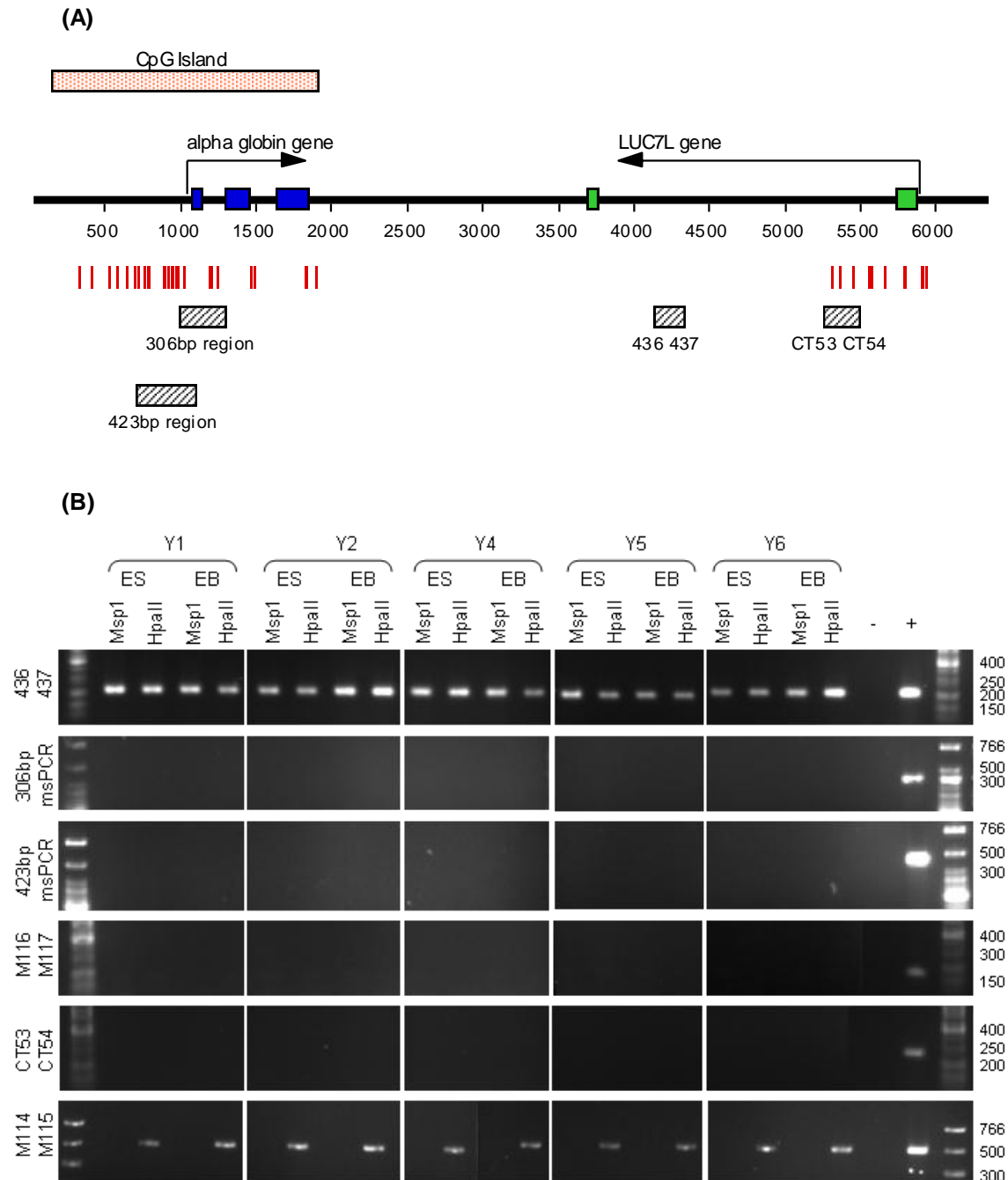
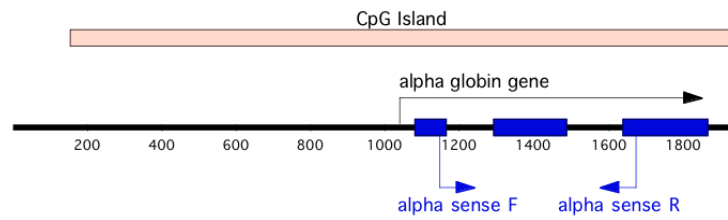
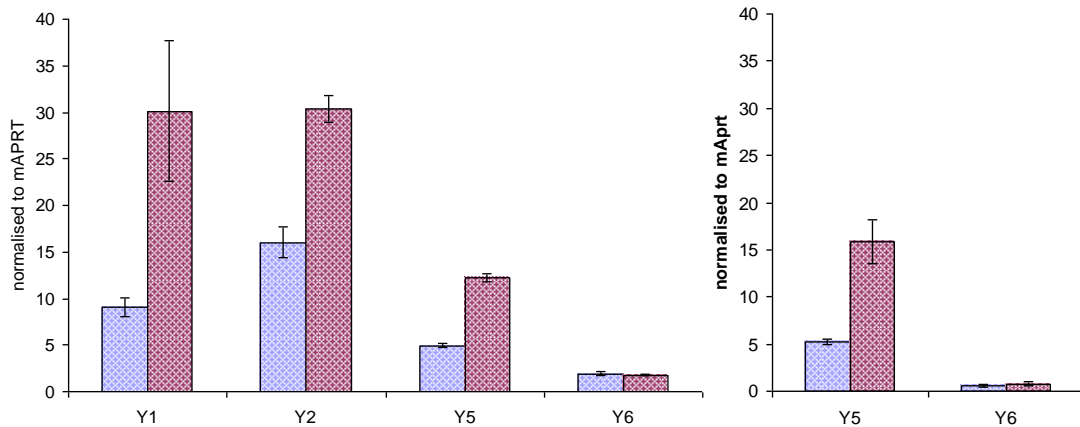


Figure 3.12 Methylation sensitive PCR within the alpha globin CpG Island. (A) shows a schematic of the antisense stop construct with Msp1/HpaII sites marked as vertical red lines and regions amplified by PCR primers represented as hatched grey boxes. Within the LUC7L CpG island; CT53 CT54 amplify across 3 restriction sites (248bp), and 436 437 primers amplify a region with no restriction sites (208bp). The 306bp msPCR product encompasses 3 restriction sites, and the 423bp msPCR product encompasses 4 restriction sites. (B) M116 M117 amplify a region within the 5' end of the APRT gene that contains 4 restriction sites and is devoid of DNA methylation (164bp), M114 M115 primers amplify within the promoter of AIR, which is stably methylated on the maternal copy in ESCs and EBs (489bp).

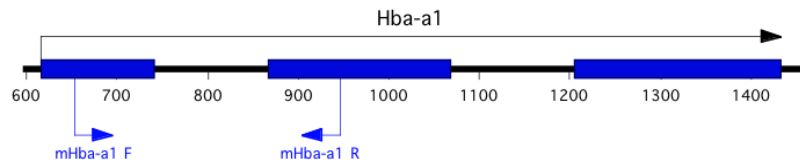
(A) Position of human alpha globin primers



(B) Expression of human alpha globin in antisense stop cell lines



(C) Position of mouse alpha globin primers



(D) Expression of mouse alpha globin in antisense stop cell lines

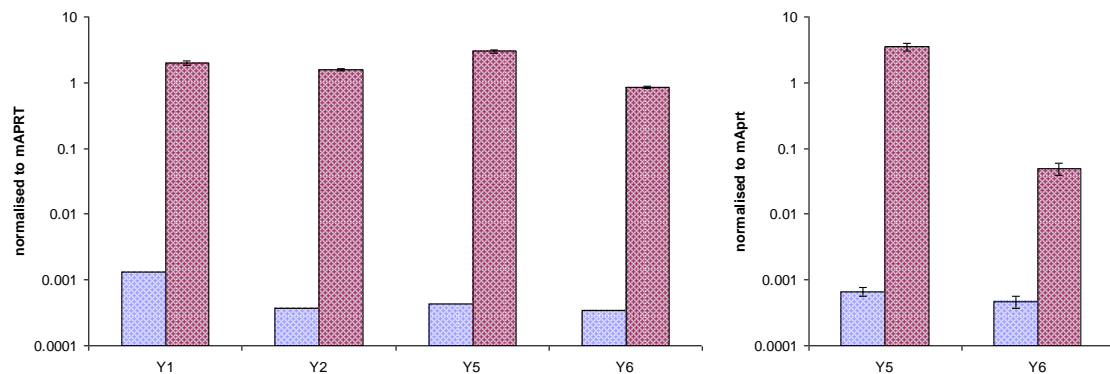


Figure 3.13 Expression of human and mouse alpha globin in antisense stop cell lines. (A) Schematic diagram showing the location of the primers used to amplify alpha globin cDNA for SYBR green realtime PCR. (B) Expression of human alpha globin. Cell lines Y5 and Y6 were re-grown for chIP experiments. (C) Schematic diagram showing the location of mouse alpha globin primers (D) Expression of mouse alpha globin shown with a logarithmic scale

Histone modifications were examined in the cell lines Y5 and Y6 (Figures 3.14 and 3.15 respectively). H3K9me3 levels for both cell lines showed no increase in enrichment upon differentiation, and the level of enrichment at all four points assayed was equivalent to or less than the level of enrichment seen at the endogenous control Actb. A very similar pattern of enrichment was observed for the repressive histone modification H3K27me3, where levels of enrichment were equivalent to Actb levels in both ESCs and EBs. H3Ac levels showed only minor fluctuations in enrichment levels upon differentiation, with most regions examined showing either no decrease or a slight increase in levels. Positions alpha 2 and alpha 3 showed some reduction in H3Ac levels in the cell line Y6 but reductions were small. H3K4me3 levels were only examined in the cell line Y6, and were seen to remain steady at three out of the four positions examined with no notable decrease upon differentiation, while a decrease in H3K4me3 levels was observed at position alpha 2.

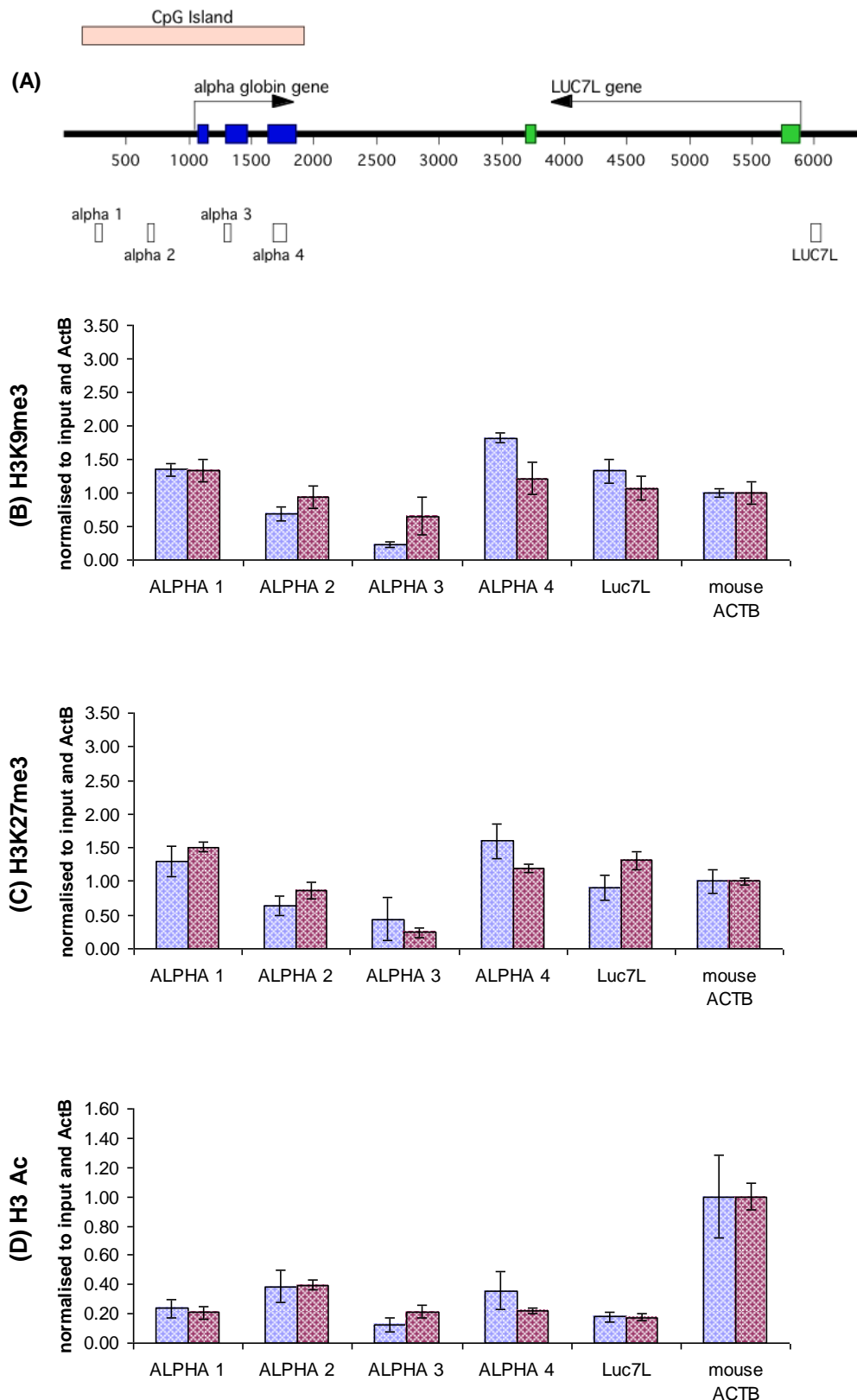


Figure 3.14 Histone modifications to the antisense stop construct in the cell line Y5. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the HBA2 gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

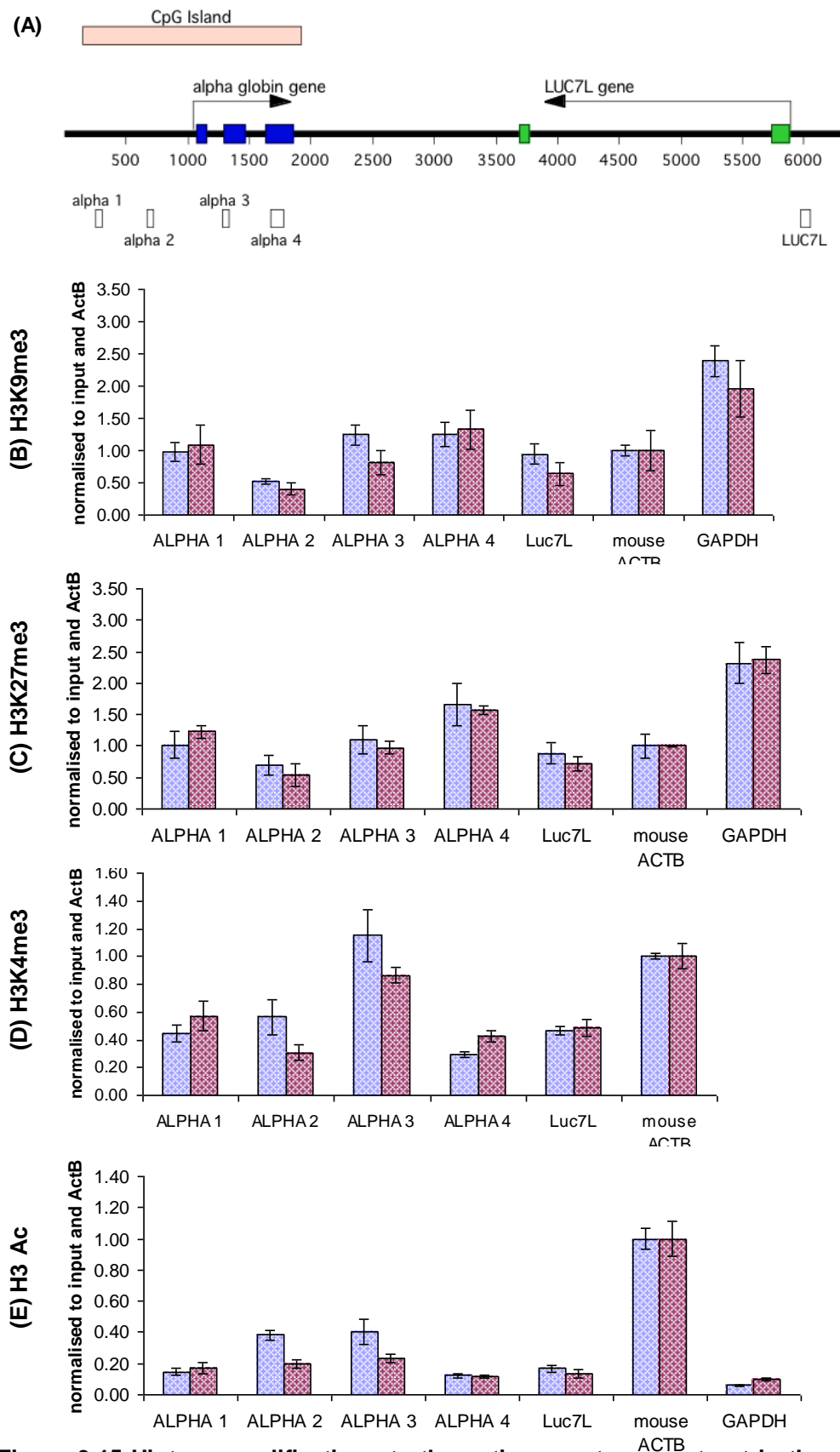


Figure 3.15 Histone modifications to the antisense stop construct in the cell line Y6. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the HBA2 gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

3.4 Discussion

3.4.1 Silencing at the alpha globin gene

The results described in this chapter show that onset of silencing at the alpha globin gene in the 2A3 cell line was a sudden event with the gene unrepressed and effectively unmarked by repressive modifications on d3, but with high levels of DNA methylation, and enrichment for the repressive histone modification H3K9me3 by d4 of differentiation with reciprocal loss of histone modifications associated with gene expression. Detectable levels of alpha globin mRNA, for which the most detailed time point data was available for, showed stable expression up to d3.75 but with levels then being dramatically reduced to less than a third within just 12hrs. This sudden act of repression suggests a close coordination of both DNA methylation and histone modification machinery to the extent that de-lineation of repressive events was not possible. The slight lag in alpha globin mRNA reduction compared to repressive events at the gene itself most likely represent the time taken for loss of expression to become apparent while mRNA transcribed prior to the silencing event reached its natural half life.

The pluripotency genes *Oct4* and *Nanog* are highly expressed in ESCs, but become repressed upon differentiation, associated with DNA methylation and H3K9me3 enrichment, mediated by the methyltransferase G9a (Epsztejn-Litman *et al.* 2008). Over one hundred genes are silenced at the same time as *Oct4* and *Nanog*, in a great wave of repression of genes involved in regulation and development of the pre-implantation embryo. Examination of expression by a multiplex PCR approach revealed that *Oct4* and *Nanog* exhibit identical repression between d4 and d5, approximately 24hrs later than repression of the

alpha globin gene, though this is sufficiently similar to prevent ruling out of a link. However, repression of the genes characterized by Epsztejn-litman *et al.* is thought to be mediated through transcription factor binding, with as yet no suggested role for a ncRNA.

Antisense RNA is expressed in ESCs and EBs at comparative levels (Tufarelli *et al.* 2003) by the ubiquitously expressed gene LUC7L. However; silencing of the alpha globin gene occurs only upon differentiation, and the silencing effect is rapid, with gene repression, loss of active chromatin marks, and enrichment for H3K9me3 and DNA methylation occurring within a 24hr time period between d3 and d4. Both X inactivation and somatic imprinting involve DNA methylation, H3K9me enrichment, with silencing dependent upon ncRNAs, and repression in both cases occurs in a developmentally linked manner, and occurs *in cis*. Differentiating ESC models of both of these processes have shown that repression occurs within the first few days of differentiation in a manner that has many parallels with the silencing effect at the alpha globin gene.

In mice, X inactivation occurs two days after the implantation of the blastocyst, between 5.5 and 6.5d.p.c (Wutz 2007). X inactivation has been investigated in mESCs differentiated by both EB formation in suspension (Keohane *et al.* 1996), as used in this study, and by differentiation using Retinoic Acid (Chaumeil *et al.* 2006, Heard *et al.* 2001), however; despite these differences in methodology very similar results were obtained pertaining to the timing of X inactivation in *in vitro* differentiated ESCs. EBs grown in suspension showed silencing of one copy of the X chromosome beginning at d2/d3 of differentiation

as evidenced by reduction in H4Ac levels and late replication of the affected X chromosome due to chromosomal condensation. This was corroborated by gene expression data, showing repression starting at d2 of differentiation and being complete by d4 of differentiation, also; the ncRNA Xist was expressed at d2, with further upregulation shown by d4 (Keohane *et al.* 1996). Studies that induced differentiation of mESC monolayers using RA showed accumulation of Xist by d2, with gene repression generally occurring within 24hrs, and with inactivation virtually complete by d4 (Chaumeil *et al.* 2006, Heard *et al.* 2001).

Imprinting of the paternal copy of the Igf2r gene, through the expression of the ncRNA Airn, occurs in post-implantation mouse embryos between 4.5 and 6.5d.p.c. Igf2r imprinting was investigated using mESCs differentiated using RA, and EBs formed using the hanging drop method then cultured in suspension (Latos *et al.* 2009). In both differentiation methods, expression of Airn, and expression of maternal Igf2r - up regulated by Airn, was shown to occur between d2 and d4, with further upregulation 24 to 48hrs later. There were differences in when DNA methylation became observable; low levels were observable by d3 in RA differentiated cells, but equivalent levels of DNA methylation took longer to develop in EBs differentiated using the hanging drop method, being detectable by d6. H3K9me3 levels were not examined in detail, but were shown to increase significantly at the Igf2r gene after differentiation (Latos *et al.* 2009). Latos *et al.* also examined imprinting at the Kcnq1 locus, where they found a similar upregulation of maternal Kcnq1 and the ncRNA Kcnqt1 between d2 and d5 using RA differentiation, but upregulation occurring much later in EBs, at around d10.

X inactivation, and imprinting at the *Kcnq1* and *Igf2r* loci has been extensively studied, with a model involving higher order chromatin organization leading to the development of a silencing compartment devoid of pol II and H3K4me3, and enriched for H3K27me3, and in some cases H3K9me3, being advanced as a common theme in repression in all three instances (Pandey *et al.* 2008, Nagano *et al.* 2008, Terranova *et al.* 2008, Chaumeil *et al.* 2006). Development of the silencing domain relies upon coating of the region of the chromosome by ncRNA; Xist in the case of X inactivation, *Kcnq1ot1* and *Airn* at the two imprinted clusters. ncRNA mediated repressive domains develop early in embryogenesis prior to the main silencing event and rapidly exclude Pol II and other components of the transcription machinery. For example, total exclusion of the transcription machinery from Xist RNA domains occurs by d3 in X inactivation (Chaumeil *et al.* 2006), with a similarly early development of repressive domains thought to occur at the imprinted clusters examined (Terranova *et al.* 2008). Physical re-localization of genes from the still active edges of the repressive domain to inside the silencing compartment then occurs (Terranova *et al.* 2008, Chaumeil *et al.* 2006), resulting in gene repression. Development of the repressive domain relies upon expression of long ncRNAs from within the repressed locus. Terranova *et al.* has shown that a region of transcriptionally permissive chromatin remains at the *Kcnq1* promoter, enriched with the active histone modification H3K4me3 despite repression in the surrounding regions (Terranova *et al.* 2008).

Development of these repressive domains is associated with repressive histone modifications; H3K9me3 becomes steadily enriched on the X chromosome undergoing silencing within the first few days of differentiation, and becomes increasingly enriched concordant with deepening gene repression. H3K27me3, mediated by PRC2, is also present and shows greater levels of enrichment than H3K9me3, though follows a near identical enrichment as differentiation continues (Chaumeil *et al.* 2006). Both DNA methylation and H3K9me3 become enriched at the paternal Igf2r gene following differentiation (Latos *et al.* 2009), though DNA methylation is partial, with complete methylation only occurring after gestation (Sleutels *et al.* 2003). DNA methylation along with H3K9me3 is also enriched at the paternally silenced Kcnq1 gene in embryonic tissues (Pandey *et al.* 2008, Lewis *et al.* 2004).

In both X inactivation and somatic imprinting there is no repression in pre-implantation embryos, or gametes, with repression rapidly established after implantation and maintained in nearly all adult tissues (Li 2002, Reik *et al.* 2001, Reik 2007). This pattern of repression is also true of silencing at the alpha globin gene, with no repression in spermatozoa but repression seen in all adult tissues examined (Barbour *et al.* 2000). Taken together with the obvious parallels in timing in repression at the alpha globin gene, of genes on the repressed X chromosome, and imprinted genes like Igf2r, this raises the possibility of silencing of the alpha globin gene involving development of repressive domains. However; repression at the alpha globin gene does not result in either spreading of repression to neighboring regions, or in late replication (Barbour *et al.* 2000, Tufarelli *et al.* 2003), and the silenced alpha

globin locus does not associate with centromeric heterochromatin (Brown *et al.* 2001), while spreading of repression is in many ways a hallmark of X inactivation and imprinting, while late replication is a feature of X inactivation.

There also appears to be differences in the kinetics of repression. Silencing at imprinted genes and during X inactivation is spread out over at least a 48hr period, likely a direct consequence of a requirement for the initial development of the repressive domain, then the subsequent sequestration of genes within that results in the observed steady, linear, repression of genes and build-up of repressive epigenetic modifications. Repression of the alpha globin gene is startlingly rapid by comparison, being concluded within a 24hr period. Additionally, DNA methylation in X imprinting occurs slowly and progressively, suggesting a maintenance role, whereas DNA methylation at the alpha globin gene shows a sudden increase of almost 50% in a day, within the same time period at H3K9me3 enrichment, suggesting that in this case it may have a greater role than purely maintenance.

While the specifics of imprinting and X inactivation show deviation from the observations of the silencing effect at the alpha globin gene, the general mechanism of a long ncRNA recruiting repressive cellular machinery, which includes HDACs, HMTs, and DNMTs, and implementing silencing via establishment of a repressive domain within a narrow developmental window, warrants further study. Immunofluorescence using antibodies against key components of the transcriptional machinery, such as RNAP II and TBP, coupled with RNA FISH against the antisense RNA produced by the *LUC7L*

gene, using cells from time points covering repression during d3 and d4 of differentiation would indicate if a repressive, though limited, repressive domain devoid of transcriptional machinery was developing in a manner analogous to the events observed in X inactivation or imprinting.

3.4.2 Prevention of antisense RNA through the alpha globin gene

Introduction of the terminator from the beta globin gene reduced the level of antisense transcripts through the alpha globin CpG island to a nearly undetectable level and resulted in the complete abrogation of the silencing effect, with no detectable enrichment of H3K9me3 upon differentiation, no silencing of alpha globin transcription, and no detectable DNA methylation, clearly demonstrating the essential role played by the antisense transcript in the silencing effect.

Examination of histone modifications across the alpha globin gene showed no increase in enrichment for H3K9me3 upon differentiation, with enrichment levels generally lower than those detectable at the ubiquitously expressed gene Actb. H3K27me3 levels showed a similar pattern of enrichment with no real change upon differentiation, and while H3K4me3 levels showed some minor fluctuations they also remained fairly constant. These changes were all in stark contrast to the observed histone modification patterns observed in the 2A3 cell line where antisense RNA is present across the length of the alpha globin gene. DNA methylation was undetectable by either southern blotting or msPCR.

Two groups of cell lines carrying the ZF α antisense Stop construct were analyzed in the course of this study, the second group, analyzed as an

experimental repeat of two cell lines, were grown for analysis of histone modifications. Results between these biological repeats were effectively identical. One cell line, Y6, showed a higher level of expression of pluripotency factors, though well formed EBs had developed, and expression of mouse alpha globin indicated differentiation had occurred, albeit not complete suggesting partial differentiation of some cells. Histone modification analyzed for this cell line matched those observed for the cell line Y5.

These data confirm that antisense RNA or the process of antisense transcription through the tissue specific alpha globin CpG island is necessary for silencing to occur. The next chapter will address some of the molecular mechanisms that may contribute to repression, while later chapters will investigate whether other tissue specific or ubiquitously expressed genes associated with CpG islands are sensitive to this form of silencing.

Chapter 4 Investigation of molecular mechanisms potentially involved in the antisense RNA initiated silencing effect

4.1 Introduction

The silencing effect observed at the alpha globin gene in the ZF patient, and recapitulated in the mouse ES cell system, has several key defining features. The repressive histone modification H3K9me3 becomes enriched indicating the presence of at least one HMT, and dense promoter methylation occurs indicating a requirement for *de novo* methyltransferases. Antisense RNA, transcribed through the alpha globin gene, has also been shown to be prerequisite for silencing to occur. Histone methyltransferases have been shown to interact directly and indirectly with DNA methyltransferases (Epsztejn-Litman *et al.* 2008, Lehnertz *et al.* 2003) and form multimeric complexes involving various repressive factors such as HP1, MeCP2, HDACs and PRC components (Lehnertz *et al.* 2003, Vaute *et al.* 2002, Ogawa *et al.* 2002). While such groupings of repressive factors naturally exist within the cell, they require to be targeted in a specific manner to their sites of action, and RNA has been shown to play an essential role in targeting transcriptional repression in a wide range of organisms through interaction with members of the argonaute family (Corey 2005, Grimaud *et al.* 2006, Sigova *et al.* 2004).

Argonaute proteins are central and ubiquitous components of the diverse systems involved in gene silencing, being essential components of the RNAi complex - implicated in both PTGS and TGS. They also associate with RNAPII and PcG proteins (Kim *et al.* 2006, Tolia & Joshua-Tor 2007, Parker & Barford 2006, Janowski *et al.* 2006). As discussed in the introduction to this thesis,

AGO1 has been implicated as the argonaute family protein responsible for TGS in mammals, with several important similarities in the manner of mammalian AGO1 silencing when compared to the role of AGO1 in *S.pombe*, the ancestral argonaute protein that gave rise to all four AGO proteins that are present in mouse and humans.

4.1.1 Structure of AGO proteins

AGO proteins comprise of two domains, with the PIWI domain divided into two subdomains with distinct roles in siRNA mediated repression (Figure 4.1). Upon binding of an siRNA duplex, the AGO protein separates the two strands, with the guide strand being bound by the PAZ domain. The PAZ domain, structurally similar to the PAZ domain of Dicer proteins, has a high affinity for single stranded 3' end overhangs of duplex siRNAs. The passenger strand of the siRNA is destroyed by Slicer activity of the PIWI subdomain.

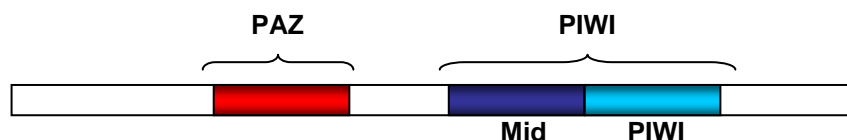


Figure 4.1 schematic representation of the primary structure of a canonical Ago protein Ago proteins contain two main domains, Paz and Piwi. The PAZ domain binds RNA. The PIWI domain is subdivided into the Mid and PIWI sub domains. The phosphorylated 5' end of the guide strand is bound in the positively charged pocket created by this domain. The PIWI sub-domain contains the catalytic 'Slicer' activity. (Information from Parker & Barford 2006; Tolia & Joshua-Tor 2007).

In mammals it has been shown that it is the siRNA strand antisense to the target that is maintained as the guide (Weinberg *et al.* 2006). The PIWI subdomain contains the RnaseH-like 'slicer' nuclease function that degrades targeted transcripts. The PIWI and Mid subdomains form the PIWI domain, together creating a positively charged pocket that binds the phosphorylated 5'

end of the guide strand of the small RNA (Peters & Meister 2007, Parker & Barford 2006, Liu *et al.* 2004).

4.1.2 AGO1 and TGS in *S.pombe*

The AGO containing silencing complex in *S.pombe*, the RNAi-induced Initiation of Transcriptional gene Silencing complex (RITS) contains AGO1, the chromodomain containing protein Chp1, and Tas3 of unknown function. RITS silences heterochromatic repeats and transposable elements, as well as some transgenes, by implementing repressive histone modifications. RITS mediates gene repression at a local level, *in cis* (Sigova *et al.* 2004, Verdel *et al.* 2004, Irvine *et al.* 2006, Noma *et al.* 2004).

AGO1, as a component of the RITS complex, combines with siRNA processed from dsRNA by Dcr1, the *S.pombe* Dicer homolog. RNA originating from heterochromatic loci is converted to dsRNA by the action of RNA-dependent-DNA-polymerase (Rdp1). Through base pair complementarity, either with nascent RNAs or possibly DNA (Irvine *et al.* 2006), the RITS complex relocates to the siRNAs site of origin and recruits additional silencing factors (Figure 4.2A) (Noma *et al.* 2004). It is thought that the RITS complex interacts with Pol II, as mutation of key residues within Pol II abolishes silencing to the same extent as mutants lacking Dcr1, suggesting a coupling of transcription with siRNA production and ultimately silencing (Kato *et al.* 2005).

The initial stages of silencing occur *in trans*, and are thought to be fairly low-key, with silencing only becoming established if the loci that the RITS complex relocates to is being transcribed (Hall *et al.* 2002, Volpe *et al.* 2002). Without

transcription, no additional siRNAs are forthcoming, the silencing event fails to initialize on a great enough scale, and any silencing effect is transitory. Upon relocating, the RITS complex recruits multiple factors, including the histone methyltransferase Clr4 (Homolog of *suv39h* in mammals) and the deacetylating complex SHREC (Noma *et al.* 2004, Sugiyama *et al.* 2007).

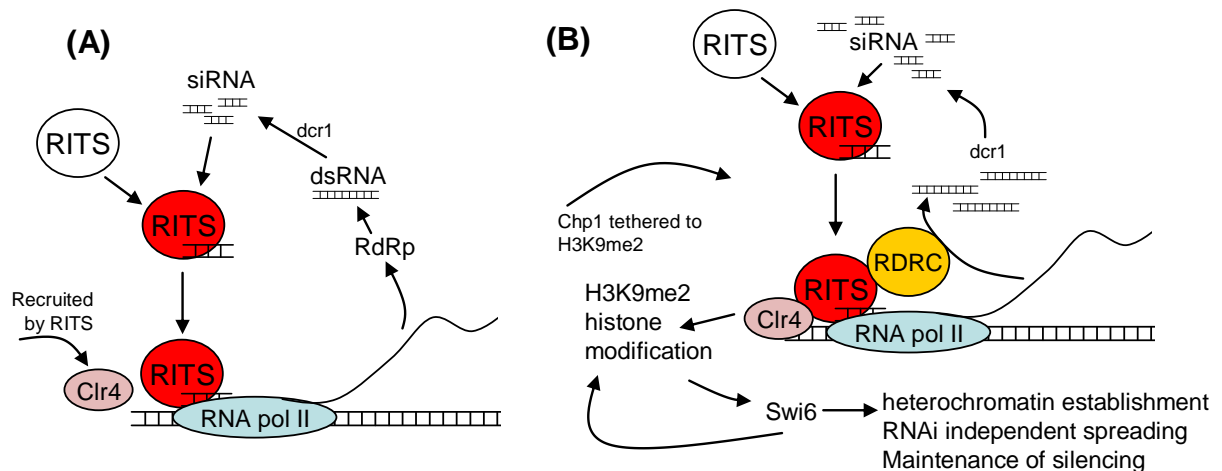


Figure 4.2 Transcriptional gene silencing in *S.pombe*. (A) siRNA originating at a heterochromatic loci are loaded into RITS and result in RITS relocation to the loci, and the recruitment of further silencing factors. (B) RITS becomes tethered to the loci through interaction between Chp1 and H3K9me2. This initiates further siRNA production, RITS recruitment, and the establishment of silencing by Swi6. This occurs in a local manner, *in cis* (Grewal & Elgin 2007, Hall *et al.* 2002, Irvine *et al.* 2006, Motamedi *et al.* 2004, Noma *et al.* 2004, Volpe *et al.* 2002)

Fission yeast do not methylate DNA as a form of control, but do modify histones in a manner conserved in higher eukaryotes. Clr4 catalyses H3K9me2, which serves to recruit the remainder of the silencing machinery and provides a means to tether the RITS complex to the specific loci via the chromodomain of Chp1. Tethering of RITS ensures the continued association of the silencing complex with the region, allows positive feedback, and ultimately results in silencing (Figure 4.2B) (Hall *et al.* 2002, Volpe *et al.* 2002).

Once RITS is tethered it recruits additional Dcr1, and Rdp1 as part of the RNA directed RNA polymerase (RDRC) complex. The RDRC physically interacts with Dcr1 thereby bringing the required cellular machinery for siRNA production into a local context (Colmenares *et al.* 2007). siRNAs produced cause recruitment of further RITS complexes, escalating the silencing effect. H3K9me2 recruits the chromodomain protein Swi6 (homolog of HP1) that establishes and then maintains silencing. The production of further siRNAs from the locus is essential for further silencing and, in conjunction with the recruitment of Swi6 results in the *in cis* silencing effect where the silencing complex is tethered to a specific region and acts in a local manner (Hall *et al.* 2002). Once Swi6 has been recruited and silencing established, repression can spread along the chromosome to nearby regions. The maintenance and spreading of silencing occurs in an RNAi independent manner and is maintained even in strains defective for RNAi components, but new regions of heterochromatin cannot be established (Irvine *et al.* 2006, Noma *et al.* 2004).

4.1.3 AGO1 and transcriptional gene silencing in mammals

The role of AGO proteins in higher eukaryotes is not as well characterized as in *S.pombe*, however recent studies have demonstrated that a similar silencing complex may exist in mammals. There are four AGO proteins in mammals, but only AGO2, the core component of the RNA Induced Silencing Complex (RISC) has 'slicer' activity. AGO2, as part of the RISC, is responsible for PTGS in mammals (Sasaki *et al.* 2003, Liu *et al.* 2004). Mammalian AGO1 is more widely expressed than AGO2, and has been shown to be present in the nucleus (Janowski *et al.* 2006) where it has been shown to localize to promoters targeted by siRNAs, and this colocalization coincides with the appearance of

repressive histone modifications, and histone methyltransferases (Hawkins *et al.* 2009, Kim *et al.* 2006) in a similar manner to the Ago1 protein in *S.pombe*.

The ability of small RNA duplexes targeted against the promoter region of a gene to cause gene repression involving either H3K9me2, H3K27me3, DNA methylation, or a combination of the three has been demonstrated by various studies (Hawkins *et al.* 2009, Castanotto *et al.* 2005, Suzuki *et al.* 2005, Ting *et al.* 2005, Morris *et al.* 2004, Weinberg *et al.* 2006, Kim *et al.* 2007), but the identification of the enzymatic components underlying the chromatin changes has occurred only recently.

In an initial study into the involvement of AGO1 in TGS by Kim *et al.*, it was shown that enrichment of AGO1 at promoter regions occurred shortly after treatment of human cell lines with siRNAs against the promoter, prior to the enrichment of both H3K9me2 and H3K27me3 and establishment of DNA methylation. H3K9me2 was highly enriched within the siRNA target region, and continued to be enriched, to a lesser extent, 300bp downstream. Knockdown of AGO1 reduced the silencing effect, demonstrating its role in the silencing process. Additionally, it was demonstrated that AGO1 and RNAPII were interacting directly, and both were present at the repressed loci, though AGO1 levels reduced sharply after the initial silencing event (Kim *et al.* 2006). The same group later demonstrated a very similar endogenous silencing effect, where a promoter region transcribed miRNA caused silencing of the POLR3D gene through recruitment of AGO1 (Kim *et al.* 2008). In the groups' most recent study, siRNA mediated knockdown of potential components of the silencing

complex identified AGO1, Dnmt3a, HDAC1, and to a lesser extent G9a, as essential components in initiation of silencing at a siRNA targeted promoter. AGO1 was not required for maintenance of silencing (Hawkins *et al.* 2009).

While these, and some previous studies, have focused on the histone modification H3K9me2 in connection with G9a, this histone methyltransferase is also capable of catalyzing the addition of H3K9me3 as a form of repression, as was demonstrated in a recent study where a network of over a hundred genes was shown to be controlled by G9a mediated DNA methylation and H3K9me3 (Epsztejn-Litman *et al.* 2008). Therefore the indications that G9a is a component of the mammalian TGS machinery, and the hypothesis that silencing at the alpha globin gene, involving H3K9me3, could be mediated by this complex is entirely consistent.

Suv39h mediated silencing at regions of pericentric heterochromatin, involving H3K9me3 and DNA methylation (Lehnertz *et al.* 2003, Peters *et al.* 2001), is at least partly reliant upon Dicer (Kanellopoulou *et al.* 2005). Dicer is essential for siRNA mediated PTGS and TGS, as it cleaves long dsRNAs into siRNAs. The link between siRNAs and Suv39h is reminiscent of TGS in budding yeast, and suggests that Suv39h could also be component of an AGO containing complex, targeted by siRNAs to pericentric regions.

The similarity in components required for silencing in *S.pombe* and those shown to be involved in mammalian TGS is striking. The core silencing components in *S.pombe* include Clr4, a homolog of the mammalian HMT SUV39H, HDACs,

and the Swi6 protein, homolog of HP1 (Volpe *et al.* 2002, Noma *et al.* 2004), which is almost identical to components identified in Hawkins *et al.* 2009. There is also similarity in the method of silencing, with repression coupled to transcription through direct interaction between AGO1 and RNAPII. Additionally; establishment requires AGO1, but maintenance is independent of the RNAi machinery, with repressive histone modifications spreading from their initial nucleation site (Hawkins *et al.* 2009, Kim *et al.* 2006, Weinberg *et al.* 2006, Irvine *et al.* 2006).

The AGO1 paralog in *S.pombe* causes transcriptional gene silencing *in cis* as siRNAs are both produced and used in a local context (Hall *et al.* 2002), and while this is not a feature yet associated with mammalian AGO1 silencing, the investigation into the silencing complex is very much in its nascency. Experiments to date have focused on transient siRNA where excess levels of siRNAs are introduced to cells in their growth media. Therefore; the potential for an *in cis* system, where low level siRNA production from a single locus orchestrate a local silencing effect, cannot be discounted.

The alpha globin gene is expressed in ESCs prior to differentiation. With both sense and antisense transcription is occurring, this raises the possibility of dsRNA formation that could then be processed by Dicer into siRNAs, subsequently targeting an AGO containing silencing complex to cause both H3K9 and DNA methylation.

This chapter examines the potential for dsRNA formation resulting in siRNA production, and assesses the potential role that AGO1 could play in the alpha globin silencing event. First a modified northern blotting protocol was used to determine if any siRNA corresponding to the alpha globin CpG island could be detected in differentiating ESCs. Second, the 2A3 cell line was stably transfected with Ago1 shRNA constructs to produce Ago1 knockdown ES cell lines and test what effect this has on alpha globin silencing. The preexisting ESC line 2A3, containing the ZfαAS construct was used to screen for small RNAs, and when creating stable knockdown cell lines for Ago1.

4.2 Materials and Methods

Primers used at the alpha globin gene are shown in Table 3.1, Chapter 3.

Primers used in this chapter are in Table 4.1

In general, expression of antisense RNA and the presence of DNA methylation was examined twice for each cell line. Gene expression analysis and Chromatin IP experiments were only carried out once for each cell line, unless otherwise stated.

4.2.1 Screening for small RNAs

If an siRNA mediated silencing mechanism were responsible for the silencing effect, then siRNAs would be most likely to be detectable prior to day 6 of differentiation, when expression of alpha globin is heavily repressed (see the previous chapter). Therefore; total RNA from a differentiation timecourse from ESCs to day 6 EBs was used in screening for small RNAs. RNA from the ESC cell line 2A3 was provided by Dr Cristina Tufarelli. Total RNA was run on a polyacrylamide gel, transferred to a nylon membrane, and hybridized with oligonucleotide probes to detect small RNA species in accordance with published techniques (Pall *et al.* 2007).

4.2.1.1 RNA preparation

10µg total RNA was combined with an equal volume of formamide loading buffer (95% Deionised Formamide, 1mM EDTA (pH8), 0.01% Bromophenol blue) and denatured at 95°C for 5min, then kept on ice. A Polyacrylamide gel (15% Polyacrylamide, 7M Urea, 1xMOPS, 3.5mM ammonium persulphate,

0.05M TEMED) was prepared and loaded into a vertical gel tank (Fisher Scientific) containing 1xMOPS buffer (0.1M MOPS (pH7), 10mM EDTA(pH8) 20mM sodium acetate). The gel was pre-run at 200v for 30mins, then the wells were washed out with 1xMOPS and the denatured RNA samples loaded. In wells at either end of the gel, a microRNA Marker (NEB N2102) was loaded. Electrophoresis was carried out at 200v for 1 hour, then at 500v for 1½ hours.

Primer	Sequence, 5' to 3'	PCR buffer	Annealling temp °C	Notes
PAC F	GTCTCCAGGAAGGCGGGCAC	Standard NEB buffer	60	Used to amplify the PAC gene in the Ago Kd construct to confirm insertion of construct into cell lines
PAC R	CGACATCGGCAAGGTGTGGG			
pSuperior F	TGGATGTGGAATGTGTGCGAG	(Sequencing)	-	Used for sequencing of shRNA insert to check for correct insertion into plasmid
m13Rev	CAGGAAACAGCTATGAC	(Sequencing)	-	Used for sequencing of shRNA insert to check for correct insertion into plasmid

Table 4.1. Primers used on Ago Kd construct

4.2.2.2 Transfer by Electroblothing

The polyacrylamide gel was removed from the tank and the glass plates separated. The gel was stained in 4µg/ml EtBr in 1xMOPS for 5min, and photographed. A stack of three layers of 3MM whatmann paper, wetted in DEPC treated water (0.1% DEPC final concentration in deionised water, incubated 1hr at 37C, autoclaved on standard cycle) was built up on the positive plate of the semi-dry electroblotter (Fisher Scientific), then a section of nylon membrane (Amersham RPN203T) was wetted in DEPC water and placed on top. The Polyacrylamide gel was transferred on to the nylon membrane and covered with a stack of three layers of 3MM whatmann paper, wetted in DEPC treated water. The negative plate of the semi-dry electroblotter was then added, and any excess moisture removed with a paper towel. The semi-dry electroblotting unit was transferred onto a bed of ice and run at 20v for 1 hour.

4.2.2.3 Chemical Crosslinking

Crosslinking solution was prepared by adding 245µl of 12.5M 1-methylimidazole stock solution to 9ml DEPC treated water, the pH of the solution was adjusted to 8.0 with 1M HCL. 0.753g of EDC (1-ethyl-3-dimethylaminopropyl carbodimide) was added and the final volume adjusted to 24ml using DEPC treated water. A section of 3MM whatmann paper just larger than the nylon membrane was soaked in the crosslinking solution, and the membrane was transferred on top orientated so that the RNA side of the membrane did not come into direct contact with the whatmann paper. Excess fluid was removed from around the nylon membrane, then the membrane/Whatmann paper sandwich was wrapped in clingfilm and incubated at 60°C for 2hrs. The

membrane was washed in DEPC water, dried between two sheets of 3mm Whatmann paper and stored at room temperature in the dark.

4.2.2.4 Hybridisation of membrane

The membrane was pre-hybridized with 15ml of Hyb solution (5xSSPE, 7.5xDenhardt's, 0.1% SDS, 0.05mg/ml yeast tRNA) for 3hrs at 37°C with constant lateral rotation. The Hyb solution was discarded and a fresh 15ml added. An oligonucleotide probe was end labeled (50pMol Oligonucleotide Probe, 1X T4 Polynucleotide Kinase, 10U T4 Polynucleotide Kinase, 25µCi γ-32P-ATP) by Incubation for 30 minutes at 37°C, then purified using a purification spin column (Amersham Pharmaceuticals). The probe was incubated with the membrane overnight at 37°C with constant lateral rotation. The membrane was washed twice for 10min at room temperature in Wash Buffer1 (5x SSPE, 0.1% SDS), then twice for 15min at 37°C in Wash Buffer2 (1x SSPE, 0.1% SDS). The membrane was then wrapped in clingfilm and exposed.

