The NuRD complex has a role in the specification of DNA methylation patterns in pluripotent cells

Christine Powell

The Institute for Stem Cell Research

The University of Edinburgh

Thesis presented for the Degree of Doctor of Philosophy

2007
Acknowledgements

There are lots of people who have helped me over the last three years and without whom I no doubt would not have got to this stage and I would like to thank all of them. Firstly my supervisor Brian Hendrich for his advice, support and encouragement when I needed it most. Also thanks to the rest of the lab – Kei for his invaluable help and advice and for never getting annoyed with the questions, Ita for making me laugh and helping me keep it all in perspective, and Isabel for being the first student in the lab to finish and proving to all of us that it could be done! Thanks are due also to Jen Back for her help at the start of my PhD, and the other members of the lab who have come and gone during my time here. I must also mention other members of the institute who have been helpful, in particular my second supervisor Andrew Smith for useful and positive feedback, Jenny Nichols and Susan Tweedie for their positive encouragement, and members of Meng’s lab, especially Suling and Sarah for their advice and support.

My time in Edinburgh has been great, and I must thank all the friends I have made up here who have been vital in helping make it that way. Thanks to my family for always encouraging me to do my best and supporting me through the hardest times. Last but not least I cannot thank my husband-to-be Pete enough for having such faith in me and being so supportive and encouraging all the way through. It would have been so much more difficult without you.
Abstract

The NuRD (Nucleosome Remodelling and Deacetylation) complex is a highly conserved and abundant transcriptional repression complex. Embryonic Stem (ES) cells lacking a central structural component of the complex, Mbd3, proliferate slowly and are unable to differentiate in the absence of LIF. We show here that in addition they show significant demethylation of DNA at a number of both repetitive sequences, such as the centromeric repeats and IAP elements, as well as some, but not all, single-copy DMRs and CpG islands throughout the genome. This demethylation is associated with histone hyperacetylation and aberrant transcription of these sequences or, in the case of regulatory regions, genes controlled by them. Many of the sequences demethylated here are also demethylated in Dnmt3a, 3b−/− ES cells, and interestingly Mbd3−/− ES cells contain significantly reduced levels of the methyltransferase Dnmt3b. However Dnmt3b−/− ES cells are demethylated to a lesser extent than Mbd3−/− cells, suggesting that a decrease in Dnmt3b levels is not the sole cause of the demethylation observed here. Moreover, Dnmt3b cannot be ectopically expressed in an Mbd3−/− background – attempts to do so result in growth arrest and significant differentiation. An interaction has been detected between Dnmt3b and the NuRD components Mta2 and Mbd3 in wild type cells, leading to speculation that they may co-operate to bring about a repressive chromatin structure at target sites, but chromatin immunoprecipitation of a tagged version of Mbd3 and another NuRD component, Mi-2β, does not reveal binding of the complex at affected sequences. However immunofluorescence of Mbd3, Mta2 and Mi-2β reveals localisation to centromeric regions in approximately 30% of wild type cells, suggesting localisation is cell-cycle dependent and therefore not detectable by ChIP. Importantly, the degree of centromeric localisation of Mta2 is significantly reduced in Mbd3−/− cells. I propose that NuRD binds to target sequences in a cell cycle dependent manner, where it participates with DNA methyltransferases in the formation and/or maintenance of a repressive chromatin structure.
## Table of Contents