4.2.2 Engineering of Ago1 Kd cell lines

4.2.2.1 Selection of siRNA target sequences

Predesigned shRNAs for a large range of different mouse and human genes were available from the Broad institute, a collaboration between Harvard University, its affiliated hospitals, the Massachusetts Institute of Technology, and the Whitehead Institute for Biomedical Research (<http://www.broad.mit.edu/node/297>). The Broad institutes' mouse database (http://www.broad.mit.edu/genome_bio/trc/publicSearch_ForHairpinsForm.php)

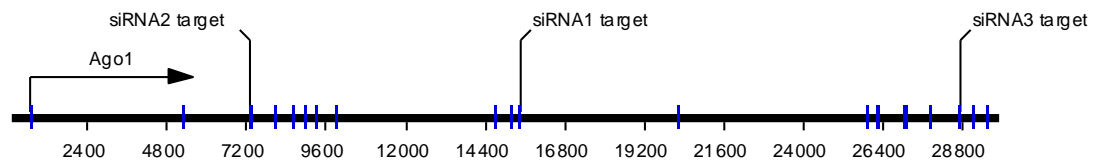
contained sequences for five shRNAs that targeted Ago1 mRNA. Three shRNAs were chosen that covered the extent of the main Ago1 transcript, including the catalytic PAZ and PIWI domains (Figure 4.3). The sequences consist of a siRNA 21 nucleotides long, a stem loop of 9 nucleotides, then the complementary 21 nucleotides of the siRNA, and finally the Pol III transcriptional terminator sequence TTTT (Figure 4.4). ShRNAs in the database conform to recommended siRNA design for maximum efficiency in PTGS. Briefly; all siRNAs contain an A/U at the 5' end of the antisense strand, a G/C at the 5' end of the sense strand; contain at least 5 A/U residues in the 5' terminal one-third of the antisense strand, and do not contain GC stretches of more than 9nt in length (Ui-Tei *et al.* 2004). The Stem loop structure is based upon the *C.Elegans* shRNA Let-7, has been experimentally tested in mammalian cells, and shown to be readily processed to an siRNA for efficient PTGS (Brummelkamp *et al.* 2002).

4.2.2.2 Cloning of target sequences into expression vector

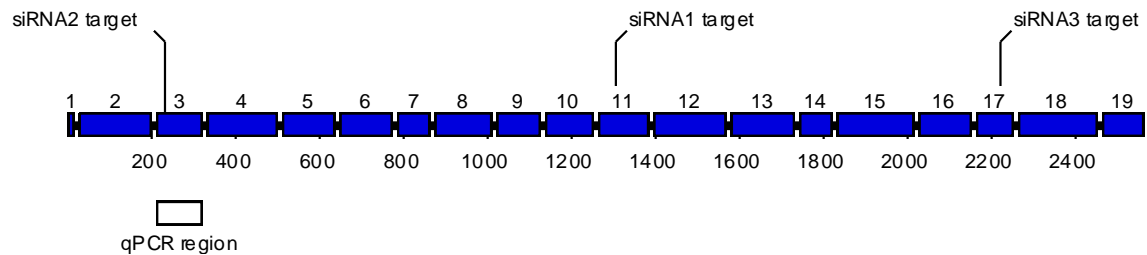
The three shRNA sequences were adapted for cloning into the vector pSuperior.puro, which contains the H1 RNA pol III promoter, by the addition of complementary sticky ends to the sites created by the restriction enzymes Bgl II and XhoI (Figure 4.4A). The shRNA sequences were purchased as individual oligonucleotides then annealed together. 500pMols of each oligonucleotide was combined with 5µl of oligo annealing buffer (200mM Tris HCL pH8, 10mM EDTA, 500mM NaCl) in a total volume of 50µl dH₂O, vortexed briefly then denatured and annealed by heating at 95°C for 5min, 70°C for 10min, then 37°C for 20min. After the final heating step, the reaction was left to cool to RT. The reaction was cleaned up by passing through a Qiagen PCR purification kit

and eluted in 50µl. Restriction digest of the plasmid pSuperior.puro was carried out using Bgl II and XhoI (Figure 4.4B) and the restricted plasmid purified by gel electrophoresis and gel extraction. A proportion of the prepared plasmid and the annealed oligonucleotides were quantified by gel electrophoresis and a ligation reaction set up, using approximately 60ng of vector and 30pMol insert. Ligation was carried out at 16°C O.N (2.4.5), then half the ligation mix was transformed into Dh5α cells. Briefly; the ligation mix and Dh5α cells were incubated together on ice for 30min, then heatshocked at 42°C for 35s, placed on ice for 2min, then luria broth (warmed to 37°C) was added and the cells incubated for 40mins. Cells were then plated out on Ampicillin containing agar plates at 20µl, 100µl and 200µl and incubated O.N at 37°C. Colonies grew at a ratio of 3:1 ligation: control plates and 16 colonies for each shRNA plasmid was selected and grown for plasmid extraction by miniprep. Purified plasmid DNA was restricted with BamHI and Bgl II. Successful ligation of the oligonucleotide would have destroyed the Bgl II site, therefore successful ligations would result in linearised plasmid 4391bp in size, while failed ligations produced two fragments 2937bp and 1417bp in size. The majority of selected colonies contained the insert (Figure 4.4D) and glycerol stocks were prepared accordingly. Sequencing of selected plasmids was carried out to confirm correct integration.

(A) Ago1 gene with siRNA locations marked



(B) Ago1 mRNA showing siRNA targets and location of taqman assay



(C) Alternative Ago1 transcripts and locations of catalytic domains

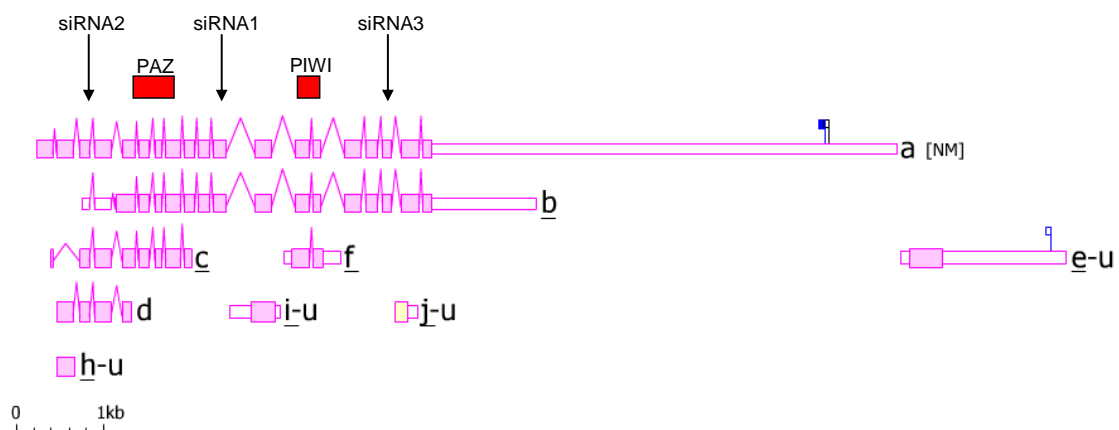


Figure 4.3 Ago1 gene and siRNA target locations. Schematic diagram of the Ago1 genomic (A) and mRNA (B) sequences showing the target regions of the three different siRNAs, and the position of the taqman probe used for assaying AGO1 mRNA levels by realtime PCR. (C) shows the different AGO1 transcripts [image adapted from NCBI Aceveiw March 2009] and the locations of the siRNA targets in relation to the PAZ and PIWI domains. Transcript A has been detected a total of 250 times in most tissue types, while other displayed transcripts have been detected less than 5 times, therefore transcript A represents approx. 90% of detectable Ago1 transcripts.

(A)

Target1

5' GATC CCCCGCGGGAAACAGTTCTACAATTCAAGAGAATTGTAGAACTGTTTCCGCGT TTTTC
3' GGGGCGCCCTTTGTCAAGATGTTAAAGTTCTCTTAACATCTTGACAAAGGGCGCAAAAAGAGCT

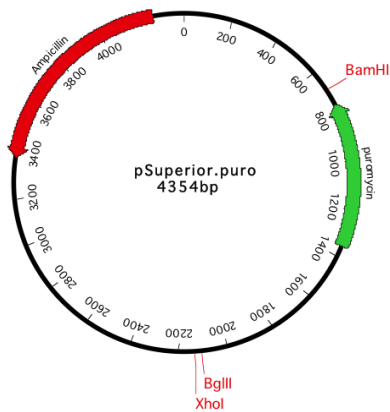
Target2

5' GATC CCCCGAGCATTTCAAACCGCAGATTCAAGAGAATCTGCGGTTTGAAATGCTGGT TTTTC
3' GGGGGTTCGTAAAGTTTGCGCTCTAAAGTTCTCTTAGACGCCAAACTTTACGACC AAAAAGAGCT

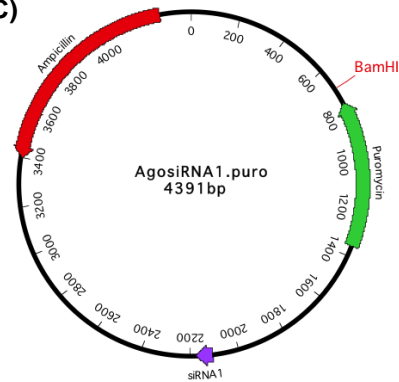
Target3

5' GATC CCCCGATTTCAGTTTCGACTTCTATTCAAGAGAATAGAAGTCGAACTCAAATGGT TTTTC
3' GGGGGTAACTCAAGCTGAAGATAAAGTTCTCTTATCTTCAGCTTGAGTTTACC AAAAAGAGCT

(B)



(C)



(D)

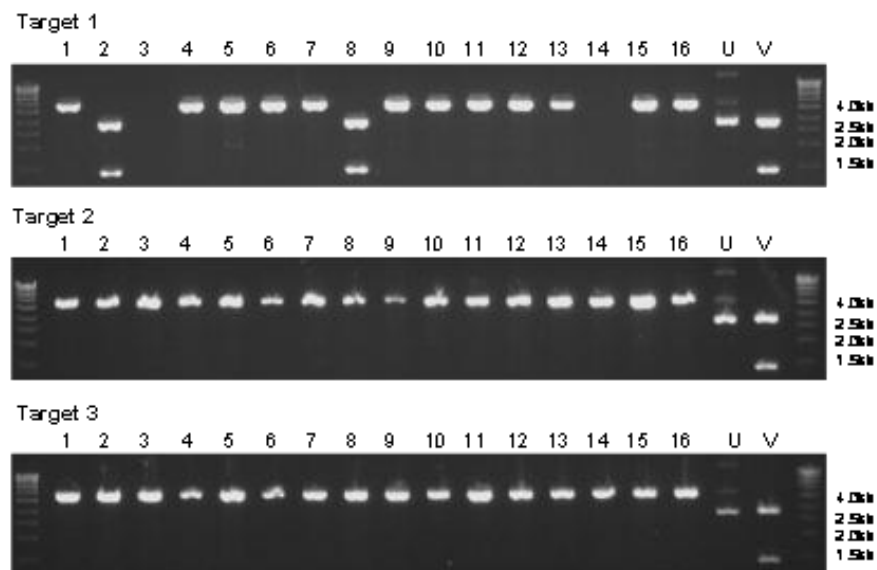


Figure 4.4 Engineering of siRNA construct into vector. (A) shows the aligned primer sequences after annealing. Shown in blue are the overhangs that act as sticky ends for Bgl II (GATC) and XhoI (AGCT). The Bgl II site is destroyed upon ligation into pSuperior while the XhoI site is maintained. Shown in red are the siRNA sequences that will anneal upon transcription to produce an shRNA. (B) and (C) show pSuperior before and after ligation of the siRNA fragment. (D) Screening by BamHI and Bgl II double digest of plasmid DNA extracted from bacterial colonies obtained after transformation of ligations for the three different shRNA oligonucleotides. Successful ligation produced a linearised plasmid of size 4391bp, while vector alone produced 2937bp and 1417bp fragments. U; Uncut, V; Vector.

4.2.4.3 Electroporation and selection of stable clones

The ESC line 2A3, containing a single copy of the ZfaAS construct, previously created by Dr Cristina Tufarelli, was used to create stable AGO1 kd cell lines. Electroporation using the Eppendorf multiporator was initially carried out at 300v, and with a time constant of 100 μ s using Eppendorf Hypoosmolar buffer as recommended. This produced very few colonies, therefore electroporation conditions were optimized by testing a variety of settings for electroporation of a GFP expressing construct. Successful integration of construct was detectable six days after electroporation by fluorescence microscopy (Figure 4.5A, and G). Electroporation in hypoosmolar buffer was tested at 450v, 100 μ s (Figure 4.5C), and 300v, 500 μ s (Figure 4.5D) and in PBS at 450v, 100 μ s (Figure 4.5D) and 450v, 500 μ s (Figure 4.5E). This demonstrated that 450v, 100 μ s in PBS was most efficient for electroporation, and an order of magnitude better than when hypoomolar buffer was used (Figure 4.5B to D). These settings were used for electroporation of the shRNA constructs into the 2A3 ESC line.

ShRNA containing cell lines were cultured and expanded in tissue culture. ESC material was collected from all cell lines for RNA and DNA analysis, and crosslinked pellets for ChIP for two cell lines were prepared. ESCs were differentiated into EBs over a period of seven to eight days, and more material prepared for RNA, DNA, and chromatin analysis. Preparation of DNA, RNA and ChIP pellets is described in 2.5.7, 2.5.8, and 2.8.1/2.8.2 respectfully. Differentiation of ESCs into EBs is described in 2.5.5

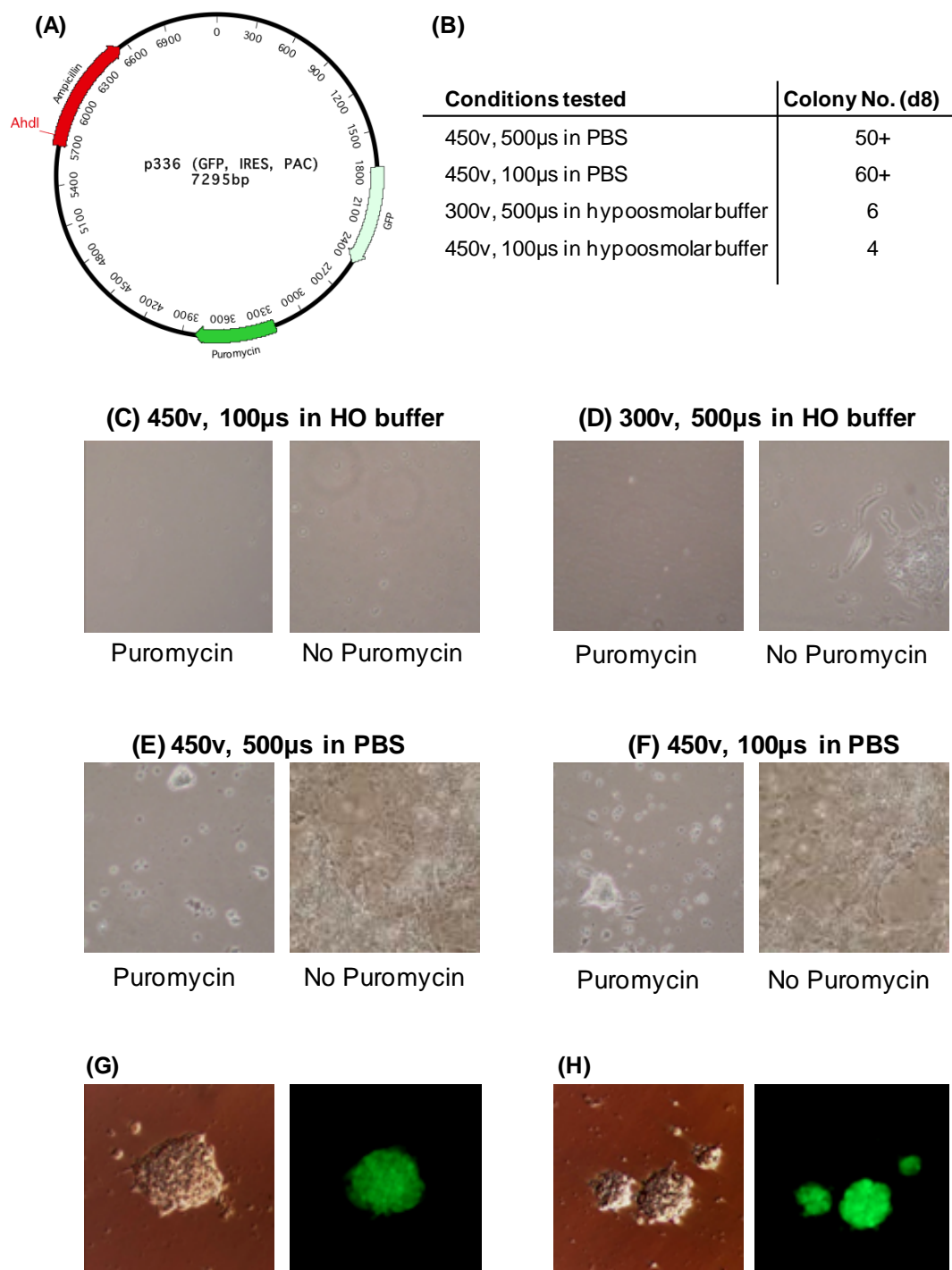


Figure 4.5 Optimisation of Electroporation conditions using Eppendorf multiporator. (A) Schematic diagram of the mammalian expression vector p336 (kind gift of Dr Denning) containing a GFP-IRES-Puro cassette. Indicated is the Adh1 site used for linearization prior to electroporation. (B) Table summarising the conditions tested and the survival on day 8 after electroporation. (C) to (F) Images of cell growth on day 8 after electroporation under the different conditions described in (B), cultured with (left panel) or without (right panel) puromycin selection. (G) and (H) phase contrast (left panel) or fluorescence (right panel) microscopy images of selected clones with good morphology expressing GFP following electroporation with p336.

4.2.2.4 Analysis of Ago1 shRNA stable clones

Expression of antisense RNA was tested for in all cell lines by strand specific RT-PCR as described in 2.7. The levels of the pluripotency factors *Oct4* and *Nanog* was also examined in ESCs and EBs as markers for successful differentiation of cells, this was carried out by multiplex PCR with primer pairs for *Oct4*, *Nanog* and *Aprt* followed by quantification (2.6.5). DNA methylation was assessed by a methylation sensitive multiplex PCR approach using primers that spanned a region of the alpha globin genes promoter containing methylation sensitive HpaII sites previously shown to become methylated upon differentiation in the presence of antisense RNA (2.6.4). *Ago1* mRNA levels were assessed by Taqman realtime PCR using a probe binding within exon 3 of the *Ago1* transcript (Figure 4.3) with *Gapdh* as the endogenous control (2.6.6.2). Alpha globin expression levels were initially assessed by standard PCR, then by SYBR green real-time PCR, as detailed in 2.6 and 2.6.6.3 respectively. Histone modification changes between ESCs and EBs were examined as described in 2.8.

4.3 Results

4.3.1 Detection of small RNAs in differentiating 2A3 cells

As described in chapter 3, 2A3 ES cells in the undifferentiated state express RNAs both sense and antisense to the alpha globin gene. To determine if the potential formation of dsRNA gives rise to siRNAs, total RNA from 2A3 cells either undifferentiated or from different differentiation time points (d1, d2, d2.5, d2.75, d3, d3.25, d3.5, d3.75, d4, d4.25, d4.5, d5, and d6) were analyzed for the presence of small RNAs. First, Northern blots were created using mESC and d8 EB RNA to optimize the conditions for the detection of endogenously expressed miRNA controls. Expression profiles by Northern blotting for the miRNAs used were published previously for J1 ESCs grown on a 0.1% gelatin layer, and differentiated for 4 days as a monolayer using Retinoic acid (Houbaviy *et al.* 2003). Published expression profiles are summarized in Figure 4.6A. miR301, miR92, miR16, and miR301 miRNAs were chosen as representatives of the four major patterns described in Houbaviy *et al.* 2003. Following electrophoresis and transfer to membranes of total RNA from undifferentiated E14Tg2a or d8 EBs (Figure 4.6B), oligonucleotide probes for miR301, miR92, miR16, and miR301 were hybridized to the membranes. Expression of miR301, miR92 and miR16 all matched previously observed expression profiles in the J1 ESCs, while expression of miR302 showed increased levels of expression in the EB sample (Figure 4.6C).

To identify the best probe to use for the detection of siRNAs at the human alpha globin gene promoter, the online siRNA search facility provided by the

Whitehead Institute, MIT (Yuan *et al.* 2004) was interrogated. The results showed that a ~200bp region immediately upstream of the TSS of the alpha globin gene contained over 60 sequences that matched the criteria for a functional siRNA, therefore a 222bp probe was chosen that covered this region (Figure 4.7A). Total RNA from a differentiation timecourse of the ESC 2A3 cell line was run on a polyacrylamide gel and transferred to a nylon membrane in preparation for siRNA screening (figure 4.7B and C). The timecourse covered 14 different points from undifferentiated ESCs to day6 EBs.

(A)

microRNA	expected expression profile
miR301	Low in ES, stonger in EB
miR92	low expression in ES and EB
miR16	low in ES, High in EB
miR302	low expression in both

(information from Houbaviy *et al.* 2003)

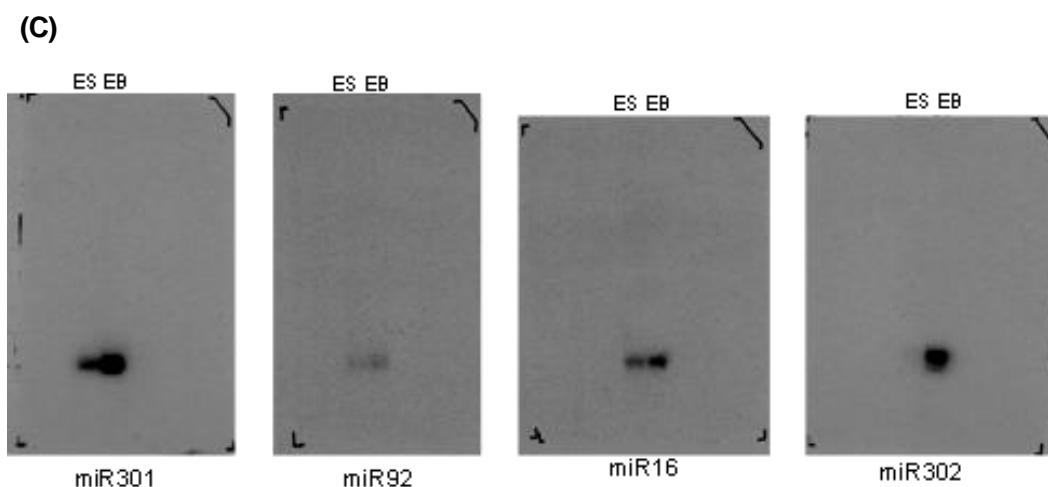
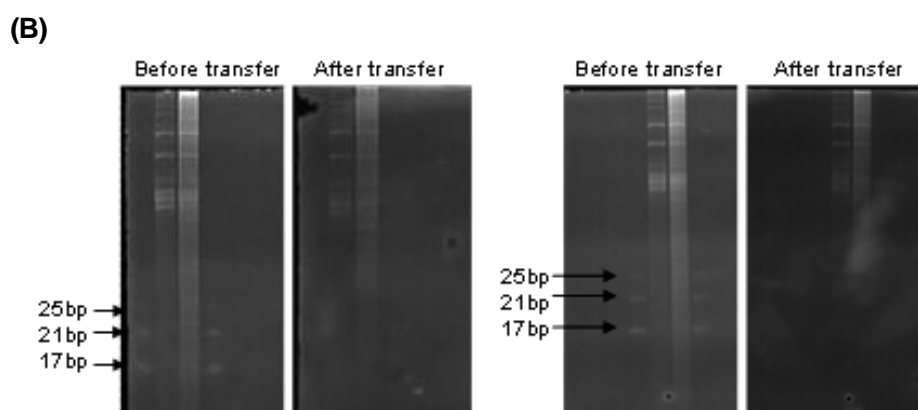


Figure 4.6 Testing of endogenous miRNA controls.

A) expression patterns of the four chosen miRNA controls as observed by Houbaviy *et al.* using retinoic acid differentiation of mESCs. (B) Electrophoresis of total RNA from ESC and EBs run on Polyacrylamide gels and post-stained with EtBr (left). Gels were re-stained with EtBr following transfer of small RNA species to nylon membrane by semi-dry electroblotting (right). (C) Autoradiograph of membrane after hybridisation of oligonucleotide probes for four miRNAs with different ESC and EB expression patterns.

No small RNA species were detected within the 222bp region (Figure 4.7D), but control miRNAs were detectable for each of the time points (4.7E), indicating that the electrophoresis and transfer of RNA had occurred correctly. The probe for miR16 revealed low level of expression in ESCs, with expression steadily increasing during differentiation, in line with previous observations. A probe covering a 994bp region containing approximately 500bp of the alpha globin promoter and the first two exons of the gene was used to expand the region screened for small RNAs (Figure 4.8A), but again no small RNAs were detectable. The blot was then hybridized using a probe for miR302 to ensure that subsequent rounds of hybridization and stripping had not damaged the membrane (Figure 4.8B). Small RNAs were readily detectable up until the day 6 time point, when levels appeared dramatically reduced. Finally a 1489bp probe covering the entire alpha globin gene (probe position is shown in Figure 4.7A) was used to probe for small RNAs, but this also failed to detect any small RNAs (Figure 4.8C). The probe used was shown to very effectively label a southern blot of alpha globin DNA that was hybridized in the same tube (Figure 4.8D).

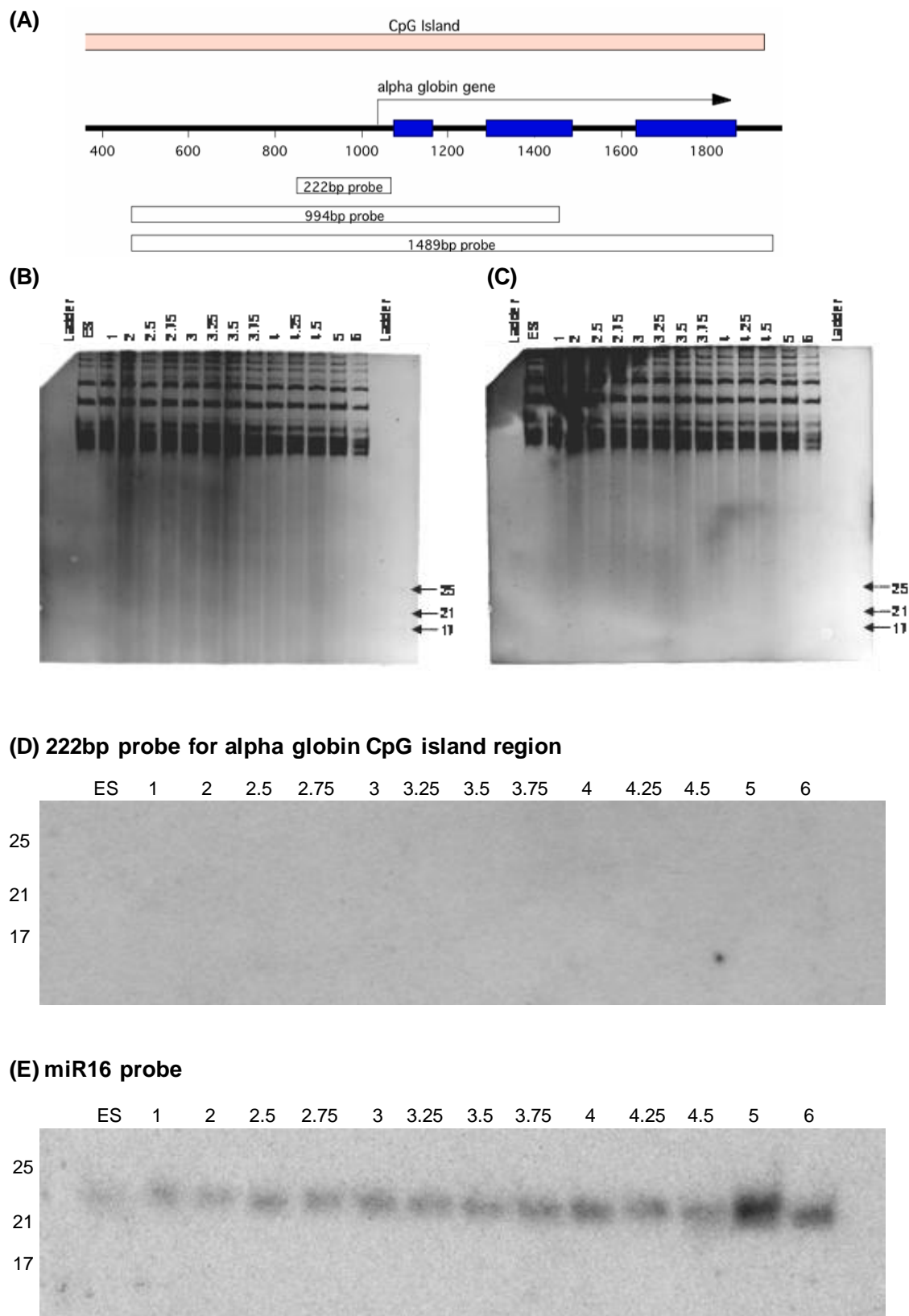


Figure 4.7 Screening for siRNAs on 2A3 timecourse (A) Schematic of the alpha globin gene showing the position of different probes. (B) Timecourse northern blot after electrophoresis and (C) after transfer. (D) screening for small RNAs using 222bp probe. (E) screening for miR16 expression.

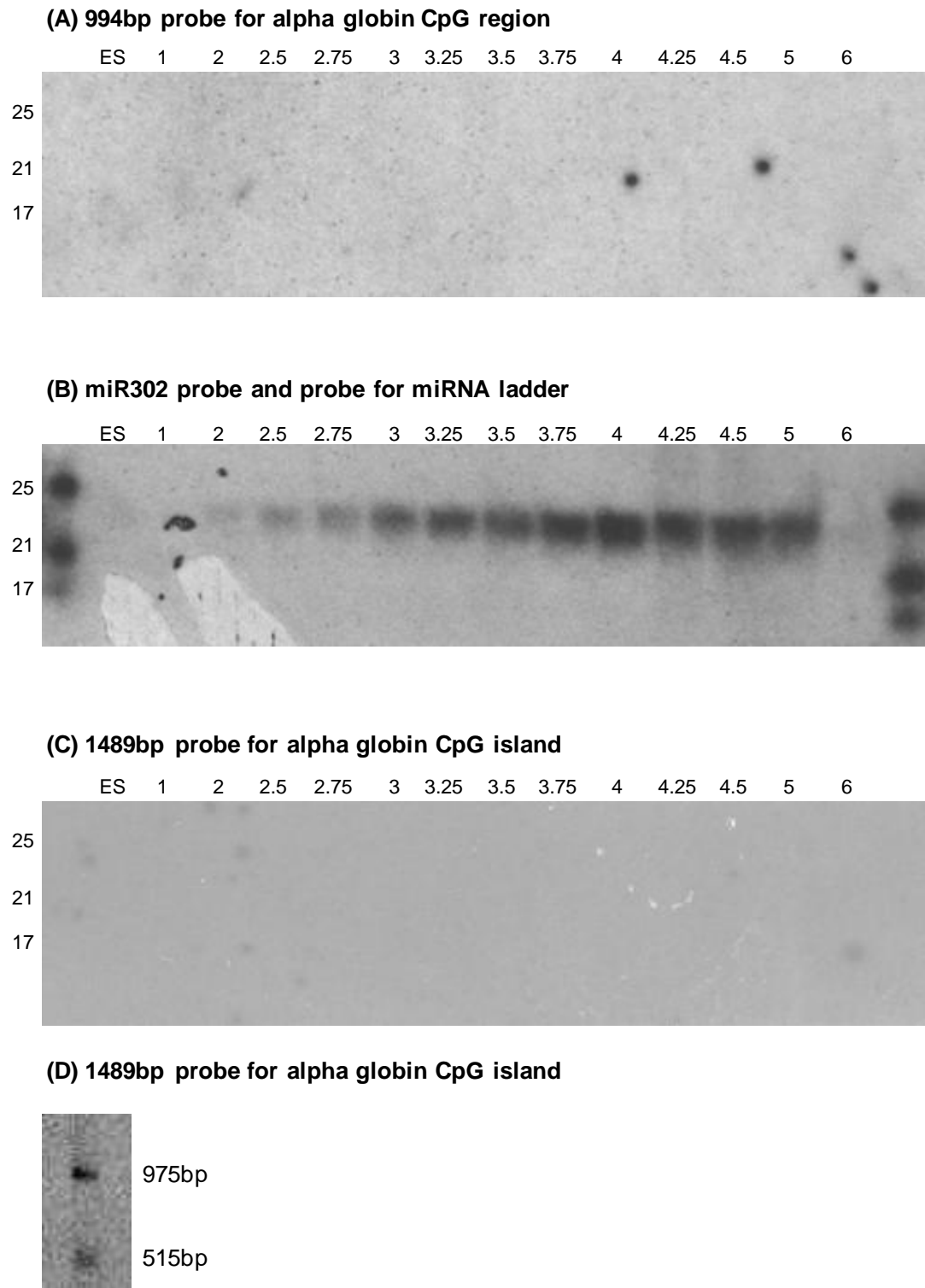
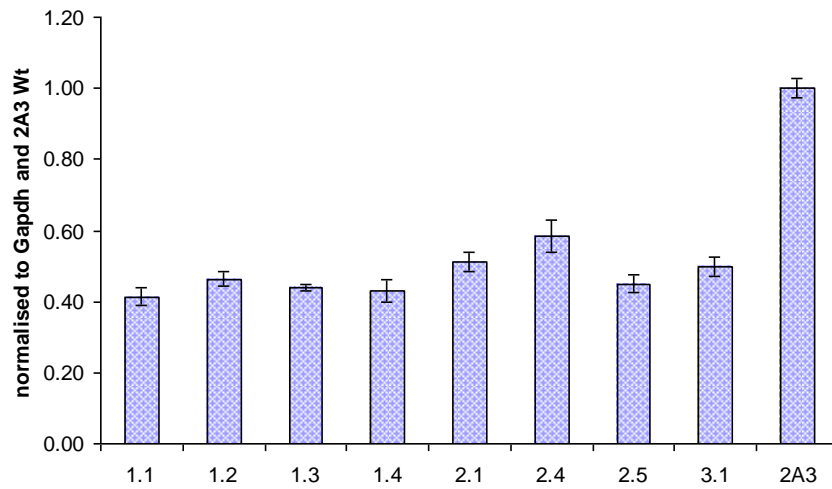


Figure 4.8 Screening for siRNAs on 2A3 timecourse. (A) screening for alpha globin small RNAs using a 994bp probe used to screen for small RNAs. (B) screening for expression of miR302 and labelling of miRNA ladder. (C) screening for small RNAs using 1489bp probe, and (D) showing the binding of the 1489bp probe to ESC DNA on membrane hybridized in same tube as C.

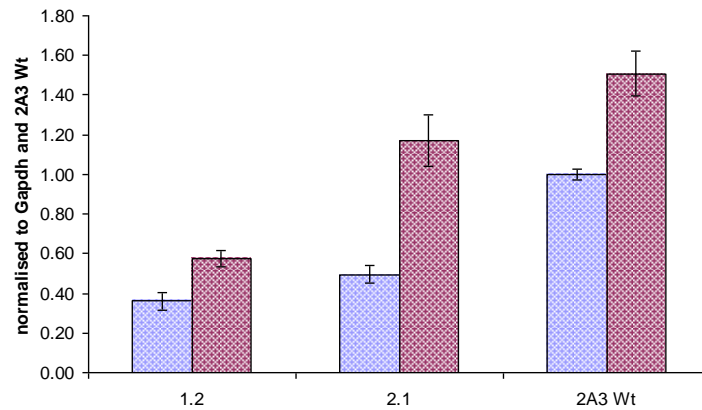
4.3.2 Analysis of Ago1 knockdown cell lines in the 2A3 background

To investigate whether Ago1 is involved in the ZF silencing event, 2A3 cell lines stably transfected with a construct expressing one of three different Ago1 shRNAs were obtained in separate electroporations. Individual clones were selected in puromycin and analyzed for the presence of the PAC gene by PCR. A total of eight clones were selected based upon their morphology and examined for the expression of Ago1 by Taqman realtime PCR. Expression was normalized to the endogenous control Gapdh, and to expression of Ago1 in the ESC line 2A3. The level of detectable Ago1 transcripts were, on average, half that of the 2A3 ES cell line with a range of between 40 to 60% knockdown observed (Figure 4.9A). There was no discernable difference between the three different shRNAs used, with all three associated with a similar reduction in detectable Ago1 transcripts. The cell lines 1.2 and 2.1 were grown and differentiated for analysis of histone modifications. The level of Ago1 transcripts in ESCs was equivalent to what had already been observed, with an increase in the detectable level of transcripts upon differentiation (Figure 4.9B). This increase was mirrored in the 2A3 cell line, suggesting that this represented increased Ago1 expression. Ago1 mRNA levels observed in the Ago1 Kd ESCs and EBs were normalized to levels in 2A3 ESCs and EBs to allow a more direct comparison (Figure 4.9C). This revealed that knockdown levels were remaining steady in the cell line 1.2 before and after differentiation, with a suggested reduction in knockdown in the cell line 2.1 in EBs.

(A) Ago1 expression in knockdown ESCs compared to 2A3



(B) Ago1 expression in cell lines 1.2 and 2.1 in ESCs and EBs normalised to 2A3 ES expression levels



(C) Ago1 expression in cell lines 1.2 and 2.1 in ESCs and EBs normalised to ES and EB expression levels

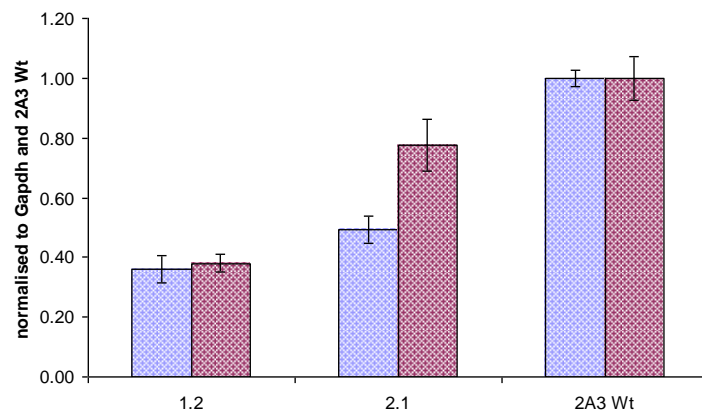
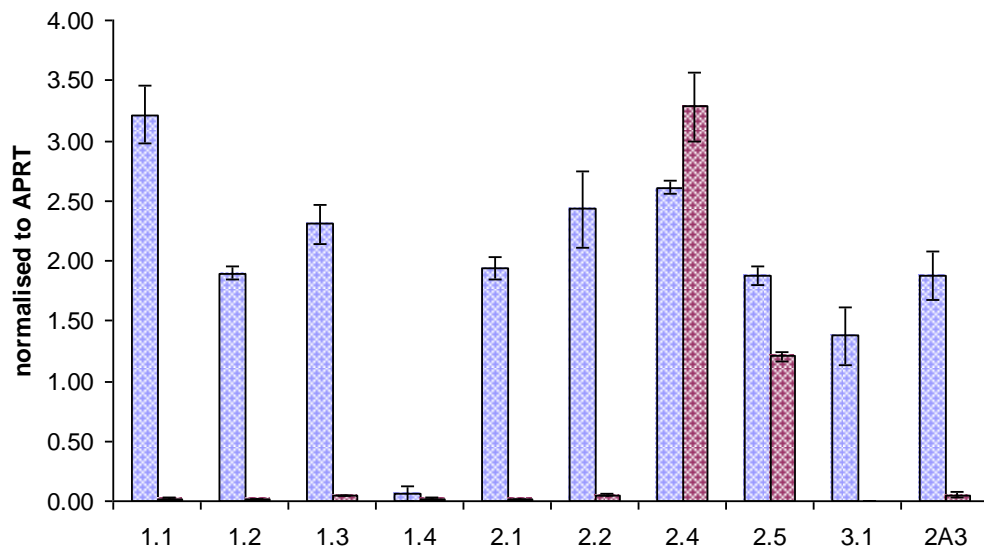


Figure 4.9 Ago1 expression in cell lines with siRNA constructs. (A) Expression of Ago1 in ESCs expressing either siRNA1 (1.1 to 1.4), 2 (2.1 to 2.5), or 3 (3.1). Ago1 levels were normalised to Gapdh, then to Ago1 expression levels in the parental 2A3 ESC cell line (B) Ago1 expression in both ESCs and EBs for the cell lines 1.2 and 2.1, normalised to both Gapdh and 2A3 ES Ago1 expression levels. (C) Shows the same data as (B) except ES values are normalised to 2A3 ES (in blue) and EB values are normalised to 2A3 EB (in red).

Eight cell lines were differentiated into EBs. To assess the extent of differentiation, expression levels of the pluripotency factors *Oct4* and *Nanog* were compared to the ubiquitously expressed gene *Aprt* in a multiplex PCR, carried out in triplicate. Following gel quantification, expression levels in ESCs and EBs were compared. This is shown in Figure 4.10. While expression of both pluripotency factors were greatly reduced upon differentiation for most cell lines, the cell line 2.4 continued to show high expression of *Nanog* suggesting incomplete differentiation, while the cell line 2.5 continued to show high expression of both *Oct4* and *Nanog* indicating a complete lack of differentiation for this cell line. The ZFaAS construct expresses antisense RNA through the alpha globin CpG island, and this is essential for the silencing effect, as discussed in the previous chapter. Antisense RNA was screened for by strand specific RT-PCR, and was detectable in all cell lines (Figure 4.11).

Alpha globin gene expression was initially examined in the Ago1 kd cell lines by standard 30 cycle PCR (Figure 4.12). This indicated that, in most cases, expression of alpha globin was becoming repressed in EBs in a comparable manner to the 2A3 cell line. Several cell lines, most notably 2.1 and 2.4, appeared to continue to express alpha globin at a higher level in EBs. This was further assessed by analysis of gel band intensity, normalizing alpha globin expression in ES and EBs to *Aprt* expression levels, then to 2A3 ES levels (Figure 4.13A), and also looking at expression in EBs as a proportion of ES expression (Figure 4.13B). This suggested that expression of alpha globin was marginally higher in EBs of Ago1 Kd cell lines, compared to 2A3, and again this was most noticeable in cell lines 2.1 and 2.4.

(A) Nanog expression



(B) Oct4 expression

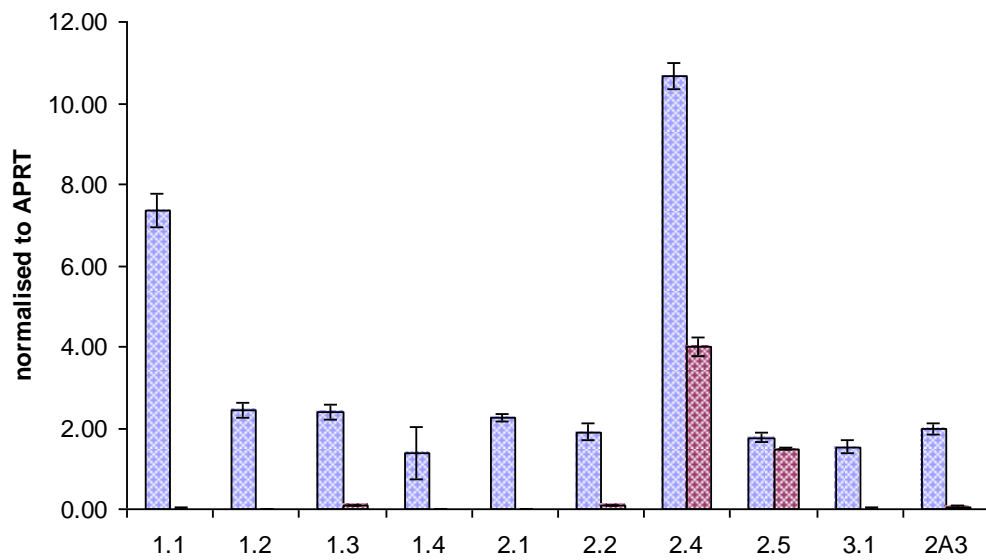


Figure 4.10 Expression of pluripotency factors. Expression of *Nanog* (A) and *Oct4* (B) in undifferentiated ESCs (blue) and EBs (red) of Ago1 shRNA expressing cells and 2A3 parental line normalised to levels of *Aprt*.

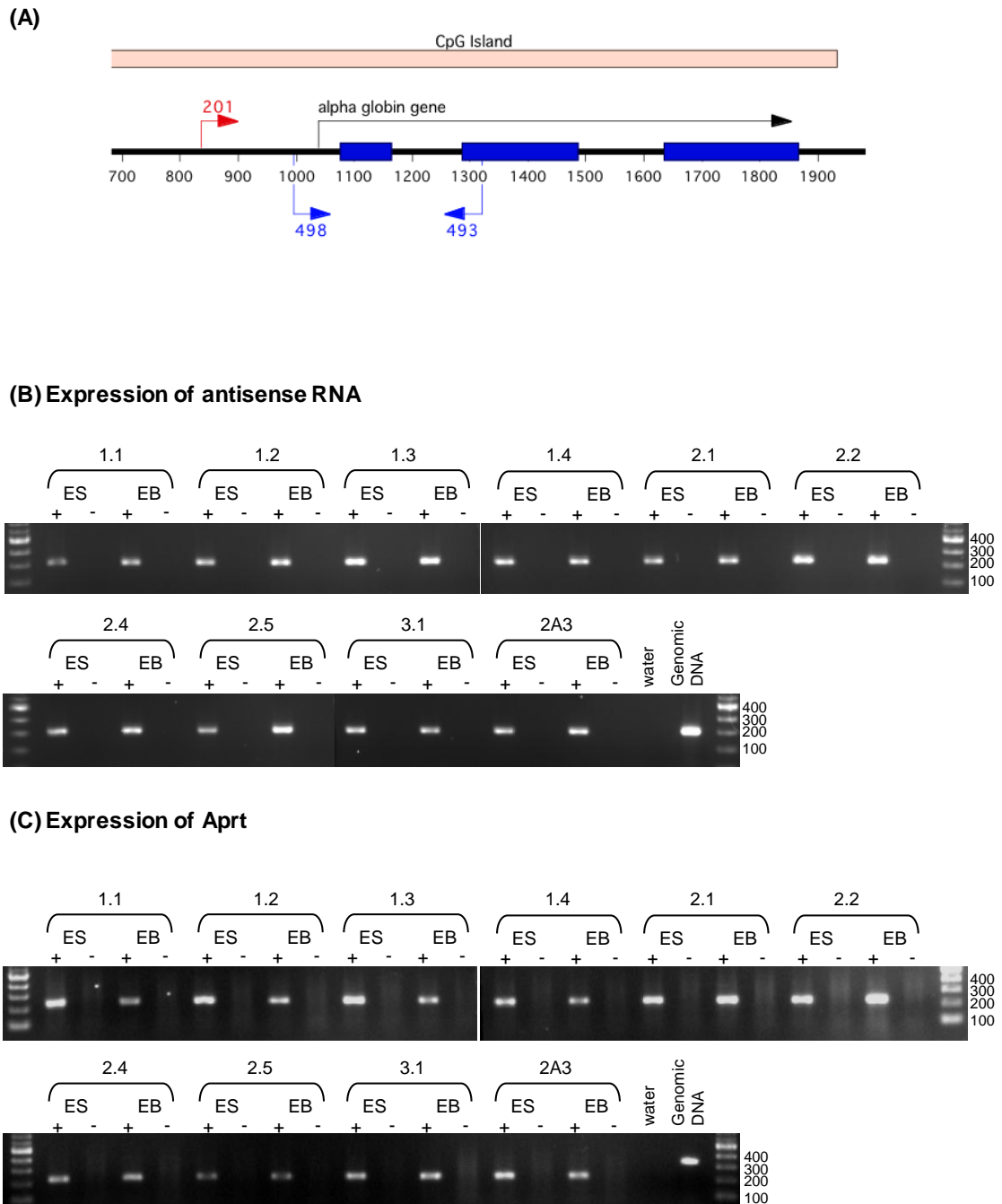


Figure 4.11 antisense RNA expression in the Ago1 cell lines (A) schematic representation of the alpha globin construct with primer locations marked. (B) strand specific PCR amplification of antisense RNA by 493 and 498 (306bp). (C) PCR amplification of control using mAPRT2 and mAPRT3 (225bp cDNA, 335bp genomic DNA). The primers 201 and mAPRT1 were used to prime a strand-specific RT reaction. Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA.