ACKNOWLEDGEMENTS 3

ABSTRACT 4

TABLE OF CONTENTS 5

LIST OF TABLES AND FIGURES 10

COMMON ABBREVIATIONS 12

CHAPTER ONE 14

1. INTRODUCTION 14

1.1 Epigenetics 14

1.2 Chromatin 14

1.1.1 Features of heterochromatin 15

1.1.2 Constitutively heterochromatic repetitive sequences 16

1.1.3 Histone acetylation 18

1.1.4 Histone methylation 19

1.1.5 Phosphorylation, ubiquitination, ADP-ribosylation and sumoylation 20

1.1.6 Chromatin remodelling 20

1.3 DNA methylation 21

1.3.1 DNA methylation and transcriptional repression 21

1.3.2 Connections between DNA methylation, histone modifications and chromatin remodelling 23

1.3.3 DNA methylation and chromatin remodelling 25

1.3.4 Replication of heterochromatin 26

1.4 DNA methylation in development 26

1.4.1 Imprinting 27

1.4.2 X-Chromosome Inactivation 29

1.4.3 DNA methylation is required for the differentiation but not self-renewal of ES cells 31

1.4.4 Protection of genome stability 31

1.4.5 Epigenetic reprogramming in development 33

1.4.6 DNA methylation and disease 36

1.5 The DNA methyltransferases 36

1.5.1 Dnmt1 37

1.5.1.1 Dnmt1 in vivo 40

1.5.2 Dnmt2 41

1.5.3 The Dnmt3 family 41

1.5.3.1 Dnmt3a and 3b – structural features 42

1.5.3.2 In vivo function of Dnmt3a and 3b 42
1.5.3.3 Dnmt3a and imprinting 43
1.5.3.4 Dnmt3b, ICF and cancer 44
1.5.3.5 Dnmt3a and 3b in ES cells – a contribution to maintenance methylation 45
1.5.4 Regulation of Dnmt3b transcript levels 46
1.5.4.1 Dnmt3a and 3b isoforms 47
1.5.4.2 Dnmt interactions 50
1.5.4.3 Targeting of the Dnmts to DNA 50
1.5.4.4 Dnmt3l 51

1.6 Embryonic Stem Cells 53
1.6.1 ES cells and the Inner Cell Mass (ICM) 53
1.6.2 Self-renewal of ES cells – a dependence on cytokine signalling 55
1.6.3 Pluripotency 55
1.6.4 ‘Master regulators’ of pluripotency? 56
1.6.4.1 Oct4 56
1.6.4.2 Sox2 57
1.6.4.3 Nanog 57
1.6.4.4 Co-operativity between Oct4, Nanog and Sox2 in gene regulation 57
1.6.4.5 Other regulators of ES cell fate 59
1.6.5 Polycomb group proteins in ES cells 60
1.6.6 Epigenetic features of ES cells 61
1.6.7 A unique permissive chromatin structure in pluripotent cells? 62

1.7 The NuRD complex 65
1.7.1 Function of NuRD 65
1.7.2 Composition of the complex 66
1.7.3 Mechanism of action 71
1.7.4 Mbd3 targeting strategy 72
1.7.5 The Mbd3\(^{(-/-)}\) phenotype 75

1.8 Thesis Aims 80

2 MATERIALS AND METHODS 81

2.1 Common solutions 81

2.2 Bacterial techniques 81
2.2.1 LB Media and agar 81
2.2.2 Production of agar plates 81
2.2.3 Production of competent cells 82
2.2.4 Glycerol storage of bacterial stocks 82
2.2.5 Transformation of bacterial competent cells 82
2.2.6 Miniprep procedure 83
2.2.7 Midi/maxi-preparation of plasmid DNA 83

2.3 DNA techniques 83
2.3.1 TA cloning 83
2.3.2 Ligations 84
2.3.3 Sequencing 84
2.3.4 UV spectrophotometry 84

2.4 Bisulfite Mutagenesis of genomic DNA 84
2.4.1 Nested PCR amplification of bisulfite modified DNA 85
2.5 Reverse-Transcription PCR
2.5.1 RNA isolation
2.5.2 DNaseI treatment
2.5.3 Reverse Transcription (oligo dT)
2.5.4 Reverse Transcription (random primers)
2.5.5 PCR amplification of cDNA

2.6 Chromatin immunoprecipitations
2.6.1 Bead blocking
2.6.2 Fixation, lysis, sonication and pre-clearing
2.6.3 Antibody incubation and binding to beads
2.6.4 Elution from beads and reversal of crosslinks
2.6.5 Purification and precipitation of chromatin fragments
2.6.6 Input samples
2.6.7 Chromatin immunoprecipitation of biotin-tagged proteins