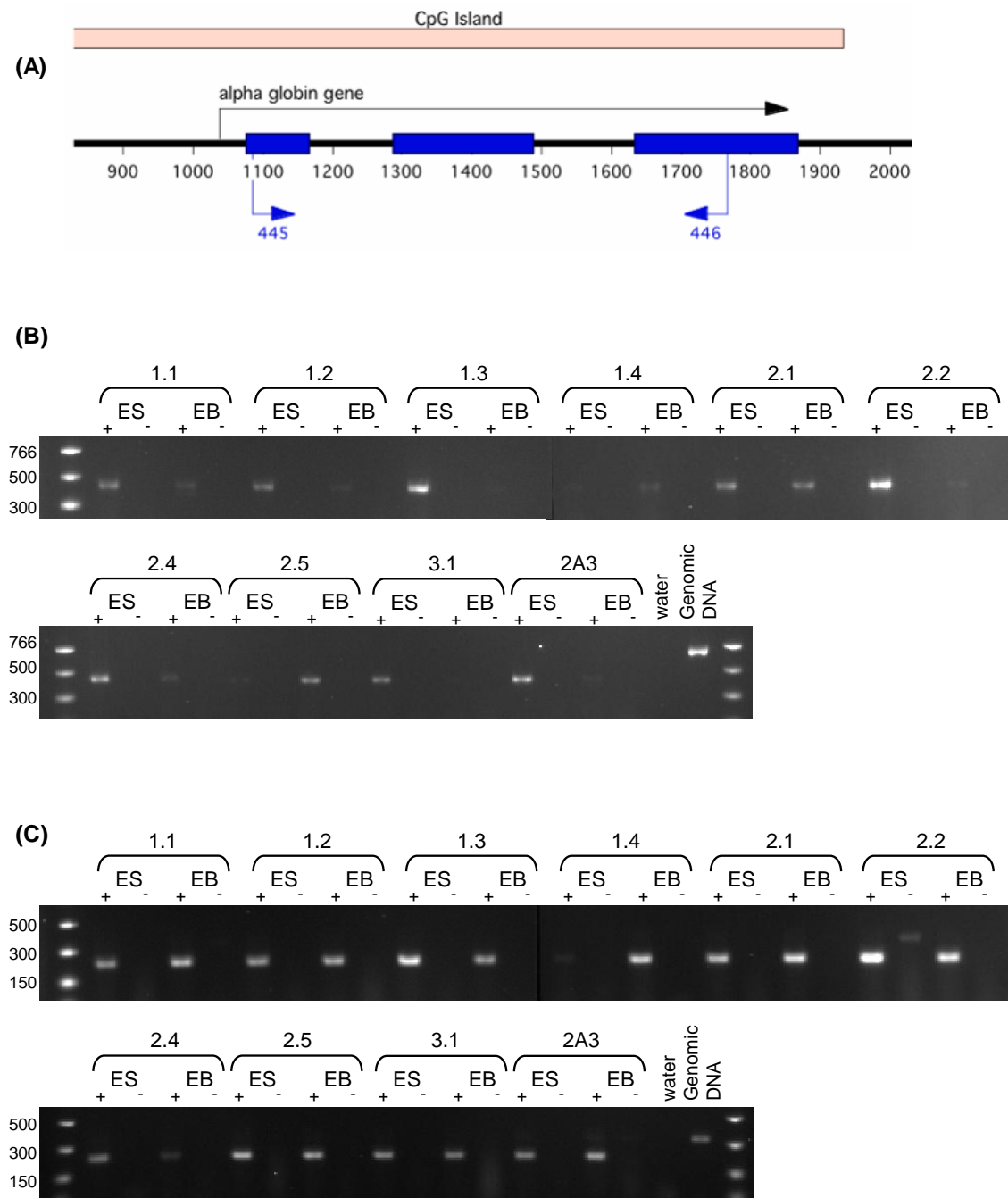
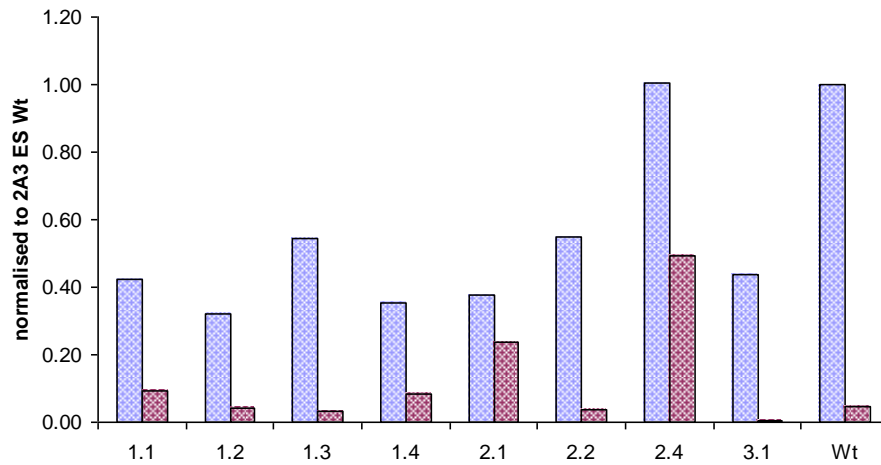


Figure 4.12 Expression of alpha globin in the Ago1 Kd cell lines. (A) Schematic representing the alpha globin construct and the position of the primers 445 and 446, amplifying a region of 426bp on cDNA. (B) expression of human alpha globin. (C) expression of mouse Aprt. RT-PCR reactions were primed using random hexamers. Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA. Standard 30 cycle PCR was carried out prior to gel electrophoresis.

(A) Alpha globin expression by gel quantification



(B) Expression of alpha globin in EBs as a proportion of ES alpha globin expression

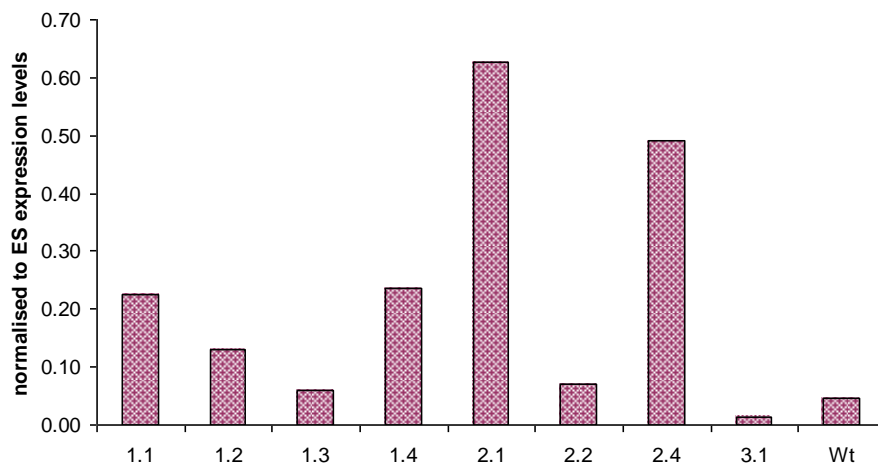


Figure 4.13 Expression of alpha globin in the Ago1 Kd cell lines. (A) Band intensity from Figure 4.6 was analysed using AIDA software and alpha globin levels were normalised to Aprt levels, then to 2A3 wt ESC levels. (B) Alpha globin levels were normalised to Aprt levels then EB expression levels were normalised to ESC expression levels to show the levels of expression in EBs compared to ESCs.

Expression of alpha globin in the cell lines 1.2 and 2.1, and in the 2A3 cell line was then examined by SYBR green realtime PCR for a more accurate assessment of gene expression (Figure 4.14), as a standard PCR approach can be inaccurate. This showed that any de-repression of alpha globin that might be occurring in EBs was far more modest than the standard 30 cycle PCR had suggested, though alpha globin expression was elevated to a minor degree in both cell lines compared to 2A3.

Real-time PCR examines the amount of PCR product during the exponential phase of PCR, when the amount of material is increasing in a linear manner in the log scale allowing very accurate determination of amount of product. With standard PCR, products have generally reached the plateau phase of amplification, whereby nucleotides and Taq have become exhausted in the reaction and little amplification is occurring. Under such conditions, the endogenous control (ENDO) reaction, which generally has starting material in excess of the gene of interest (GOI) reaches the plateau phase of the PCR reaction first and ceases to amplify in an exponential manner whereas the PCR reaction with the GOI, generally containing less starting material, can continue to amplify until it too has utilized all available raw materials in the reaction. Unless the number of PCR cycles and the amount of material for both the GOI and the ENDO are well balanced and primer pairs have roughly equal efficiency, then the true difference between the two samples may be obscured as the GOI effectively 'catches up' with the ENDO. Under such circumstances, real-time PCR offers far greater accuracy in determining gene expression levels.

DNA methylation of the alpha globin gene contained in the ZFaAS construct occurs upon differentiation. DNA methylation was examined in the Ago Kd cell lines using a methylation sensitive multiplex PCR approach, whereby a region of the alpha globin CpG island, containing 3 HpaII sites, and a region of the Luc7L gene containing no HpaII sites, were amplified in the same PCR reaction (Figure 4.15A). Multiplex PCR was carried out in triplicate, with gel analysis used to quantify the intensity of the band corresponding to PCR across the alpha globin CpG island, normalized to the control band. DNA methylation levels observed in Ago1 Kd cell line EBs was normalized to DNA methylation levels observed in 2A3 EBs. This is shown in Figure 4.15B. DNA methylation in the Ago1 Kd cell lines appears to be very close to the levels observed in the 2A3 cell line, with the exception of the cell line 2.1 where DNA methylation levels were observed to be approximately 30% of 2A3 levels.

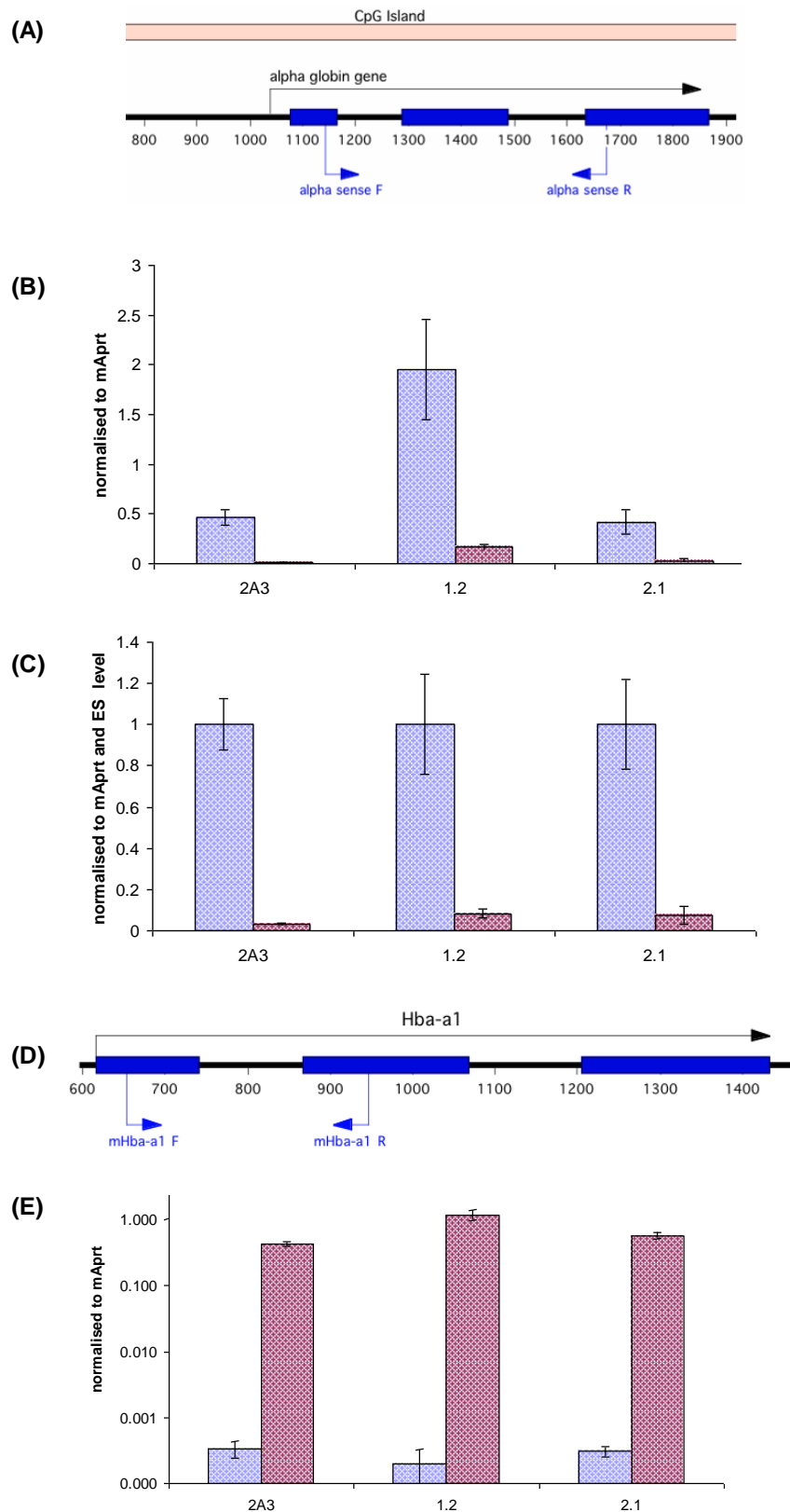


Figure 4.14 Expression of alpha globin assayed by realtime PCR. (A) Schematic representing the alpha globin construct and the position of the primers Alpha sense F and Alpha sense R used for realtime PCR. (B) Expression of alpha globin in ESCs and EBs for the 2A3 cell line, and the Ago Kd cell lines 1.2 and 2.1. (C) Expression of alpha globin in EBs normalised to alpha globin expression in ESCs for each cell line. (D) Schematic diagram showing the location of mouse alpha globin primers (E) expression of mouse alpha globin shown with a logarithmic scale

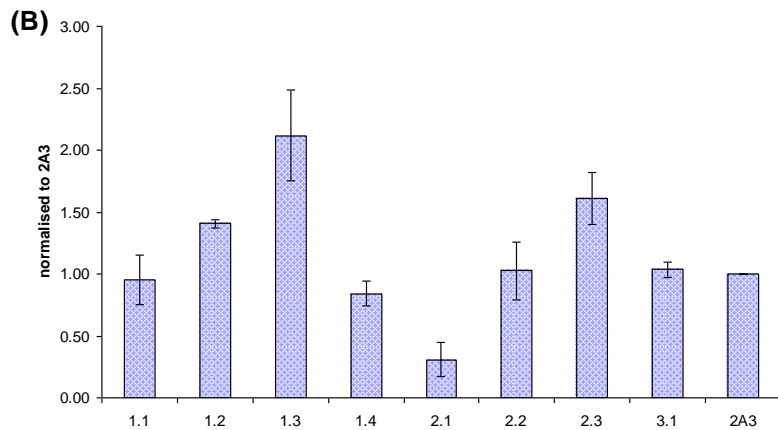
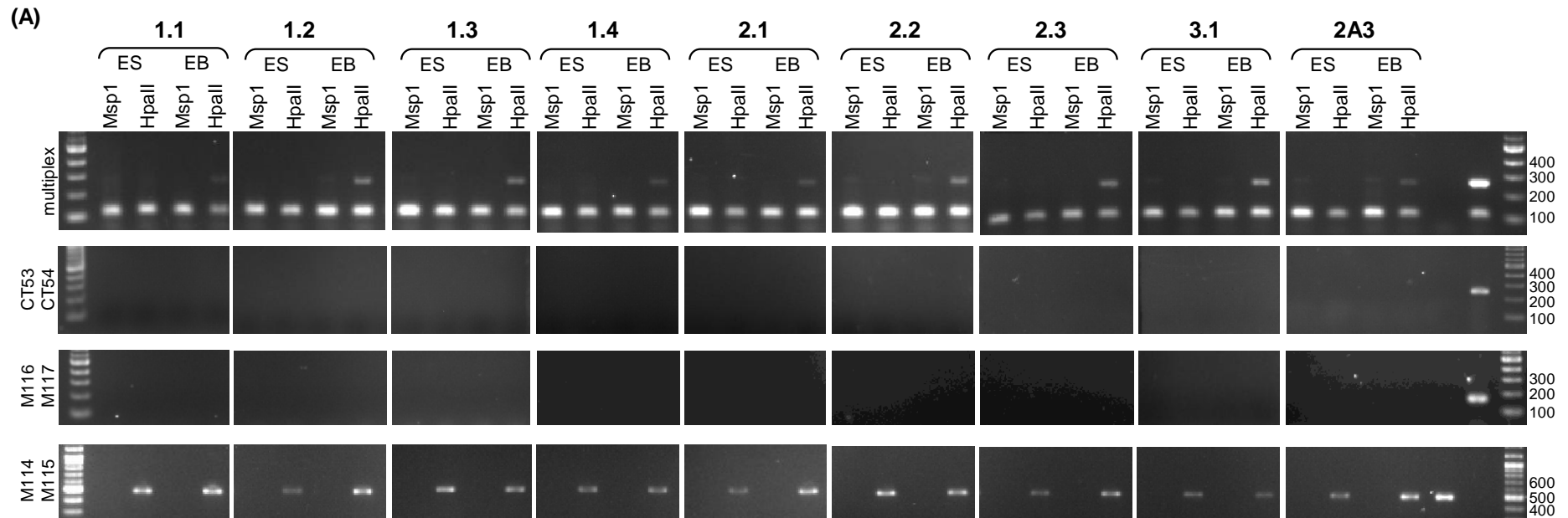


Figure 4.15 DNA methylation of the alpha globin CpG island in Ago1 Kd cell lines. (A) multiplex PCR of alpha globin CpG island combining a primer pair that amplifies a region with no HpaII cut sites (144bp), and a primer pair that amplify a region with 3 HpaII sites (295bp). CT53 CT54 amplify across 3 restriction sites (248bp), M116 M117 amplify a region within the 5' end of the APRT gene that contains 4 restriction sites and is devoid of DNA methylation (164bp), and M114 M115 primers amplify within the promoter of AIR, which is stably methylated on the maternal copy in ESCs and EBs (489bp). (B) The multiplex PCR was carried out in triplicate, band intensity analyzed by AIDA software, with DNA methylation being normalized to the control band and then normalized to 2A3 methylation levels.

Histone modifications in the cell lines 1.2 and 2.1 were examined by chromatin immunoprecipitation and SYBR green realtime PCR quantification. Enrichments for different histone modifications at the alpha globin gene are shown in Figures 4.16 and 4.17 for the cell lines 1.2 and 2.1 respectively.

For the cell line 1.2, upon differentiation H3K9me3 levels more than doubled across the alpha globin gene at all four regions examined, while the levels of H3K4me3 were reduced to less than a third of ES cell levels. H3K27me3 levels in ESCs appeared elevated compared to the endogenous control *ActB*, and continued to show enrichment in EBs. H3K27me3 data was hard to interpret due to the level of variation between realtime PCR technical replicates, though H3K27me3 did appear elevated compared to *Actb*, with a trend for minor increases in the levels of H3K27me3 at all four locations upon differentiation. Acetylation levels were low in ESCs and become further reduced upon differentiation.

For the cell line 2.1, H3K9me3 levels increased upon differentiation into EBs, though this increase was not as severe as observed in the 1.2 cell line. At the position alpha 2 within the promoter region of the alpha globin gene the increase was barely perceptible, in contrast to both the 2A3 cell line (Chapter 3 Figure 3.8) and the cell line 1.2, which showed methylation more than doubling in this region. In ESCs, H3K27me3 levels were low but, upon differentiation, enrichment levels more than doubled. Both H3K4me3 and H3 Ac levels were low in ESCs and EBs, though levels show a modest increase at position alpha1 within the promoter region and at position alpha 4 within the body of the gene.

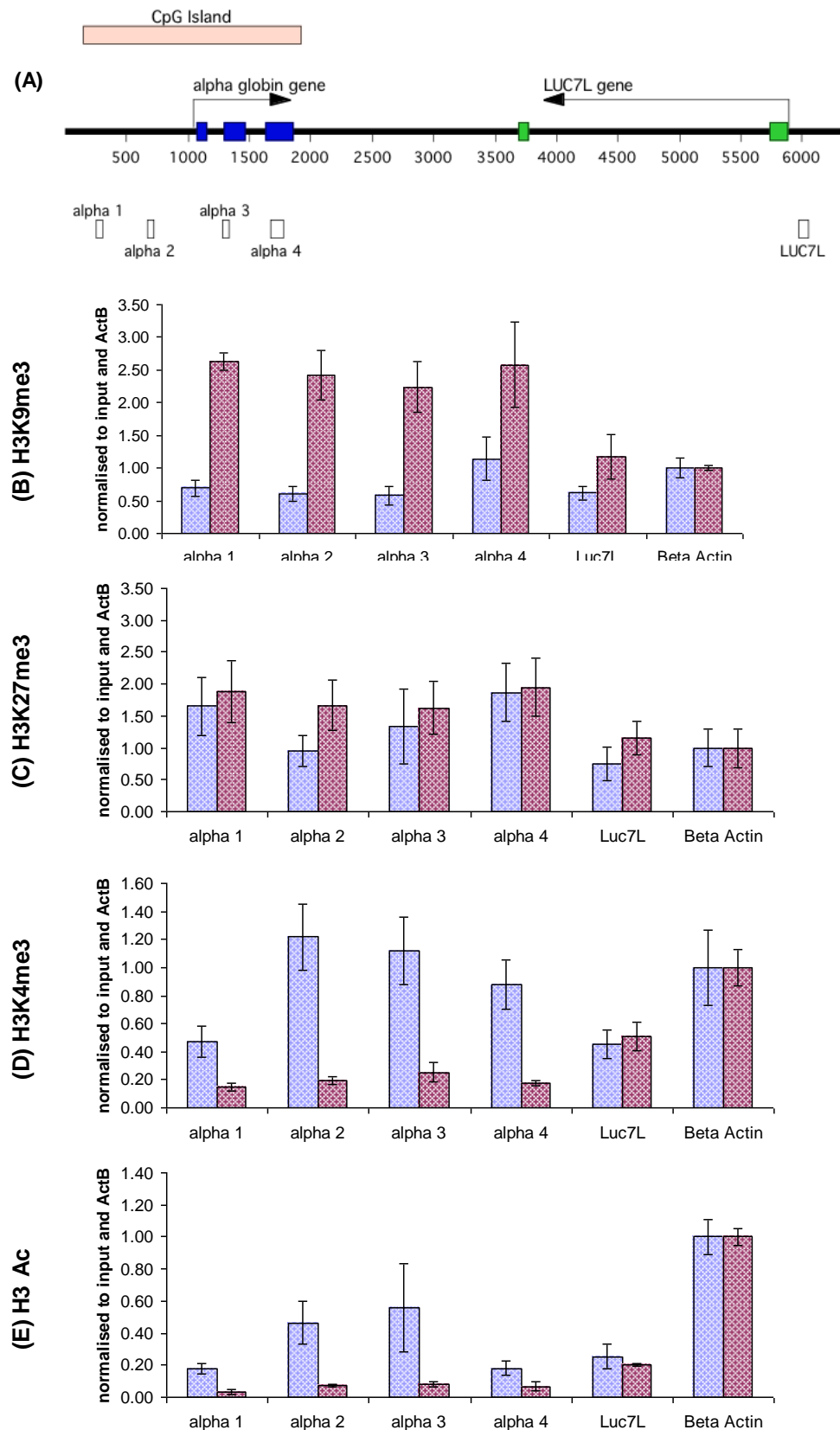


Figure 4.16 Histone modifications to the *alpha* globin construct in the cell line 1.2. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *HBA2* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

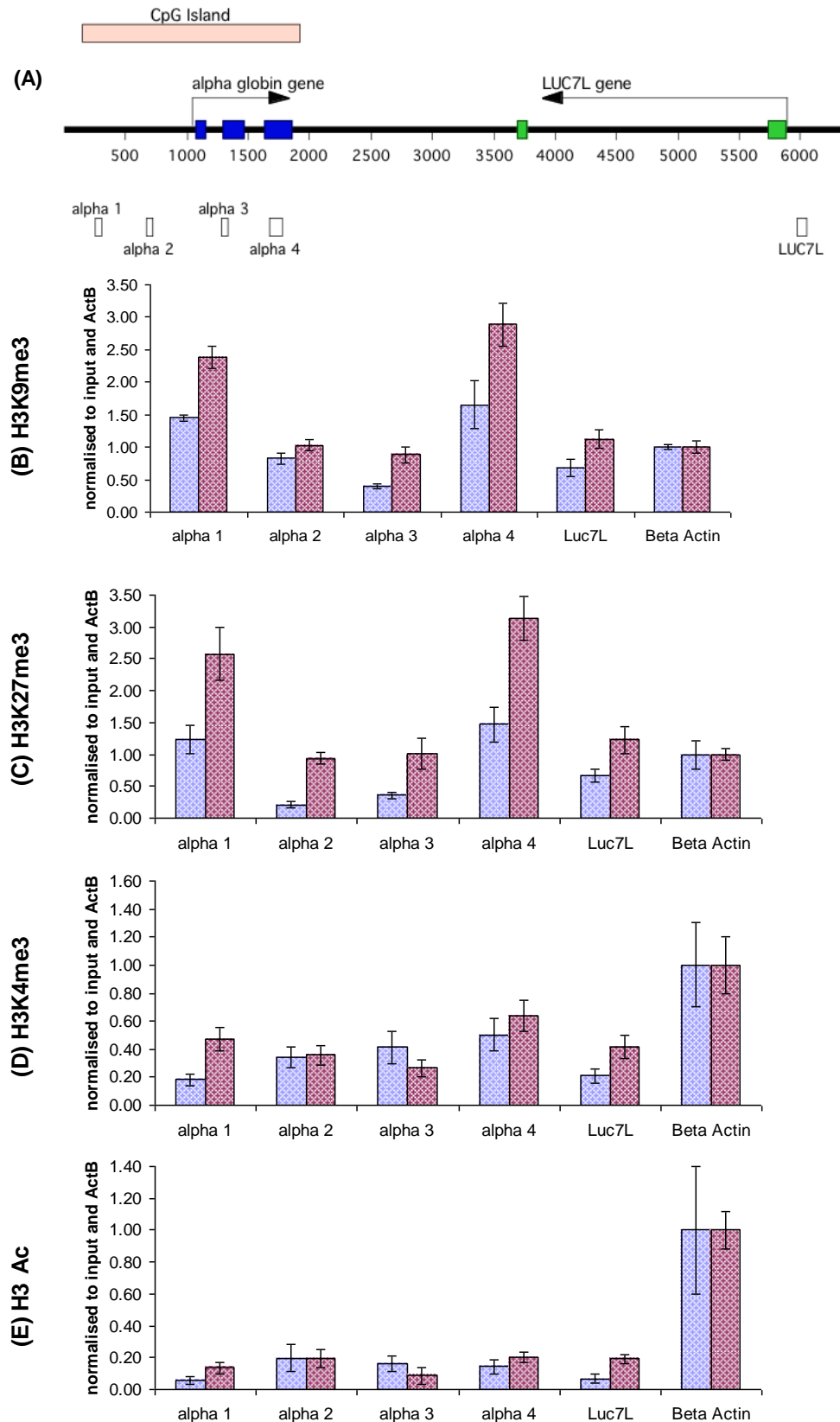


Figure 4.17 Histone modifications to the alpha globin construct in the cell line 2.1. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *HBA2* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

4.4 Discussion

siRNAs corresponding to the alpha globin CpG island were not detectable, despite the readily observable presence of other small RNA species. It is possible that siRNAs, if present, may be at too low a level to detect. The required threshold in siRNA levels to trigger silencing may be quite low, if the system is acting *in cis* in a similar manner to that observed in *S.pombe* where repression is linked to transcription and freshly produced siRNAs are immediately incorporated into silencing complexes in a local manner. Due to the rapid turnover in siRNA usage, there may not be many intact siRNA duplexes left to detect. There are also conserved mechanisms that are responsible for destruction of siRNAs; the exoribonuclease Eri1 is an evolutionarily conserved protein that reduces the levels of siRNAs corresponding to heterochromatic repeats in *S.pombe* (Buhler *et al.* 2006), and has been shown to have conserved function in *C.elegans*, while the human ortholog has been demonstrated to degrade siRNAs *in vitro* (Kennedy *et al.* 2004). In *S.pombe*, Eri1 is thought to destroy siRNAs that have escaped incorporation into RITS complexes, thereby ensure silencing remains *in cis* (Buhler *et al.* 2006). Such a mechanism might well reduce the potentially already low levels of siRNAs to a point where they are undetectable by northern blotting, even with the adaptations to the Northern blot protocol for better small RNA retention and transfer of small RNAs.

Ago1 mRNA levels were reduced by just over 50% by shRNA mediated PTGS, with no apparent differences in efficiency of the three different siRNAs used. Eight cell lines were differentiated into EBs, and showed DNA methylation of the

alpha globin CpG island in a manner similar to the 2A3 cell line expressing normal levels of Ago1. Alpha globin gene expression, assayed by standard PCR, showed four cell lines with slightly elevated levels of alpha globin mRNA compared to the 2A3 cell line, however when alpha globin mRNA levels were assessed by realtime PCR for the cell lines 1.2 and 2.1, they showed no real discernable increase in expression compared to 2A3. For the cell line 1.2, analysis of histone modifications showed a clear pattern of repression, with high levels of H3K9me3 enrichment upon differentiation, and loss of the active marks H3K4me3 and H3 Ac. The cell line 2.1 showed a distinct loss in DNA methylation levels and a lower level of enrichment for H3K9me3, though this was clearly sufficient for gene repression as the level of detectable transcripts upon differentiation were comparable to 2A3 levels. Taken together, halving of the Ago1 mRNA levels has had no clear phenotypic effect on repression at the alpha globin gene. While the cell line 2.1 does show some loss of repression, this is not mirrored in other cell lines incorporating the same shRNA construct and showing very similar levels of AGO1 Kd.

The inability to detect siRNAs, and the lack of effect that Ago1 Kd has on silencing may be because there are no siRNAs to detect, and that Ago1 is not involved in the silencing pathway responsible for repression of the alpha globin gene. It is becoming clearer that long single stranded ncRNA mechanisms are capable of repression, with one such example recently characterized silencing the p15 gene, where expression of a 34.8kb antisense transcript, normally expressed at low levels but unregulated in cancer and with high antisense expression associated with DNA methylation and H3K9 methylation of the p15

promoter leading to gene repression in an RNAi independent manner (Yu *et al.* 2008).

Yu *et al.* demonstrated that the AS RNA transcript was responsible for initiating the silencing effect, but was not required for maintenance. Gene silencing was strongly *in cis*, with only a weak *in trans* effect, when a construct bearing a copy of the p15 gene and expressing AS RNA was strongly repressed, but with only weak repression of the endogenous p15 occurring. H3K9me2 enrichment was examined and shown to occur across a 6Kb region upon induction of silencing, while H3K4me2 levels were reduced but remained detectable. When examined in ESCs, the introduced construct became repressed in a similar manner, but DNA methylation was seen to occur only upon differentiation into EBs (Yu *et al.* 2008).

Finally, Yu *et al.* examined gene repression in Dicer knockout mESCs. Dicer is essential for siRNA production and KO would prevent any potential dsRNA from being processed into siRNAs and used to guide silencing complexes (Kanellopoulou *et al.* 2005, Murchison *et al.* 2005). P15 gene repression continued to occur in mESCs, suggesting that Dicer, and by extrapolation the siRNA pathway, was not involved in p15 silencing. This study suggests that gene repression is mediated by a long AS transcript, in a manner reminiscent of silencing that occurs during imprinting and X inactivation, and further suggests the presence of an as yet uncharacterized silencing pathway utilizing long AS RNAs.

Knockout of Dicer would theoretically allow a means to differentiate between an siRNA or long AS RNA mediated silencing mechanism in the case of repression of the alpha globin gene, however such an approach is not suitable as alpha globin gene repression occurs upon differentiation, and knockout of Dicer has been shown to prevent ESC differentiation, indeed; proper EBs do not develop at all but rather resemble ESC aggregates still expressing high levels of pluripotency factors (Kanellopoulou *et al.* 2005). Differentiation has also been attempted with retinoic acid, but, while some differentiation did occur, pluripotency factors were still expressed at higher than normal levels (Kanellopoulou *et al.* 2009). Additionally; ESCs KO for Dicer exhibit reduced proliferation, abnormal cell cycles, and an unstable karyotype in female cells (Kanellopoulou *et al.* 2009, Kanellopoulou *et al.* 2005, Murchison *et al.* 2005), which has cast doubts upon the veracity of the results obtained with their use.

An alternative approach to the screening for siRNAs would be to screen for large RNAs, both single and double stranded, originating from the alpha globin/*LUC7L* region. Prepared RNA could be alternatively digested with Rnase A, which selectively digests ssRNA, and Rnase V1, which digests dsRNA without sequence bias. This could reveal the presence of long dsRNA transcripts and therefore indicate whether it is likely that siRNA production is occurring. However, similar to caveats mentioned for siRNA screening, it is possible that dsRNA formation and rapid processing to siRNAs may make long dsRNAs equally elusive to detect.

In classical RNAi mediated PTGS, siRNAs are included in the growth media, resulting in high levels of siRNAs within the cell, corresponding to high levels of knockdown. This approach was incompatible with the model system used to examine the silencing effect as there would be incomplete penetration of siRNAs into developing EBs resulting in an unacceptably high variation in the level of knockdown within the cell population. Stable incorporation of shRNA expressing vectors ensures uniform levels of knockdown within the population but can be limiting in the extent of knockdown. Knockdown of Ago1 reduced mRNA levels to, on average, just under 50% of wt mRNA levels, with no observable difference in the efficiency of the three different siRNAs used. To enhance the level of knockdown two or more of the siRNAs could be combined together and expressed either from the same construct (Hinton & Doran 2008, Song *et al.* 2008) or combined together on an extended hairpin (e-shRNA), which is readily processed by Dicer to release the individual siRNAs (Liu *et al.* 2007).

Combining several siRNAs together as an e-shRNA has been shown to produce high levels of knockdown, however there was some indication of bias for one siRNA sequence over the other, suggesting that the siRNA closest to the hairpin loop was less efficiently processed (Liu *et al.* 2007). Creating a construct containing two or more shRNA sequences has been shown to result in very high levels of PTGS, with a synergistic effect observed with knockdown twice that of the individual shRNAs (Song *et al.* 2008).

This chapter has focused upon the molecular mechanism that may play a role in repression of the alpha globin gene, initially observed in the ZF patient. The following chapters investigate whether repression is specific to the *HBA2* gene, or is more widely applicable and can effect other tissue specific or ubiquitous genes.

Chapter 5 Role of antisense RNA in silencing of the tissue specific gene MYOD1

5.1 Introduction

The CpG island of the tissue specific alpha globin gene has been shown to be susceptible to antisense RNA mediated DNA methylation and repressive histone modifications resulting in gene silencing. To test if this form of silencing involving both DNA methylation and H3K9me3 was unique to the alpha globin gene, or whether it was applicable to other tissue specific genes, Myogenic Differentiation factor 1 (*MYOD*) was cloned into a construct that contained an antisense-orientated RNA driven by the *LUC7L* promoter. The status of the *MYOD* gene was then examined using the differentiating mouse ES cell model system.

In mammalian ESCs, *MYOD* is bivalently marked with H3K27me3 and H3K4me3, and is also associated with the histone variant H2A.Z and PRC2 components, but not PRC1 (Ku *et al.* 2008, Creyghton *et al.* 2008). The bivalent domain resolves to a stable H3K27me3 repressive domain in mouse embryonic fibroblasts (MEFs), while both marks are lost in neural progenitor cells (NPCs) (Mikkelsen *et al.* 2007). Extensive analysis of the *MYOD* CpG island in adult tissues shows no DNA methylation, demonstrating that this repressive mark is not normally associated with this gene, even in tissues where it is not expressed (Jones *et al.* 1990). *MYOD* contains a CpG island approximately 3Kb in length with a GC content of 63.7%. The gene has one characterized transcript consisting of three exons, all of which are encompassed by the CpG island.

The tissue specific gene *MYOD* encodes for a transcription factor that is a key regulator of myogenesis in vertebrates. It belongs to a group of basic Helix-Loop-Helix (bHLH) containing myogenic regulatory factors that are critical for differentiation of skeletal muscle, and together coordinate a multitude of activators and repressors to control gene expression during embryogenesis (Berkes & Tapscott 2005). *MYOD* can form complexes with both HATs and HDACs, and binds within the regulatory regions of target genes via its bHLH domain and causes either gene activation or repression through changes in chromatin structure (Tapscott 2005). *MYOD* has been shown to affect the expression of over 300 genes, and directly binds to the regulatory regions of over 70 of these (Bergstrom *et al.* 2002).

In mice, expression of *Myod1* occurs in early somites 10.5 days post conception and, along with expression of *Myf5*, another myogenic regulatory factor. Expression of these factors induce the multipotent somites to commit to the myogenic lineage, which gives rise to skeletal muscle (Berkes & Tapscott 2005, Buckingham 1992).

5.2 Methods

All primers used in the investigation of the *MYOD* gene are listed in Table 5.1.

Standard primers used are listed in Materials and Methods.

In general, expression of antisense RNA and the presence of DNA methylation was examined twice for each cell line. Gene expression analysis and Chromatin IP experiments were only carried out once for each cell line, unless otherwise stated.

5.2.1 Construction of the pZERO MYOD + LUC7L plasmid

Growth of bacterial cultures and purification of PAC DNA, the engineering of the construct, long range PCR, cloning, and transformation into ultracompetent cells is described Chapter 2 Materials and Methods section 2.

The pZERO MYOD + LUC7L plasmid was engineered as shown in Figure 5.1. DNA was extracted from *E.coli* bearing the PAC RPC15-1082L12, and used as a template for long accurate PCR of a 7523bp region containing the *MYOD1* gene. The PCR fragment was restricted with Afl II, blunt ended using klenow then subsequently digested using EcoR1 (Figure 5.1A). The plasmid pZero ha +80AS was restricted with Pac1, blunt ended with klenow, then sequentially digested with AatII to destroy an unwanted fragment, and with EcoRI, then finally treated with SAP (Figure 5.1B). The Qiagen PCR kit was used to clean up in between reactions, with DNA being eluted in 30µl. Concentration of PCR product and linearised plasmid were assessed by gel electrophoresis and a ligation reaction set up (see 2.4.5). The ligation product was transformed into

XL-10 gold ultracompetent cells, using the antibiotic Kanamycin. Successful ligation of the MYOD1 fragment into the pZERO vector created the plasmid pZERO MYOD + LUC7L (Figure 5.1C).

Primer	Sequence, 5' to 3'	PCR buffer	Annealling temp °C	Notes
MYODF2	TACTGTGTGACTGCCACTGTGATG	x11 Buffer as detailed in 2.4.2	56	7523bp product. Primer pair used for long range PCR as described in 2.4.2. PCR program: 94°C 2min, (94°C 10s, 56°C 30s, 68°C 8min) 30x cycles, 68°C 20min, 15° 1min.
MYODR2	TTCTATGGCGTTCCCCTGCTAC			
MYOD RT	TGATTCTACAGCCGCTCTACCC		-	Primer used to prime reverse transcription reaction
MYODmspcrF	GGGTGTTGGAGAGGTTTGGAAAGGGC	2xDMSO	60	Primer pair used for 192bp msPCR and screening for antisense RNA. MYOD msPCR primer F and MYOD SYBR R used to amplify 629bp region for msPCR, 59°C annealling with 40s elongation.
MYODmspcrR	ACACTGCGGGGCGGGGCTTG			
MYODsense F	CGGCGGAAGTCTACGAAG	2xDMSO	60	174bp product. Also used for SYBR green with PCR program: 95°C 10min, (95°C 15s, 60°C 15s, 72°C 15s) x40 cycles, expert mode 1+4. 25ng input. For standard PCR, 28 cycles were used
MYODsense R	AGGCGACTCAGAAGGCACG			
mMyod F	CGGCGGCAGAATGGCTAC	2xDMSO	60	607bp product (DNA), and 282bp product (cDNA). Elongation time of 35s in standard PCR reaction.
mMyod R	CGCATTGGGGTTTGAGCC			
MYOD SYBR F	CGGCTCTCTGCTCCTTTG	ABI SYBR green	65	Referred to as 'MYOD2' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
MYOD SYBR R	TCGAAACACGGGTCGTCAT			
MYOD SYBRF2	TGTGGAAACGTGCAGATTTAGATG	ABI SYBR green	60	Referred to as 'MYOD1' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles
MYOD SYBRR2	GCGCCAATCTCTCCAAACTC			
MYOD SYBRF3	CCCTCCCAACAGCGCTTTA	ABI SYBR green	60	Referred to as 'MYOD3' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
MYOD SYBRR3	CAGTTCTCCCGCTCTCCTA			

Table 5.1. Primer used for the MYOD gene

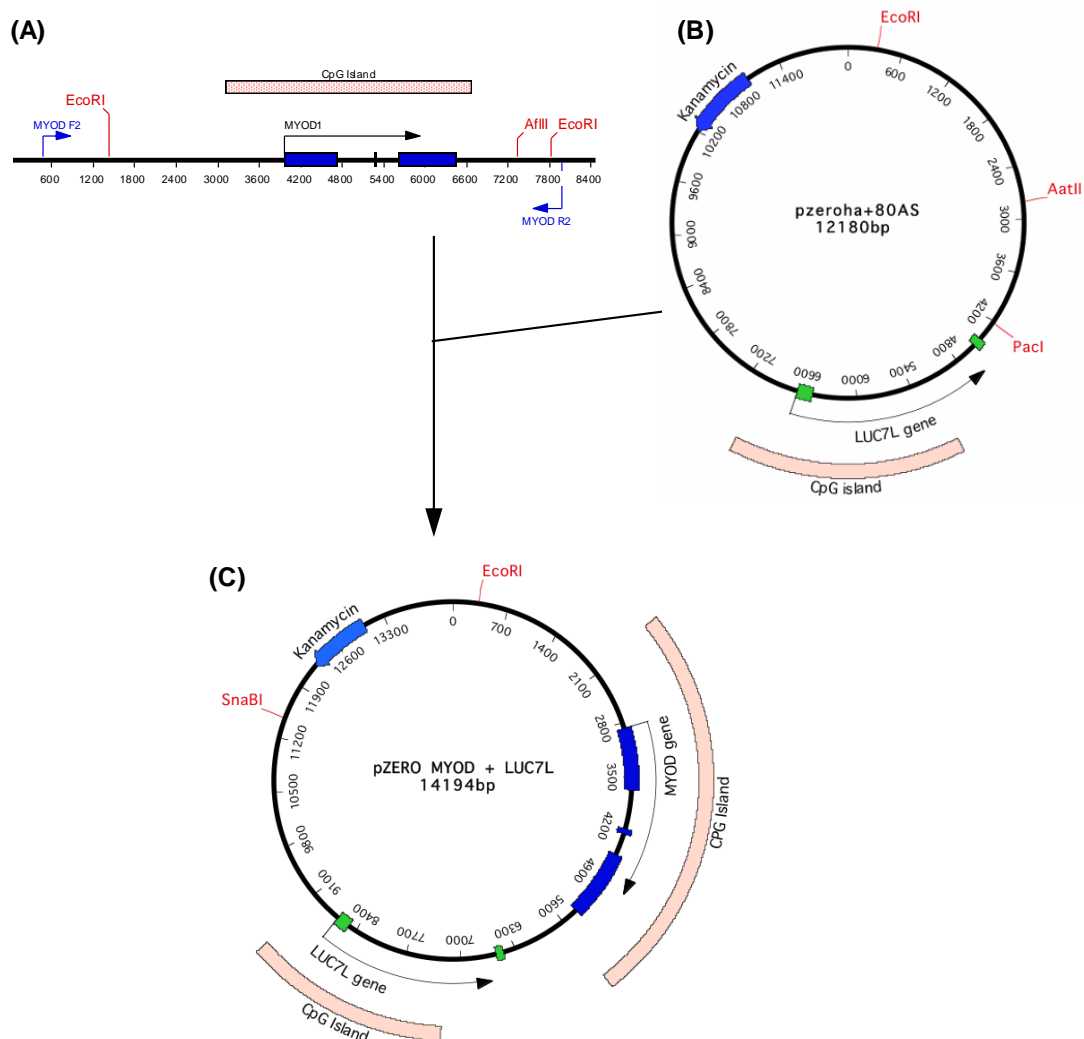


Figure 5.1 Creation of MYOD construct (A) human MYOD gene cloning strategy and digestion for ligation into pzeroha+80AS. (B) Restriction digest of pzeroha+80AS to remove unwanted portion of the plasmid (C) The completed construct on the PZERO MYOD + LUC7L plasmid. EcorR1 and SnaB1 were used for the excision of the construct from the plasmid.

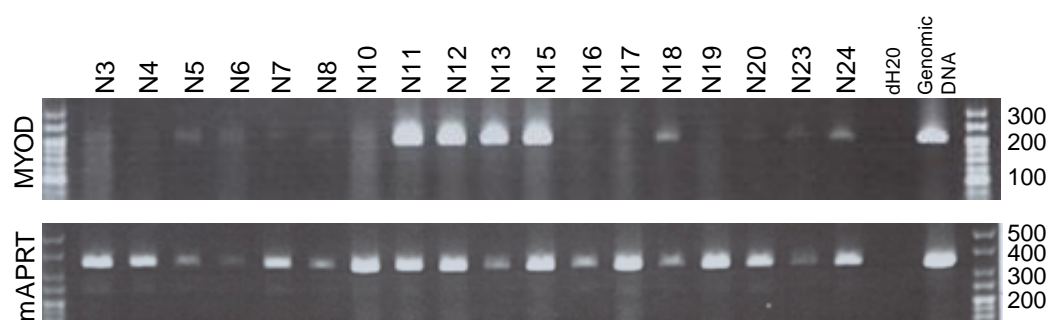


Figure 5.2 Screening of ESC lines for incorporation of the MYOD construct. PCR screening of 18 cell lines using the MYOD msPCR primer pair that amplifies are region of the human MYOD CpG island. Cell lines N5, N11, N12, N13, N15, N18, N23, N24 all tested positive. mAPRT primers were used as a positive control for amplification from extracted DNA.

5.2.2 Electroporation of MYOD construct into ESCs and screening for successful incorporation

Purification of the MYOD construct from its plasmid backbone, and its integration into mESCs by electroporation using the BioRad Gene Pulser Xcell, growth of cells following electroporation, and extraction of DNA is described in section 2.5.6 of the Materials and Methods chapter.

The MYOD construct was excised from the pZERO MYOD + LUC7L plasmid by double restriction digest with EcoR1 and SnaB1, shown in Figure 5.1C. Gel electrophoresis was used to separate the 11133bp band containing the MYOD construct from the remaining 3061bp band. The construct was purified by band excision and electroelution and electroporated into ESCs (2.5.6.2) along with the geneticin (G418) resistance gene fragment. After 8 days, 24 colonies were selected based upon their size and morphology, and cultured in 24 well plates until there were sufficient cells for DNA extraction, and for cryogenic storage. 18 Cell lines were established, with 6 cell lines exhibiting insufficient growth, or poor morphology and not being retained. Extracted DNA was screened by PCR for the presence of the MYOD construct using MYOD msPCRF/R primers (Figure 5.2).

5.2.3 Analysis of cell lines

Five cell lines that had tested positive for each construct were cultured and expanded in tissue culture. ESC material was collected from all cell lines for RNA and DNA analysis, and crosslinked pellets for ChIP for two cell lines were prepared. ESCs were differentiated into EBs over a period of seven to eight days, and more material prepared for RNA, DNA, and chromatin analysis.

Preparation of DNA, RNA and ChIP pellets is described in 2.5.7, 2.5.8, and 2.8.1/2.8.2 respectively. Differentiation of ESCs into EBs is described in 2.5.5

Expression of antisense RNA was tested for in all cell lines by strand specific RT-PCR as described in 2.7 The levels of the pluripotency factors Oct4 and Nanog was also examined in ESCs and EBs as markers for successful differentiation of cells, this was carried out by multiplex PCR with primer pairs for *Oct4*, *Nanog* and *Aprt* (2.6.5). This is detailed in 2.6.5. DNA methylation was assessed by msPCR using primers that spanned various regions of the gene's promoter that contained methylation sensitive HpaII sites, as described in 2.6.3. Gene expression levels were initially assessed by standard PCR, then by SYBR green real-time PCR, as detailed in 2.6 and 2.6.6.3 respectively. Histone modification changes between ESCs and EBs was examined using the protocol described in 2.8.