2.7 Real-time PCR
2.7.1 Standard curves

2.8 Protein Techniques
2.8.1 Extraction of nuclear proteins from ES cells
2.8.2 Quantification using the Bradford Assay method
2.8.3 Separation of proteins on acrylamide gels and detection by antibodies (SDS-PAGE and Western Blotting)
  2.8.3.1 Gel preparation
  2.8.3.2 Sample preparation
  2.8.3.3 Wet transfer onto nitrocellulose membranes
  2.8.3.4 Blocking, antibody incubations and washing
  2.8.3.5 Signal detection
  2.8.3.6 Membrane stripping and re-probing
2.8.4 Immunoprecipitations
  2.8.4.1 Preclearing
  2.8.4.2 Immunoprecipitation
  2.8.4.3 Washing, sample preparation and gel loading
  2.8.4.4 Immunoprecipitation of biotin-tagged proteins
2.8.5 Immunofluorescence
  2.8.5.1 Fixing and permeabilisation
  2.8.5.2 Blocking, antibody incubation and washing
  2.8.5.3 Mounting and visualisation

2.9 In vitro measurement of de novo DNA methyltransferase activity

2.10 Tissue Culture
2.10.1 Thawing of ES cells
2.10.2 Passaging of ES cells
2.10.3 Freezing of ES cells
2.10.4 Transient transfection of ES cells using Transfast
  2.10.4.1 Preparation of Transfast reagent
  2.10.4.2 Preparation of DNA to be transfected
  2.10.4.3 Preparation of cells and transfection procedure
2.10.5 Stable transfection of ES cells using Transfast
2.10.6 Stable transfection of ES cells by electroporation
  2.10.6.1 Preparation of DNA
  2.10.6.2 Preparation of cells and electroporation
2.10.7 Picking of Colonies
CHAPTER THREE

3 PERTURBED DNA METHYLATION AND HISTONE MODIFICATION PATTERNS IN Mbd3\(^{(−/−)}\) ES CELLS

3.1 Introduction

3.2 Results

3.2.1 Reduced DNA methylation in Mbd3\(^{(−/−)}\) ES cells

3.2.2 Demethylation is associated with histone hyperacetylation

3.2.3 No significant alterations to histone methylation in null cells

3.2.4 No changes to the localisation of HP1\(\alpha\) or Lsh

3.2.5 Aberrant transcription of demethylated, hyperacetylated sequences

3.3 Discussion

3.3.1 How is NuRD involved in DNA methylation?

3.3.2 Do the epigenetic aberrations cause transcriptional misregulation or vice-versa?

3.3.3 Increased transcription of repetitive sequences

3.3.4 A contribution by transcription misregulation towards the Mbd3 phenotype?

CHAPTER FOUR

4 Mbd3\(^{(−/−)}\) ES CELLS CONTAIN REDUCED LEVELS OF Dnmt3b

4.1 Introduction

4.2 Results

4.2.1 Mbd3\(^{(−/−)}\) cells have reduced levels of Dnmt3b mRNA

4.2.2 Dnmt3b(1) is the main isoform expressed in mES cells

4.2.3 Mbd3\(^{(−/−)}\) cells have reduced levels of Dnmt3b protein

4.2.4 The decrease in Dnmt3b(1) transcript levels is not caused by decreased stability of the transcript in Mbd3\(^{(−)}\) cells

4.2.5 An in vitro assay reveals no detectable decrease in de novo methyltransferase activity in Mbd3\(^{(−/−)}\) cells

4.2.6 Ectopic expression of Dnmt3b(1) in Mbd3\(^{(−/−)}\) ES cells

4.2.6.1 Ectopic expression of Dnmt3b(1) leads to growth arrest and differentiation in Mbd3\(^{(−/−)}\) ES cells but not Mbd3\(^{(−/−)}\) ES cells

4.2.6.2 Increased Gata-4, and decreased Oct4 expression in Mbd3\(^{(−/−)}\) ES cells overexpressing Dnmt3b(1)
4.3 Discussion
4.3.1 The NuRD complex has a role in the control of Dnmt3b transcript levels
4.3.2 A reduction in Dnmt3b protein levels cannot solely explain the demethylation observed in Mbd3\(^{(-/-)}\) ES cells
4.3.3 Differentiation of Mbd3\(^{(-/-)}\)/Dnmt3b OE cells towards the endoderm lineage?

5 NURD AND Dnmt3b INTERACT AND CO-LOCALISE IN A SUBSET OF ES CELLS
5.1 5.1 Introduction
5.2 Results
5.2.1 NuRD binding is not detectable at any of the demethylated, hyperacetylated sequences by chromatin immunoprecipitation
5.2.1.1 Biotinylation strategy
5.2.2 Immunofluorescence reveals localisation of the NuRD complex to centromeric regions in a subset of wild type cells in a Mbd3 dependent manner
5.2.2.1 Mta2 and Mbd3 localise to centromeric regions in a subset of wild type cells
5.2.3 The NuRD complex interacts with Dnmt3b
5.2.3.1 Dnmt3b interacts with biotin-tagged Mbd3b
5.2.3.2 Mta2 and Mbd3 interact with Dnmt3b

5.3 Discussion
5.3.1 NuRD binds to centromeric regions in only a subset of wild type ES cells
5.3.2 NuRD is involved in the maintenance of DNA methylation patterns by Dnmt3b
5.3.3 A role for NuRD in re-establishing a heterochromatic structure following DNA replication?

6 GENERAL DISCUSSION AND FUTURE DIRECTIONS
6.1 The NuRD complex and DNA methylation
6.2 A role in de novo methylation?
6.3 DNA methylation, NuRD, and the differentiation of pluripotent cells
6.4 Mbd3 and DNA methylation in non-pluripotent cells?