5.3 Results

The expression of the pluripotency factors *Oct4* and *Nanog* was investigated using a multiplex PCR approach that included primers for the ubiquitously *Appt* gene. Multiplex PCRs were carried out in triplicate and, following band intensity analysis using AIDA software, *Oct4* and *Nanog* expression was normalized to *Appt* expression. This is shown in Figure 5.3. Cell lines N11, N12, N13, and N15 showed a large decrease in expression of both *Oct4* and *Nanog*, indicating that the majority of the ESC population has differentiated as expected. Expression of both pluripotency factors were higher than expected in the N18 cell line after differentiation, though did show a modest decrease in EBs, indicating a high level of undifferentiated cell within the cell population.

Antisense RNA expression was examined in the cell lines by PCR analysis of cDNA from a strand specific RT-PCR reaction (Figure 5.4) using primers that amplify a region within the CpG island of the *MYOD* gene upstream of the TSS. Antisense RNA was detectable in all cell lines in ESCs and EBs.

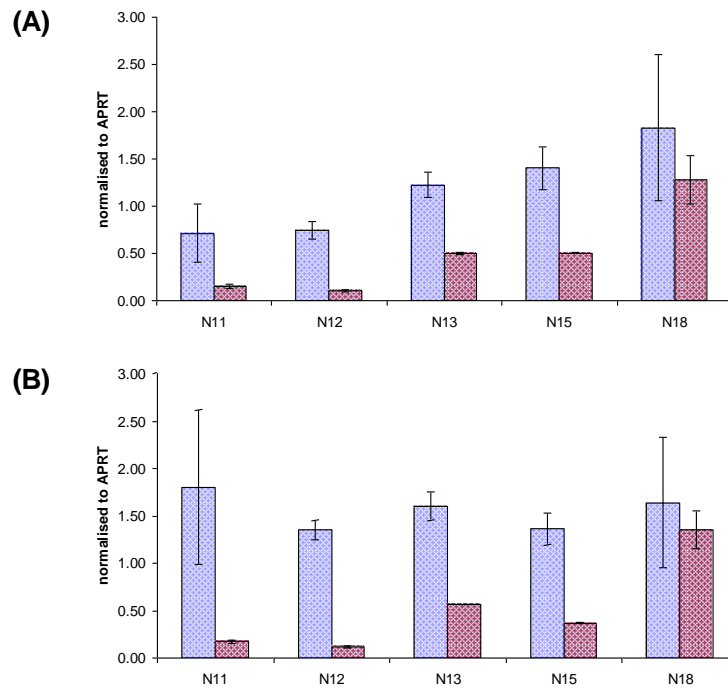


Figure 5.3 Expression of pluripotency factors. Relative expression of *Nanog* (A) and *Oct4* (B) in ESCs (blue bars) and d8 EBs (red bars). Note retention of *Oct4* and *Nanog* expression in the cell line N18

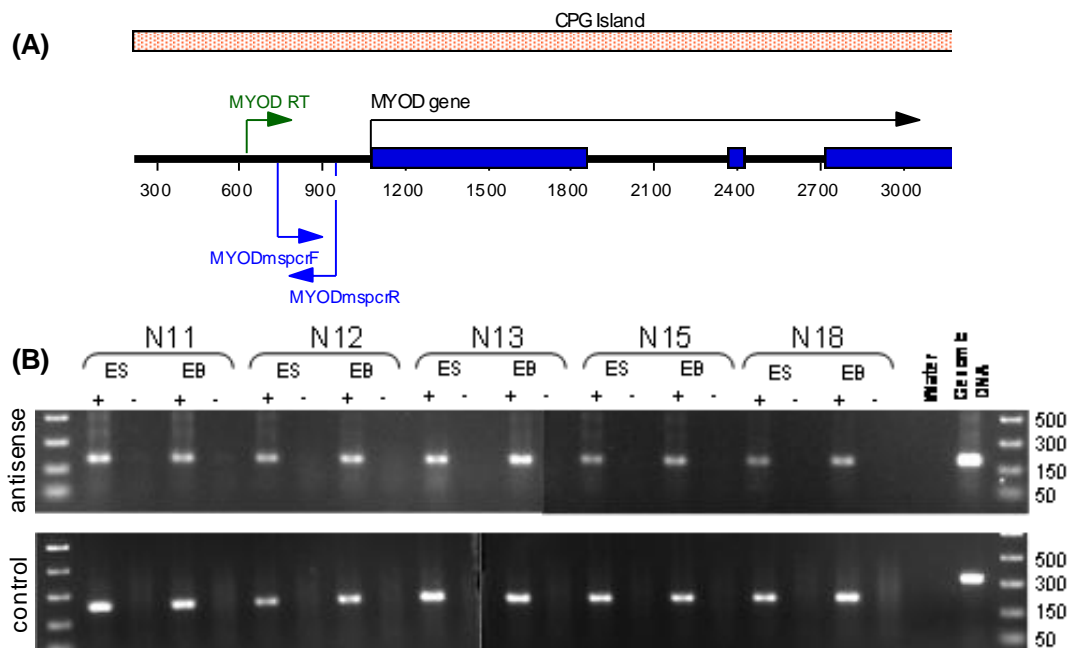


Figure 5.4 Expression of antisense RNA. (A) schematic representation of the promoter region of the MYOD construct with primer locations marked. (B) The primers MYOD RT and mAPRT1 were used to prime a strand-specific RT-PCR reaction. PCR amplification of antisense RNA by MYOD msPCR (191bp), amplification of control using mAPRT2 and mAPRT3 (225bp cDNA, 335bp genomic DNA). Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA.

MYOD gene expression was examined by standard PCR. Random-primed cDNA was used as the template for primers for both human and mouse *Myod* genes, and for the endogenous control, the ubiquitously expressed gene *Appt* (Figure 5.5). Human *MYOD* was expressed at a low level in ESCs for all 5 cell lines examined, with little or no expression detectable in N11, N12, N13, and N15 upon differentiation. Expression of human *MYOD* was detectable in N18 EBs at a higher level than in the other cell lines (Figure 5.5B). A low level of expression of mouse *Myod* was detectable in ESCs for 3 cell lines tested, with no expression detectable in EBs (Figure 5.5D). There was insufficient RNA to examine *Myod* expression for the cell lines N11 and N12. Gel analysis of human *MYOD* and *Appt* PCRs was carried out using AIDA software, then expression of *MYOD* was normalized to *Appt* expression (Figure 5.6). Reduction in gene expression is seen in all cell lines, though reduction in expression is modest for the N18 cell line. Using the same primers designed for standard PCR, SYBR green real-time PCR was carried out to more accurately assess gene expression of the human *MYOD* gene. This is shown in Figure 5.7. This confirmed the pattern of expression that had been observed by normal PCR, with expression detectable in ESCs but barely detectable in EBs. Again, the N18 cell line showed a similar level of expression in ESCs and EBs, consistent with the incomplete level of differentiation.

Methylation sensitive PCR was initially carried out a 192bp region of the *MYOD* CpG Island located upstream of the TSS containing 4 HpaII sites (Figure 5.7A). Four of the cell lines examined showed DNA methylation across this region in EBs, with none detectable in ESCs, and with controls indicating complete

digestion and no detectable DNA methylation within the LUCL CpG island (Figure 5.7C). The cell line N11 had detectable DNA methylation in both ESCs and EBs within the MYOD CpG island, and also had detectable DNA methylation within the Luc7L CpG island at a low level in ESCs, and at a high level in EBs - indicating that a small proportion of the ESC population were methylated in this region, with DNA methylation of LUC7L becoming widespread within the EB cell population upon differentiation. msPCR was then carried out over a larger region covering a total of 9 HpaII sites, including the original for sites examined. DNA methylation across this larger region exactly matched the initial observations, with DNA methylation detectable in N12, N13, N15, and N18 upon differentiation, and DNA methylation in both ESCs and EBs for the cell line N11 (Figure 5.7B and C).

The pattern of histone modifications for the repressive marks H3K9me3, H3K27me3, and for the active marks H3K4me3 and H3Ac were examined at different locations across the MYOD gene by SYBR green real-time PCR for the cell lines N15 and N18 in ESCs and EBs. Results for N15 and N18 are shown in Figure 5.8, and 5.9 respectively.

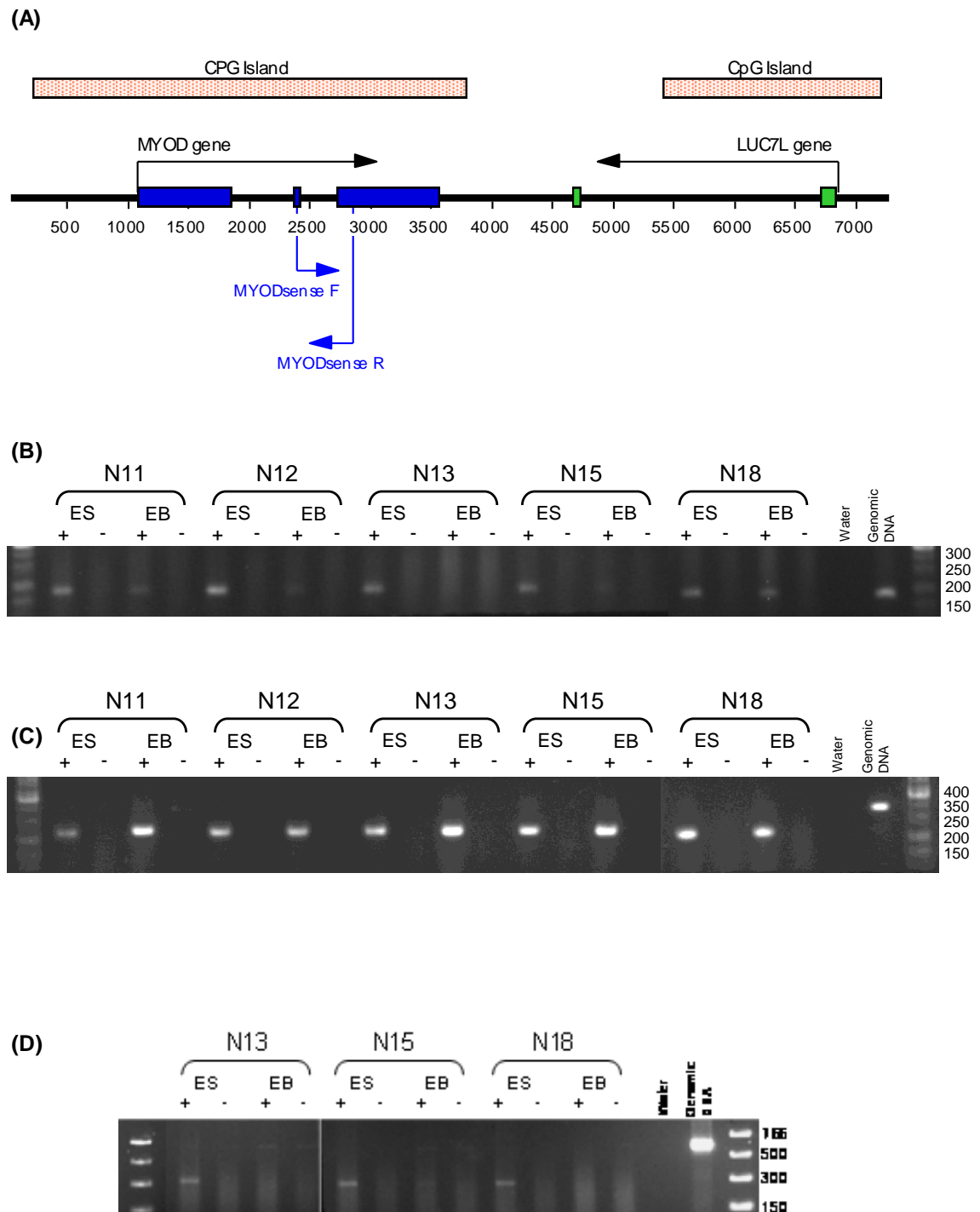


Figure 5.5 Gene expression in the MYOD cell lines analysed by PCR. (A) Schematic representing the MYOD construct and the position of the primers. (B) expression of human MYOD in cell lines. Human ESC cDNA used as positive control. (C) expression of mouse Aprt. (D) expression of mouse MYOD. RT-PCR reactions were primed using random hexamers. Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA.

Both cell lines showed very similar enrichments for H3K9me3, H3K4me3, and H3Ac. H3K9me3 levels increased across the *MYOD* gene upon differentiation. Within the body of the *MYOD* gene, enrichment levels more than doubled (MYOD2 and MYOD3), while upstream of the TSS (MYOD1), levels doubled for the cell line N15, and showed a slightly lesser increase for the N18 cell line. H3K9me3 levels also increased upon differentiation within the LUC7L region, immediately prior to the TSS. This increase was more pronounced in the N15 cell line. H3K4me3 levels remained fairly steady upon differentiation, with a slight decrease in enrichment levels within the body of the gene for both cell lines. Acetylation levels were low in ESCs in comparison to the active mouse *Actb* gene, and decreased further within the body of the gene upon differentiation. Acetylation also decreased within the promoter region of N18, while N15 promoter acetylation levels remained constant.

H3K27me3 levels in the N15 cell line increased upon differentiation to just under double the levels observed in ESCs. This occurred at all three points assayed within the *MYOD* gene as well as the LUC7L region. Contrary to this, the N18 cell line showed only a modest enrichment upon differentiation, and only within the body of the gene. H3K27me3 levels remained constant at the promoter region, and showed a minor decrease at the *LUC7L* promoter.

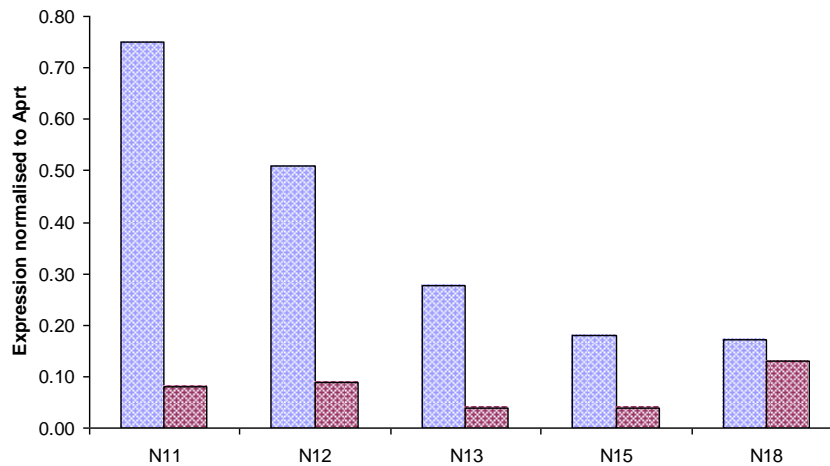


Figure 5.6 Gene expression in the MYOD cell lines quantified by gel analysis. Gel images shown in Figure 5b and 5c were quantified using AIDA software, and intensity of bands for *MYOD* expression were normalised to bands for mouse *Aprt*. ESC samples are blue, EB samples are red.

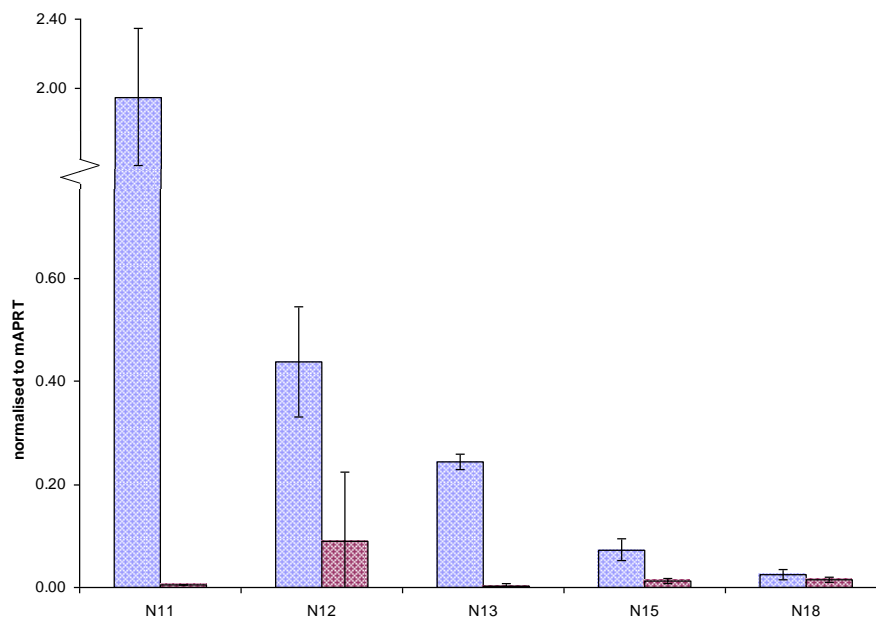


Figure 5.7 Gene expression in the MYOD cell lines quantified by realtime PCR. The graph shows Human *MYOD* expression normalised to *Aprt*. The position of the primers used to detect human *MYOD* are shown in Figure 5.5. ESC samples are blue, EB samples are red.

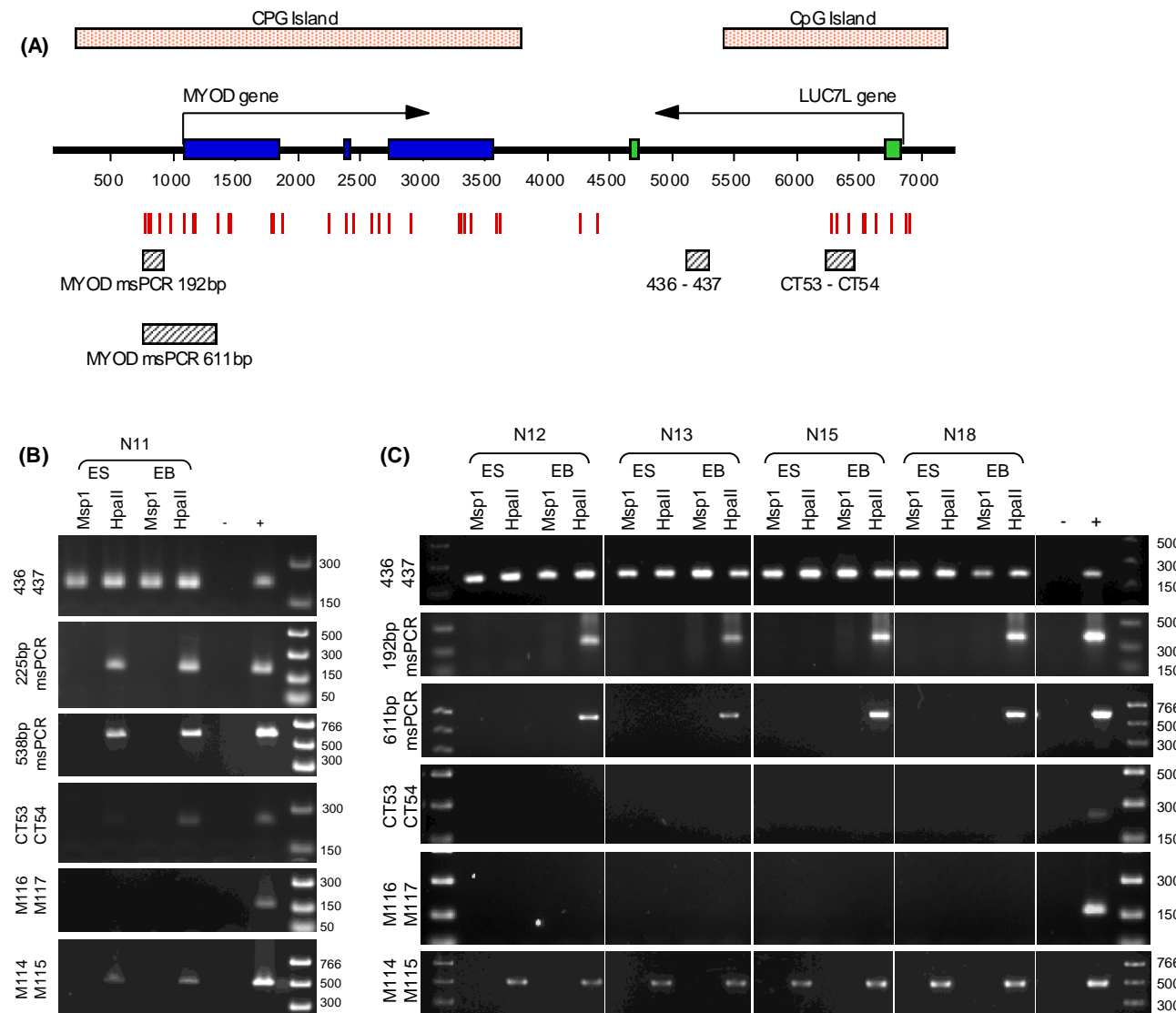


Figure 5.7 Methylation sensitive PCR within the ACTB CpG Island. (A) shows a schematic of the MYOD construct with Msp1/HpaII sites marked as vertical red lines and regions amplified by PCR primers represented as hatched grey boxes. Within the LUC7L CpG island; CT53 CT54 amplify across 3 restriction sites (248bp), and 436 437 primers amplify a region with no restriction sites (208bp). The 192bp PCR product encompasses 4 restriction sites, and the 611bp product encompasses 9 restriction sites. (B) and (C) M116 M117 amplify a region within the 5' end of the APRT gene that contains 4 restriction sites and is devoid of DNA methylation (164bp), M114 M115 primers amplify within the promoter of AIR, which is stably methylated on the maternal copy in ESCs and EBs (489bp).

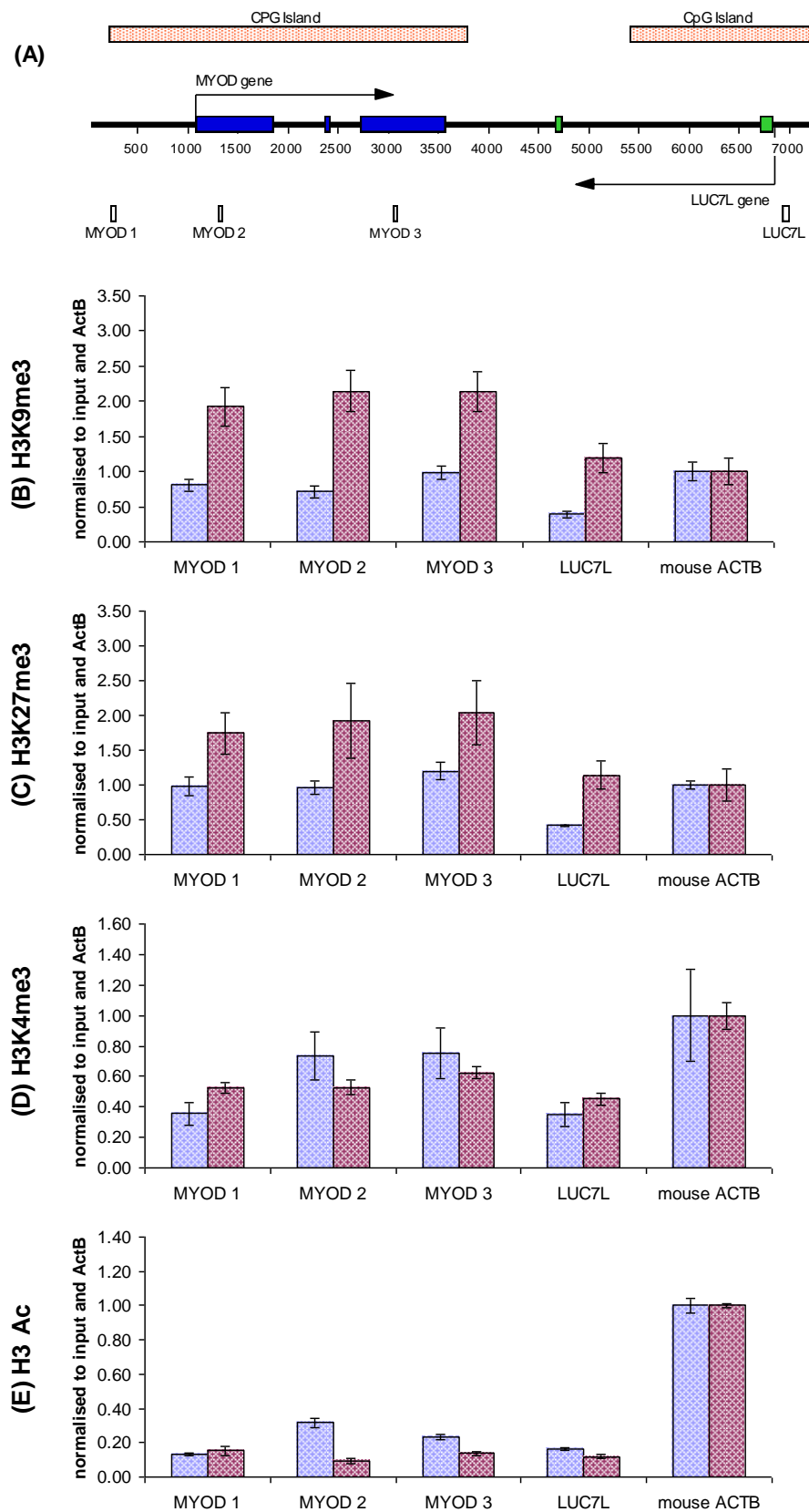


Figure 5.8 Histone modifications to the MYOD construct in the N15 cell line. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *MYOD* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

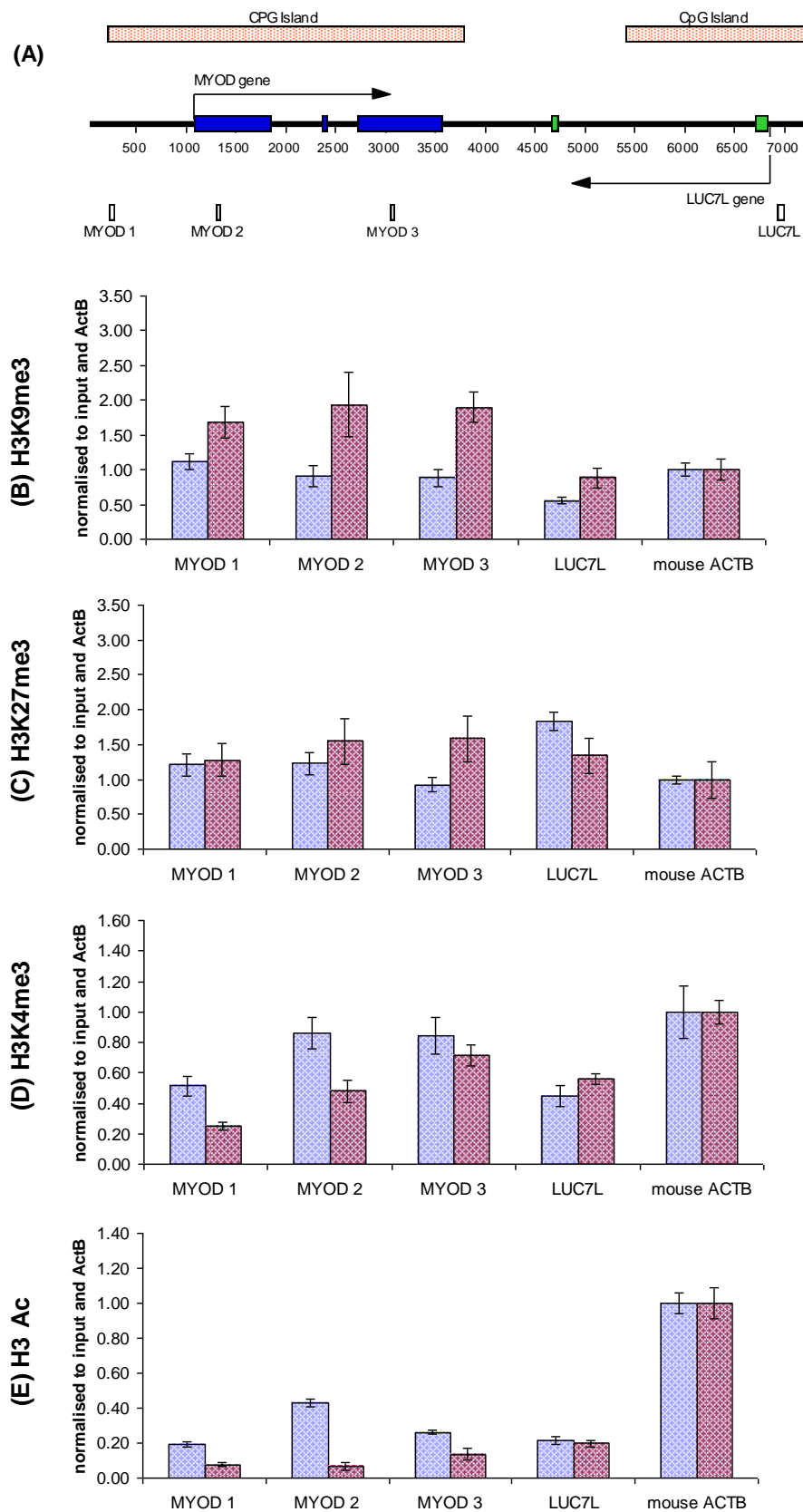


Figure 5.9 Histone modifications to the MYOD construct in the N18 cell line. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *MYOD* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

5.4 Discussion

The results presented in this chapter show that upon differentiation of mESCs antisense transcription through the *MYOD* CpG island leads to an increase in H3K9me3 and DNA methylation in a near identical manner to the alpha globin gene. DNA methylation occurred upon differentiation in four of the cell lines analyzed, and H3K9me3 became enriched throughout the *MYOD* gene upon differentiation in both of the cell lines examined. DNA methylation is not a normal feature of repression of the *MYOD* gene, regardless of its transcriptional state (Jones *et al.* 1990) but dense DNA methylation within the CpG island occurring during differentiation is one of the hallmarks of the silencing effect observed at the alpha globin gene.

MYOD gene expression is regulated by the PRC/TRX repressor/activator complexes that are responsible for marking developmentally important tissue specific genes with the bivalent mark H3K27me3/H3K4me3 in ESCs. This maintains the gene in a repressive state, but poised for activation in the correct tissues (Ku *et al.* 2008, Bernstein *et al.* 2006a). Upon commitment to a lineage where the gene will not be expressed, the H3K4me3 mark is lost and the gene is marked by H3K27me3 alone. In *D. melanogaster* this switch to a monovalent repressive domain is associated with both H4K20me3 and H3K9me3 at the *ubx* gene (Papp & Muller 2006), however the association with H3K9me is not observed in mammals (Mikkelsen *et al.* 2007, Bernstein *et al.* 2006a). This suggests that the development of H3K9me3 at the *MYOD* gene upon differentiation is directly attributable to the presence of antisense RNA. Together with the development of DNA methylation, this strongly suggests that

the silencing effect originally characterized at the alpha globin gene is equally applicable to other tissue specific genes containing CpG islands.

The MYOD gene is bivalently marked in mouse (Mikkelsen *et al.* 2007) and human ESCs (Ku *et al.* 2008), but while enrichment of H3K4me3 in ESCs is detectable in both cell lines analyzed, enrichment for H3K27me3 is far less obvious, with only very minor levels of enrichment above that of the endogenous control Actb detectable, and then only at some sites within the MYOD gene (Figures 5.8 and 5.9), Enrichment for H3K27me3 becomes much more apparent upon differentiation, with levels increasing throughout the MYOD gene and the LUC7L gene in the N15 cell line, while there are more modest increases in the N18 cell line - possibly as a result of the higher proportion of undifferentiated cells in the N18 cell line that was detectable when screening for Oct4 and Nanog expression levels. An increasing level of H3K27me3 enrichment, coupled with decreasing levels of H3K4me3 upon differentiation suggests that the bivalent domain is resolving into a repressive domain, at least for a proportion of cells within the EB population.

Enrichment for H3K27me3 at inactive loci is described as 'modest' in terminally differentiated cells (Barski *et al.* 2007), however the level of ESC H3K27me3 observed in the N15 and N18 cell lines is barely greater than that for an active gene not associated with H3K27me3. The H3K27me3 mark is added by the HMT EZH2, a subunit of the PRC2 complex. In *D. melanogaster* PRCs are recruited to distinct DNA sequences referred to as Polycomb Recruiter Elements (PREs) ((Schuettengruber *et al.* 2007)), but PREs are not present in

mammals and the PRC recruitment mechanism has yet to be elucidated, though it has been demonstrated that DNA sequence alone is sufficient for PRC recruitment in mammals (Zhou and Bernstein personal communication). In *D. melanogaster* there are several PREs present both upstream and downstream of the TSS, and PRCs can mediate silencing over large distances (Papp & Muller 2006). Therefore, if the DNA elements responsible for PRC recruitment in mammals follow a similar pattern it is possible that upstream regulatory factors were excluded when *MYOD* was cloned, even though a 1.5Kb region of the *MYOD* gene upstream of the TSS was included. This could weaken PRC binding, and result in a reduced H3K27me3 levels.

Of the five cell lines examined, four developed DNA methylation upon differentiation, while one cell line, N11, was methylated in ESCs as well as EBs, and methylation was not restricted to the *MYOD* gene but was also detectable within the CpG island of the *LUC7L* gene. DNA methylation within the *LUC7L* CpG island is not a feature observed in the other cell lines analyzed, therefore this would suggest that the silencing observed the ESCs was not a consequence of a high level of spontaneous differentiation within the population. Additionally, *Oct4* and *Nanog* expression levels are comparable to other cell lines and are not suggestive of differentiation.

There are several possible explanations for the methylation observed in ESCs. Firstly; it is possible that the construct incorporated into a heterochromatic locus already associated with repressive DNA and histone modifications that subsequently spread throughout the incorporated DNA construct. However, if

this were the case the expression of *LUC7L* has not been sufficiently affected to be detectable by standard 30 cycle PCR, as antisense transcripts are readily detectable in both ESCs and EBs. It could be possible that the silencing is not uniform within the cell population with silencing occurring in some cells but not others, thereby providing both a detectable methylation signal and maintained expression of *LUC7L*. An alternative possibility regards copy number of the construct in the cell, as there is some evidence that high copy numbers entering the same locus as a concatameric array result in repression of the array by both DNA methylation and a condensed chromatin state (Garrick *et al.* 1998). While the copy number is unknown for the *MYOD* cell lines, Garrick *et al.* demonstrated a silencing state occurring when the number of copies was approaching one hundred, while five copies in series did not result in repression. While such high copy numbers are not generally observed using the electroporation technique, which normally results in between one and four copies per cell, this does remain a possible explanation for the DNA methylation observed in the N11 cell line.

Bivalently modified genes with CpG rich promoters are known to be expressed at low levels in ESCs (Mikkelsen *et al.* 2007, Bernstein *et al.* 2006a), which matches the expression levels of mouse and human *MYOD* observed in this study, with expression detectable in ESCs, before becoming repressed upon differentiation. The cell line N18 exhibits a higher level of expression in EBs than the other cell lines examined, again this is likely to be due to incomplete differentiation as indicated by *Oct4* and *Nanog* expression levels.

Myod has been shown to be enriched for H2AZ in mouse ESCs (Creyghton *et al.* 2008). H2AZ is a variant of H2A that has been shown to have a potential DNA methylation protection effect in *Arabidopsis thaliana* (Zilberman *et al.* 2008) and it is thought that this role may be conserved in mammals where in ESCs it is involved with the PRC2 complex in maintaining developmentally important genes in a silent state, but protected from DNA methylation and 'poised' for activation (Creyghton *et al.* 2008). In differentiated tissues, it is thought to have a different but related role where it localizes to the promoters of active genes (Barski *et al.* 2007). If this histone variant is also present at the human *MYOD* gene, it is interesting to note that it was incapable of preventing DNA methylation in the N11 ESCs.

In summary: the data described in this chapter indicates that antisense RNA driven silencing of a CpG island, as observed in the ZF patient, is not restricted to the alpha globin CpG island, but can also occur at the CpG island of another tissue specific gene, *MYOD*. The next chapter will address whether CpG islands associated with ubiquitously expressed genes are also sensitive to this form of silencing.

Chapter 6 Role of antisense RNA in silencing ubiquitously expressed genes

6.1 Introduction

As described in chapter 5, the tissue specific gene *MYOD* is susceptible to the same form of silencing originally observed at the alpha globin gene in the ZF patient. Tissue specific genes are normally silenced by repressive histone modifications, and in some cases DNA methylation (Shen *et al.* 2007), in tissues where they are not expressed, and therefore are already subject to repressive control. Ubiquitously expressed genes are expressed in all tissue types, normally producing proteins essential for the cell to function, and are not subject to repression. To investigate whether ubiquitously expressed genes containing CpG islands can become silenced by the expression of an antisense RNA in the same manner as has been observed for tissue specific genes, the ubiquitously expressed genes *UBC* and *ACTB* were cloned into constructs containing an antisense orientated RNA, expressed by the *LUC7L* gene.

The *ACTB* gene is one of six actin isoforms in humans and codes for the β actin protein, a cytoskeletal component that forms long helical polymers. Actin filaments, together with microtubules and intermediate filaments make up the cellular cytoskeleton that is essential for maintenance of cell shape and motility. *ACTB* is highly conserved among eukaryotes, and protein levels are normally high within the cell (Creighton 1999, Nelson 2000). The *ACTB* gene has a CpG island with a GC density of 71.5%, which covers approximately a 2Kb region that includes the TSS and the first two exons. There are three main mRNA splice variants transcribed, each consisting of six exons. The *ACTB* gene is

expressed at very high levels in mammalian cells (Mikkelsen *et al.* 2007), and the *ACTB* gene locus is marked with H3K4me3, and other histone modifications associated with high levels of gene expression (Mikkelsen *et al.* 2007, Pan *et al.* 2007, Guenther *et al.* 2007).

The *UBC* gene is highly conserved, and encodes for a protein involved in ubiquitin conjugation. A multi-enzyme cascade is involved in the addition of ubiquitin to proteins to target them for destruction in the proteasome in eukaryotic organisms. Ubiquitin conjugating proteins bind C-terminal modified ubiquitin from ubiquitin activating proteins and attach it to the target protein through interaction with a ubiquitin ligase (Creighton 1999, Colowick *et al.* 2005, Verma R & Deshaies RJ 2005, Pickart & Eddins 2004). The *UBC* gene is transcribed at a medium levels (Mikkelsen *et al.* 2007) and contains two exons, the first exon is small and contained within the CpG island, while the second exon is approximately 2Mb in size, and highly repetitive in sequence. This is shown in Figure 6.1A, and a sequence comparison with mouse *Ubc* is shown in Figure 6.1B. Exon2 is highly conserved between species. The CpG island, which is just under 2Kb in size, has a CpG density of 60.7%. There are seven *UBC* mRNA splice variants, two of which are more commonly transcribed. All splice variants consist of the first exon and alternative splice forms of exon 2. The *UBC* gene has been observed to be marked by a high level of H3K4me3 in both ESCs and in a variety of differentiated tissues in mammals (Mikkelsen *et al.* 2007, Pan *et al.* 2007, Guenther *et al.* 2007).

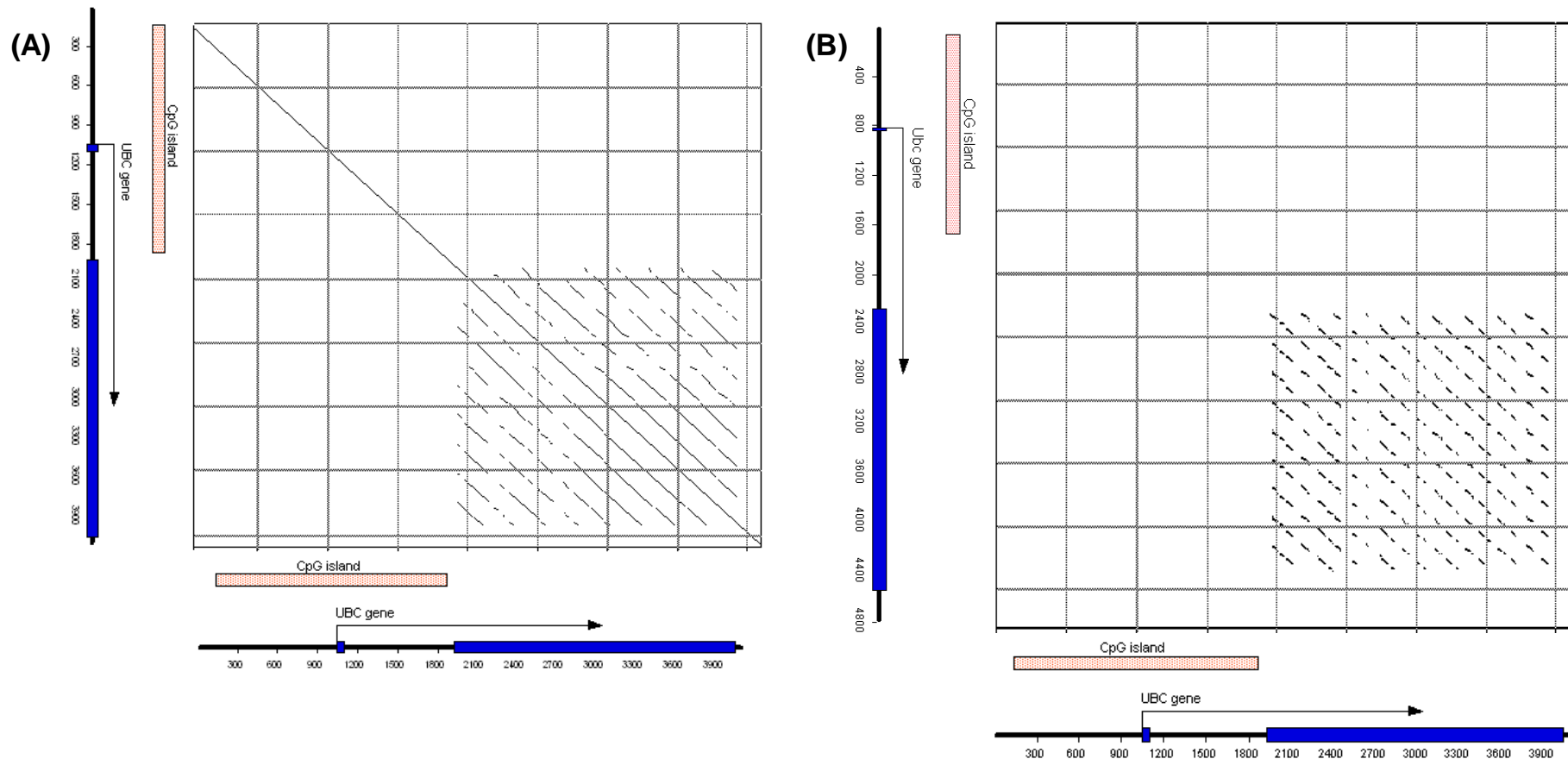


Figure 6.1 Sequence alignments for the UBC gene. Alignments shown have a minimum of 90% sequence identity. (A) Alignment of the UBC gene against its own sequence. Exon2 contains a highly repetitive sequence from 1939bp onwards. (B) Alignment of the human UBC gene (X axis) against the mouse Ubc gene (Y axis). Exon2 shows a high level of sequence conservation between the species

6.2 Methods

All primers used in the investigation of the *ACTB* gene and *UBC* gene are listed in Table 6.1 and Table 6.2 respectively. Standard primers used are listed in Materials and Methods.

In general, expression of antisense RNA and the presence of DNA methylation was examined twice for each cell line. Gene expression analysis and Chromatin IP experiments were only carried out once for each cell line, unless otherwise stated.

6.2.1 Engineering of constructs for electroporation into ESCs

Any techniques specific to this chapter are detailed below, while general methods are detailed in Chapter 2. Growth of bacterial cultures and purification of BACs is described in the Section 2.2. The engineering of constructs, including long range PCR, cloning, and transformation into ultracompetent cells is described in section 2.4.

6.2.1.1 Construction of the *ACTB* + *LUC7L* plasmid

The *ACTB* + *LUC7L* plasmid was engineered as shown in Figure 6.2. DNA extracted from *E.coli* transformed with the BAC CTB-161C1 was used as the template for long accurate PCR of a 4306bp fragment containing the *ACTB* gene, as shown in Figure 6.2A. Purified PCR product was digested with the REs Bsp1 and Sbf1 (restriction double digest). The restricted PCR product was blunt ended using klenow, then cleaned up by P/C. The pSL301 plasmid was digested with EcoRV (Figure 6.2B). Concentration of PCR product and

linearised plasmid were assessed by gel electrophoresis and a ligation reaction was set up. The ligation product was transformed into XL-10 gold ultracompetent cells, using the antibiotic Ampicillin. Control plates contained 4 colonies, and ligation plates contained 18 colonies. Plasmid DNA was extracted and restriction digest using Nco1 carried out to identify successful ligations. Two plasmids were obtained, those containing the *ACTB* gene in the orientation left to right, as shown in Figure 6.2C, named pSL301*ACTB*, and those with the *ACTB* gene in the opposite orientation, named pSL301*ACTB*Rev. An additional digest using Xmn1 was used to excise the region of the plasmid containing the insert to confirm it was of the correct size. Restriction sites are shown in Figure 6.2C.