APPENDICES

BIBLIOGRAPHY

WEB LINKS
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Epigenetic reprogramming during development</td>
<td>35</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>The DNA methyltransferases</td>
<td>39</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Dnmt3a and Dnmt3b isoforms</td>
<td>49</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>The ‘poised’ chromatin state in pluripotent cells</td>
<td>64</td>
</tr>
<tr>
<td>Figure 1.5.a</td>
<td>The NuRD complex</td>
<td>70</td>
</tr>
<tr>
<td>Figure 1.5.b</td>
<td>Mbd3 isoforms</td>
<td>70</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Mbd3 targeting strategy</td>
<td>73</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>Demethylation of repetitive sequences in Mbd3&lt;sup&gt;-/-&lt;/sup&gt; ES cells</td>
<td>78</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>Demethylation of single copy sequences in Mbd3&lt;sup&gt;-/-&lt;/sup&gt; ES cells</td>
<td>79</td>
</tr>
<tr>
<td>Table 2.1</td>
<td>Expression primer sequences</td>
<td>101</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>ChIP primer sequences</td>
<td>102</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Bisulfite sequencing primer sequences</td>
<td>102</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Schematic representation of the differentially methylated regions at the H19/Igf2 and Dlk1/Gtl2 loci</td>
<td>115</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Demethylation of various single copy sequences in Mbd3&lt;sup&gt;-/-&lt;/sup&gt; ES cells</td>
<td>116</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Summary of bisulfite sequencing data for imprinted regions</td>
<td>117</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Changes to histone acetylation and methylation</td>
<td>118</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Immunofluorescence of Me3H3K9, HP1 and Lsh</td>
<td>119</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Transcriptional misregulation in Mbd3 null ES cells</td>
<td>121</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Reduced Dnmt3b transcript levels in Mbd3 null ES cells</td>
<td>139</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Schematic diagram of Dnmt3b(1), Dnmt3b(1) is the main isoform expressed in WT ES cells</td>
<td>140</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Reduced amount of Dnmt3b protein in Mbd3 null ES cells</td>
<td>141</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Mbd3&lt;sup&gt;-/-&lt;/sup&gt; ES cells do not have an increased rate of Dnmt3b transcript turnover compared to Mbd3&lt;sup&gt;(flo&lt;sup&gt;x&lt;/sup&gt;)&lt;/sup&gt; cells</td>
<td>141</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>No decrease in de novo methyltransferase activity in Mbd3&lt;sup&gt;(flo&lt;sup&gt;x&lt;/sup&gt;)&lt;/sup&gt; ES cells</td>
<td>142</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Abnormal morphology of Mbd3&lt;sup&gt;-/-&lt;/sup&gt; ES cells overexpressing Dnmt3b</td>
<td>143</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Level of Dnmt3b overexpression in three Mbd3&lt;sup&gt;(flo&lt;sup&gt;x&lt;/sup&gt;)&lt;/sup&gt; cell lines</td>
<td>143</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Reduced number of colonies following deletion of Mbd3 from Dnmt3b overexpressing cells compared to wild type cells</td>
<td>144</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Abnormal morphology of Dnmt3b overexpressing cells following deletion of Mbd3</td>
<td>144</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Reduced alkaline phosphatase staining</td>
<td>145</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>PCR genotyping confirms <em>Mbd3</em> deletion</td>
<td>146</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Decreased <em>Oct4</em> and increased <em>Gata4</em> expression in <em>Mbd3</em> ES cells overexpressing <em>Dnmt3b</em></td>
<td>147</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Chromatin immunoprecipitation of biotin-tagged Mbd3 and Mi2β</td>
<td>163</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Immunofluorescence of NuRD components Mbd3 and Mta2</td>
<td>164</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>Immunoprecipitation of Dnmt3b with both biotin-tagged and endogenous NuRD components</td>
<td>165</td>
</tr>
<tr>
<td>Appendix I</td>
<td>Map of Dnmt3b(1) expression vector</td>
<td>179</td>
</tr>
<tr>
<td>Appendix II</td>
<td>Map of Cre expression vector</td>
<td>180</td>
</tr>
</tbody>
</table>
List of Common Abbreviations

Amp: Ampicillin
APS: Ammonium Persulfate
ATP: Adenosine Triphosphate
Bp: Base pairs
BSA: Bovine Serum Albumin
C-: Carboxyl
ChIP: Chromatin Immunoprecipitation
DAPI: 4',6-Diamidino-2-phenylindole
DEPC: Diethyl pyrocarbonate
DMSO: Dimethyl Sulfoxide
DNase: Deoxyribonucleoside
Dnmt: DNA methyltransferase
dNTP: Deoxyribonucleoside triphosphate
DTT: Dithiothreitol
EDTA: Diaminoethanetetraacetic acid
ES: Embryonic stem
g: Relative centrifugal force
HCl: Hydrogen Chloride
Hdac: Histone deacetylase
Hepes: N-2-hydroxyethylpeperazine-N'-2-ethanesulfonic acid
HMT: Histone methyltransferase
IF: immunofluorescence
IP: Immunoprecipitation
IPTG: Isopropyl-β-D-1-thiogalactopyranoside
kb: kilobase
kDa: Kilodalton
l: litre
LIF: Leukocyte Inhibitory Factor
Chapter One

1. Introduction

1.1 Epigenetics

How is it that during the development of a multicellular organism, a single fertilized egg gives rise to a plethora of specialized cell types, all of which contain the same genetic information, but which differ in how they read this information? This is a question that has puzzled developmental biologists for decades. It was in the 1950’s that the term ‘Epigenetics’ was first introduced by Conrad Waddington, when he introduced the idea of an ‘epigenetic landscape’ to help explain how decision making by genes is controlled differently in different cells of the developing embryo (Waddington, 1957). Epigenetics is currently defined as ‘a mitotically and/or meiotically heritable change in gene function that cannot be explained by changes in DNA sequence’ (Wu and Morris, 2001). These days, the field of epigenetics is primarily concerned with changes to the structure and/or chemical modifications of chromatin, the nucleoprotein filamentous structure contained within eukaryotic nuclei that allows the DNA to be packaged tightly yet accessed when required.