Primer	Sequence, 5' to 3'	PCR buffer	Annealling temp °C	Notes
ACTB F2	GAGTGTGGTCCTGCGACTTCTAAG	x11 Buffer as detailed in 2.4.2	56	Primer pair used for long range PCR as described in 2.4.2. PCR program: 94°C 2min, (94°C 10s, 56°C 30s, 68°C 8min) 30x cycles, 68°C 20min, 15° 1min.
ACTB R2	CCTCTAAGGCTGCTCAATGTCAAG			
ACTB RT2	TGCTTTCAGGGCAGTTGCTC	-	-	Primer used to prime reverse transcription reaction
ACTBmsPCR F	AGTGCCCAAGAGATGTCCACA	2xDMSO	60	Primer pair used for 225bp msPCR and screening for antisense RNA
ACTBmsPCR R	AAAAGAGCGAGAGCGAGATTGAG			
ACTBmsPCR F2	CGCCTCCGACCAGTGTTTG	2xDMSO	60	this primer and the ACTB SYBR R used for 538bp msPCR. 40s annealling time used in PCR.
ACTBsenseF2	CCAGTCCTCTCCCAAGTCCACAC	2xDMSO	60	Also used for SYBR green with PCR program: 95°C 10min, (95°C 15s, 60°C 15s, 72°C 15s) x40 cycles, expert mode 1+4. 25ng input. For standard PCR, 28 cycles were used
ACTBsenseR2	TGCCTCCACCCACTCCCAG			
mACB R	GGGGGGACAAAAAAGGGAG	2xDMSO	60	Also used for SYBR green with PCR program: 95°C 10min, (95°C 15s, 60°C 15s, 72°C 15s) x40 cycles, expert mode 1+4. 25ng input. For standard PCR, 28 cycles were used.
mACTB F	CGGGGAAGGTGACAGCATTG			
ACTB SYBR F	ACGAGGCCCAGAGCAAGA	ABI SYBR green	65	Referred to as 'ACTB2' in text. realtime program used; 95°C 10min, (95°C 15s, 65°C 1min) x40 cycles.
ACTB SYBR R	GACGATGCCGTGCTCGAT			
ACTB SYBR F2	CGGCCAACGCCAAACT	ABI SYBR green	60	Referred to as 'ACTB1' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles
ACTB SYBR R2	TCCCCTCCTTTTGCAGAAA			
ACTBSYBR 5F	CCACCCCACTTCTCTCTAAGGA	ABI SYBR green	60	Referred to as 'ACTB3' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
ACTBSYBR R5	ACCTCCCCTGTGTGGACTTG			

Table 6.1. Primer used for the ACTB gene

Primer	Sequence, 5' to 3'	PCR buffer	Annealling temp °C	Notes
UBC F	TTTTTCTCTGGGAGTGACCATACG	x11 Buffer as detailed in 2.4.2	56	Primer pair used for long range PCR as described in 2.4.2. PCR program: 94°C 2min, (94°C 10s, 56°C 30s, 68°C 8min) 30x cycles, 68°C 20min, 15° 1min.
UBC R	TGTAGAGAAACATCAGACACCTGCG			
UBC msPCR F	GAACAGGCGAGGAAAAGTAGTCC	2xDMSO	60	Primer pair used for 255bp msPCR
UBC msPCR R	TTGGCGGTCTCTCCACACG			
UbiC1	CCCACCTTGTTTCAACGACCTC	NEB PCR buffer	60	Primer pair used to screen for antisense RNA
UbiC2	TAGTCCCTTCTCGGCGATTCTG			
UBC RT	TCAGCAGAAGGACATTTTAGGACG		-	Primer used to prime reverse transcription reaction
UBC Sense F	TTGGGTCGCAGTTCTTGTGTTG	2xDMSO	60	Also used for SYBR green with PCR program: 95°C 10min, (95°C 15s, 60°C 15s, 72°C 15s) x40 cycles, expert mode 1+4. 25ng input.
UBC sense R	AGGGATGCCTTCCTTATCTTGG			
mUbc F	GCAAGACCATCACCTGGACG	ABI SYBR green	60	Reatime PCR program: 95°C 10min, (95°C 15s, 60°C 15s, 72°C 15s) x40 cycles, expert mode 1+4. 25ng input.
mUbc R	CACCCAAGAACAAGCACAAGGAG			
UBC SYBR F	TTTCATGCCTCCCTGTTG	ABI SYBR green	60	Referred to as 'UBC1' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
UBC SYBR R	TGTACATTCTGAGGGCCAGGTA			
UBC SYBRF2	GTGAAGTTTGTCAGTACTGGAGAA	ABI SYBR green	60	Referred to as 'UBC2' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
UBC SYBRR2	CGGCACCGCCATAACTG			
UBC SYBRF3	GCCGAGTGACACCATTGAGA	ABI SYBR green	60	Referred to as 'UBC3' in text. Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles. Amplifies four regions within Exon 2 of the UBC gene. Also amplifies mouse UBC.
UBC SYBRR3	GGGATGCCTTCCTTGCTTG			

Table 6.2. Primers used for the UBC gene

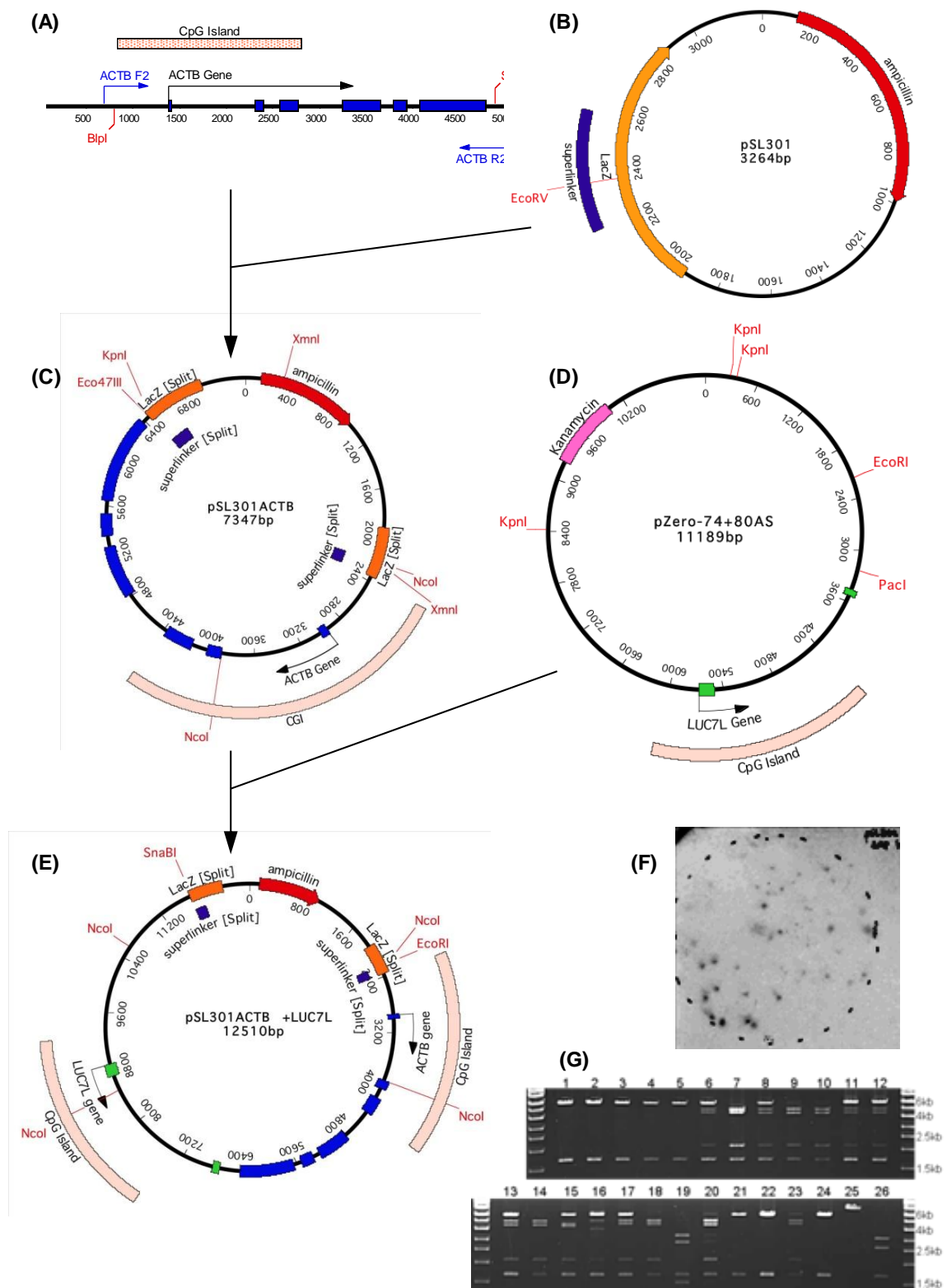


Figure 6.2 Creation of ACTB construct (A) human ACTB gene cloning strategy and digestion for ligation into pSL301(B) to create pSL301ACTB (C). The LUC7L gene fragment including promoter and the first two exons (D) was then ligated into pSL301ACTB to create the completed construct (E). (F) colony screening of transformed ultracompetent cells by radioactive probe. (G) Nco1 restriction digest, with lanes 9,10,14 and 18 with the expected digest sizes of 4611, 4168, 2078, and 1653bp. lane 25; vector uncut, lane 26; vector Nco1 digest.

The plasmid pSL301*ACTB* was digested with the RE Eco47III, cleaned up by P/C, and digested by Kpn1 (Figure 6.2C). This produced a fragment with one blunt end and one “sticky” Kpn1 end. The plasmid pZero -74 +80AS was digested with the Pac1, and then the linearised plasmid was blunt-ended using klenow. Digestion was carried out with Kpn1 and EcoR1, then half the product was treated by SAP, and half remained untreated (Figure 6.2D). The 5196bp fragment was then gel purified. Vector and insert were then ligated O.N at 16°C. Following transformation, control plates contained in excess of a 1000 colonies while the colony numbers on ligation plates were too many to accurately count.

The plates were screened by colony screening with a probe specific for the *LUC7L* insert. Hybond-N nylon membrane (Amersham pharmaceuticals) was briefly submerged in LB media, with excess blotted off using 3MM whatmann paper. The prepared membrane was then lowered on top of the colonies to be screened and lightly pressed down to ensure contact. The membrane was then transferred, colony-side facing up, to a fresh petri dish containing LB agar and Ampicillin, and incubated O.N. The original agar plate containing the colonies being screened was stored at 4°C.

The membrane was transferred onto a section of 3MM Whatmann paper soaked with Denaturing solution (200mM NaOH, 1% SDS), and left for 5min. It was then transferred to a section of 3MM whatmann paper soaked with Neutralisation buffer (3M potassium acetate pH5.5) and left for 5min. The membrane was then submerged in 2xSSC and cell debris was removed. The

membrane was carefully blotted between sheets of 3MM whatmann paper to remove excess moisture, baked for 30min at 80°C, then crosslinked in a U.V transiluminator.

The membrane was prehybridised at 42°C for 2 hours with constant rotation in Oligo hyb buffer (7% PEG 8000, 10% SDS), pre-warmed at 42°C. The Oligo hyb buffer was removed and replaced with fresh Oligo hyb buffer, warmed to 42°C, and an end labelled Oligonucleotide probe, created using the primer 436 specific for the human *LUC7L* region of the construct (Probe preparation is described in 2.10.3). The membrane was incubated O.N at 42°C with constant rotation.

The membrane was washed twice at room temperature in 2xSSC, 0.1% SDS with radiation levels monitored between washes. The membrane was then exposed to film for 20min at RT before the film was developed.

Positive colonies could be identified as dark spots on the film, as shown in Figure 6.2F. After alignment of the film with the original plate, colonies were selected and minipreped. Following restriction digest screening, positive cultures were selected. Figure 6.2G shows restriction digest by Nco1, with positives in lanes 9,10,14, and 18 producing the expected restriction fragment sizes of 4611, 4168, 2078, and 1653bp. Vector, uncut and cut is shown in lanes 25 and 26 respectively. Following further restriction digest to confirm correct ligation glycerol stocks were made and sequencing carried out.

6.2.1.2 Construction of the *UBC* + *LUC7L* plasmid

The *UBC* + *LUC7L* plasmid was engineered as shown in Figure 6.3. DNA was extracted from *E.coli* transformed with the BAC RPC1 11-59202 and purified BAC DNA used as the template for long accurate PCR of a 6983bp fragment containing the *UBC* gene as shown in Figure 6.3A. Following restriction digest to confirm PCR amplification of the correct region, the PCR product was restricted in a double digest with BamH1 and Nde1, then gel purified to produce a 4536bp fragment containing the *UBC* gene (Figure 6.3A). The plasmid pSL301 was restricted with Nde1 and BamH1 in a double digest (Figure 6.3B), ligated with the *UBC* gene fragment O.N, then transformed. Following O.N incubation on plates containing the antibiotic Ampicillin, colonies grew in a ratio of 4:1 on ligation plates vs. control plates. Following miniprep of 24 colonies, correctly ligated plasmids were selected by restriction digest.

Available restriction sites on pZero -74 +80AS (Figure 6.3D) and pSL301*UBC* (Figure 6.3C), allowed for a partially sticky ended ligation of the *UBC* gene into the pZero plasmid, containing the promoter and the first three exons of the *LUC7L* gene. The plasmid pSL301*UBC* was restricted with Nde1, blunt ended with klenow, then sequentially digested, first with Acl1 to destroy an unwanted fragment that could ligate, then with EcoR1. The sample was cleaned up using a Qiagen mini-elute kit. The plasmid pZero ha +80AS was restricted with Pac1, blunt ended with klenow, then sequentially digested with AatII to destroy an unwanted fragment, and with EcoRI (Figure 6.3D), and finally treated with SAP. A ligation reaction was set up and transformed into XL-10 gold cells, which were grown O.N on agar plates containing Kanamycin. This produced colonies in

ratio of 5:1 on ligation plates vs. control plates. Following miniprep of 24 colonies, correctly ligated plasmids were selected by restriction digest and sequenced. A Schematic of the final plasmid is shown in 6.3E.

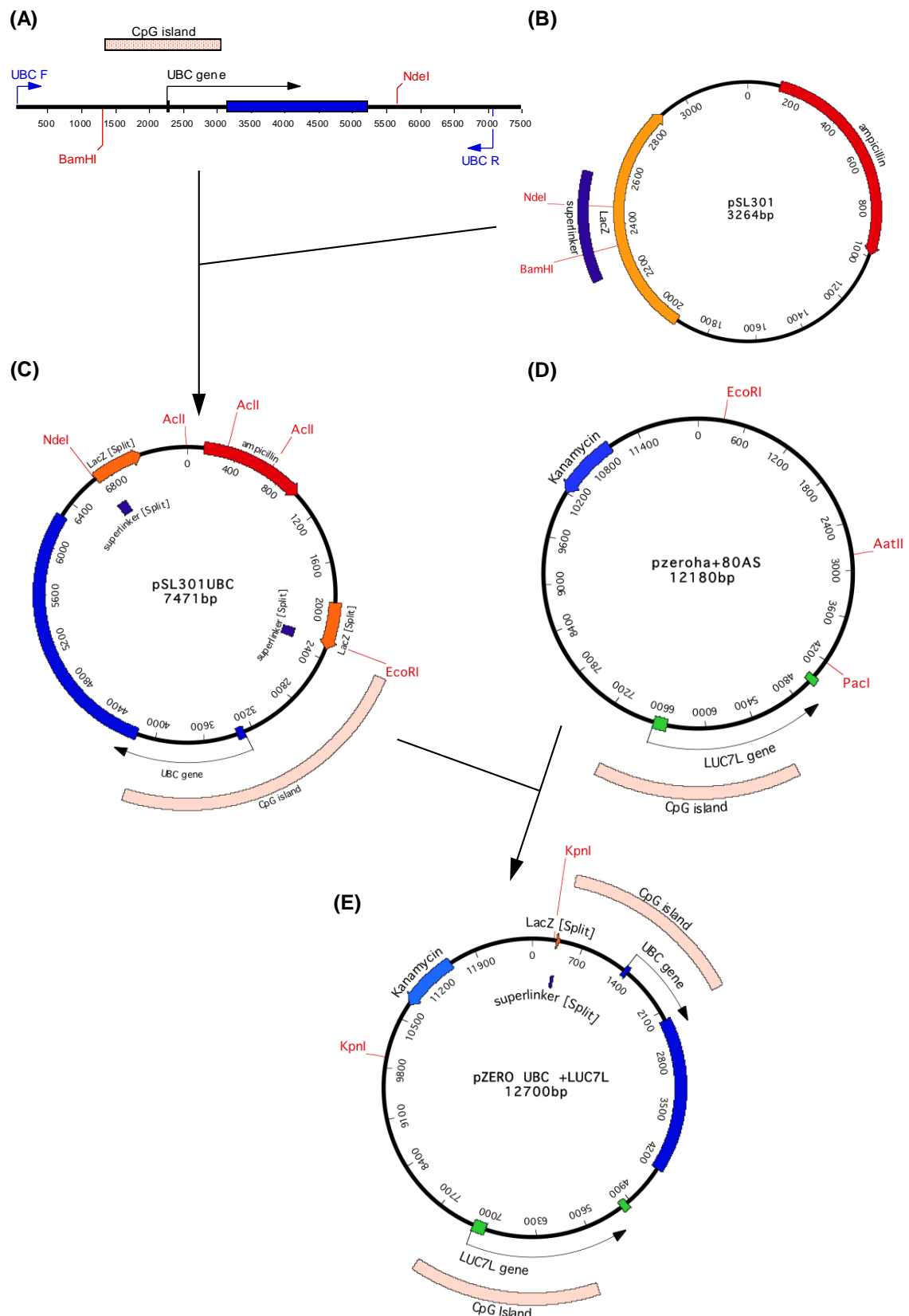


Figure 6.3 Creation of UBC construct (A) human UBC gene cloning strategy and digestion for ligation into pSL301(B) to create pSL301UBC (C). The plasmid pzeroha+80AS had the human alpha gene removed by restriction with EcoRI and Pac1 (D), then the UBC gene was ligated to create the completed construct (E).

6.2.2 Electroporation of Constructs into ESCs and screening for successful incorporation

The techniques used in the purification of DNA constructs from their plasmid backbones, their integration into mESCs by electroporation using the BioRad Gene Pulser Xcell, growth of cells following electroporation, and extraction of DNA are described in section 2.5.6 of chapter 2.

6.2.2.1 Electroporation of the *ACTB* construct into mESCs

The *ACTB* construct was excised from pSL301*ACTB+LUC7L* by double digest with EcoRI and SnaB1, as shown in Figure 6.2E. The desired 9422bp fragment was separated from the 3088bp fragment by gel electrophoresis and the construct purified by band excision and electroelution. Purified construct and the geneticin (G418) resistance gene fragment was then electroporated (2.5.6.2). After 8 days, plates contained on average 20 colonies per plate. 24 colonies were selected based upon their undifferentiated morphology, and cultured until there were sufficient cells to store an aliquot in LN₂ (2.5.4), as well as extract DNA for analysis (2.5.7). 10µg of DNA was digested with Pst1 and digested samples run out on a 0.8% TBE gel before being transferred to a nylon membrane for Southern blotting (2.10). Approximately 35ng of the 1252bp product of the primer pair CT29/CT20 was labeled by random priming (2.10.4) and used to probe for the presence of *LUC7L*. Membrane hybridization was carried out at 58°C. The Southern blot is shown in Figure 6.4, Cell lines 1,4,6,8,9,17, and 18 were selected as clear positives for the construct.

6.2.2.2 Electroporation of the *UBC* construct into ESCs

The *UBC* construct was excised from the pZero*UBC+LUC7L* by restriction digest with Kpn1, as shown in Figure 6.3E. The 9654bp fragment bearing the construct was purified from the remaining 3046bp fragment by electroelution. 20µg of purified construct and 2µg of G418 resistance gene fragment were electroporated. After 8 days, the plates contained a total of 24 colonies, all were transferred to a 24 well plate to be expanded. 14 cell lines showed good cell morphology and growth and were expanded into 2 wells of a 24 well plate. One well of cells was stored in LN₂ (2.5.4), while the other was used to extract DNA to screen for incorporation of the construct by PCR. This is shown in Figure 6.5. Cell lines U3, U4, U7, U10, U13, and U17 all tested positive.

6.2.3 Analysis of cell lines

Five cell lines that had tested positive for each construct were cultured and expanded in tissue culture. ESC material was collected from all cell lines for RNA and DNA analysis, and crosslinked pellets for ChIP for two cell lines were prepared. ESCs were differentiated into EBs over a period of seven to eight days, and more material prepared for RNA, DNA, and chromatin analysis. Preparation of DNA, RNA and ChIP pellets is described in 2.5.7, 2.5.8, and 2.8.1/2.8.2 respectfully. Differentiation of ESCs into EBs is described in 2.5.5

Expression of antisense RNA was tested for in all cell lines by strand specific RT-PCR (2.7). The levels of the pluripotency factors *Oct4* and *Nanog* were also examined in ESCs and EBs as markers for successful differentiation of cells, this was carried out by multiplex PCR with primer pairs for *Oct4*, *Nanog* and *Aprt*. The PCR products were subject to gel electrophoresis and quantified

using gel analysis software (2.6.5). DNA methylation was assessed by msPCR using primers that spanned various regions of the gene's promoter that contained methylation sensitive HpaII sites, as described in 2.6.4. Gene expression levels was initially assessed by standard PCR, then by SYBR green real-time PCR, as detailed in 2.6 and 2.6.6.3 respectively. Histone modification changes between ESCs and EBs was then examined using the protocol described in 2.8.

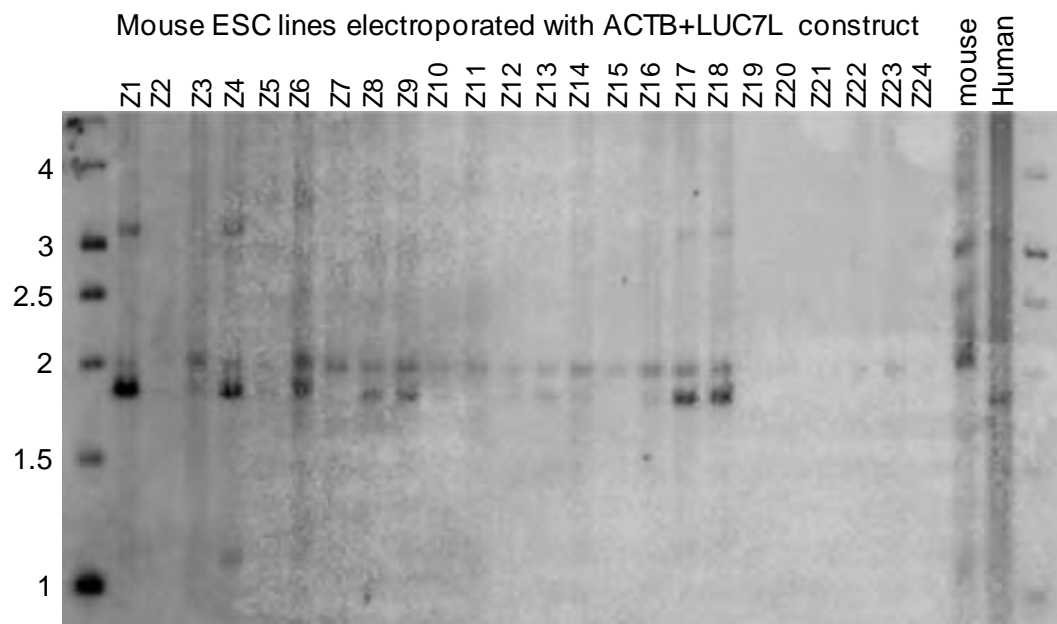


Figure 6.4 Screening of ESC lines for incorporation of the ACTB construct. Pst1 digest of DNA produces a 1.84kb fragment for human LUC7L, and a 1.88Kb fragment for mouse Luc7L. The 1252bp product of the PCR primer pair CT29/CT20 was used as a probe. Cell lines 1,4,6,8,9,17, and 18 are clear positives. Ladder sizes are in Kb.

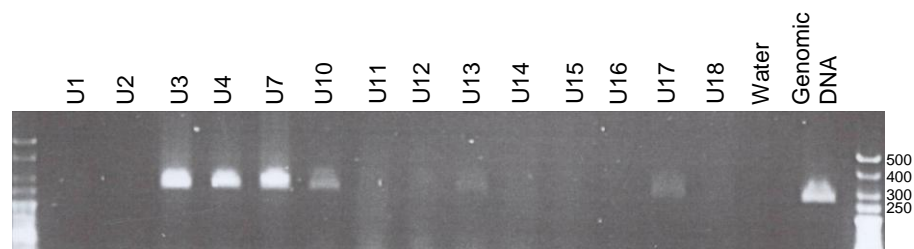


Figure 6.5 Screening of ESC lines for incorporation of the UBC construct. PCR screening of the 14 cell lines that grew well following colony selection using UBC msPCR primers, with product size 255bp amplifying a region of the human *UBC* CpG island. Cell lines U3, U4, U7, U10, U13, and U17 all tested positive.

6.3 Results

6.3.1 Analysis of ACTB cell lines

As has been previously described, transcriptional repression involving both DNA methylation and the histone modification H3K9me3, occurs between day three and day four of differentiation. The level of differentiation within the EB population was assessed by examining the transcription levels of the pluripotency factors *Oct4* and *Nanog* in comparison to the ubiquitously expressed gene *Aprt* in a multiplex PCR (Figure 6.6). Z4, Z6, Z8, and Z18 all demonstrated a marked decrease in expression levels of both of these factors, indicating a high level of differentiation within these EB cultures. Cell line Z9 shows a decrease in the levels of *Oct4*, but no discernable decrease in the expression levels of *Nanog* indicating that not all the ESCs in the differentiation culture had differentiated, and that some remained undifferentiated and expressing *Nanog* - and to a lesser extent *Oct4* - pluripotency factors.

To confirm expression of antisense RNA transcribed from the *LUC7L* promoter, strand-specific RT-PCR was carried out. The antisense-specific primer ACTB RT 2 was used to prime synthesis of cDNA from within the *ACTB* promoter CpG island as shown in Figure 6.7A, while the primer mAPRT1 was used as a positive control for the reaction to prime cDNA synthesis from mRNA of the *Aprt* gene. PCR was carried out using the primer pairs ACTBmsPCR F and R, and mAPRT2 and mAPRT3 for the positive control. Results of the PCR are shown in Figure 6.7B. All Five *ACTB* cell lines expressed antisense RNA in both ESCs and EBs.

To examine whether the presence of antisense RNA, driven by the *LUC7L* promoter, was capable of initiating gene silencing of the *ACTB* gene, gene expression was initially examined by standard PCR. Primers were designed and tested for specificity for human *ACTB* and for mouse *ACTB* genes, and subsequently used in PCR reactions using random-primed cDNA. Primers for Human *ACTB* were designed within exon 5 of the *ACTB* gene, as shown in Figure 6.8A. While 11 of 19 known splice-forms include exon 5, this includes all major splice-forms observed and as such these primers represent over 97% of detectable *ACTB* transcripts and therefore provide an accurate portrayal of the transcriptional activity of this gene. When a 28 cycle PCR was carried out for human *ACTB* expression, there was no observable difference between expression levels in ESCs and EBs when compared to the expression of mouse *ACTB* and the ubiquitously expressed gene *Aprt*. (Figure 6.8).

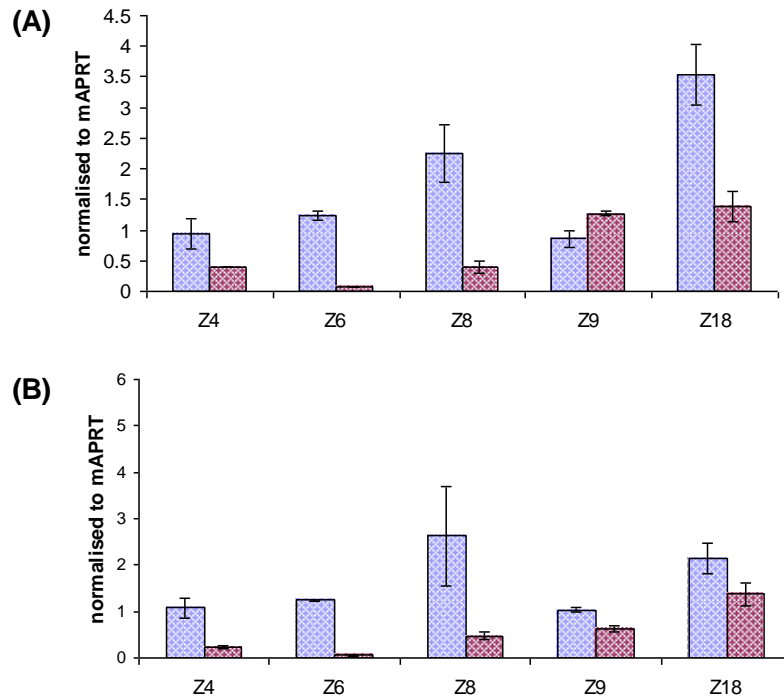


Figure 6.6 Expression of pluripotency factors. Relative expression of *Nanog* (A) and *Oct4* (B) in ESCs (blue bars) and d8 EBs (red bars).

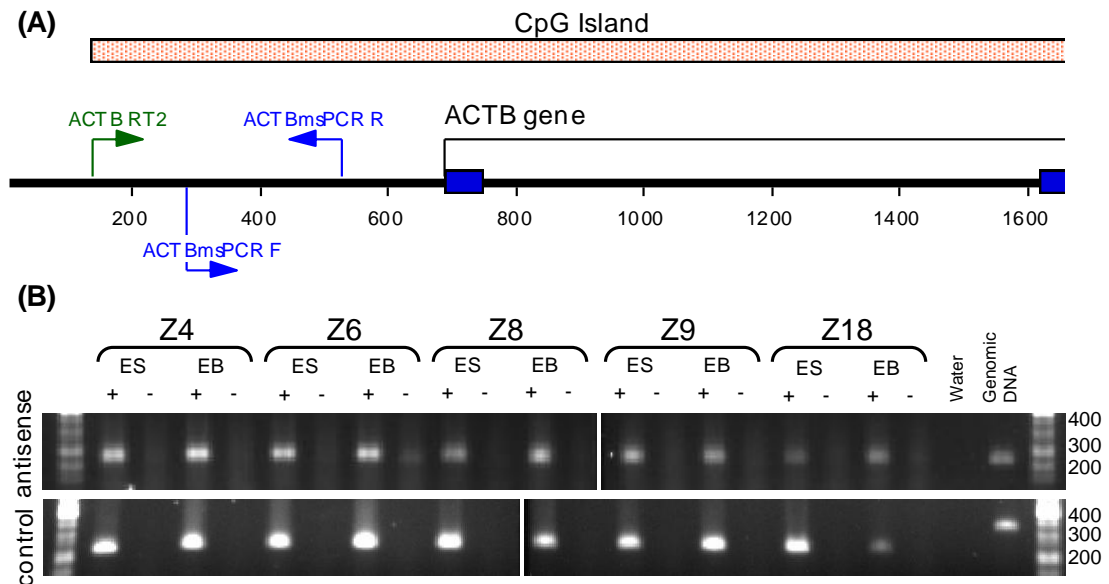


Figure 6.7 Expression of antisense RNA. (A) schematic representation of the ACTB construct with primer locations marked. (B) The primers ACTB RT2 and mAPRT1 were used to prime a strand-specific RT-PCR reaction. PCR amplification of antisense RNA by ACTB msPCR F and ACTB msPCR R (225bp), amplification of control using mAPRT2 and mAPRT3 (225bp cDNA, 335bp genomic DNA). Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA.

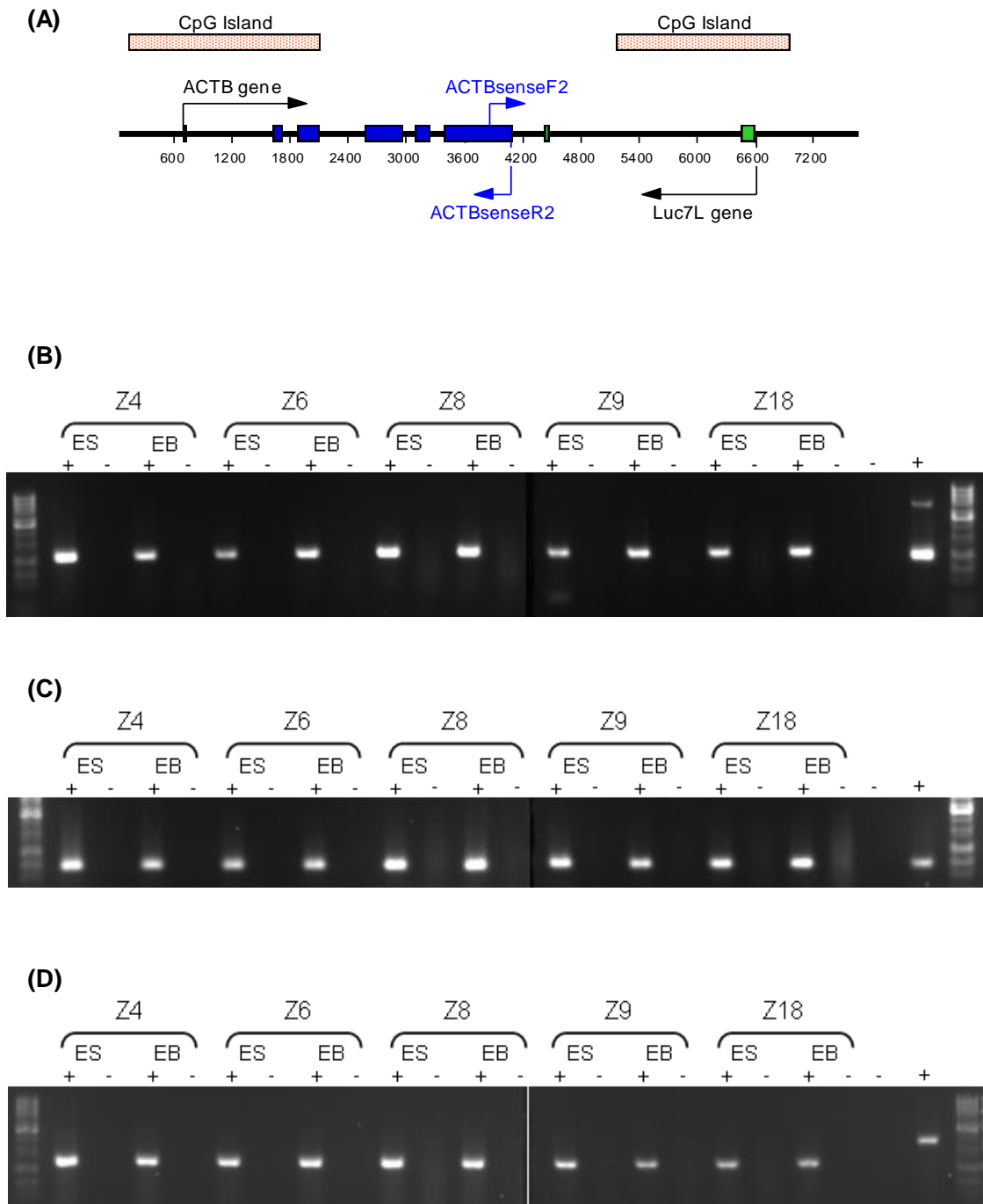


Figure 6.8 Gene expression in the ACTB cell lines analysed by PCR. (A) Schematic diagram of the ACTB construct with the position of the primers indicated. (B) expression of human ACTB. (C) expression of mouse *Actb*. (D) expression of mouse *Aprt*. RT-PCR reactions were primed using random hexamers. Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA. 28 PCR cycles were used for mouse and human ACTB, 30 cycles for mouse *aprt*.

Expression levels of human *ACTB*, mouse *Actb*, and *Aprt* were then examined by real-time PCR, using the same primers as used for normal PCR. This revealed that for the cell lines Z6, Z8, Z9, and Z18 *ACTB* gene expression, compared to *Aprt*, remained either steady or increased slightly as ESCs were differentiated into EBs, this is shown in 6.9A. Expression of *ACTB* was compared to expression of mouse *ACTB*. Again, for the cell lines Z6, Z8, Z9, and Z18 *ACTB* gene expression remained either steady or increased as shown in figure 6.9C. In stark contrast to the other cell lines examined, Z4 showed decreasing levels of human *ACTB* expression comparing ESCs to EBs, when normalized to either *Aprt* or to mouse *Actb*. Interestingly, while expression levels of human *ACTB* was high in the Z4 cell line compared to other cell lines, this was mirrored by an equally high expression of mouse *ACTB* within the Z4 cell line. This resulted in the levels of human *ACTB* being among the lowest observed in both ESCs and EBs, when compared to mouse *ACTB* levels.

Methylation sensitive PCR was carried out to determine the methylation status of the *ACTB* CpG Island. Initially, a region of 225bp encompassing three HpaII sites was amplified by the PCR primers ACTBmsPCRF and ACTBmsPCRR. This is shown in Figure 6.10A and B. No methylation was observed in the cell lines Z6, Z8, Z9 or Z18. The positive control for methylation M114 M115 amplified correctly, and no methylation was observed at the regions amplified by M116 M117, a region of the *Aprt* CpG island that does not become methylated, or within the *LUC7L* CpG Island amplified by the primer pair CT53 CT54. The cell line Z4 showed some DNA methylation in the cell population in ESCs within the 225bp region, with a higher level of DNA methylation

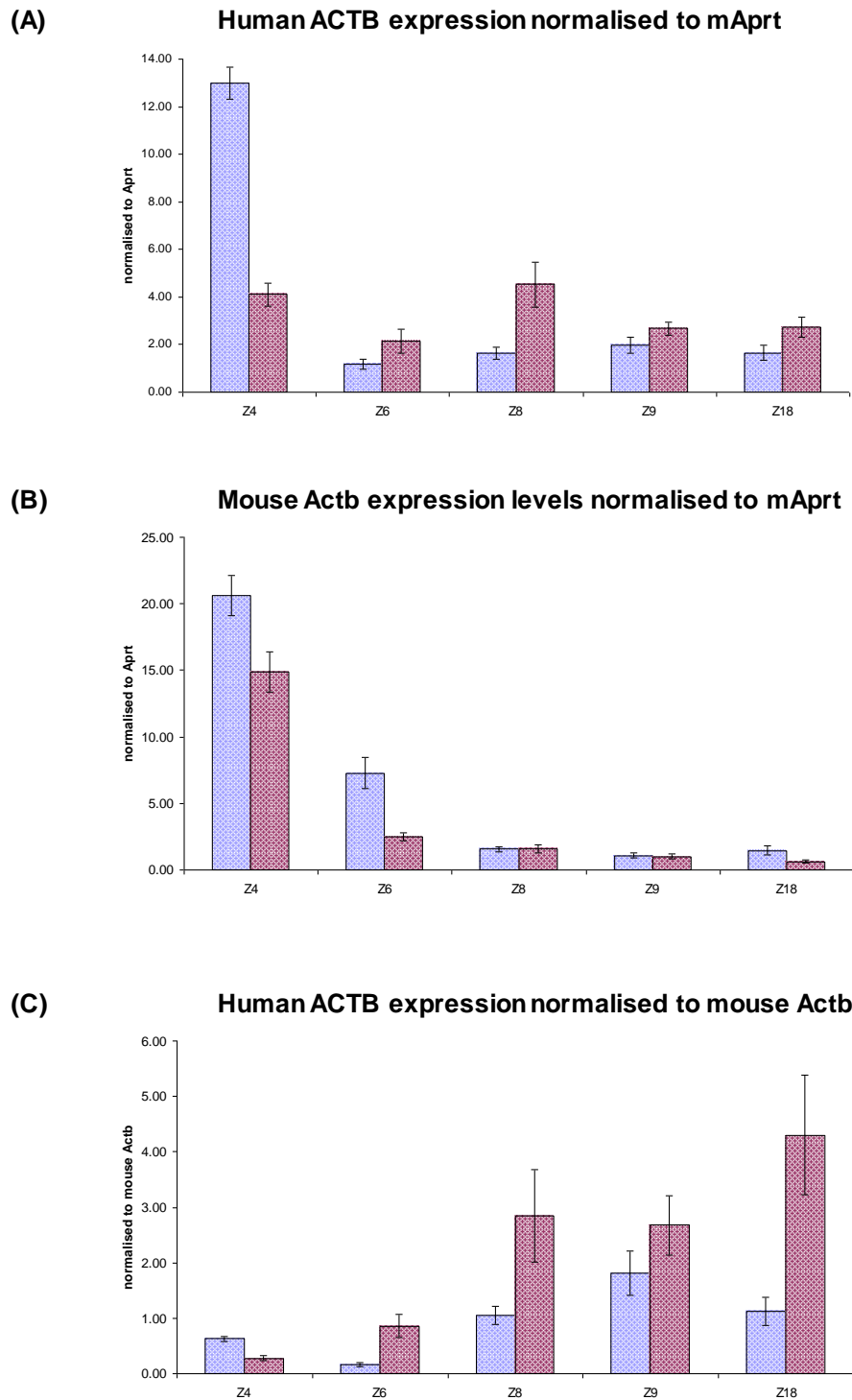


Figure 6.9 Gene expression in the ACTB cell lines quantified by realtime PCR. (A) shows Human ACTB expression normalised to mAprt. (B) shows Mouse Actb expression levels normalised to mAprt. (C) shows Human ACTB expression normalised to mouse Actb. The position of the primers used to detect human ACTB are shown in figure 3.7

detectable in the EB population. DNA methylation is also detectable within the *LUC7L* CpG Island in EBs, while M116 M117, the control for complete digestion, show no product indicating that digestion ran to completion. A larger methylation sensitive PCR assay was also carried out, with a product of 538bp and encompassing 11 HpaII sites (Figure 6.10A and B). As with the smaller methylation assay, no product was detected for the cell lines Z6, Z8, Z9 or Z18, however a PCR product in ESCs and EBs could be detected for the cell line Z4, indicating DNA methylation of all 11 CpG sites within this region.

Chromatin immunoprecipitation was carried out on material from the cell lines Z4 and Z6. The histone modifications H3K9me3, H3K27me3, H3K4me3, and H3Ac were examined at three locations by real-time PCR across the body of the *ACTB* gene over 1Kb apart, and also within the *LUC7L* CpG island. Results for Z4 are shown in Figure 6.11, and results for the cell line Z6 are shown in Figure 6.12. Sonication efficiency was tested as detailed in 2.9.5, and analysis of sonication efficiency is shown in Appendix B.

The Z4 cell line shows enriched levels of H3K9me3 in ESCs and EBs, with levels highest within the body of the gene as shown by real-time primer positions *ACTB2* and *ACTB3*. There was an increase in levels of H3K9me3 as ESCs were differentiated to EBs for the real-time primer position *ACTB1* upstream of the TSS of the gene from 1.5x to 2x, and a similar change observable within the *LUC7L* CpG island.

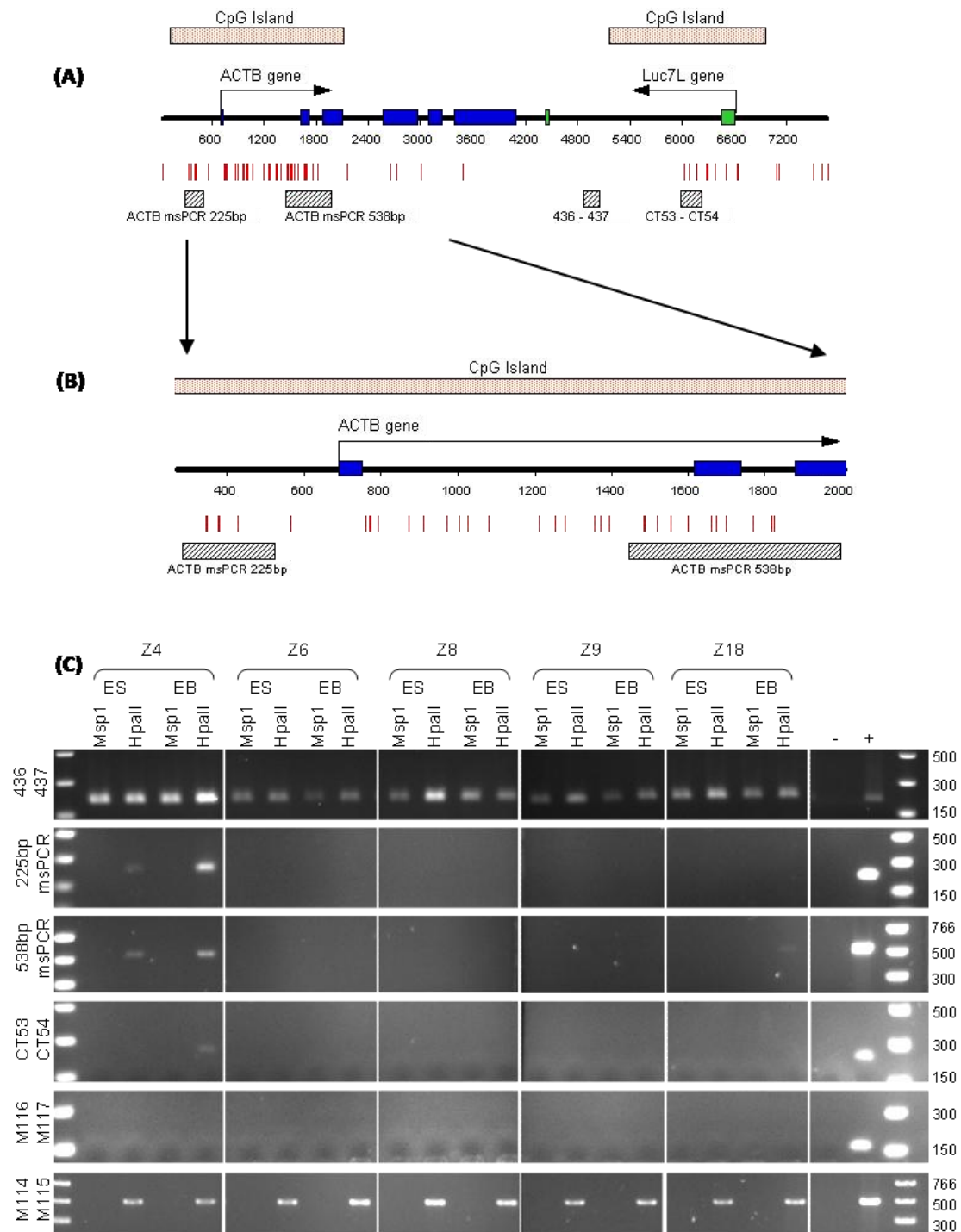


Figure 6.10 Methylation sensitive PCR within the ACTB CpG Island. **(A)** Schematic diagram of the ACTB construct with Msp1/HpaII sites marked as vertical red lines and regions amplified by PCR primers represented as hatched grey boxes. Within the LUC7L CpG island; CT53 CT54 amplify across 3 restriction sites (248bp), and 436 437 primers amplify a region with no restriction sites (208bp). **(B)** shows a enlarged view of the ACTB CpG island. The 225bp PCR product encompasses 3 restriction sites, and the 538bp sites encompasses 11 restriction sites. **(C)** M116 M117 amplify a region within the 5' end of the APRT gene that contains 4 restriction sites and is devoid of DNA methylation (164bp), M114 M115 primers amplify within the promoter of AIR, which is stably methylated on the maternal copy in ESCs and EBs (489bp). Water (-) and genomic DNA (+) were used as PCR controls.

H3K27me3 levels barely altered between ESCs and EBs, though were enriched across the *ACTB* gene and, as with the histone mark H3K9me3, this was most noticeable within the body of the gene at positions *ACTB2* and *ACTB3*. H3K4me3 levels were low in ESCs across the *ACTB* gene, but became enriched to the level of the mouse *ACTB* control upon differentiation into EBs. For all four regions examined, enrichment for this mark more than doubled upon differentiation into EBs. H3Ac levels were very low across the *ACTB* gene and did not alter upon differentiation.

Histone modifications observed within the *ACTB* gene for the cell line Z6 were in stark contrast to the results obtained for the cell line Z4. Levels of enrichment for both H3K9me3 and H3K27me3 remained low, comparable to the levels of enrichment observable at the mouse *Actb* gene.

H4K4me3 and H3Ac levels were both high around the TSS of the *ACTB* gene at positions *ACTB1* and *ACTB2* in ESCs, and this region continued to be enriched for H3K4me3 in EBs. At position *ACTB3* within the last exon of the *ACTB* gene, and within the *LUC7L* CpG island, H3K4me3 levels increased modestly in EBs.