1.2 Chromatin

It is the presence of a chromatin structure that allows approximately twoM of DNA to be packaged into a human cell nucleus with an average diameter of 10μM (Mohd-Sarip and Verrijzer, 2004). The structure of chromatin varies throughout the nucleus, and impacts strongly on nuclear processes such as transcription and DNA replication (Wolffe, 1998). The basic structural unit of chromatin is the nucleosome, which consists of a core histone octamer containing dimers of H2A/H2B and H3 and H4, with 146bp of DNA wrapped around it (Luger et al., 1997). Histone H1 is considered to be a ‘linker’ histone because it binds to nucleosome core particles and organises the linker DNA on either
side of the nucleosome to facilitate the folding of chromatin into a 30nm fibre *in vitro* (Bednar et al., 1998). The positioning of the nucleosomes on the DNA backbone can be altered in a process termed chromatin remodelling. This allows the chromatin structure to be made more or less condensed and for particular sequences to be made accessible or inaccessible (Workman and Kingston, 1998). Chromatin can be broadly divided into two types; transcriptionally active euchromatin, and repressed heterochromatin (Cooper, 1959).

Importantly, the N-terminal tails of all the core histones can be chemically modified, and this impacts strongly on transcription and other nuclear processes such as DNA replication and repair (Strahl and Allis, 2000). These modifications exert their effect both by changing the physical characteristics of the chromatin structure, and by providing a specialised docking site for other proteins (Turner, 2000). Modifications include acetylation, methylation, phosphorylation, ubiquitination, ADP-ribosylation (Wolffe and Hayes, 1999) and sumoylation (Nacerddine et al., 2005), and different combinations of modifications define the ‘histone code’, which has been suggested to specify the transcriptional status of a genomic region (Strahl and Allis, 2000).

1.1.1 *Features of heterochromatin*

Repressed chromatin, or heterochromatin, is characterised by a sparseness of genes, histone hypoacetylation, inaccessibility to nucleases, methylation of H3K9 and H3K27, binding of HP1, DNA methylation and late replication (Friedman et al., 1996; Grewal and Moazed, 2003; Lima de Faria and Jaworska, 1968; Schotta et al., 2004; Solage and Cedar, 1978; Wakimoto, 1998; Wallace and Orr-Weaver, 2005). Constitutive heterochromatin i.e. regions which are invariably in a heterochromatic state regardless of cell type or context, also tends to contain regions of repetitive sequence e.g. the centromeric major and minor satellites, and the telomeric repeats (reviewed by (Arney and Fisher, 2004; Lachner et al., 2003). Facultative heterochromatin refers to genomic regions which are found in either a euchromatic or a heterochromatic state depending on
the cell type or context; the inactive X chromosome is an excellent example of this (Chadwick and Willard, 2004).

There is a strong correlation between heterochromatin and transcriptional repression (for review see (Grewal and Moazed, 2003)); regions of heterochromatin tend to be gene-poor and translocation of a gene that normally resides in euchromatin, where it is active, into heterochromatin results in repression of the gene. This repression is often unstable, leading to expression in some clonal cells but not others, a phenomenon referred to as position-effect variegation, or PEV (Reuter and Spierer, 1992). The D.melanogaster White and Rolled genes are interesting exceptions to this; they rely on a heterochromatic environment for their expression (Devlin et al., 1990; Eberl et al., 1993). It is not yet clear why this is so, but unique promoter sequences have been ruled out, and it is thought that specialised enhancers and/or higher order interactions are necessary for the correct transcription of these genes (Yasuhara et al., 2005). In contrast, euchromatin tends to be hyperacetylated, possess demethylated DNA and to possess histone methylation at H3K4. All these features are associated with transcriptional activation and early replication (reviewed by (Arney and Fisher, 2004; Lachner et al., 2003).