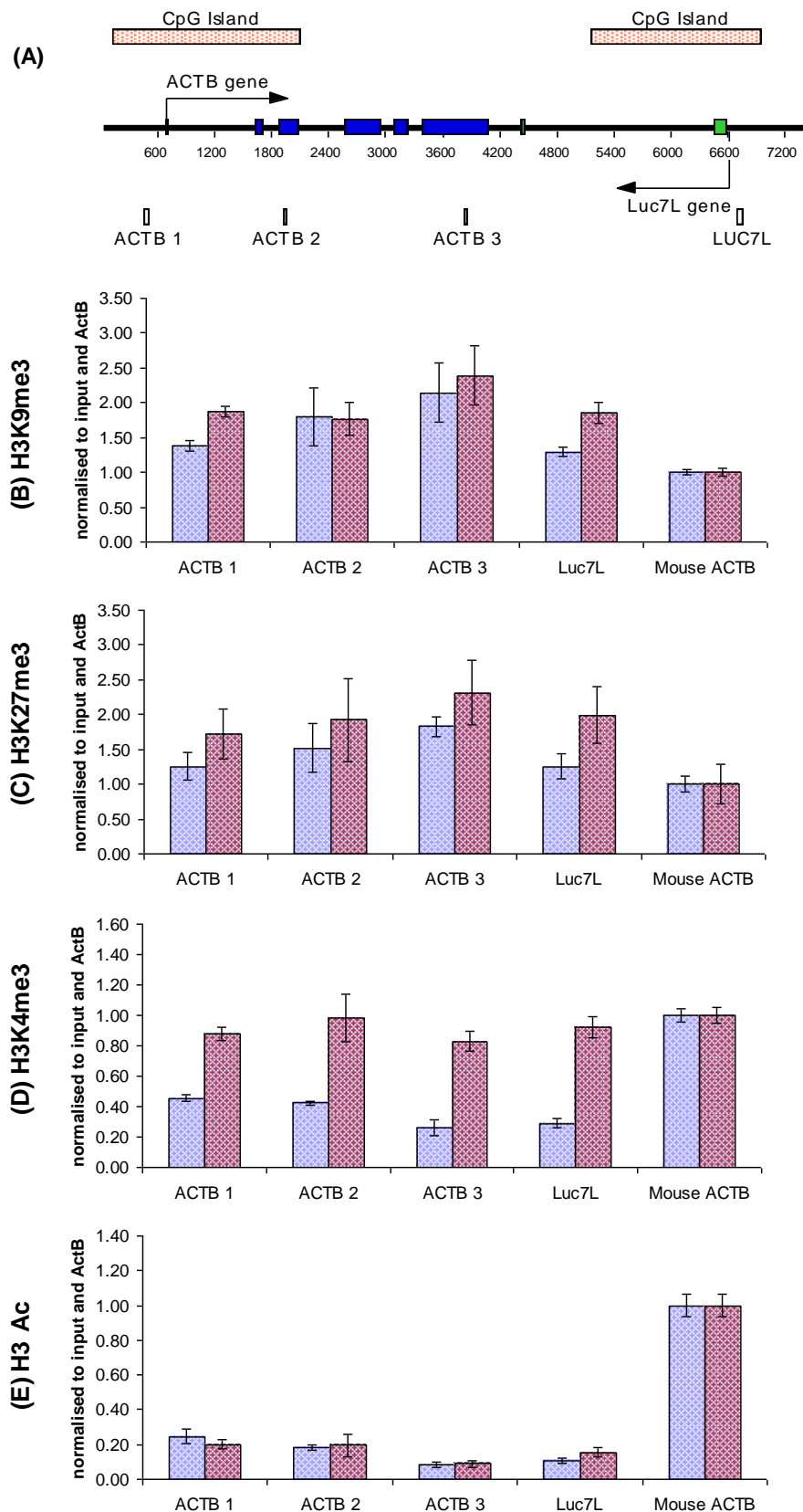


Figure 6.11 Histone modifications to the ACTB construct in the Z4 cell line. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *ACTB* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

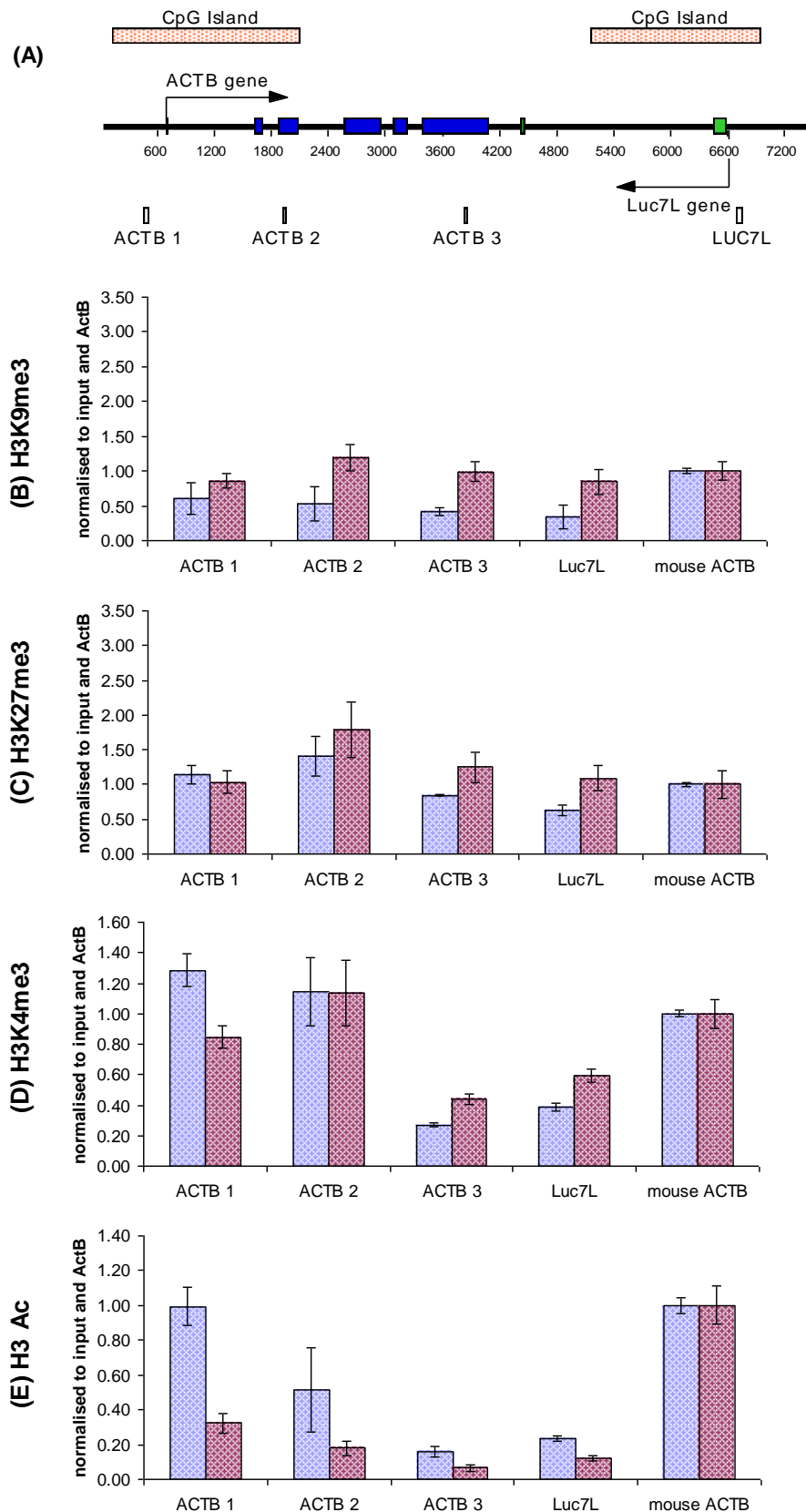


Figure 6.12 Histone modifications to the ACTB construct in the Z6 cell line. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *ACTB* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

6.3.2 Analysis of *UBC* cell lines

Expression levels of *Oct4* and *Nanog* in ESCs and EBs for *UBC* cell lines is shown in Figure 6.13. All five cell lines showed a decrease in expression of *Nanog* following differentiation, while this drop in expression levels was large for the cell lines U4, U10, and U17 – indicating high levels of differentiation from a mostly ESC population - the expression levels in ESCs observed in cell lines U3 and U7 was lower than the other cell lines indicating a higher level of spontaneously differentiated ESCs in these cell lines. A very similar pattern of expression was observed for *Oct4*, with the exception of U17 that appeared to have elevated levels of *Oct4* in EBs.

Strand-specific RT-PCR was carried out to confirm expression of antisense RNA transcribed from the *LUC7L* promoter. The antisense-specific primer *UBC* RT was used to prime synthesis of cDNA from within the *UBC* promoter CpG island as shown in Figure 6.14A, while the primer mAPRT1 was used as a positive control for the reaction to prime cDNA synthesis from mRNA of the Aprt gene. PCR was carried out using the primer pairs *UBCi1* and *UBCi2* (375bp), and mAPRT2 and mAPRT3 (225bp cDNA, 335bp genomic DNA) for the positive control. Results of the PCR are shown in Figure 6.14B. All Five *UBC* cell lines express antisense RNA in both ESCs and EBs.

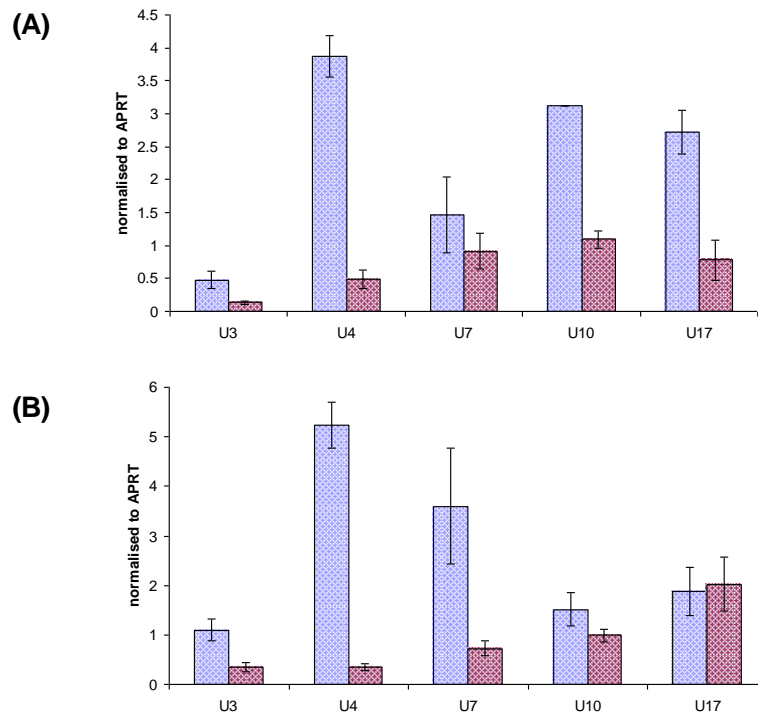


Figure 6.13 Expression of pluripotency factors. Relative expression of *Nanog* (A) and *Oct4* (B) in ESCs (blue bars) and d8 EBs (red bars).

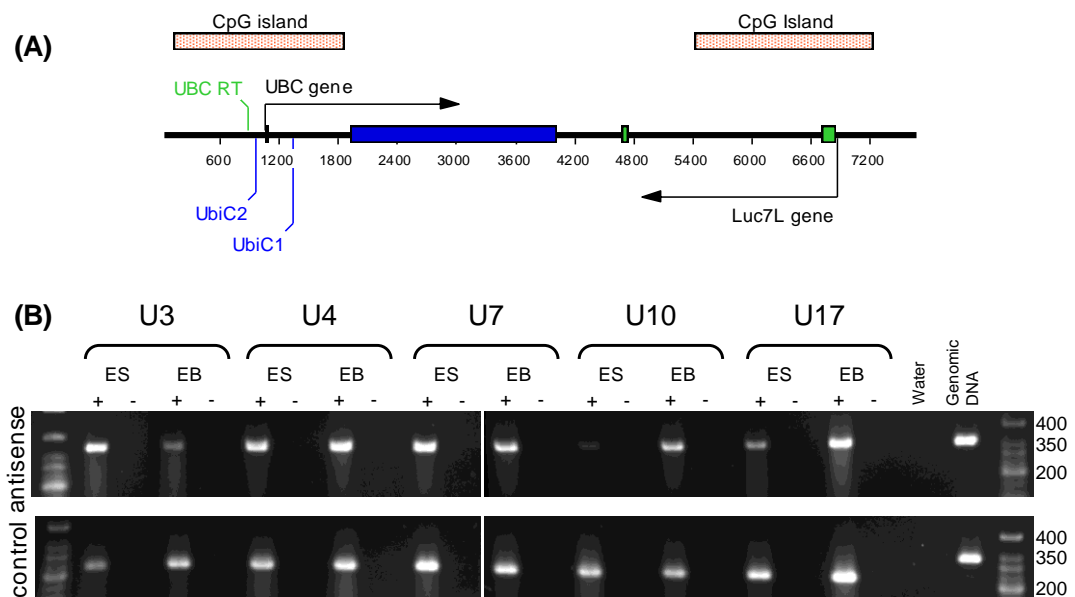


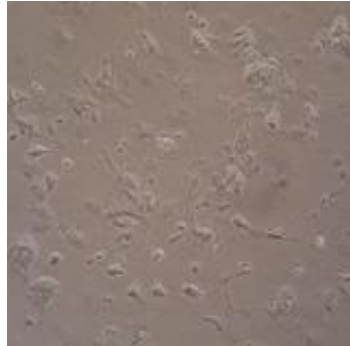
Figure 6.14 Expression of antisense RNA. (A) schematic representation of the UBC construct with primer locations marked. (B) The primers UBC RT and mAPRT1 were used to prime a strand-specific RT-PCR reaction. PCR amplification of antisense RNA was by the primers UBCi1 and UBCi2 (375bp), amplification of control used the mAPRT2 and mAPRT3 primer pair (225bp cDNA, 335bp genomic DNA). Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA.

The five cell lines U3, U4, U7, U10, and U17 were cultured for more than a week prior to differentiation into EBs. When the cell lines were initially revived after storage in LN₂ it was observed that the cell lines U3, U7, and U10 showed a greater proportion of ESC differentiation within the culture than was normal for a healthy ESC line (Figure 6.15A). To discourage growth of differentiated cells and encourage the proliferation of undifferentiated ESCs, the cell lines were subject to passaging every day, including disaggregation with an 18G needle in addition to treatment with 0.25% trypsin EDTA. As shown in Figure 6.15B, after at least five passages, cell morphology within the ESC lines had improved, though there was still a greater proportion of differentiated cells than normal observable within the U3 and U7 cell lines.

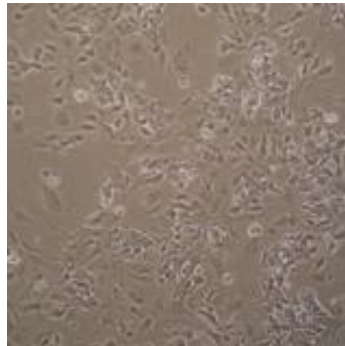
To investigate whether the presence of antisense RNA, driven by the *LUC7L* promoter, could cause gene silencing of the *UBC* gene, levels of sense expression were initially examined by standard PCR on random primed cDNA. Primers were designed and tested for specificity for the human *UBC* gene. The *UBC* gene coding region comprises just two exons, the first exon is small while exon 2 covers more than 2Kb. A primer pair was designed that amplified from exon 1 to exon 2, with a product size of 158bp on cDNA but 970bp on genomic DNA. (Figure 6.16A).

Standard PCR analysis detected *UBC* transcripts in both ESCs and EBs (6.16B), though band intensity was suggestive of a decrease in expression in EBs, when levels were compared to the Aprt controls (Figure 6.16C).

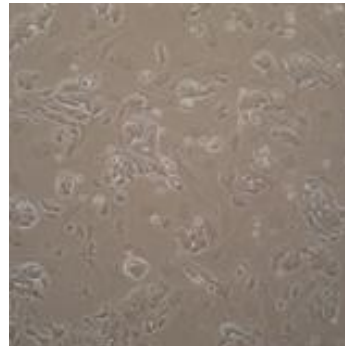
(A) ESC morphology growing in culture



U3



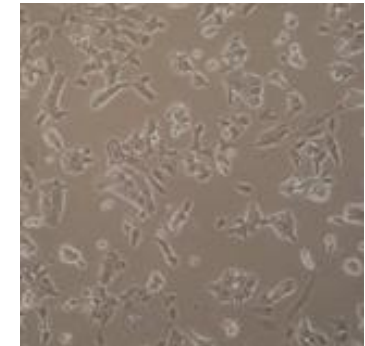
U4



U7

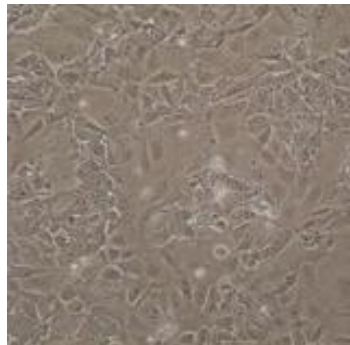


U10

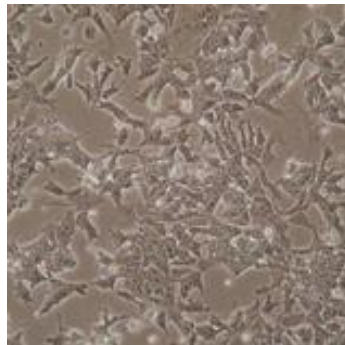


U17

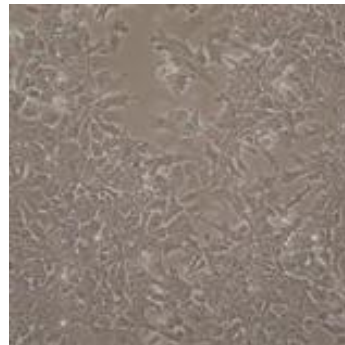
(B) ESC morphology after five or more passages



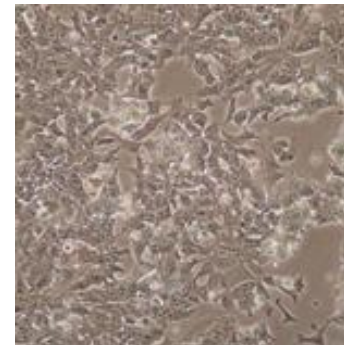
U3



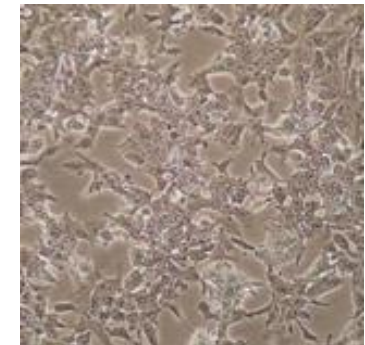
U4



U7



U10



U17

Figure 6.15 UBC ESC line morphology prior to differentiation into embryo bodies. (A) shows cells after they have been recovered from liquid nitrogen and passaged for several days. (B) shows cell morphology after an 8 day period, with cells being passaged most days using 0.25% trypsin EDTA and an 18G needle to ensure single cell suspension for replating.

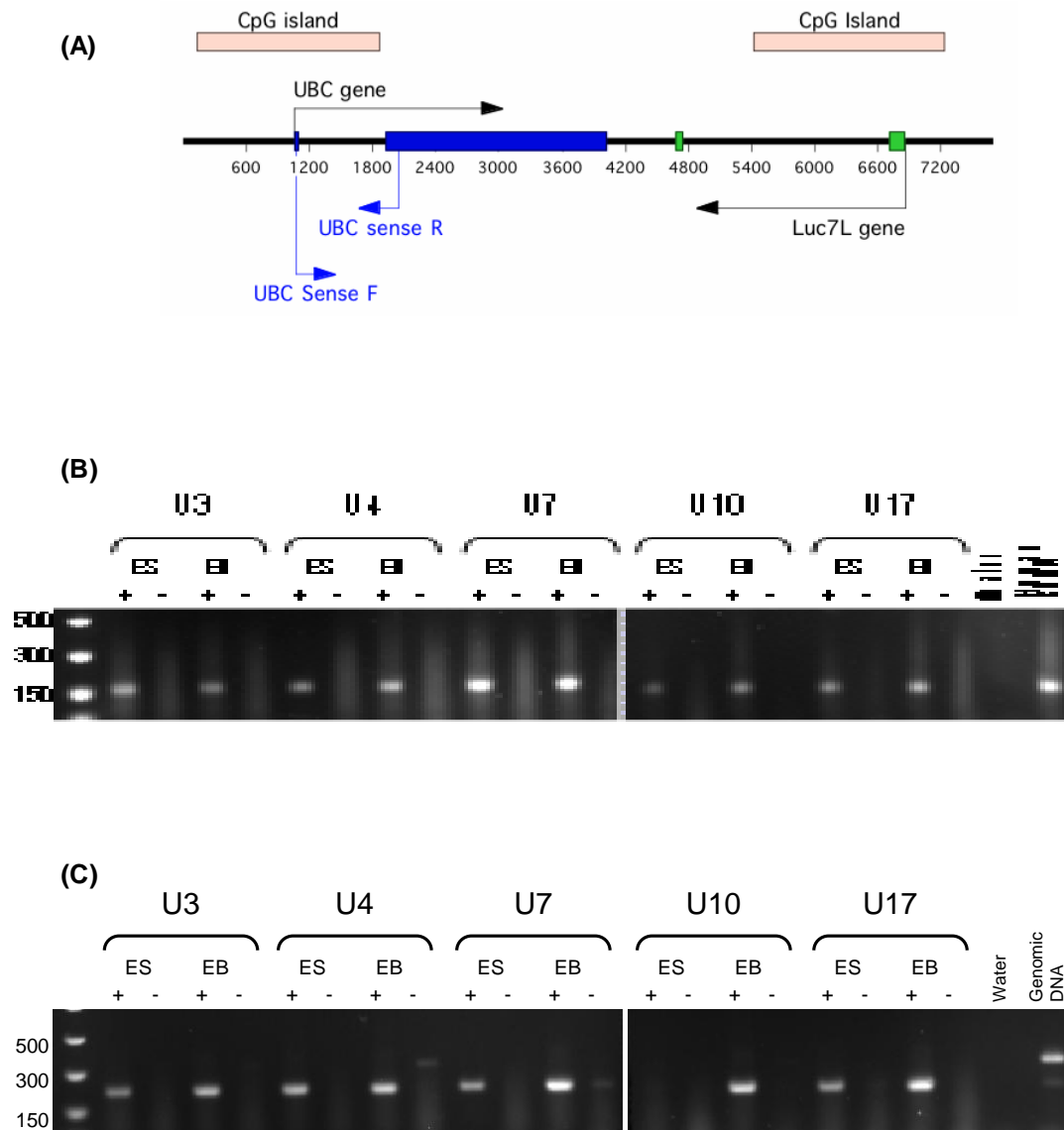


Figure 6.16 Gene expression in the UBC cell lines analysed by PCR. (A) Schematic representing the UBC construct and the position of the primers, amplifying a region of 158bp on cDNA. (B) expression of human UBC. (C) expression of mouse Aprt. RT-PCR reactions were primed using random hexamers. Negative controls (-) are reverse transcription reactions set up without the addition of the reverse transcriptase enzyme as controls for contaminating genomic DNA.

Gene expression of human *UBC*, mouse *Ubc*, and *Aprt* was then examined by real-time PCR to accurately quantify expression changes between ESCs and EBs. For Human *UBC* expression, the same primers that had been designed for normal PCR were used for real-time PCR. New primers were designed for mouse *Ubc*, as primers designed for standard PCR showed non-specificity when used with the SYBR green real-time PCR master mix.

Human *UBC* levels decreased in all five cell lines upon differentiation of ESCs into EBs (Figure 6.17A). Expression of *UBC* in cell line U4 was high in ESCs, but this cell line also showed high expression levels of mouse *Ubc* in ESCs and decreased in EBs (Figure 6.17B). For the cell lines U3, U7, U10, and U17 mouse *UBC* expression altered very little from ESCs to EBs. When expression of human *UBC* in the five cell lines was compared to expression of mouse *Ubc*, the trend of the data did not alter, with decreased expression upon differentiation into EBs in all cell lines clearly indicating that the *UBC* gene is becoming silenced in this context (Figure 6.17C).

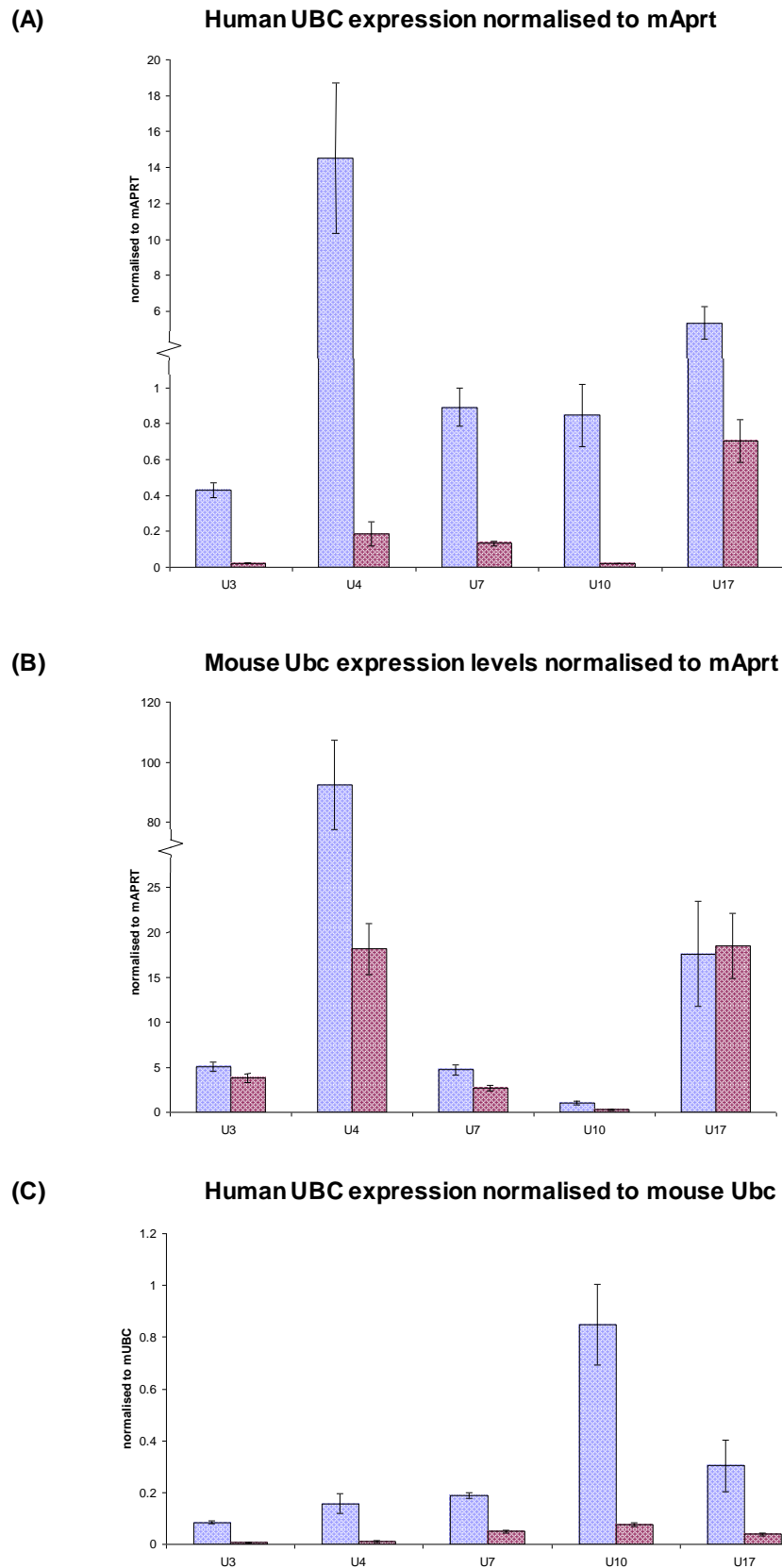


Figure 6.17 Gene expression in the UBC cell lines quantified by realtime PCR. (A) shows human *UBC* expression normalised to mAprt. (B) shows mouse *Ubc* expression levels normalised to mAprt. (C) shows human *UBC* expression normalised to mouse *Actb*. The position of the primers used to detect human *UBC* are shown in Figure 3.7. ESC samples are in blue, EB samples are in red.

Methylation sensitive PCR was carried out to investigate the methylation status of the *UBC* CpG island within the region of the genes TSS. The primer pair UBCmsPCR F and R was used to amplify a region containing 4 HpaII cut sites (figure 6.18A). A PCR product was detectable in the HpaII digest for all five cell lines in EBs, indicating that the *UBC* CpG island has become methylated in this region upon differentiation, this is shown in figure 6.18B. The cell lines U3 and U7 also showed some DNA methylation in ESC samples. As previously mentioned, these cell lines had a higher proportion of differentiated cells within the ESC culture, and the detection of DNA methylation is not unexpected in the partially differentiated ESC samples since it was detectable in EBs. The faintness of the bands in the U3 and U7 ESC HpaII digests clearly suggests that only a proportion of the population had become differentiated. Band intensity increases in EBs as the majority of the cell population had differentiated and the DNA methylation mark had become established. Controls for digestion functioned as expected.

The diagram illustrates the genomic organization of the UBC and Luc7L genes on chromosome 12p11.2. The UBC gene is transcribed from left to right, while the Luc7L gene is transcribed from right to left. CpG islands are located upstream of both genes. The UBCmsPCR region is located within the UBC gene. The 436-437 and CT53-CT54 regions are located downstream of the Luc7L gene.

	U3				U4				U7				U10				U17						
	ES		EB		ES		EB		ES		EB		ES		EB		ES		EB		-	+	
	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII	Msp1	HpaII			
436 437																							500 300 150
255bp msPCR																							766 500 300 150
CT53 CT54																							500 300 150
M116 M117																							300 150 50
M114 M115																							766 500 300

264

Histone modifications in the cell lines U4 and U17 were examined by ChIP for the repressive marks H3K9me3, H3K27me3 and the activating marks H3K4me3 and H3Ac. Real-time PCR primer pairs were designed for three regions of the *UBC* gene (Figure 6.19A). Due to the highly repetitive nature of exon 2 of the *UBC* gene, primer pairs designed in this region bind at more than one location, meaning that the primer pair *UBC3* binds four times in close proximity. The primer pairs *UBC1*, *UBC2*, and the annealing sites for *UBC3* are spaced over 1Kb apart (Figure 6.19A). Sonication efficiency was tested as detailed in 2.9.5, and analysis of sonication efficiency is shown in Appendix B.

Results for cell lines U4 (Figure 6.19), and U17 (Figure 6.20) were very similar. Enrichment for the repressive mark H3K9me3 at least doubled at positions *UBC1* and *UBC2* upon differentiation into EBs, while the cell line U17 showed a similar increase at position *UBC3*, but this was not reproduced clearly in the U4 cell line. An increase in H3K27me3 levels occurred within promoter regions for both cell lines upon differentiation, with enrichment also occurring within the body of the gene, though this was less obvious in the U17 cell line, though the trend was for H3K27me3 enrichment.

H3K4me3 levels were slightly less than those observed at the mouse *ACTB* gene, and did not alter significantly after differentiation in EBs. H3Ac levels were uniformly low across the *UBC* gene and, again, did not alter significantly upon differentiation.

H3K9me3 and H3K27me3 levels at the *LUC7L* CpG island were analogous to those observed at the mouse *ACTB* gene in both ESCs and EBs. H3Ac and H3K4me3 levels were low in both ESCs and EBs.

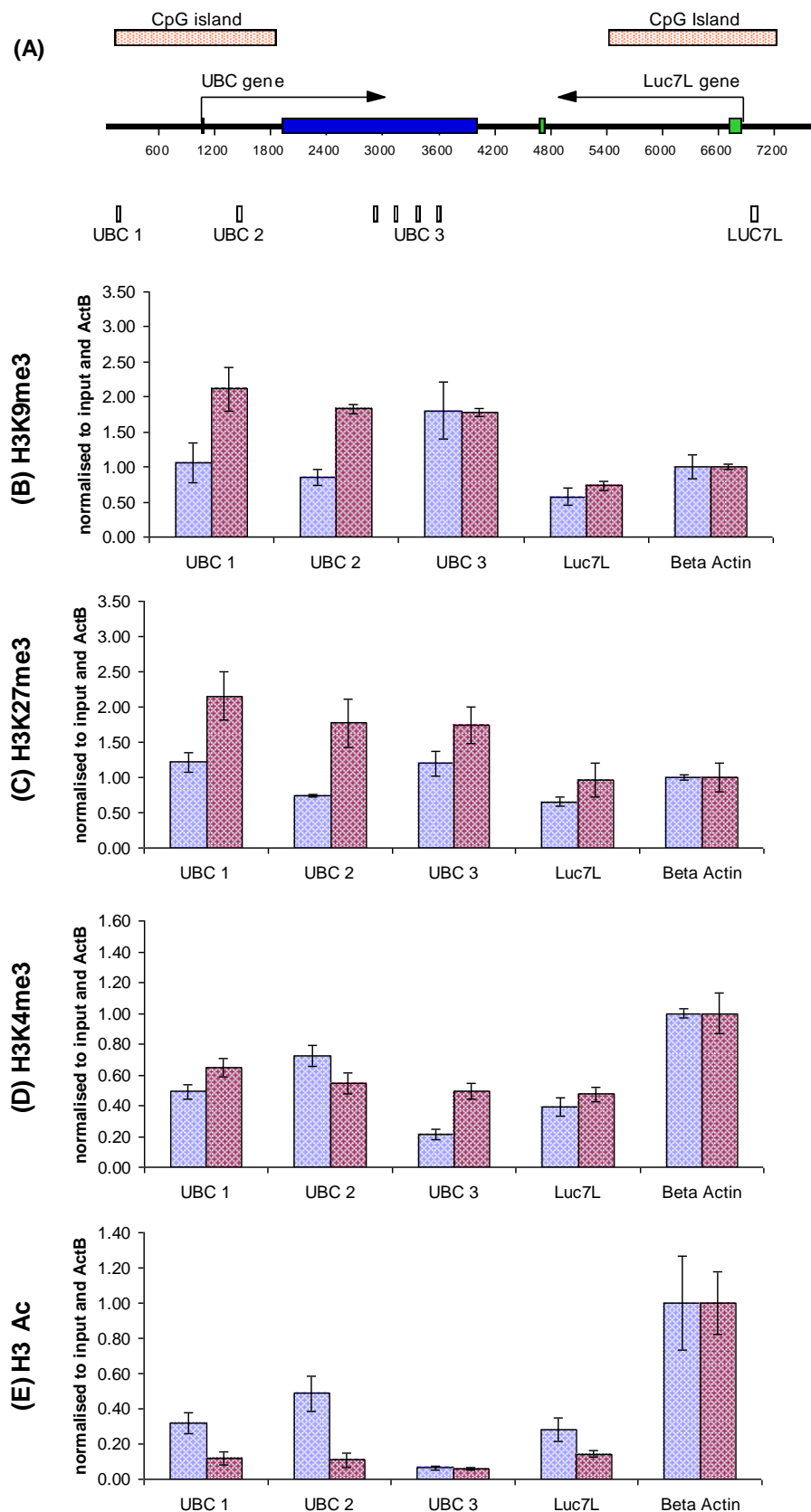


Figure 6.19 Histone modifications to the UBC construct in the U4 cell line. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *UBC* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

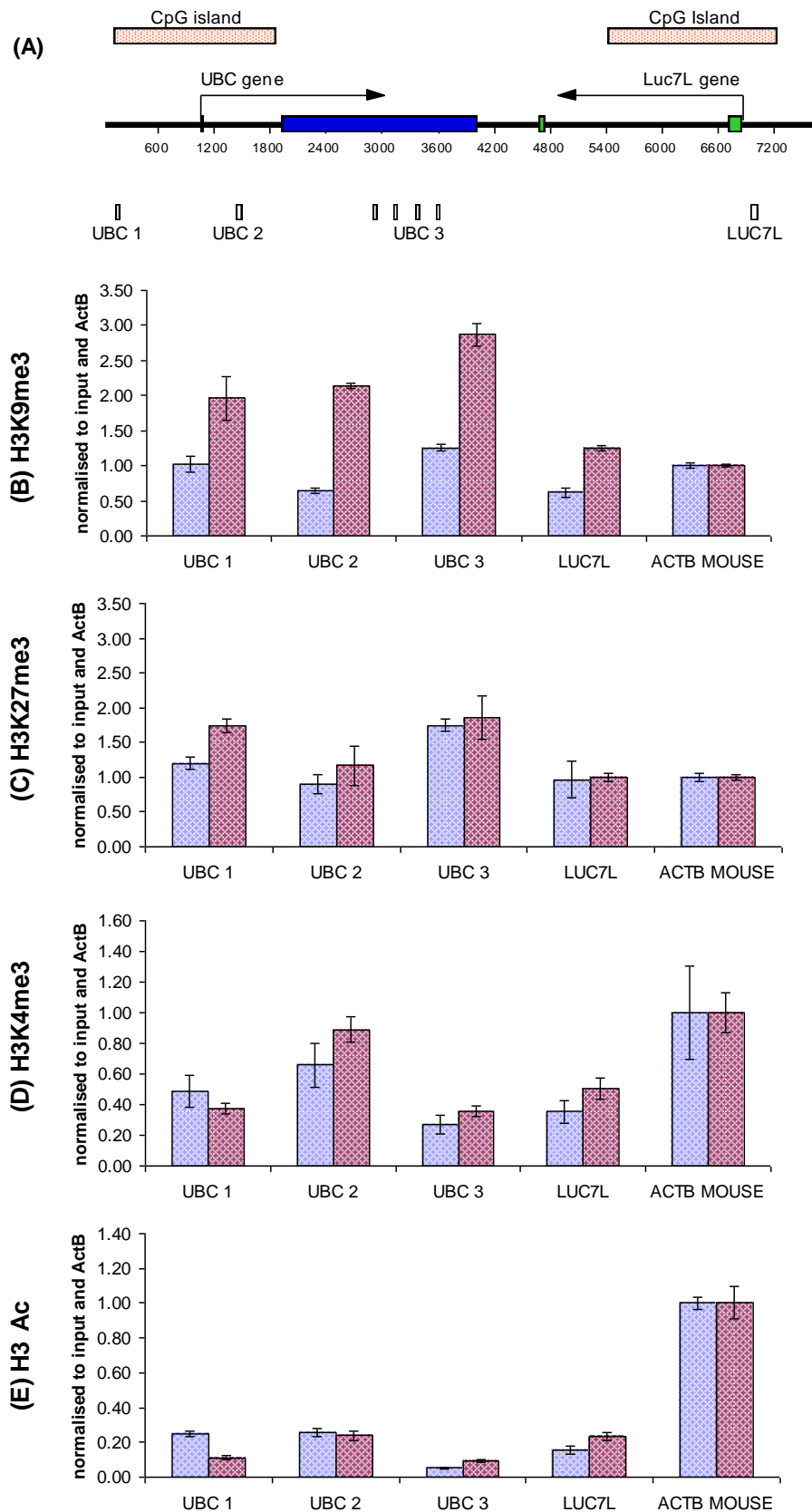


Figure 6.20 Histone modifications to the UBC construct in the U17 cell line. (A) depicts the position of the realtime PCR primers in relation to the promoter and coding regions of the *UBC* gene. Enrichment levels in ESCs (blue bars) and EBs (red bars) for H3K9me3, H3K4me3, H3K37me3 and H3Ac are shown in B, C, D and E respectively.

6.4 Discussion

The results presented in this chapter show that in the presence of antisense RNA, the ubiquitously expressed gene *UBC* becomes repressed in a similar manner as observed for the tissue specific *HBA* and *MYOD* genes. In all five cell lines examined, DNA methylation was detectable within the CpG island upon differentiation. DNA methylation was also detectable in the cell lines U4 and U7 in ESCs, most likely as a result of the higher than normal degree of differentiation that was observed in these ESC cell lines, as was indicated by the reduced expression of *Nanog* in both cell lines, and *Oct4* in the U3 cell line

Enrichment for H3K9me3 was detectable throughout the body of the *UBC* gene upon differentiation in both cell lines analyzed, and with H3K27me3 also shown to increase upon differentiation. Both histone modifications are associated with repression of gene expression, and H3K9me3 enrichment is a key feature associated with repression of the alpha globin gene in the presence of AS RNA. Interestingly, enrichment for H3K4me3 did not appear reduced in EBs despite the presence of repressive histone marks and DNA methylation. Indeed H3K4me3 around the TSS remained approximate to levels observed at the mouse *Actb* gene. Despite the continued presence of this histone mark, the *UBC* gene became repressed in EBs with gene expression being sharply reduced. Acetylation of lysine residues on histone tails is associated with a relaxed open chromatin structure conducive to gene expression, and acetylation reduced around the TSS in EBs, though this was less pronounced in the U17 cell line.

UBC gene expression was shown to occur at a lower level than expression of mouse *Ubc* in ESCs. However, gene expression studies in mouse and humans indicate that human *UBC* expression levels are generally about half that of its mouse counterpart (NCBI Aceveiw gene expression data). Upon differentiation into EBs, human *UBC* expression became clearly repressed in all cell lines, whereas expression of mouse *UBC* was seen to remain fairly constant.

Taken together, the parallel with the AS RNA-dependent silencing effect observed at tissue specific genes is clear, demonstrating that in a differentiating mouse ESC model system a ubiquitously expressed gene can become targeted for repression by DNA methylation and repressive histone modifications. However, it will be important to understand why H3K4me3 is retained at the repressed *UBC* locus, and could reflect its ubiquitous nature.

The *ACTB* gene demonstrated no such clearly defined susceptibility to repression. In four of the five cell lines studied, DNA methylation was not detectable upon differentiation, and gene expression levels were seen to remain at approximately equal levels in ESCs and EBs, or even increased in level upon differentiation, in comparison to endogenous *Actb* expression. The Z4 cell line was a clear exception to this, with a low level of expression in ESCs that became further reduced upon differentiation. Also DNA methylation was detectable across a considerable proportion of the *ACTB* CpG island in ESCs and EBs, with methylation becoming detectable within the CpG island of the normally unmethylated *LUC7L* gene, downstream of the TSS, upon differentiation. Cell morphology was not indicative of a high level of

spontaneously differentiated cells, and expression of *Oct4* and *Nanog* in ESCs was comparable to the other four *ACTB* cell lines analyzed. Taken together with the lack of silencing observed in the other cell lines in either ESCs or EBs, this would suggest that Z4 DNA methylation observed in ESCs was not the result of partial differentiation, but rather indicates a form of constitutive repression in ESCs and EBs, disparate from the developmentally linked repression observed at the alpha globin, *MYOD* and *UBC* genes.

It is possible for a proportion of a construct to be lost during electroporation and integration, with exonuclease activity causing loss from either end of the construct, and in extreme cases larger sections can be lost, possibly due to stress-shearing. In the case of Z4, the primers used in the investigation of enrichment for different histone modifications suggest that the *ACTB* promoter region is intact, with one pair of primers (ACTBSYBRF4/R4, labeled “ACTB1” in Figure 6.11) successfully amplifying a region just 30bp from the beginning of the *ACTB* gene, as it was cloned into the pSL301 vector following restriction with Blp1 (Figure 6.2). This would argue against deletion of a region of the promoter. As has already been described in the MYOD chapter, high copy number has been shown to induce repression of integrated constructs (Garrick *et al.* 1998), and the Z4 copy number may well be the highest out of the cell lines analyzed, as the ratio of human to mouse LUC7L appears large when analyzed by southern blot (Figure 6.4), though background staining within the lanes prevent formal quantification.

For the Z4 cell line, enrichment for both H3K9me3 and H3K27me3 occurred at all three regions examined, and also within the promoter of the *LUC7L* gene. While levels were high in ESCs for both modifications, enrichment tended to increase further in EBs suggesting a deepening repressive state as cells differentiated. While H3 Acetylation levels remained low across the region in both ESCs and EBs, H3K4me3 levels significantly increased, more than doubling in most cases so that they were approaching endogenous *ACTB* levels in EBs. This increase did not reduce the repression of transcription observed in ESCs, rather transcriptional repression increased in line with the observed increases of DNA methylation and repressive histone modifications, suggesting that together these repressive marks exert greater influence on chromatin structure than H3K4me3, and far outweigh any positive influence it could have on gene expression levels.

Silencing of the *HBA*, *MYOD* and *UBC* genes occurred upon differentiation, whereas silencing of the *ACTB* gene in Z4 occurs in ESCs before increasing in EBs. This suggests that an alternative mechanism may be responsible. If the construct integrated into a region of the genome that containing constitutive heterochromatin, associated with both H3K9me3 and DNA methylation, then spreading of heterochromatin from either end would have significantly impaired transcription. Additionally, histone modifications associated with enhanced transcription such as H3K79me3 and H3K36me3 would have been excluded, further reducing the genes transcriptional potential.

The pattern of histone modifications observed across the *ACTB* gene in the Z6 cell line are in stark contrast to the repressive chromatin state observed in the Z4 cell line. Histone modifications across the *ACTB* gene in the Z6 cell line did not indicate the development of a repressive histone state, with enrichment of neither H3K9me3 or H3K27me3 at levels greater than that for endogenous *ACTB*. Additionally; enrichment of H3K4me3 remained high around the TSS in both ESCs and EBs, following the classical pattern of enrichment for an actively expressed gene where high levels of enrichment occur as a sharp 'spike' in a region of ~1Kb centered on the TSS before declining sharply in the body of the gene (Barski *et al.* 2007). Curiously, the levels of H3 acetylation were observed to decrease in EBs around the TSS, though this did not cause a reduction in the level of transcription suggesting that H3 acetylation was not playing a pivotal role in maintaining an open chromatin state conducive to gene expression.

The high level of expression of the *ACTB* gene, as observed in this study and by others (Mikkelsen *et al.* 2007), may be a factor in the lack of repression observed in the Z6, Z8, Z9 and Z18 cell lines, as it has been indicated that high levels of transcription can protect a CpG island-containing promoter from *de-novo* DNA methylation (Stricker *et al.* 2008). Stricker *et al.* demonstrated that replacing the promoter of the Airn ncRNA with another strong promoter resulted in imprinting at the Igf2r locus as normal, but when the weaker tetracycline-inducible promoter was used, it became marked by DNA methylation and repressed in ESCs, with methylation levels increasing upon differentiation. The high levels of *ACTB* expression may therefore confer protection from silencing.

Genes have also been shown to be protected from DNA methylation by binding of the Sp1 transcription factor to specific binding sites within the promoter, where point mutations at binding sites eliminate Sp1 binding and result in DNA methylation (Brandeis *et al.* 1994, Han *et al.* 2001, Macleod *et al.* 1994, Hsieh 1999). While Sp1 binding sequences are diverse, there is an agreed consensus sequence (Gumucio *et al.* 1991), and this was not observed within the *ACTB* genes' CpG island or promoter region. However; due to the highly variable nature of Sp1 binding sequences in both composition and reported length, it cannot be discounted entirely, and if a slightly reduced consensus sequence is excepted then a possible binding site can be detected within the promoter region of the *ACTB* gene that could represent a weakened but still viable site for Sp1 binding. Indeed, this potential binding site retains the core sequence of three different Sp1 sites that have been shown to protect the 5' region of the Aprt gene from DNA methylation (Brandeis *et al.* 1994).

ACTB has been shown to be enriched for a variety of histone modifications that are absent from the *UBC* gene, and could play a role in the insusceptibility of *ACTB* to antisense-mediated silencing. Within the region of the TSS, the *ACTB* gene is enriched for H3K79me3 in human ESCs, while within the body of the gene its has been shown to be enriched for H3K36me3, while *UBC* has neither of these marks (Guenther *et al.* 2007). Global histone modification studies in both mice and humans have shown that H3K36me3 is sharply elevated downstream of TSSs of active genes, and that levels of enrichment are highly correlated with gene expression levels (Mikkelsen *et al.* 2007, Barski *et al.* 2007, Guenther *et al.* 2007). H3K79me3 has a more complex enrichment

pattern, and has been suggested to be moderately correlated with gene silencing. The exception to this is a small group of highly active genes that were observed to be enriched with high levels of this mark surrounding the TSS (Barski *et al.* 2007). The potentially important roles played by these different histone modifications, the potential for a binding site for an 'anti-silencing' factor, and the link between promoter strength and protection from DNA methylation could all have roles to play in maintaining the *ACTB* gene in an active and unrepressed state.

As the silencing mechanism responsible for AS RNA mediated silencing at the alpha globin and other genes has yet to be elucidated, it increases the difficulty in accurately discerning the cause of the lack of *ACTB* silencing, and the explanation of the intense repression observed in the Z4 cell line. Additionally, the investigation into the interplay between histone modifications is in many respects still in its nascency, but has demonstrated that histone modifications can have wildly different roles in gene expression depending upon cellular context, and other histone modifications with which they are combined. Also; new histone modifications continue to be discovered, whose roles in modifying chromatin states remain unknown.

In summary, it seems that ubiquitous CpG islands can be silenced by antisense RNA. However this sensitivity may be dependent on the level of expression, with highly expressed genes such as *ACTB* being insensitive and genes expressed at lower levels, such as *UBC* being susceptible.

Chapter 7 Final Discussion

The work described in this thesis has provided a deeper understanding of the mechanistic details of antisense RNA mediated transcriptional silencing as observed at the alpha globin gene in a case of α thalassaemia, and has started to investigate whether this mechanism is more generally applicable to other genes. In particular, it has clarified the role of antisense RNA expression, further determined the timing of repression during *in vitro* differentiation, and demonstrated that such silencing is not an event unique to the alpha globin gene but is applicable to other CpG island containing genes.

7.1 Antisense RNA mediated silencing

Expression of antisense RNA through the alpha globin gene was found to be essential for repression to occur, with gene silencing occurring rapidly between d3 and d4 of *in vitro* differentiation. Gene silencing by both H3K9me3 and DNA methylation, involving the transcription of an antisense RNA through the promoter region of a gene was not restricted to the alpha globin gene but appears to be equally applicable to other tissue specific or ubiquitously expressed genes. These genes have entirely disparate sequences, regulatory mechanisms, and expression profiles compared to the alpha globin gene. However; expression of antisense RNA alone is clearly not the only factor in determining whether gene silencing will occur. The ubiquitously expressed gene *ACTB* appears unaffected by antisense transcription displaying neither DNA methylation or enrichment for H3K9me3, suggesting that other factors, such as the expression level of the gene or its associated histone modifications, may well be of equal importance in determining whether a gene is vulnerable to repression by this RNA-mediated mechanism.

7.2 Involvement of H3K27me3 in gene silencing

The polycomb repressor complexes have been shown to play a role in RNA mediated repression. While The PRC1 complex cooperates with ncRNAs preferentially during X inactivation (Schoeftner *et al.* 2006), the PRC2 complex has a widespread association with ncRNAs, and has been implicated in siRNA mediated TGS (Hawkins *et al.* 2009) as well as in the process of X inactivation (Zhao *et al.* 2008). PRC2 has also been shown to interact with ncRNAs at the HOX loci (Rinn *et al.* 2007), and its action at some imprinted loci is inferred by the enrichment in H3K27me3, a mark that is laid out by a PRC2 component (Terranova *et al.* 2008). In many of these cases of RNA mediated repression, PRC components have been observed interacting with RNA, leading some to suggest that these interactions may be responsible for recruitment of the PRC complexes (Bernstein *et al.* 2006b, Zhao *et al.* 2008).

RNA-ChIP has demonstrated co-localisation of both Ezh2 and Suz12 with the Kcnqot1 ncRNA during imprinting in the placenta (Pandey *et al.* 2008), with similar observations of enrichment for the ncRNA HOTAIR upon native immunoprecipitation of Suz12 at the HOX loci (Rinn *et al.* 2007) clearly demonstrating that, at the very least, these PRC proteins occur in the same complex as ncRNAs. Direct interaction between polycomb repressors and ncRNA has been demonstrated during X inactivation, where the RepA repeats within the first exon of the Xist ncRNA form a dual stem loop structure that is bound by Ezh2 (Zhao *et al.* 2008). Direct interactions have also been demonstrated for the Cbx component of PRC1, with four members of the Cbx

family shown to directly bind oligomeric RNA in a sequence independent manner (Bernstein *et al.* 2006b).

Due to the widespread involvement of H3K27me3 in repression, and the indications that the PRC2 complex interacts with ncRNAs, H3K27me3 enrichment were examined at the *HBA2*, *UBC*, *MYOD* and *ACTB* genes to assess whether it had a functional role in repression at these genes upon antisense RNA expression. The tissue specific genes *MYOD* and *HBA2* are bivalently marked with H3K4me3 and H3K27me3 in human ESCs (Pan *et al.* 2007), with mouse *Myod* also bivalently marked in mESCs (Mikkelsen *et al.* 2007), while neither of the two ubiquitously expressed genes, *UBC* and *ACTB*, normally have any H3K27me3 enrichment in both mouse and human ESCs as well as differentiated tissues (Mikkelsen *et al.* 2007, Pan *et al.* 2007). In the work presented in this thesis, all cell lines were differentiated in media containing various factors that promote development of hematopoietic lineages. In their endogenous context and under these culturing conditions, it would be expected that the bivalent domain at the alpha globin gene would become resolved with loss of H3K27me3 and maintenance of H3K4me3, while the reverse would be true at the *MYOD* gene, with either continued bivalency or loss of H3K4me3 and maintenance of H3K27me3 enrichment.

In order to calibrate the level of enrichment that represents a functional level of H3K27me3, the level of enrichment at several genes shown by other studies to be unmarked by this modification (Mikkelsen *et al.* 2007) was examined. For two genes expressed in mESCs and for a intergenic region of the genome,

enrichment when normalised to both input and *Actb* showed an enrichment of between 1 and 1.7 (described in detail in Appendix A), thereby establishing a threshold for background levels of H3K27me3. The *Pcdh8* gene, demonstrated by Mikkelsen *et al.* to be bivalently marked in mESCs with very high levels of H3K27me3, was found in this study to have an enrichment of 6.5, normalized to input and *Actb* levels, thereby establishing a value for high levels of H3K27me3.

An issue concerning the definition of relevant levels of H3K27me3 centres upon the perceived role of both the PRC2 and H3K27me3 in the silencing process. At bivalent domains H3K27me3 is the sole repressive mark present to maintain the gene in a repressed state, counterbalancing H3K4me3 enrichment. In cases of transcriptional gene silencing, H3K27me3 is just one of several epigenetic marks that mediate repression - in the case of the silencing effect at the *HBA2*, *MYOD* and *UBC* genes, DNA methylation and H3K9me3 are both present. Additionally; though the manner of PRC2 recruitment is unknown in both normal H3K27me3 enrichment and in the TGS mechanism under investigation, targeting on PRC2 in the 'classical' context is deliberate and H3K27me3 alone must be sufficient for repression whereas in the context of the form of repression under investigation in this thesis, PRC2 recruitment could be a result of its inclusion in a multimeric complex due to the high level of protein interaction observed between the PRC2 and other components of transcriptionally repressive pathways, which could result in lower PRC2 levels than at 'classical' H3K27me3 domains, with PRC2 present due to a role in targeting of a multimeric silencing complex, with H3K27 methylation a subsidiary effect.

Cell lines bearing the ZF α Antisense Stop construct, where antisense RNA is prevented from reaching the alpha globin gene, showed no H3K27me3 enrichment in ESCs or EBs, with levels generally equal to or less than those observable for endogenous *Actb*. Across the *MYOD* gene, H3K27me3 levels doubled to twice those observed at the *Actb* gene upon differentiation, while H3K4me3 levels showed a trend to decrease, though this was slight. The *ACTB* gene was highly expressed in both ESCs and EBs in the cell line Z6, and H3K27me3 levels remained at similar levels upon differentiation, with enrichment in the range of 1 to 1.7 clearly indicating that the K27 modification was not present at biologically significant levels. Therefore; the trends for H3K27me3 enrichment are in line with the predicted epigenetic landscape for these genes. However, the low levels of H3K27me3 observed in mESCs for both *HBA2* and *MYOD* were surprising given the widely characterised H3K27me3 enrichment of these genes in hESCs (Guenther *et al.* 2007), with bivalency also shown to occur at *Myod* in mESCs (Mikkelsen *et al.* 2007), suggesting that bivalent domain formation may have been inhibited to some degree due to placement of the genes outside of their normal chromosomal environment.

The manner in which PRC complexes are guided to target loci in mammals has yet to be elucidated, and the region covered by H3K27me3 at bivalent domains tend to be large, around 3Kb, with medium to low levels of H3K27me3 throughout, with a small region of H3K4me3 centred on the TSS (Ku *et al.* 2008, Bernstein *et al.* 2006a). It is possible that the correct formation of this large

H3K27me3 domain requires genomic sequences or the presence of long range elements, not all of which are included in the constructs incorporated into mESCs so, outside of these genes normal chromosomal context, proper formation of the H3K27me3 domain is impaired in undifferentiated cells.

In contrast to the normal epigenetic state of a ubiquitously expressed gene, cell lines containing the *UBC* construct showed H3K27me3 enrichment upon differentiation. The cell line U4 was marked with H3K27me3 within the promoter and the body of the gene, with enrichment levels within the promoter region and the first intron downstream of the TSS doubling to approximately twice the levels observed at the *Actb* endogenous control. Similar H3K27me3 enrichment was also observed for the cell line U17, though enrichment was not as marked as observed for the U4 cell line. Analysis of H3K27me3 enrichment across the alpha globin gene for cell lines containing the ZF α Antisense construct also showed enrichment for this repressive mark upon differentiation, with enrichment clearest within the promoter region of the gene. Enrichment upon differentiation was not so clear cut as was observed for the *UBC* gene, though the trend for increasing levels was clear at three of the four regions assayed. H3K27me3 levels also increased within the two Ago1 Kd cell lines examined, with increasing levels clear within the 2.1 cell line, but harder to discern within the 1.2 cell line due to the higher than usual disparity between the real-time PCR replicates.

H3K27me3 levels showed an increase in enrichment upon differentiation at the alpha globin gene in an antisense RNA dependent manner. Enrichment levels

were greater than those observed for controls used to establish background H3K27me3 levels, but less than the enrichment observed for the *Pcdh8* gene, a positive control for high H3K27me3 enrichment. Alongside the observed H3K27me3 enrichment at the *UBC* gene, this suggests that PRC2 mediated H3K27me3 is indeed present as a result of antisense expression. Such an observation inevitably raises further questions concerning whether the PRC2 could have a role in initiation of gene repression, or whether recruitment is due to its involvement with other repressive factors, such as H3K9 methyltransferases, with which it has been suggested to interact (Pandey *et al.* 2008).

While some components of the PRC2 have been shown to directly interact with ncRNAs, this requires the RNA to form a specific secondary structure with even single base pair changes resulting in abrogation of these interactions (Zhao *et al.* 2008), with no such secondary structure predicted for the various antisense RNA sequences that occur across the *HBA2*, *UBC*, *MYOD*, or *ACTB* genes such an interaction seems unlikely. This suggests that PRC2 involvement with other components of the silencing machinery and the observed inclusion of PRC2 within complexes that contain ncRNAs is the more likely explanation for the presence of H3K27me3 at these genes, and may also offer an explanation as to why only medium to low levels of enrichment were observed if the PRC2 is indeed recruited at low levels due to its inclusion in a large multimeric complex. ChIP at different time points for EzH2, the catalytic component of PRC2, would confirm its involvement in silencing and could also establish when the H3K27me3 mark is added in comparison to H3K9me3 and DNA methylation that would indicate whether it could have a role in initiation, or only exists in a

maintenance capacity, while RNA ChIP would clarify its potential involvement with any antisense RNA containing complexes. Such an experiment would best be carried out in *UBC* cell lines to avoid detection of endogenous repression, as PRC2 recruitment and H3K27me3 are not features normally associated with the *UBC* gene, while both are enriched at the *HBA2* and *MYOD* genes when not expressed (Schuettengruber *et al.* 2007, Garrick *et al.* 2008).

Given the indicated involvement of the PRC2 complex in repression, it would also be interesting to examine if the PRC1 complex is involved, given that its core catalytic component Cbx binds to H3K27me3 and has also been demonstrated to be capable of binding H3K9me3 (Bernstein *et al.* 2006b).

7.3 Small RNAs as mediators of transcriptional gene silencing

TGS by the siRNA pathway has recently been the subject of intense study, and results in gene repression that bears more than a passing resemblance to the form of silencing observed at the alpha globin gene. In the most recent model, a promoter derived transcript acts as the target for the antisense strand of an siRNA, which recruits a complex composed of Ago1, Dnmt3a, HDACs, and potentially histone methyltransferases for H3K9 and H3K27 (Hawkins *et al.* 2009, Han *et al.* 2007). The authors also suggest that antisense RNA transcripts could produce endogenous siRNA substrates for targeted repression (Han *et al.* 2007). The initial investigation into the potential role of siRNAs in repression at the alpha globin locus presented in this thesis proved inconclusive, with no siRNAs detected or any phenotypic effect of Ago1 Kd shown but with the caveats, as detailed in Chapter 3, that siRNA levels may be

below the detection threshold, and that the level of Ago1 Kd achieved could still have allowed functional Ago1 complexes to form.

Recently published work examining Ago1 mediated TGS highlighted several important differences in the kinetics of silencing between siRNA mediated repression and the form of silencing investigated in this thesis. Serendipitously, the study by Hawkins *et al.* chose to examine TGS of the *UBC* gene, using an shRNA targeted at the genes promoter. They observed enrichment for H3K9me2, and modest enrichment for H3K27me3, coupled with very high levels of Ago1 within the first 48hrs of repression. Gene silencing took several days to occur, with gene expression levels halving after three days of shRNA treatment. Histone modifications became enriched during the initial stages of silencing, while DNA methylation occurred after histone methylation had become established. DNA methylation remained partial, with only around 60% of sites assayed showing dense methylation after ten days of high shRNA expression (Hawkins *et al.* 2009). This rate of repression is at odds with the startlingly rapid rate of antisense RNA mediated silencing described in this work, with gene repression occurring within 24hr, with concurrent histone and dense DNA methylation occurring within the same timeframe.

Such differences do not rule out siRNA involvement in gene repression, and indeed; some variation is only to be expected between such disparate model systems. As discussed in chapter 4, a more robust Ago1 Knockdown should clarify what, if any, role siRNA mediated TGS plays in the silencing mechanism first characterised at the alpha globin gene. Additionally, while the current

dearth in commercially available Ago1-specific antibodies has hampered any direct investigation of Ago1 enrichment during initiation of silencing by ChIP, this does represent another viable route of investigation for the future.

7.4 An inducible silencing system

Gene silencing by targeting of small RNAs against promoter regions is initiated immediately upon expression of, or treatment with, the small RNA. Gene silencing as characterised in this thesis is dependent upon *in vitro* differentiation, with gene silencing unable to occur in ESCs despite antisense RNA expression. However; small RNA mediated gene silencing has yet to be attempted in ESCs and so far has only been carried out in immortalized cell lines such as HEK293 or HeLa (Hawkins *et al.* 2009, Kim *et al.* 2006, Castanotto *et al.* 2005), or in cancer cell lines such as the colorectal cancer cell line HCT116 (Ting *et al.* 2005). Therefore, such an apparent disparity in repression may be due to the different cell types used.

Silencing as characterised at the alpha globin, *MYOD* and *UBC* genes upon differentiation may rely upon a critical factor that is not present in ESCs, but becomes expressed only upon differentiation, perhaps linked to regulatory control of pluripotency factors. The process of X inactivation also only occurs upon differentiation, however it must occur within a set developmental window, as when Xist is placed under the control of an inducible promoter and expression is triggered later in differentiation, X inactivation can no longer occur (Chow & Heard 2009). Therefore; modification of the original alpha globin silencing construct, ZF α Antisense, so as to place *LUC7L* expression under the control of an inducible promoter, such as TET, and incorporation into a

hematopoietic cell line would reveal whether silencing can occur outside of the d3/d4 differentiation time period, and whether silencing is immediate upon expression of antisense RNA in this cell line. The K562 erythroid-like cell line expresses alpha globin, with the HBA genes demonstrated to be unmarked by repressive factors (Garrick *et al.* 2008), and would therefore provide a good cell line to test an inducible silencing system.

If silencing were revealed to occur upon antisense RNA expression, such a silencing model would greatly facilitate identification of components involved in silencing though both ChIP for likely candidates upon initiation of silencing, as well as siRNA mediated Kd of possible components through introduction of siRNAs directly in the media.

7.5 A useful silencing model

The current pace of regulatory RNA discovery will only accelerate, as the role played by these ncRNAs in genome regulation becomes more apparent. Alongside the well characterised ncRNAs involved in imprinting and X inactivation, many thousands of conserved non-coding RNAs are now being annotated (Guttman *et al.* 2009). While some ncRNAs have already been demonstrated to play roles in gene regulation (Rinn *et al.* 2007), others have been linked to gene repression though aberrant expression in cancer (Yu *et al.* 2008). All of the ncRNA mediated silencing mechanisms characterised rely upon RNA:protein interactions for the targeting of repressive complexes. This interaction forms the heart of RNA mediated repression, and yet the proteins that bind long ncRNAs and so target DNA and histone methyltransferases remain mostly unidentified. A host of proteins, including Dnmt3a, G9a, Suz12,

Ezh2, and Cbx, have all been shown to occur within the same complex as ncRNAs (Pandey *et al.* 2008, Bernstein *et al.* 2006b, Rinn *et al.* 2007, Zhao *et al.* 2008, Sun *et al.* 2006, Schoeftner *et al.* 2006) and yet, with the notable exception of the PRC proteins, have not been shown to be capable of direct RNA binding, suggesting the presence of an as yet unidentified protein that acts as an intermediate between the RNA and components of the silencing complex.

The silencing mechanism responsible for repression at the alpha globin gene has now been demonstrated to be capable of gene repression in a more general way, silencing both tissue specific and ubiquitously expressed genes in an identical manner, and if it is found that silencing is the result of a long RNA, it offers the chance to study a repressive mechanism that appears to operate in a distinct developmental window during early development. This mechanism, although aberrantly recruited in the case of the alpha globin gene, must naturally exist within early pluripotent cells, contains both DNA and histone methylase ability, and may therefore have a role to play at this early stage of differentiation. The presence of such a complex at a point when gene expression patterns are fixed through epigenetic modifications, and cell fate decisions are made, can only suggest the potential role that the suggested silencing mechanism may fulfil.

Knockout, or Knockdown of the major histone methyltransferases G9a and Suc39h1/h2 would allow identification of the histone methyltransferase responsible for H3K9me3, while RNA ChIP could be used to examine

enrichment for likely components of the silencing complex containing the antisense RNA. An electrophoretic mobility shift assay (EMSA) could be initially used to examine crude nuclear extract for its ability to bind RNA oligomers of the antisense RNA, and identify the period of differentiation when binding can occur thereby defining when this complex is present. Subsequently, once components of the complex have been identified, EMSA could be used to identify which components are directly capable of binding to the antisense RNA. Once a component of the silencing complex has been identified, pull-down of the entire complex followed by mass spectrometry would help to identify the individual components. If a silencing complex were to be identified and characterised, it would be of great interest to ascertain the biological role it normally plays during early differentiation. To such an end, ChIP-chip for the various components of any identified silencing complex on material from early differentiating ESCs would identify regions bound by the various components, which could then be subsequently assessed for ncRNA expression.

7.6 Concluding remarks

Precise control of gene transcription is essential for a cell to function correctly. Repressive epigenetic modifications such as methylation of H3K9 and H3K27, along with DNA methylation, provide powerful repressive mechanisms to ensure transcriptional silencing - with H3K9me3 and DNA methylation in particular associated with stable long term repression. The manner in which such repressive marks are targeted to their sites of action is a vitally important part of epigenetic regulation, and in recent years non coding RNAs have taken centre stage as highly specific coordinators of repression, targeting repressive complexes and defining regions to be repressed through RNA:DNA interactions.

The importance and scale of RNA involvement in epigenetic regulation within the nucleus cannot be overstated, as the rate of discovery of regulatory RNAs gathers pace, redefining the way in which genomic regulation is viewed.

The work in this thesis has focussed upon a silencing event that occurred as a result of a genomic deletion, to further understand the role played by expression of antisense RNA in this silencing event, and to investigate whether such a repressive mechanism could be more widely applicable. As a result of these investigations, antisense RNA has been shown to be an integral component of this silencing mechanism, and other tissue specific as well as ubiquitously expressed genes have been demonstrated to be susceptible to antisense mediated repression.

Such an experimental model system provides an excellent opportunity to further characterise a developmentally activated repressive complex containing both DNA and histone methyltransferases, and recruited by an antisense RNA mechanism. Knowledge of how such repressive complexes are recruited, and how they function, will be essential to understand the full role played by the many regulatory RNAs that are only now beginning to be characterised.

Appendix A: Testing of antibodies for specificity and enrichment in mESCs and EBs

Antibodies were tested against a selection of endogenous controls chosen from published data. The majority of controls were selected from an investigation into global histone modification patterns in mESCs, MEFs and NPCs by Mikkelsen *et al.* ostensibly as enrichment data for many of the controls was available for more than one histone modification, and ChIP-seq data was confirmed using ChIP and real-time PCR (Mikkelsen *et al.* 2007). H3K9me3 was also examined for enrichment of a region of the *Igf2r* gene stably methylated in oocytes with methylation maintained at a stable level throughout development (Regha *et al.* 2007).

H3K9me3

Enrichment for the *Air* promoter was almost three times that observed at the *Actb* gene, and was shown to be at near equal levels in ESCs and EBs as expected. H3K9me3 at the *Polrmt* gene was enriched 5 fold compared to *Actb* in ESCs, and was enriched to near equal levels in EBs. The ubiquitously expressed gene *Gapdh* was also enriched compared to *Actb* and showed equal levels of enrichment in ESCs and EBs indicating its suitability as control for this histone modification. These observations matched previously observed H3K9me3 enrichment for these regions Results are shown in Figure A1 (A).

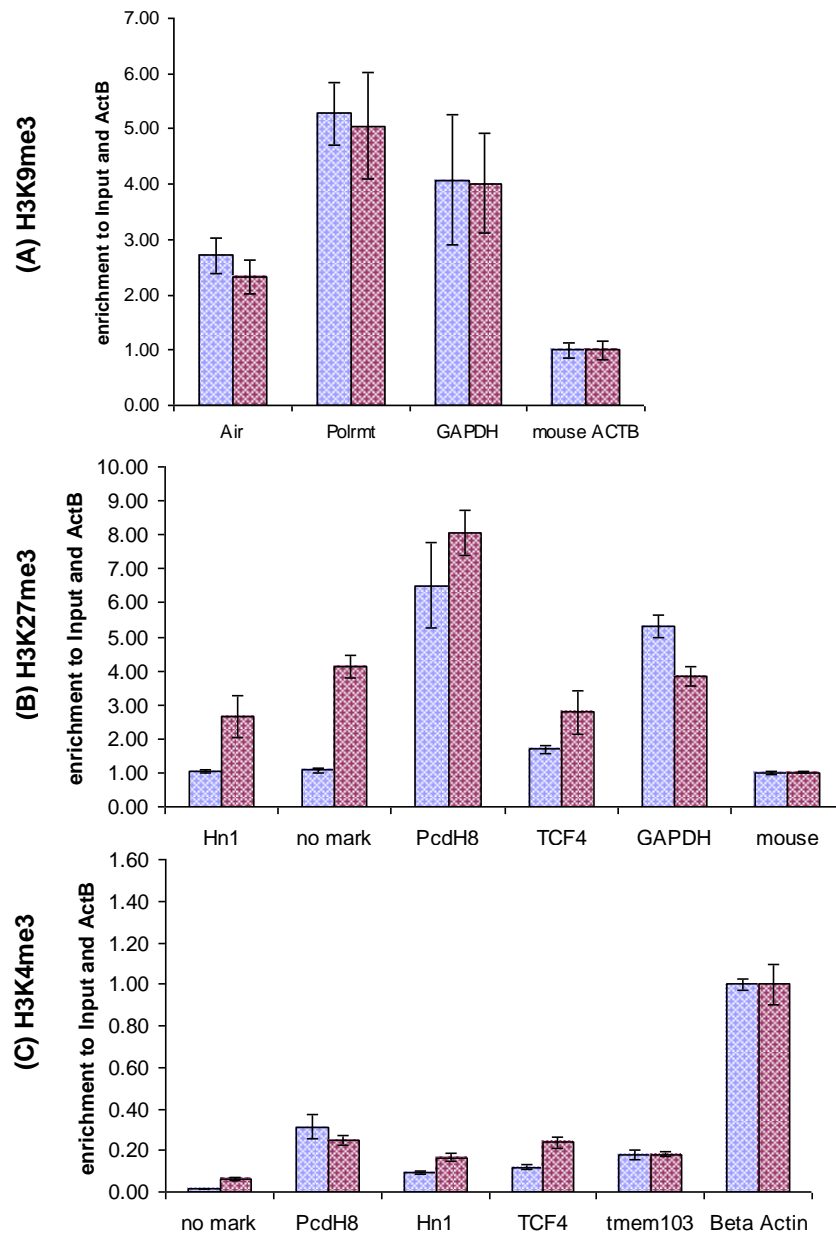
H3K27me3

Enrichment for H3K27me3 was examined using a range of controls with known histone modification profiles in ESCs, NPCs, and MEFs. Results are shown in

Figure A1 (B). Hn1, 'no mark' (referring to a gene devoid region of the genome unmarked for either histone modification), and Tcf4 are all reported as lacking H3K27me3 enrichment in mouse ESCs (Mikkelsen *et al.* 2007). Hn1 and no mark both had an enrichment level equal to Actb - which is also known to be unmarked by H3K27me3 (Figure 3.9B). Tcf4 showed a slightly greater level of enrichment, though this was still far lower than for Pcdh8, a gene shown to be enriched for H3K27me3 in ESCs that was enriched 7 fold compared to Actb. Gapdh also showed a higher than background level of enrichment, with less enrichment in EBs.

H3K4me3

Mikkelsen *et al.* showed that Pcdh8, tmem103, Tcf4 and Hn1 are enriched for H3K4me3 in mouse ESCs, while 'no mark' lacks this histone modification. Enrichment for the four genes was detectable in ESCs, though at low levels of enrichment compared to Actb, while enrichment for 'no mark' verged on the undetectable (Figure A1 (C)).



(D)

Control	H3K27me3 enrichment	H3K4me3 enrichment	Gene details
Hn1	None in ESC, NPC, MEF	ES, NPC, MEF	Very highly expressed, protein coding (unknown function)
No mark	None in ES	None in ES	Intergenic region of Chr11 (11E2).
PcdH8	High in ES, NPC, MEF	ES, NPC, none in MEF	Low expression levels, codes for cell adhesion factor
TCF4	None in ESC, NPC, MEF	ES, MEF	High expression in NPCs. Codes for transcriptional regulator
tmem103	unknown	ES	Widely expressed. Codes for transmembrane protein

Figure A1. Endogenous controls for H3K4me3 and H3K9me3 methylation. (A) to (C) shows enrichment levels for different histone modifications for different endogenous controls. Blue; ESCs, red; EBs. Enrichment is normalised to Input and to mouse ACTB. (D) known enrichment profiles for endogenous controls (information from Mikkelsen *et al.* 2007)

Primer	Sequence, 5' to 3'	PCR buffer	Annealing temp °C	Notes
tcf4 qF	CGGATGTGAATGGATTACAATGTATC	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
Tcf4 qR	CCCCGAGGAGTCACATTGA			
tmem103 qF	CCATTGCCCCGAATGAGAGAT	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
tmem103 qR	CGCGTGAGCACGTATGCT			
air qF	CTGAGCTTTCCCTTCCCTTTC	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
air qR	TGCCGTGATCCTTGTTGT			
Pcdh8 qF	TGCACCAGGCGGAAGTGT	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
Pcdh8 qR	TGCTGGCCTTCGATGTTGT			
Polrmt qF	CCGTGGCACGCATTAGTTT	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
Polrmt qR	CTGGACGTCGCTGCATTCT			
neg qF	GCCTAAGTGAGGAAGGCTCATC	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
neg qR	AGGCCAAGCCCTCTCTTACAC			
hn1 qF	GGCCACCACGAGGAAAAGA	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
hn1 qR	CGCTGCGAGCAAATCTGA			
HoxaA3 qF	GCTTTTGCTGTGGGAAGCA	ABI SYBR green	60	Standard realtime program used; 95°C 10min, (95°C 15s, 60°C 1min) x40 cycles.
HoxaA3 qR	GCCTTGAAGAGTCTGCCTCAGT			

Table A.1. ChIP control primers

Antibodies show specific enrichment

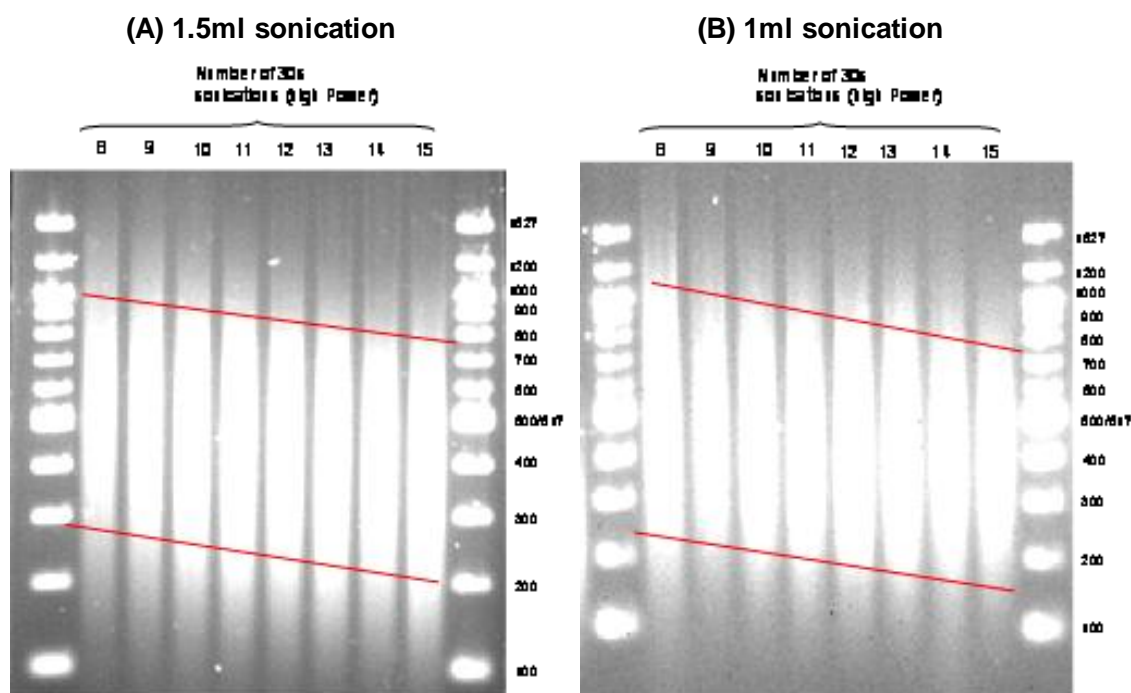
Enrichment for a region identified by Mikkelsen *et al.* as being unmarked by H3K4me3 or H3K27me3 in mESCs was examined in order to establish the level of background enrichment for each antibody, and also as a check against non-specific enrichment. For H3K27me3 this produced an enrichment value of 'one' when normalised to *Actb*, thereby showing no enrichment for H3K27me3, as *Actb* is essentially unmarked by this histone modification. The gene *Pcdh8* was chosen as it was highly enriched for H3K27me3 in the three cell types examined by Mikkelsen *et al.*, and indeed showed high - and nearly equal - level of enrichment in ESCs and EBs, thereby establishing a marker for biologically relevant level of H3K27me3 enrichment. H3K4me3 methylation was also verging on undetectable at this region, with a far lower level of enrichment compared to endogenous controls positive for the mark, again confirming that non-specificity was not occurring.

Enrichment in ESCs and EBs was examined for the different endogenous controls, though there was no data available for known enrichment in EBs and, in various cases, enrichment varied in EBs from levels observed in MEFs and NPCs. This is perhaps not unexpected due to the heterogeneous mix of cell types that have been shown to be present in EBs, including hematopoietic cell lineages (erythroid, myeloid and lymphoid lineages) as well as endothelial cells and cardiac muscle precursors (Keller 1995). Controls, such as *Tmem103*, *Polrmt*, and potentially *Pcdh8*, showed steady levels of enrichment in both ESCs and EBs and could therefore potentially offer alternative or complementary controls to *Actb* in future studies.

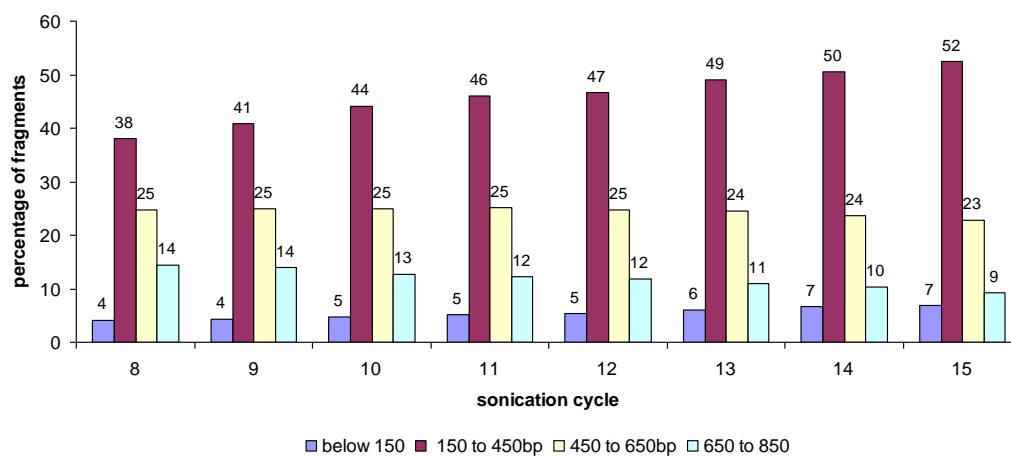
Appendix B: Calibration of sonication for ChIP

Cell samples were prepared as described in section 2.8 and 2.9.1, and DNA was sheared to the size of mono or dinucleosomes (150bp to 450bp) by sonication using a Diaganode BiorupterTM 200. Different volumes of sample (at a concentration of 2×10^6 cells/ml), and range of sonication pulses were tested at high power (30s on/30s off) to optimize sonication for fragments in the range of 150bp to 450bp.

Three different ESC sample volumes, 1.5ml, 1ml, and 500 μ l were used during sonication optimization, with fifteen 30s cycles carried out on each volume, and samples taken after each cycle for analysis. Gel electrophoresis of the samples was carried (2% TBE agarose gel), run at low voltage to ensure fragment separation. The same conditions were used for all samples to allow comparison between gels. Gel lane quantification was carried out using AIDA software, with fragment sizes correlated to the DNA ladder. Volumes of both 1.5ml and 1ml proved unsuitable as a quarter of the sample remained above 450bp in size after 15 cycles of sonication (Figure B1). Sonication using a ESC sample volume of 500 μ l resulted in, on average, 65% of the sample in the desired size range of 150 to 450bp, with only around 15% of the sample consisting of larger fragments. Sonication of EB material displayed a near identical pattern of sonication efficiency (Figure B2). Sonication of all samples was therefore carried out in a 500 μ l volume, for 14 sonication cycles, with a cell concentration of 2×10^6 cells/ml.



(C) 1.5ml sonication fragment analysis



(D) 1ml sonication fragment analysis

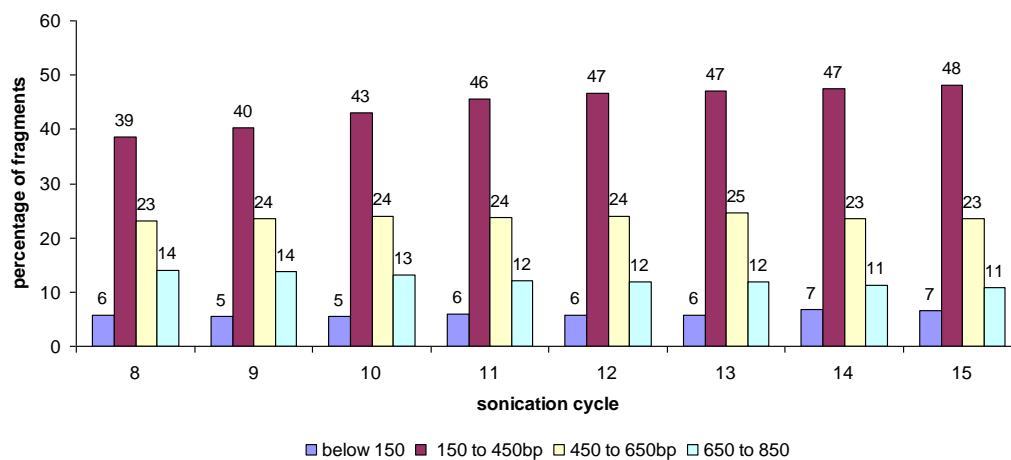
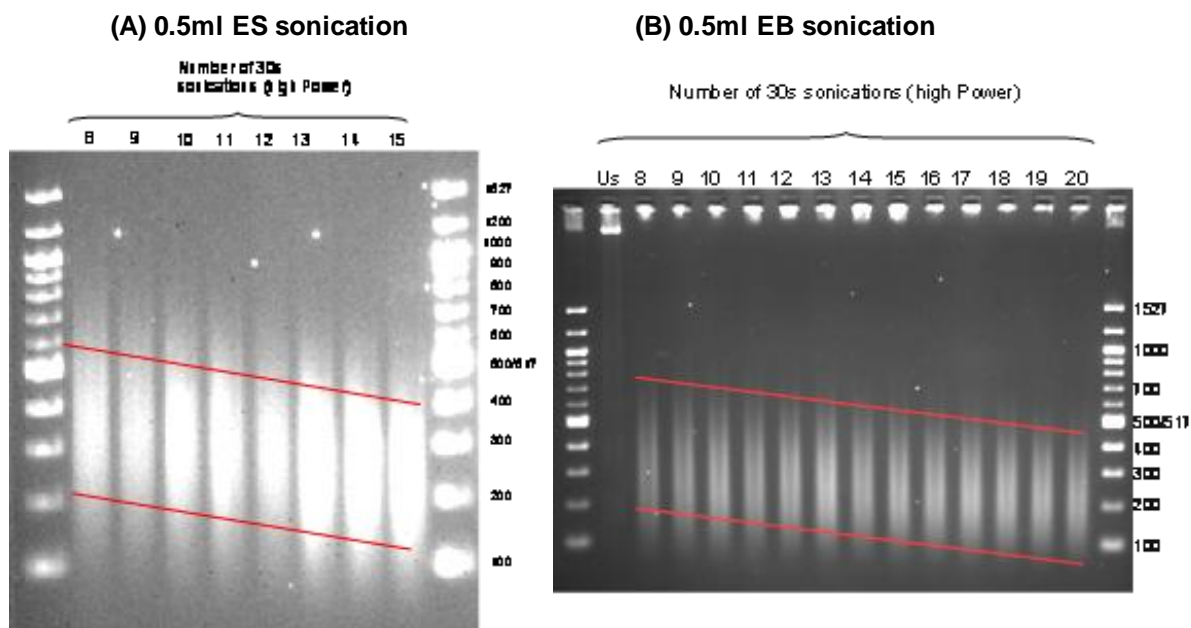
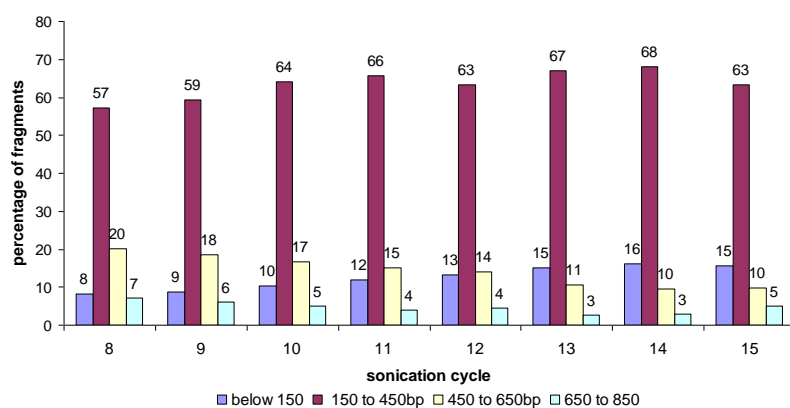


Figure B1 optimisation of sonication efficiency. 1.5ml and 1ml ESC samples sonicated for 30s on/30s off.



(C) 0.5ml ESC sonication fragment analysis



(D) 0.5ml EB sonication fragment analysis

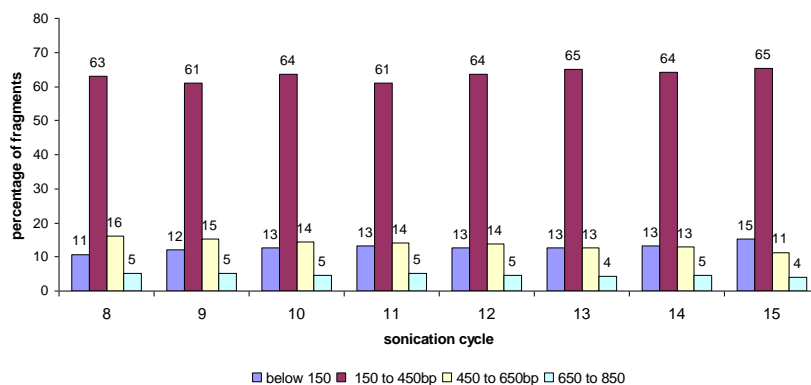


Figure B2 optimisation of sonication efficiency. 500ul ESC and EB samples sonicated for 30s on/30s off

Testing of sonication efficiency of ChIP samples.

Following sonication of all ChIP samples, 5µl of supernatant was removed for analysis and prepared as described in 2.9.5. Gel electrophoresis of samples was carried out at low voltage on a 2% TBE gel, and gel lanes analysed using AIDA software. Analysis results are shown in Figure B3 for UBC, ACTB, MYOD, and alpha globin cell lines. The data as a whole showing excellent reproducibility, with a mean value of 59% of fragments in the size range of 150bp to 450bp, and a mean of 9% of fragments in the size range of 450 to 850bp. Samples for the cell lines 2.1 and Y5 show higher than usual levels of fragments less than 150bp, but this was likely a result of incomplete RNase A digestion during sample preparation.

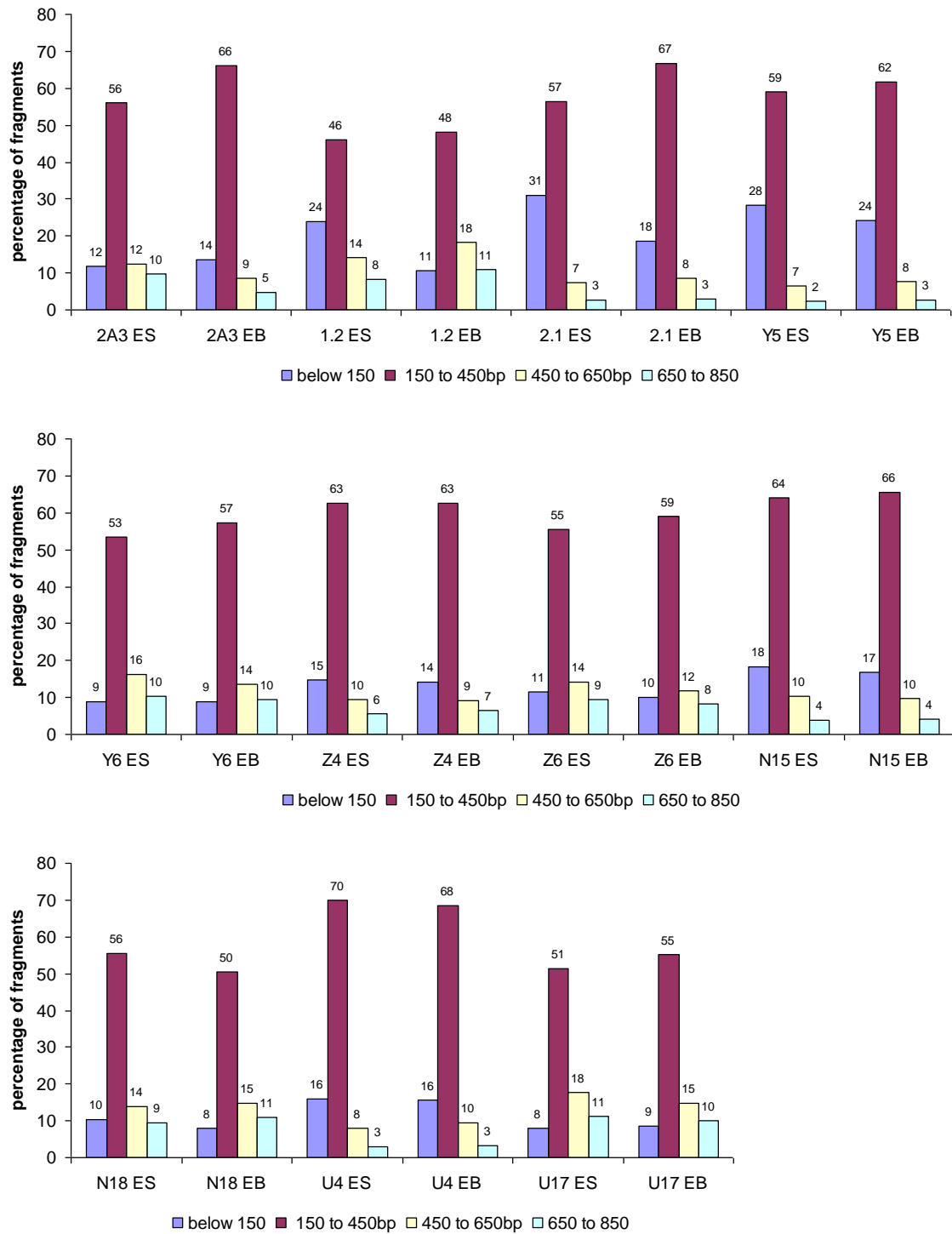


Figure B3 Examination of sonication efficiency. Size range of fragments following sonication of samples for ChIP.

Appendix C: comparing IgG enrichment to histone modification specific enrichment for the different primer pairs used during ChIP experiments

All primer pairs used in ChIP experiments were used to amplify material immunoprecipitated by native IgG in order to investigate the basal level of non-specific enrichment for each primer pair. A generic IgG was provided with histone modification specific antibodies purchased from Millipore. Individual matched IgG for each antibody was not provided.

None of the primers tested showed any significant enrichment for IgG immunoprecipitated material, indicating that none of the regions assayed have any high level of background enrichment that might bias results when comparing enrichment for a specific histone modification between ESCs and EBs. In most cases, input material for the realtime PCR was increased as little or no amplification was possible when standard dilutions were carried out in line with treatment of purified immunoprecipitated material from histone modification specific antibodies. To provide a direct comparison of the level of IgG enrichment compared to enrichment from an antibody for a specific histone mark, graphs in Figure C1 display IgG enrichment compared to Input, and enrichment for H3K9me3 compared to Input. In almost all cases, enrichment in IgG is verging on negligible and often the error between the technical repeats of the realtime PCR triplicate is larger in magnitude than any assayed level of enrichment, due to the level of enrichment being at the extent of the realtime PCR assays' sensitivity.

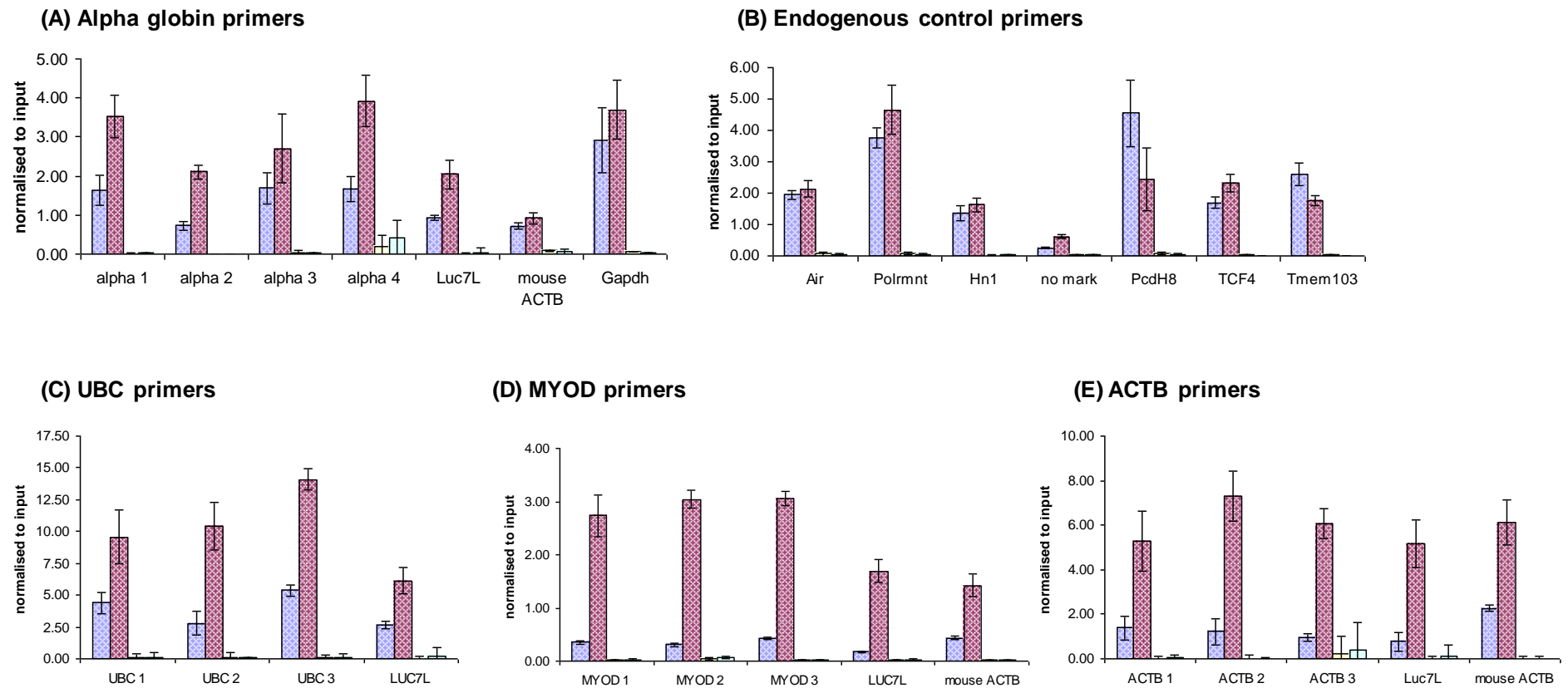


Figure C1. Comparison of IgG enrichment to antibody enrichment for the different primer pairs used. Graphs compare amplification using primer pairs on material immunoprecipitated using IgG, and material immunoprecipitated using the H3K9me3 antibody for comparison, with the exception of (B) where the majority of the primer pairs show H3K4me3 enrichment for comparison to IgG enrichment. All enrichment values are normalised to input to allow direct comparison of enrichment levels for IgG compared to H3K9me3/H3K4me3 specific antibodies. ES enrichment for comparison antibody is in blue, EB in red. ES IgG enrichment is yellow, EB IgG enrichment is turquoise.