1.1.2 Constitutively heterochromatic repetitive sequences

Mammals have a highly complex genome organisation which has resulted from the accumulation of repetitive elements and non-coding sequences, or so-called ‘junk’ DNA. These sequences form a large part of the genomic make-up; in mouse 44% of total genomic DNA is composed of repetitive elements and 52% of non-coding elements, compared to only 4% of protein-coding sequences (Lander et al., 2001; Waterston et al., 2002). A high proportion of constitutively heterochromatic regions are highly repetitive. Repetitive elements include the tandemly repeated major and minor satellite repeats, and the telomeric arrays which form the telomeres. In both cases these sequences form part of nucleoprotein structures that have functional, specialised roles in the cell (Gonzalo et al., 2006).
Together, the major and minor satellite repeats form the pericentric and centric regions respectively. The centric chromatin, or primary constriction, consists of largely uninterrupted blocks of tandem arrays of around 2000 copies of the 123bp minor satellite repeat unit. In total there are more than 50,000 copies of the minor satellite repeat unit in murine cells (Waterston et al., 2002). It is upon the primary constriction that the kinetochore forms, the protein structure to which microtubules of the mitotic spindle attach, and which ensures equal segregation of the chromosomes at cell division (Pidoux and Allshire, 2000). Surprisingly, the sequence of centromeric DNA is not evolutionarily conserved and it is thought that centromeres are regulated largely at an epigenetic level (Sullivan et al., 2001). A feature of centromeric chromatin, or ‘centrochromatin’ as it has been referred, to is the presence of the histone H3 variant CENP (CENP-A in humans and CID in Drosophila) (Blower et al., 2002). Adjacent to the primary constriction, pericentric chromatin is composed of tandem arrays of more than 10,000 copies of the 234bp A/T rich major satellite repeat unit. In total there are more than 200,000 copies of the major satellite unit in murine cells (Waterston et al., 2002).

Other repetitive elements include DNA transposons, retrotransposons, long interspersed nucleotide elements (LINEs) and short interspersed nucleotide elements (SINEs). These mobile elements have the ability to integrate into the genome at a new site within their cell of origin and as such they have the potential to be deleterious to genome stability, as their mobilisation facilitates recombination between non-homologous loci, leading to chromosomal deletions and translocations (Kazazian, 2004). A particularly common type of retrotransposon in mammalian genomes is the Intracisternal A particle (IAP) retrotransposon, which contains long repeat sequences at it’s termini. Present at 500 - 1000 copies per haploid genome (Lueders and Kuff, 1977), it is repressed by DNA methylation; demethylation caused by the loss of the maintenance DNA methyltransferase Dnmt1 is inversely correlated with expression of this element (Walsh et al., 1998).
1.1.3 Histone acetylation

Occurring post-translationally and reversibly on ε-NH$_3^+$ groups of conserved lysine and arginine residues on the N-terminal tails of all the core histones, acetylation was one of the first histone modifications to be identified and was first associated with transcriptional activation more than 40 years ago (Allfrey et al., 1964; Pogo et al., 1966). It is now well established that acetylation correlates with transcriptional activation (Kurdistani et al., 2004). It has this effect by neutralising the positive charge and increasing the hydrophobicity of the histone tails, thereby decreasing their affinity for DNA (Hong et al., 1993), and allowing increased access of transcriptional regulators and machinery to chromatin templates (Lee et al., 1993; Vettese-Dadey et al., 1996). It also acts as a binding site for a variety of proteins. In some respects acetylation can be regarded as a default or ‘ground’ state, because when nucleosomes are produced during S phase, they are initially acetylated on H3 and H4; this is later removed where appropriate (Sobel et al., 1995).