References

- Aapola, U., Liiv, I. & Peterson, P. 2002, Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity, *Nucleic acids research*, vol. 30, no. 16, pp. 3602-3608.
- Agalioti, T., Chen, G. & Thanos, D. 2002, Deciphering the transcriptional histone acetylation code for a human gene, *Cell*, vol. 111, no. 3, pp. 381-392.
- Antequera, F. & Bird, A. 1999, CpG islands as genomic footprints of promoters that are associated with replication origins, *Current biology : CB*, vol. 9, no. 17, pp. R661-7.
- Avilion, A., Nicolis, S., Pevny, L., Perez, L., Vivian, N. & Lovell-Badge, R. 2003, Multipotent cell lineages in early mouse development depend on SOX2 function, *GENES & DEVELOPMENT*, vol. 17, no. (1), pp. 126-140.
- Barbour, V.M., Tufarelli, C., Sharpe, J.A., Smith, Z.E., Ayyub, H., Heinlein, C.A., Sloane-Stanley, J., Indrak, K., Wood, W.G. & Higgs, D.R. 2000, Alpha-Thalassemia Resulting from a Negative Chromosomal Position Effect, *Blood*, vol. 96, no. 3, pp. 800-807.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. & Zhao, K. 2007, High-resolution profiling of histone methylations in the human genome, *Cell*, vol. 129, no. 4, pp. 823-837.
- Bell, A.C. & Felsenfeld, G. 2000, Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene, *Nature*, vol. 405, no. 6785, pp. 482-485.
- Bergstrom, D.A., Penn, B.H., Strand, A., Perry, R.L., Rudnicki, M.A. & Tapscott, S.J. 2002, Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression, *Molecular cell*, vol. 9, no. 3, pp. 587-600.
- Berkes, C.A. & Tapscott, S.J. 2005, MyoD and the transcriptional control of myogenesis, *Seminars in cell & developmental biology*, vol. 16, no. 4-5, pp. 585-595.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S.L. & Lander, E.S. 2006a, A bivalent chromatin structure marks key developmental genes in embryonic stem cells, *Cell*, vol. 125, no. 2, pp. 315-326.
- Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E. & Allis, C.D. 2006b, Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin, *Molecular and cellular biology*, vol. 26, no. 7, pp. 2560-2569.
- Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. & Bestor, T.H. 2001, Dnmt3L and the establishment of maternal genomic imprints, *Science (New York, N.Y.)*, vol. 294, no. 5551, pp. 2536-2539.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R. & Young, R.A. 2005, Core transcriptional regulatory circuitry in human embryonic stem cells, *Cell*, vol. 122, no. 6, pp. 947-956.
- Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., Bell, G.W., Otte, A.P., Vidal, M., Gifford, D.K., Young, R.A. & Jaenisch, R. 2006, Polycomb complexes repress developmental regulators in murine embryonic stem cells, *Nature*, vol. 441, no. 7091, pp. 349-353.
- Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. & Helin, K. 2006, Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions, *Genes & development*, vol. 20, no. 9, pp. 1123-1136.
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. & Cedar, H. 1994, Sp1 elements protect a CpG island from de novo methylation, *Nature*, vol. 371, no. 6496, pp. 435-438.
- Brena, R.M., Huang, T.H. & Plass, C. 2006, Toward a human epigenome, *Nature genetics*, vol. 38, no. 12, pp. 1359-1360.
- Brown, K.E., Amoils, S., Horn, J.M., Buckle, V.J., Higgs, D.R., Merckenschlager, M. & Fisher, A.G. 2001, Expression of alpha- and beta-globin genes occurs within different nuclear domains in haemopoietic cells, *Nature cell biology*, vol. 3, no. 6, pp. 602-606.

- Brummelkamp, T.R., Bernards, R. & Agami, R. 2002, A system for stable expression of short interfering RNAs in mammalian cells, *Science (New York, N.Y.)*, vol. 296, no. 5567, pp. 550-553.
- Buckingham, M. 1992, Making muscle in mammals, *Trends in genetics : TIG*, vol. 8, no. 4, pp. 144-148.
- Buhler, M., Verdel, A. & Moazed, D. 2006, Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing, *Cell*, vol. 125, no. 5, pp. 873-886.
- Caiafa, P. & Zampieri, M. 2005, DNA methylation and chromatin structure: the puzzling CpG islands, *Journal of cellular biochemistry*, vol. 94, no. 2, pp. 257-265.
- Cao, R. & Zhang, Y. 2004, SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex, *Molecular cell*, vol. 15, no. 1, pp. 57-67.
- Castanotto, D., Tommasi, S., Li, M., Li, H., Yanow, S., Pfeifer, G.P. & Rossi, J.J. 2005, Short hairpin RNA-directed cytosine (CpG) methylation of the RASSF1A gene promoter in HeLa cells, *Molecular therapy : the journal of the American Society of Gene Therapy*, vol. 12, no. 1, pp. 179-183.
- Cavaleri, F. & Scholer, H.R. 2003, Nanog: a new recruit to the embryonic stem cell orchestra, *Cell*, vol. 113, no. 5, pp. 551-552.
- 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*, vol. 113, no. 5, pp. 643-655.
- Chaumeil, J., Le Baccon, P., Wutz, A. & Heard, E. 2006, A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced, *Genes & development*, vol. 20, no. 16, pp. 2223-2237.
- Chedin, F., Lieber, M.R. & Hsieh, C.L. 2002, The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 26, pp. 16916-16921.
- Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. 2003, Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b, *Molecular and cellular biology*, vol. 23, no. 16, pp. 5594-5605.
- Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D.K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D.S. & Gingeras, T.R. 2005, Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution, *Science (New York, N.Y.)*, vol. 308, no. 5725, pp. 1149-1154.
- Chow, J. & Heard, E. 2009, X inactivation and the complexities of silencing a sex chromosome, *Current opinion in cell biology*, vol. 21, no. 3, pp. 359-366.
- Clouaire, T. & Stancheva, I. 2008, Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cellular and molecular life sciences : CMLS*, vol. 65, no. 10, pp. 1509-1522.
- Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M. & Moazed, D. 2007, Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi, *Molecular cell*, vol. 27, no. 3, pp. 449-461.
- Colowick, S.P., Kaplan, N.O., Abelson, J.N., Simon, M.I. & Deshaies, R.J. 2005, *Methods in enzymology*, Academic Press, San Diego, Calif. ; London.
- Core, L.J., Waterfall, J.J. & Lis, J.T. 2008, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters, *Science (New York, N.Y.)*, vol. 322, no. 5909, pp. 1845-1848.
- Corey, D.R. 2005, Regulating mammalian transcription with RNA, *Trends in biochemical sciences*, vol. 30, no. 12, pp. 655-658.
- Creighton, T.E. 1999, *Encyclopedia of molecular biology*, Wiley, New York ; Chichester.
- Creyghton, M.P., Markoulaki, S., Levine, S.S., Hanna, J., Lodato, M.A., Sha, K., Young, R.A., Jaenisch, R. & Boyer, L.A. 2008, H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment, *Cell*, vol. 135, no. 4, pp. 649-661.

- Cross, S.H., Meehan, R.R., Nan, X. & Bird, A. 1997, A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins, *Nature genetics*, vol. 16, no. 3, pp. 256-259.
- Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W. & Zhao, K. 2009, Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation, *Cell stem cell*, vol. 4, no. 1, pp. 80-93.
- da Rocha, S.T. & Ferguson-Smith, A.C. 2004, Genomic imprinting, *Current biology : CB*, vol. 14, no. 16, pp. R646-9.
- Delgado, S., Gomez, M., Bird, A. & Antequera, F. 1998, Initiation of DNA replication at CpG islands in mammalian chromosomes, *The EMBO journal*, vol. 17, no. 8, pp. 2426-2435.
- Dodge, J.E., Kang, Y.K., Beppu, H., Lei, H. & Li, E. 2004, Histone H3-K9 methyltransferase ESET is essential for early development, *Molecular and cellular biology*, vol. 24, no. 6, pp. 2478-2486.
- Dong, K.B., Maksakova, I.A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H.W., Lam, L.L., Mager, D.L., Schubeler, D., Tachibana, M., Shinkai, Y. & Lorincz, M.C. 2008, DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity, *The EMBO journal*, vol. 27, no. 20, pp. 2691-2701.
- Dye, M.J. & Proudfoot, N.J. 2001, Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II, *Cell*, vol. 105, no. 5, pp. 669-681.
- Edwards, C.A. & Ferguson-Smith, A.C. 2007, Mechanisms regulating imprinted genes in clusters, *Current opinion in cell biology*, vol. 19, no. 3, pp. 281-289.
- ENCODE Project Consortium, Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Weinstock, G.M., Gibbs, R.A., Graves, T., Fulton, R., Mardis, E.R., Wilson, R.K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D.B., Chang, J.L., Lindblad-Toh, K., Lander, E.S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B. & de Jong, P.J. 2007, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, *Nature*, vol. 447, no. 7146, pp. 799-816.
- Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H. & Bergman, Y. 2008, De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes, *Nature structural & molecular biology*, vol. 15, no. 11, pp. 1176-1183.
- Esteve, P.O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F. & Pradhan, S. 2006, Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication, *Genes & development*, vol. 20, no. 22, pp. 3089-3103.
- Evans, M.J. & Kaufman, M.H. 1981, Establishment in culture of pluripotent cells from mouse embryos, *Nature*, vol. 292, no. 5819, pp. 154-156.
- Flint, J., Thomas, K., Micklem, G., Raynham, H., Clark, K., Doggett, N.A., King, A. & Higgs, D.R. 1997, The relationship between chromosome structure and function at a human telomeric region, *Nature genetics*, vol. 15, no. 3, pp. 252-257.
- Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H. & Nakao, M. 1999, Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms, *Molecular and cellular biology*, vol. 19, no. 9, pp. 6415-6426.
- Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L. & Kouzarides, T. 2000, DNA methyltransferase Dnmt1 associates with histone deacetylase activity, *Nature genetics*, vol. 24, no. 1, pp. 88-91.
- Fuks, F., Hurd, P.J., Deplus, R. & Kouzarides, T. 2003, The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase, *Nucleic acids research*, vol. 31, no. 9, pp. 2305-2312.
- Fulka, H., St John, J.C., Fulka, J. & Hozak, P. 2008, Chromatin in early mammalian embryos: achieving the pluripotent state, *Differentiation; research in biological diversity*, vol. 76, no. 1, pp. 3-14.
- Garrick, D., De Gobbi, M., Samara, V., Rugless, M., Holland, M., Ayyub, H., Lower, K., Sloane-Stanley, J., Gray, N., Koch, C., Dunham, I. & Higgs, D.R. 2008, The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells, *Blood*, vol. 112, no. 9, pp. 3889-3899.
- Garrick, D., Fiering, S., Martin, D.I. & Whitelaw, E. 1998, Repeat-induced gene silencing in mammals, *Nature genetics*, vol. 18, no. 1, pp. 56-59.

- Geiman, T.M., Sankpal, U.T., Robertson, A.K., Zhao, Y., Zhao, Y. & Robertson, K.D. 2004, DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system, *Biochemical and biophysical research communications*, vol. 318, no. 2, pp. 544-555.
- Gowher, H., Liebert, K., Hermann, A., Xu, G. & Jeltsch, A. 2005, Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L, *The Journal of biological chemistry*, vol. 280, no. 14, pp. 13341-13348.
- Grewal, S.I. & Elgin, S.C. 2007, Transcription and RNA interference in the formation of heterochromatin, *Nature*, vol. 447, no. 7143, pp. 399-406.
- Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U. & Cavalli, G. 2006, RNAi components are required for nuclear clustering of Polycomb group response elements, *Cell*, vol. 124, no. 5, pp. 957-971.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. 2007, A chromatin landmark and transcription initiation at most promoters in human cells, *Cell*, vol. 130, no. 1, pp. 77-88.
- Gumucio, D.L., Rood, K.L., Blanchard-McQuate, K.L., Gray, T.A., Saulino, A. & Collins, F.S. 1991, Interaction of Sp1 with the human gamma globin promoter: binding and transactivation of normal and mutant promoters, *Blood*, vol. 78, no. 7, pp. 1853-1863.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Cabili, M.N., Jaenisch, R., Mikkelsen, T.S., Jacks, T., Hacohen, N., Bernstein, B.E., Kellis, M., Regev, A., Rinn, J.L. & Lander, E.S. 2009, Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals, *Nature*, vol. 458, no. 7235, pp. 223-227.
- Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A. & Grewal, S.I. 2002, Establishment and maintenance of a heterochromatin domain, *Science (New York, N.Y.)*, vol. 297, no. 5590, pp. 2232-2237.
- Han, J., Kim, D. & Morris, K.V. 2007, Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 30, pp. 12422-12427.
- Han, L., Lin, I.G. & Hsieh, C.L. 2001, Protein binding protects sites on stable episomes and in the chromosome from de novo methylation, *Molecular and cellular biology*, vol. 21, no. 10, pp. 3416-3424.
- Hawkins, P.G., Santoso, S., Adams, C., Anest, V. & Morris, K.V. 2009, Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells, *Nucleic acids research*, .
- He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N. & Kinzler, K.W. 2008, The antisense transcriptomes of human cells, *Science (New York, N.Y.)*, vol. 322, no. 5909, pp. 1855-1857.
- Heard, E., Clerc, P. & Avner, P. 1997, X-chromosome inactivation in mammals, *Annual Review of Genetics*, vol. 31, pp. 571-610.
- Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D. & Spector, D.L. 2001, Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation, *Cell*, vol. 107, no. 6, pp. 727-738.
- Hellman, A. & Chess, A. 2007, Gene body-specific methylation on the active X chromosome, *Science (New York, N.Y.)*, vol. 315, no. 5815, pp. 1141-1143.
- Hendrich, B. & Bird, A. 1998, Identification and characterization of a family of mammalian methyl-CpG binding proteins, *Molecular and cellular biology*, vol. 18, no. 11, pp. 6538-6547.
- Hernandez-Munoz, I., Taghavi, P., Kuijl, C., Neefjes, J. & van Lohuizen, M. 2005, Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1, *Molecular and cellular biology*, vol. 25, no. 24, pp. 11047-11058.
- Hinton, T.M. & Doran, T.J. 2008, Inhibition of chicken anaemia virus replication using multiple short-hairpin RNAs, *Antiviral Research*, vol. 80, no. 2, pp. 143-149.
- Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. 1987, HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells, *Nature*, vol. 326, no. 6110, pp. 292-295.
- Houbaviy, H.B., Murray, M.F. & Sharp, P.A. 2003, Embryonic stem cell-specific MicroRNAs, *Developmental cell*, vol. 5, no. 2, pp. 351-358.

- Hsieh, C.L. 1999, Evidence that protein binding specifies sites of DNA demethylation, *Molecular and cellular biology*, vol. 19, no. 1, pp. 46-56.
- Ideraabdullah, F.Y., Vigneau, S. & Bartolomei, M.S. 2008, Genomic imprinting mechanisms in mammals, *Mutation research*, vol. 647, no. 1-2, pp. 77-85.
- Ikegami, K., Iwatani, M., Suzuki, M., Tachibana, M., Shinkai, Y., Tanaka, S., Grealley, J.M., Yagi, S., Hattori, N. & Shiota, K. 2007, Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells, *Genes to cells : devoted to molecular & cellular mechanisms*, vol. 12, no. 1, pp. 1-11.
- Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J.B., Sabuncuyan, S. & Feinberg, A.P. 2009, The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores, *Nature genetics*, vol. 41, no. 2, pp. 178-186.
- Irvine, D.V., Zaratiegui, M., Tolia, N.H., Goto, D.B., Chitwood, D.H., Vaughn, M.W., Joshua-Tor, L. & Martienssen, R.A. 2006, Argonaute slicing is required for heterochromatic silencing and spreading, *Science (New York, N.Y.)*, vol. 313, no. 5790, pp. 1134-1137.
- Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D. & Corey, D.R. 2006, Involvement of AGO1 and AGO2 in mammalian transcriptional silencing, *Nature structural & molecular biology*, vol. 13, no. 9, pp. 787-792.
- Jones, P.A. 1999, The DNA methylation paradox, *Trends in genetics : TIG*, vol. 15, no. 1, pp. 34-37.
- Jones, P.A., Wolkowicz, M.J., Rideout, W.M., 3rd, Gonzales, F.A., Marziasz, C.M., Coetzee, G.A. & Tapscott, S.J. 1990, De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 16, pp. 6117-6121.
- Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. & Wolffe, A.P. 1998, Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription, *Nature genetics*, vol. 19, no. 2, pp. 187-191.
- Kanellopoulou, C., Muljo, S.A., Dimitrov, S.D., Chen, X., Colin, C., Plath, K. & Livingston, D.M. 2009, X chromosome inactivation in the absence of Dicer, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 4, pp. 1122-1127.
- Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M. & Rajewsky, K. 2005, Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing, *Genes & development*, vol. 19, no. 4, pp. 489-501.
- Kapranov, P., Willingham, A.T. & Gingeras, T.R. 2007, Genome-wide transcription and the implications for genomic organization, *Nature reviews. Genetics*, vol. 8, no. 6, pp. 413-423.
- Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., Wahlestedt, C., RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Project Core Group) & FANTOM Consortium 2005, Antisense transcription in the mammalian transcriptome, *Science (New York, N.Y.)*, vol. 309, no. 5740, pp. 1564-1566.
- Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K. & Murakami, Y. 2005, RNA polymerase II is required for RNAi-dependent heterochromatin assembly, *Science (New York, N.Y.)*, vol. 309, no. 5733, pp. 467-469.
- Keller, G.M. 1995, In vitro differentiation of embryonic stem cells, *Current opinion in cell biology*, vol. 7, no. 6, pp. 862-869.
- Kennedy, S., Wang, D. & Ruvkun, G. 2004, A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*, *Nature*, vol. 427, no. 6975, pp. 645-649.
- Keohane, A.M., O'Neill, L.P., Belyaev, N.D., Lavender, J.S. & Turner, B.M. 1996, X-Inactivation and histone H4 acetylation in embryonic stem cells, *Developmental biology*, vol. 180, no. 2, pp. 618-630.
- Kim, D.H., Saetrom, P., Snove, O., Jr & Rossi, J.J. 2008, MicroRNA-directed transcriptional gene silencing in mammalian cells, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 42, pp. 16230-16235.

- Kim, D.H., Villeneuve, L.M., Morris, K.V. & Rossi, J.J. 2006, Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells, *Nature structural & molecular biology*, vol. 13, no. 9, pp. 793-797.
- Kim, J.W., Zhang, Y.H., Zern, M.A., Rossi, J.J. & Wu, J. 2007, Short hairpin RNA causes the methylation of transforming growth factor-beta receptor II promoter and silencing of the target gene in rat hepatic stellate cells, *Biochemical and biophysical research communications*, vol. 359, no. 2, pp. 292-297.
- Kiyosawa, H., Yamanaka, I., Osato, N., Kondo, S., Hayashizaki, Y., RIKEN GER Group & GSL Members 2003, Antisense transcripts with FANTOM2 clone set and their implications for gene regulation, *Genome research*, vol. 13, no. 6B, pp. 1324-1334.
- Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I. & Bird, A.P. 2005, DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG, *Molecular cell*, vol. 19, no. 5, pp. 667-678.
- Kouzarides, T. 2007, Chromatin modifications and their function, *Cell*, vol. 128, no. 4, pp. 693-705.
- Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., Adli, M., Kasif, S., Ptaszek, L.M., Cowan, C.A., Lander, E.S., Koseki, H. & Bernstein, B.E. 2008, Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains, *PLoS genetics*, vol. 4, no. 10, pp. e1000242.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. 2001, Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins, *Nature*, vol. 410, no. 6824, pp. 116-120.
- Lachner, M., O'Sullivan, R.J. & Jenuwein, T. 2003, An epigenetic road map for histone lysine methylation, *Journal of cell science*, vol. 116, no. Pt 11, pp. 2117-2124.
- Latham, K.E. 2005, X chromosome imprinting and inactivation in preimplantation mammalian embryos, *Trends in genetics : TIG*, vol. 21, no. 2, pp. 120-127.
- Latos, P.A., Stricker, S.H., Steenpass, L., Pauler, F.M., Huang, R., Senergin, B.H., Regha, K., Koerner, M.V., Warczok, K.E., Unger, C. & Barlow, D.P. 2009, An in vitro ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias, *Development (Cambridge, England)*, vol. 136, no. 3, pp. 437-448.
- Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C.M. & Casari, G. 2004, In search of antisense, *Trends in biochemical sciences*, vol. 29, no. 2, pp. 88-94.
- Lee, J., Hagerty, S., Cormier, K.A., Kim, J., Kung, A.L., Ferrante, R.J. & Ryu, H. 2008, Monoallele deletion of CBP leads to pericentromeric heterochromatin condensation through ESET expression and histone H3 (K9) methylation, *Human molecular genetics*, vol. 17, no. 12, pp. 1774-1782.
- Lee, J.T. & Jaenisch, R. 1997, The (epi)genetic control of mammalian X-chromosome inactivation, *Current opinion in genetics & development*, vol. 7, no. 2, pp. 274-280.
- Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T. & Peters, A.H. 2003, Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin, *Current biology : CB*, vol. 13, no. 14, pp. 1192-1200.
- Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R. & Reik, W. 2004, Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation, *Nature genetics*, vol. 36, no. 12, pp. 1291-1295.
- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. & Bird, A. 1992, Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA, *Cell*, vol. 69, no. 6, pp. 905-914.
- Li, E. 2002, Chromatin modification and epigenetic reprogramming in mammalian development, *Nature reviews.Genetics*, vol. 3, no. 9, pp. 662-673.
- Li, E., Bestor, T.H. & Jaenisch, R. 1992, Targeted mutation of the DNA methyltransferase gene results in embryonic lethality, *Cell*, vol. 69, no. 6, pp. 915-926.

- Li, H., Rauch, T., Chen, Z.X., Szabo, P.E., Riggs, A.D. & Pfeifer, G.P. 2006, The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells, *The Journal of biological chemistry*, vol. 281, no. 28, pp. 19489-19500.
- Li, J.Y., Pu, M.T., Hirasawa, R., Li, B.Z., Huang, Y.N., Zeng, R., Jing, N.H., Chen, T., Li, E., Sasaki, H. & Xu, G.L. 2007, Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog, *Molecular and cellular biology*, vol. 27, no. 24, pp. 8748-8759.
- Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D. & Songyang, Z. 2008, Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells, *Nature cell biology*, vol. 10, no. 6, pp. 731-739.
- Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. & Hannon, G.J. 2004, Argonaute2 is the catalytic engine of mammalian RNAi, *Science (New York, N.Y.)*, vol. 305, no. 5689, pp. 1437-1441.
- Liu, Y.P., Haasnoot, J. & Berkhout, B. 2007, Design of extended short hairpin RNAs for HIV-1 inhibition, *Nucleic acids research*, vol. 35, no. 17, pp. 5683-5693.
- Lu, X., Simon, M.D., Chodaparambil, J.V., Hansen, J.C., Shokat, K.M. & Luger, K. 2008, The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure, *Nature structural & molecular biology*, vol. 15, no. 10, pp. 1122-1124.
- Mabuchi, H., Fujii, H., Calin, G., Alder, H., Negrini, M., Rassenti, L., Kipps, T.J., Bullrich, F. & Croce, C.M. 2001, Cloning and characterization of CLLD6, CLLD7, and CLLD8, novel candidate genes for leukemogenesis at chromosome 13q14, a region commonly deleted in B-cell chronic lymphocytic leukemia, *Cancer research*, vol. 61, no. 7, pp. 2870-2877.
- Macleod, D., Charlton, J., Mullins, J. & Bird, A.P. 1994, Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island, *Genes & development*, vol. 8, no. 19, pp. 2282-2292.
- 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*, vol. 78, no. 12, pp. 7634-7638.
- Mattick, J.S. 2005, The functional genomics of noncoding RNA, *Science (New York, N.Y.)*, vol. 309, no. 5740, pp. 1527-1528.
- Matzke, M.A. & Birchler, J.A. 2005, RNAi-mediated pathways in the nucleus, *Nature reviews. Genetics*, vol. 6, no. 1, pp. 24-35.
- Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. & Bird, A.P. 1989, Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs, *Cell*, vol. 58, no. 3, pp. 499-507.
- Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R. & Lander, E.S. 2008, Genome-scale DNA methylation maps of pluripotent and differentiated cells, *Nature*, vol. 454, no. 7205, pp. 766-770.
- Meister, G. & Tuschl, T. 2004, Mechanisms of gene silencing by double-stranded RNA, *Nature*, vol. 431, no. 7006, pp. 343-349.
- Mellor, J., Dudek, P. & Clynes, D. 2008, A glimpse into the epigenetic landscape of gene regulation, *Current opinion in genetics & development*, vol. 18, no. 2, pp. 116-122.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S. & Bernstein, B.E. 2007, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, *Nature*, vol. 448, no. 7153, pp. 553-560.
- Miranda, T.B. & Jones, P.A. 2007, DNA methylation: the nuts and bolts of repression, *Journal of cellular physiology*, vol. 213, no. 2, pp. 384-390.
- 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*, vol. 113, no. 5, pp. 631-642.

- Morris, K.V., Chan, S.W., Jacobsen, S.E. & Looney, D.J. 2004, Small interfering RNA-induced transcriptional gene silencing in human cells, *Science*, vol. 305, no. 5688, pp. 1289-1292.
- 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*, vol. 102, no. 34, pp. 12135-12140.
- Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R. & Fraser, P. 2008, The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin, *Science (New York, N.Y.)*, vol. 322, no. 5908, pp. 1717-1720.
- Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. & Bird, A. 1998, Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex, *Nature*, vol. 393, no. 6683, pp. 386-389.
- Nan, X., Tate, P., Li, E. & Bird, A. 1996, DNA methylation specifies chromosomal localization of MeCP2, *Molecular and cellular biology*, vol. 16, no. 1, pp. 414-421.
- Nelson, D.L.C., M.M. 2000, *Lehninger Principles of Biochemistry*, 3rd edn, worth publishers, new york.
- Nesterova, T.B., Johnston, C.M., Appanah, R., Newall, A.E., Godwin, J., Alexiou, M. & Brockdorff, N. 2003, Skewing X chromosome choice by modulating sense transcription across the Xist locus, *Genes & development*, vol. 17, no. 17, pp. 2177-2190.
- Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. & Bird, A. 1999, MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex, *Nature genetics*, vol. 23, no. 1, pp. 58-61.
- Nichols, J., Zevnik, B., Anastasiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H. & Smith, A. 1998, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4, *Cell*, vol. 95, no. 3, pp. 379-391.
- Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C. & Woodcock, C.L. 2007, Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin, *Molecular and cellular biology*, vol. 27, no. 3, pp. 864-877.
- Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D. & Grewal, S.I. 2004, RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing, *Nature genetics*, vol. 36, no. 11, pp. 1174-1180.
- Norris, D.P., Brockdorff, N. & Rastan, S. 1991, Methylation status of CpG-rich islands on active and inactive mouse X chromosomes, *Mammalian genome : official journal of the International Mammalian Genome Society*, vol. 1, no. 2, pp. 78-83.
- Numata, K., Okada, Y., Saito, R., Kiyosawa, H., Kanai, A. & Tomita, M. 2007, Comparative analysis of cis-encoded antisense RNAs in eukaryotes, *Gene*, vol. 392, no. 1-2, pp. 134-141.
- Nussbaum, R., McInnes, R. & Willard, H. 2007, *Thompson and Thompson Genetics in medicine*, 7th edn, Elsevier Science.
- O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A. & Jenuwein, T. 2001, The polycomb-group gene Ezh2 is required for early mouse development, *Molecular and cellular biology*, vol. 21, no. 13, pp. 4330-4336.
- O'Carroll, D., Scherthan, H., Peters, A.H., Opravil, S., Haynes, A.R., Laible, G., Rea, S., Schmid, M., Lebersorger, A., Jerratsch, M., Sattler, L., Mattei, M.G., Denny, P., Brown, S.D., Schweizer, D. & Jenuwein, T. 2000, Isolation and characterization of Suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression, *Molecular and cellular biology*, vol. 20, no. 24, pp. 9423-9433.
- Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M. & Nakatani, Y. 2002, A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells, *Science (New York, N.Y.)*, vol. 296, no. 5570, pp. 1132-1136.
- Ogawa, Y. & Lee, J.T. 2003, Xite, X-inactivation intergenic transcription elements that regulate the probability of choice, *Molecular cell*, vol. 11, no. 3, pp. 731-743.

- Ogawa, Y., Sun, B.K. & Lee, J.T. 2008, Intersection of the RNA interference and X-inactivation pathways, *Science (New York, N.Y.)*, vol. 320, no. 5881, pp. 1336-1341.
- Ohhata, T., Hoki, Y., Sasaki, H. & Sado, T. 2008, Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification, *Development (Cambridge, England)*, vol. 135, no. 2, pp. 227-235.
- Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M. & Shirakawa, M. 2001, Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA, *Cell*, vol. 105, no. 4, pp. 487-497.
- Okano, M., Bell, D.W., Haber, D.A. & Li, E. 1999, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, *Cell*, vol. 99, no. 3, pp. 247-257.
- Okano, M., Xie, S. & Li, E. 1998, Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nature genetics*, vol. 19, no. 3, pp. 219-220.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., Hayashizaki, Y., FANTOM Consortium & RIKEN Genome Exploration Research Group Phase I & II Team 2002, Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs, *Nature*, vol. 420, no. 6915, pp. 563-573.
- Okitsu, C.Y. & Hsieh, C.L. 2007, DNA methylation dictates histone H3K4 methylation, *Molecular and cellular biology*, vol. 27, no. 7, pp. 2746-2757.
- Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., Cheng, X. & Bestor, T.H. 2007, DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA, *Nature*, vol. 448, no. 7154, pp. 714-717.
- Pall, G.S., Codony-Servat, C., Byrne, J., Ritchie, L. & Hamilton, A. 2007, Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot, *Nucleic acids research*, vol. 35, no. 8, pp. e60.
- Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R. & Thomson, J.A. 2007, Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells, *Cell stem cell*, vol. 1, no. 3, pp. 299-312.
- Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D. & Kanduri, C. 2008, Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation, *Molecular cell*, vol. 32, no. 2, pp. 232-246.
- Papp, B. & Muller, J. 2006, Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins, *Genes & development*, vol. 20, no. 15, pp. 2041-2054.
- Parker, J.S. & Barford, D. 2006, Argonaute: A scaffold for the function of short regulatory RNAs, *Trends in biochemical sciences*, vol. 31, no. 11, pp. 622-630.
- Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E. & Helin, K. 2004, Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity, *The EMBO journal*, vol. 23, no. 20, pp. 4061-4071.
- Patel, S.A., Graunke, D.M. & Pieper, R.O. 1997, Aberrant silencing of the CpG island-containing human O6-methylguanine DNA methyltransferase gene is associated with the loss of nucleosome-like positioning, *Molecular and cellular biology*, vol. 17, no. 10, pp. 5813-5822.
- Patnaik, D., Chin, H.G., Esteve, P.O., Benner, J., Jacobsen, S.E. & Pradhan, S. 2004, Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase, *The Journal of biological chemistry*, vol. 279, no. 51, pp. 53248-53258.
- Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M. & Jenuwein, T. 2001, Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability, *Cell*, vol. 107, no. 3, pp. 323-337.
- Peters, L. & Meister, G. 2007, Argonaute proteins: mediators of RNA silencing, *Molecular cell*, vol. 26, no. 5, pp. 611-623.

- Peterson, C.L. & Laniel, M.A. 2004, Histones and histone modifications, *Current biology : CB*, vol. 14, no. 14, pp. R546-51.
- Pickart, C.M. & Eddins, M.J. 2004, Ubiquitin: structures, functions, mechanisms, *Biochimica et biophysica acta*, vol. 1695, no. 1-3, pp. 55-72.
- Ponjavic, J., Ponting, C.P. & Lunter, G. 2007, Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs, *Genome research*, vol. 17, no. 5, pp. 556-565.
- Ramsahoye, B.H., Biniszkiwicz, D., Lyko, F., Clark, V., Bird, A.P. & Jaenisch, R. 2000, Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 10, pp. 5237-5242.
- Redrup, L., Branco, M.R., Perdeaux, E.R., Krueger, C., Lewis, A., Santos, F., Nagano, T., Cobb, B.S., Fraser, P. & Reik, W. 2009, The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing, *Development (Cambridge, England)*, vol. 136, no. 4, pp. 525-530.
- Regha, K., Sloane, M.A., Huang, R., Pauler, F.M., Warczok, K.E., Melikant, B., Radolf, M., Martens, J.H., Schotta, G., Jenuwein, T. & Barlow, D.P. 2007, Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome, *Molecular cell*, vol. 27, no. 3, pp. 353-366.
- Rehwinkel, J., Natalin, P., Stark, A., Brennecke, J., Cohen, S.M. & Izaurralde, E. 2006, Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in *Drosophila melanogaster*, *Molecular and cellular biology*, vol. 26, no. 8, pp. 2965-2975.
- Reik, W. 2007, Stability and flexibility of epigenetic gene regulation in mammalian development, *Nature*, vol. 447, no. 7143, pp. 425-432.
- Reik, W., Dean, W. & Walter, J. 2001, Epigenetic reprogramming in mammalian development, *Science (New York, N.Y.)*, vol. 293, no. 5532, pp. 1089-1093.
- Reik, W. & Lewis, A. 2005, Co-evolution of X-chromosome inactivation and imprinting in mammals, *Nature reviews. Genetics*, vol. 6, no. 5, pp. 403-410.
- Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y. & Allis, C.D. 2003, Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains, *Molecular cell*, vol. 12, no. 6, pp. 1591-1598.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. & Chang, H.Y. 2007, Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, *Cell*, vol. 129, no. 7, pp. 1311-1323.
- Robertson, K.D., Ait-Si-Ali, S., Yokochi, T., Wade, P.A., Jones, P.L. & Wolffe, A.P. 2000a, DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters, *Nature genetics*, vol. 25, no. 3, pp. 338-342.
- Robertson, K.D., Keyomarsi, K., Gonzales, F.A., Velicescu, M. & Jones, P.A. 2000b, Differential mRNA expression of the human DNA methyltransferases (DNMTs) 1, 3a and 3b during the G(0)/G(1) to S phase transition in normal and tumor cells, *Nucleic acids research*, vol. 28, no. 10, pp. 2108-2113.
- Rossant, J. & Tam, P.P. 2009, Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse, *Development (Cambridge, England)*, vol. 136, no. 5, pp. 701-713.
- Rougeulle, C., Chaumeil, J., Sarma, K., Allis, C.D., Reinberg, D., Avner, P. & Heard, E. 2004, Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome, *Molecular and cellular biology*, vol. 24, no. 12, pp. 5475-5484.
- Sasaki, T., Shiohama, A., Minoshima, S. & Shimizu, N. 2003, Identification of eight members of the Argonaute family in the human genome small star, filled, *Genomics*, vol. 82, no. 3, pp. 323-330.
- Saxonov, S., Berg, P. & Brutlag, D.L. 2006, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 5, pp. 1412-1417.

- Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C. & Leonhardt, H. 2007, Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation, *Nucleic acids research*, vol. 35, no. 13, pp. 4301-4312.
- Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., Bergman, Y., Simon, I. & Cedar, H. 2007, Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer, *Nature genetics*, vol. 39, no. 2, pp. 232-236.
- Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T. & Wutz, A. 2006, Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing, *The EMBO journal*, vol. 25, no. 13, pp. 3110-3122.
- Schorpp, M., Jager, R., Schellander, K., Schenkel, J., Wagner, E.F., Weiher, H. & Angel, P. 1996, The human ubiquitin C promoter directs high ubiquitous expression of transgenes in mice, *Nucleic acids research*, vol. 24, no. 9, pp. 1787-1788.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. 2007, Genome regulation by polycomb and trithorax proteins, *Cell*, vol. 128, no. 4, pp. 735-745.
- Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G. & Rauscher, F.J., 3rd 2002, SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins, *Genes & development*, vol. 16, no. 8, pp. 919-932.
- Seidl, C.I., Stricker, S.H. & Barlow, D.P. 2006, The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export, *The EMBO journal*, vol. 25, no. 15, pp. 3565-3575.
- Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A. & Sharp, P.A. 2008, Divergent transcription from active promoters, *Science (New York, N.Y.)*, vol. 322, no. 5909, pp. 1849-1851.
- Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M. & Koseki, H. 2007, The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, *Nature*, vol. 450, no. 7171, pp. 908-912.
- Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A. & Issa, J.P. 2007, Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters, *PLoS genetics*, vol. 3, no. 10, pp. 2023-2036.
- Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R. & Peterson, C.L. 2006, Histone H4-K16 acetylation controls chromatin structure and protein interactions, *Science (New York, N.Y.)*, vol. 311, no. 5762, pp. 844-847.
- Sigova, A., Rhind, N. & Zamore, P.D. 2004, A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in *Schizosaccharomyces pombe*, *Genes & development*, vol. 18, no. 19, pp. 2359-2367.
- Sleutels, F., Tjon, G., Ludwig, T. & Barlow, D.P. 2003, Imprinted silencing of Slc22a2 and Slc22a3 does not need transcriptional overlap between Igf2r and Air, *The EMBO journal*, vol. 22, no. 14, pp. 3696-3704.
- Smallwood, A., Esteve, P.O., Pradhan, S. & Carey, M. 2007, Functional cooperation between HP1 and DNMT1 mediates gene silencing, *Genes & development*, vol. 21, no. 10, pp. 1169-1178.
- Smith, A.G. 2001, Embryo-derived stem cells: of mice and men, *Annual Review of Cell and Developmental Biology*, vol. 17, pp. 435-462.
- Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H. & Held, W.A. 2005, Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 9, pp. 3336-3341.
- Song, J., Giang, A., Lu, Y., Pang, S. & Chiu, R. 2008, Multiple shRNA expressing vector enhances efficiency of gene silencing, *BMB reports*, vol. 41, no. 5, pp. 358-362.
- Stavropoulos, N., Rowntree, R.K. & Lee, J.T. 2005, Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation, *Molecular and cellular biology*, vol. 25, no. 7, pp. 2757-2769.

- Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G. & Pombo, A. 2007, Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells, *Nature cell biology*, vol. 9, no. 12, pp. 1428-1435.
- Stoger, R., Kubicka, P., Liu, C.G., Kafri, T., Razin, A., Cedar, H. & Barlow, D.P. 1993, Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal, *Cell*, vol. 73, no. 1, pp. 61-71.
- Strahl, B.D. & Allis, C.D. 2000, The language of covalent histone modifications, *Nature*, vol. 403, no. 6765, pp. 41-45.
- Stricker, S.H., Steenpass, L., Pauler, F.M., Santoro, F., Latos, P.A., Huang, R., Koerner, M.V., Sloane, M.A., Warczok, K.E. & Barlow, D.P. 2008, Silencing and transcriptional properties of the imprinted Airn ncRNA are independent of the endogenous promoter, *The EMBO journal*, vol. 27, no. 23, pp. 3116-3128.
- Struhl, K. 2007, Transcriptional noise and the fidelity of initiation by RNA polymerase II, *Nature structural & molecular biology*, vol. 14, no. 2, pp. 103-105.
- Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. & Tajima, S. 2004, DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction, *The Journal of biological chemistry*, vol. 279, no. 26, pp. 27816-27823.
- Sugiyama, T., Cam, H.P., Sugiyama, R., Noma, K., Zofall, M., Kobayashi, R. & Grewal, S.I. 2007, SHREC, an effector complex for heterochromatic transcriptional silencing, *Cell*, vol. 128, no. 3, pp. 491-504.
- Sun, B.K., Deaton, A.M. & Lee, J.T. 2006, A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization, *Molecular cell*, vol. 21, no. 5, pp. 617-628.
- suzuki, K., Shijuuku, T., Fukamachi, T., Zaunders, J., Guillemin, G., Cooper, D. & Kelleher, A. 2005, prolonged transcriptional silencing and CpG methylation induced by siRNAs targeted to the HIV-1 promoter region, *Journal of RNAi and gene silencing*, vol. 1, no. 2, pp. 66-67-78.
- Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H. & Shinkai, Y. 2008, G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription, *The EMBO journal*, vol. 27, no. 20, pp. 2681-2690.
- Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H. & Shinkai, Y. 2002, G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis, *Genes & development*, vol. 16, no. 14, pp. 1779-1791.
- Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T. & Shinkai, Y. 2005, Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9, *Genes & development*, vol. 19, no. 7, pp. 815-826.
- Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K., Irvine, K., Arakawa, T., Nakamura, M., Kubosaki, A., Hayashida, K., Kawazu, C., Murata, M., Nishiyori, H., Fukuda, S., Kawai, J., Daub, C.O., Hume, D.A., Suzuki, H., Orlando, V., Carninci, P., Hayashizaki, Y. & Mattick, J.S. 2009, Tiny RNAs associated with transcription start sites in animals, *Nature genetics*, vol. 41, no. 5, pp. 572-578.
- Tapscott, S.J. 2005, The circuitry of a master switch: MyoD and the regulation of skeletal muscle gene transcription, *Development (Cambridge, England)*, vol. 132, no. 12, pp. 2685-2695.
- Terranova, R., Yokobayashi, S., Stadler, M.B., Otte, A.P., van Lohuizen, M., Orkin, S.H. & Peters, A.H. 2008, Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos, *Developmental cell*, vol. 15, no. 5, pp. 668-679.
- Tilghman, S.M. 1999, The sins of the fathers and mothers: genomic imprinting in mammalian development, *Cell*, vol. 96, no. 2, pp. 185-193.
- Ting, A.H., Schuebel, K.E., Herman, J.G. & Baylin, S.B. 2005, Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation, *Nature genetics*, vol. 37, no. 8, pp. 906-910.
- Tolia, N.H. & Joshua-Tor, L. 2007, Slicer and the argonautes, *Nature chemical biology*, vol. 3, no. 1, pp. 36-43.
- Tost, J. 2008, *Epigenetics*, 1st edn, caister accademic press.

- Tse, C., Sera, T., Wolffe, A.P. & Hansen, J.C. 1998, Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III, *Molecular and cellular biology*, vol. 18, no. 8, pp. 4629-4638.
- Tufarelli, C., Frischauf, A.M., Hardison, R., Flint, J. & Higgs, D.R. 2001, Characterization of a widely expressed gene (LUC7-LIKE; LUC7L) defining the centromeric boundary of the human alpha-globin domain, *Genomics*, vol. 71, no. 3, pp. 307-314.
- Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G. & Higgs, D.R. 2003, Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease, *Nature genetics*, vol. 34, no. 2, pp. 157-165.
- Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T., Ohki-Hamazaki, H., Juni, A., Ueda, R. & Saigo, K. 2004, Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference, *Nucleic acids research*, vol. 32, no. 3, pp. 936-948.
- Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y. & Feil, R. 2004, Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes, *Nature genetics*, vol. 36, no. 12, pp. 1296-1300.
- Vakoc, C.R., Mandat, S.A., Olenchok, B.A. & Blobel, G.A. 2005, Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin, *Molecular cell*, vol. 19, no. 3, pp. 381-391.
- Vakoc, C.R., Sachdeva, M.M., Wang, H. & Blobel, G.A. 2006, Profile of histone lysine methylation across transcribed mammalian chromatin, *Molecular and cellular biology*, vol. 26, no. 24, pp. 9185-9195.
- van der Stoep, P., Boutsma, E.A., Hulsman, D., Noback, S., Heimerikx, M., Kerkhoven, R.M., Voncken, J.W., Wessels, L.F. & van Lohuizen, M. 2008, Ubiquitin E3 ligase Ring1b/Rnf2 of polycomb repressive complex 1 contributes to stable maintenance of mouse embryonic stem cells, *PLoS ONE*, vol. 3, no. 5, pp. e2235.
- Vaute, O., Nicolas, E., Vandel, L. & Trouche, D. 2002, Functional and physical interaction between the histone methyltransferase Suv39H1 and histone deacetylases, *Nucleic acids research*, vol. 30, no. 2, pp. 475-481.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I. & Moazed, D. 2004, RNAi-mediated targeting of heterochromatin by the RITS complex, *Science (New York, N.Y.)*, vol. 303, no. 5658, pp. 672-676.
- Verma R & Deshaies RJ 2005, Assaying degradation and deubiquitination of a ubiquitinated substrate by purified 26S proteasomes. *Methods in enzymology*, vol. 398, pp. 391-9.
- Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y. & Fuks, F. 2006, The Polycomb group protein EZH2 directly controls DNA methylation, *Nature*, vol. 439, no. 7078, pp. 871-874.
- Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. & Martienssen, R.A. 2002, Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi, *Science*, vol. 297, no. 5588, pp. 1833-1837.
- Voncken, J.W., Roelen, B.A., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J. & van Lohuizen, M. 2003, Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2468-2473.
- Wade, P.A., Geronne, A., Jones, P.L., Ballestar, E., Aubry, F. & Wolffe, A.P. 1999, Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation, *Nature genetics*, vol. 23, no. 1, pp. 62-66.
- Wagschal, A., Sutherland, H.G., Woodfine, K., Henckel, A., Chebli, K., Schulz, R., Oakey, R.J., Bickmore, W.A. & Feil, R. 2008, G9a histone methyltransferase contributes to imprinting in the mouse placenta, *Molecular and cellular biology*, vol. 28, no. 3, pp. 1104-1113.
- Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J. & Wong, G.K. 2004, Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs, *Nature*, vol. 431, no. 7010, pp. 1 p following 757; discussion following 757.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L. & Schubeler, D. 2005, Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells, *Nature genetics*, vol. 37, no. 8, pp. 853-862.

- Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M. & Schubeler, D. 2007, Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome, *Nature genetics*, vol. 39, no. 4, pp. 457-466.
- Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguoui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J. & Morris, K.V. 2006, The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells, *RNA (New York, N.Y.)*, vol. 12, no. 2, pp. 256-262.
- Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I. & Laird, P.W. 2007, Epigenetic stem cell signature in cancer, *Nature genetics*, vol. 39, no. 2, pp. 157-158.
- Wiles, M.V. & Keller, G. 1991, Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture, *Development (Cambridge, England)*, vol. 111, no. 2, pp. 259-267.
- Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B. & Schultz, P.G. 2005, A strategy for probing the function of noncoding RNAs finds a repressor of NFAT, *Science (New York, N.Y.)*, vol. 309, no. 5740, pp. 1570-1573.
- Wu, H.A. & Bernstein, E. 2008, Partners in imprinting: noncoding RNA and polycomb group proteins, *Developmental cell*, vol. 15, no. 5, pp. 637-638.
- Wutz, A. 2007, Xist function: bridging chromatin and stem cells, *Trends in genetics : TIG*, vol. 23, no. 9, pp. 457-464.
- Yang, L., Mei, Q., Zielinska-Kwiatkowska, A., Matsui, Y., Blackburn, M.L., Benedetti, D., Krumm, A.A., Taborsky, G.J., Jr & Chansky, H.A. 2003, An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B, *The Biochemical journal*, vol. 369, no. Pt 3, pp. 651-657.
- Yang, L., Xia, L., Wu, D.Y., Wang, H., Chansky, H.A., Schubach, W.H., Hickstein, D.D. & Zhang, Y. 2002, Molecular cloning of ESET, a novel histone H3-specific methyltransferase that interacts with ERG transcription factor, *Oncogene*, vol. 21, no. 1, pp. 148-152.
- Yoder, J.A., Walsh, C.P. & Bestor, T.H. 1997, Cytosine methylation and the ecology of intragenomic parasites, *Trends in genetics : TIG*, vol. 13, no. 8, pp. 335-340.
- Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P. & Cui, H. 2008, Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA, *Nature*, vol. 451, no. 7175, pp. 202-206.
- Yuan, B., Latek, R., Hossbach, M., Tuschl, T. & Liewitter, F. 2004, siRNA Selection Server: an automated siRNA oligonucleotide prediction server, *Nucleic acids research*, vol. 32, no. Web Server issue, pp. W130-4.
- Zhang, L.F., Huynh, K.D. & Lee, J.T. 2007, Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing, *Cell*, vol. 129, no. 4, pp. 693-706.
- Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. 1997, Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex, *Cell*, vol. 89, no. 3, pp. 357-364.
- Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A. & Reinberg, D. 1999, Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation, *Genes & development*, vol. 13, no. 15, pp. 1924-1935.
- Zhang, Y. & Reinberg, D. 2001, Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails, *Genes & development*, vol. 15, no. 18, pp. 2343-2360.
- Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. 2008, Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome, *Science (New York, N.Y.)*, vol. 322, no. 5902, pp. 750-756.
- Zhao, X.D., Han, X., Chew, J.L., Liu, J., Chiu, K.P., Choo, A., Orlov, Y.L., Sung, W.K., Shahab, A., Kuznetsov, V.A., Bourque, G., Oh, S., Ruan, Y., Ng, H.H. & Wei, C.L. 2007, Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells, *Cell stem cell*, vol. 1, no. 3, pp. 286-298.
- Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K. & Rosenfeld, M.G. 2008, Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation, *Molecular cell*, vol. 29, no. 1, pp. 69-80.

Zilberman, D. 2007, The human promoter methylome, *Nature genetics*, vol. 39, no. 4, pp. 442-443.

Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. 2008, Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks, *Nature*, vol. 456, no. 7218, pp. 125-129.