Acetylation is catalysed by histone acetyltransferases (HATs), of which a variety have been identified in eukaryotes (Sterner and Berger, 2000). HATs are transcriptional coactivators; they do not bind directly to DNA but rather with DNA-binding activators (Utley et al., 1998). They are frequently found as components of large transcriptional activatory complexes, such as the yeast SAGA/human PCAF/Gcn5, and the yeast NuA4/human Tip60 complexes (Kuo and Allis, 1998). The reverse process, deacetylation, is carried out by a family of histone deacetylases, or HDACs. In humans, 18 have been identified and can be placed into 4 classes on the basis of their sequence and/or domain structure (Gregoretti et al., 2004). There seems to be a significant degree of redundancy between at least certain members of the Hdac family; for example ES cells lacking Hdac1 display a relatively mild phenotype of slow growth due to misregulation of certain cell-cycle control genes. Upregulation of Hdac2 and Hdac3 is seen in these cells, presumably to compensate for loss of Hdac1, and most likely explains the mild phenotype (Lagger et al., 2002).
1.1.4 Histone methylation

Occurring on conserved lysine and arginine residues on histones H1, H3 and H4, histone arginine methylation is correlated with transcriptional activation (Wang et al., 2001), whereas lysine methylation is correlated both with transcriptional activation and repression. In particular, methylation of H3K9, H3K27 or H4K20 is associated with transcriptional repression, whereas methylation of H3K4 is associated with activation (for review see (Lachner et al., 2003; Vaquero et al., 2003)). Lysine residues can be mono-, di-, or tri-methylated (DeLange et al., 1973), thereby extending the coding potential of a methylatable lysine position. For example both H3K4 and H3K9 can be di- or tri-methylated and this is biologically relevant for transcriptional regulation or Polycomb mediated gene silencing (Czermin et al., 2002; Kuzmichev et al., 2002; Santos-Rosa et al., 2002; Tamaru and Selker, 2001; Tamaru et al., 2003). Certain regions of the genome are particularly associated with certain histone methylation patterns. For example, pericentric heterochromatin is associated with tri-methylated H3K9 (me3H3K9) and me3H4K20 (Lehnertz et al., 2003; Schotta et al., 2004), whereas the inactive X chromosome is associated with both me2H3K9 and me3H3K27 (Rougeulle et al., 2004)

Histone methylation is catalysed by SET-domain containing proteins (Rea et al., 2000). The conserved SET (Su(var), E(z), Trithorax) domain was initially characterised as a common motif in several D.melanogaster PEV (position effect variegation) modifying proteins (Jones and Gelbart, 1993; Stassen et al., 1995; Tschiersch et al., 1994). Studies of the mammalian genome have led to the discovery of seven to ten gene families comprising more than 70 gene sequences containing the SET domain (Jenuwein, 2001). These include the essential genes G9a, which can methylate both H3K9 and H3K27 in euchromatic regions (Tachibana et al., 2005), ESET, which is specific for H3K9 (Yang et al., 2002), and Suv39h1/h2 which specifically methylate H3K9 in the pericentric heterochromatin (Lehnertz et al., 2003). This diversity indicates that different histone methyltransferases might be targeted to different regions of the genome to regulate chromatin structure and gene expression.
This modification was until recently considered to be a stable modification that was not as dynamic as other modifications such as histone acetylation, partly because the marks were seen to be propagated through several cell divisions and to resist reprogramming in early mammalian embryos (Lachner et al., 2004; Santos et al., 2003; Schotta et al., 2004). However the discovery by Shi and colleagues in 2004 of a histone lysine demethylase (Lsd1) that is able to demethylate H3K4 demonstrated that this modification is in fact reversible (Shi et al., 2004). Demethylation of arginine residues by PADI4-mediated deimination was demonstrated shortly after (Cuthbert et al., 2004). Demethylation of tri-methylated H3K9 at pericentric heterochromatin by Jmjd2 proteins has also been demonstrated in mammalian cells (Fodor et al., 2006).

1.1.5 Phosphorylation, ubiquitination, ADP-ribosylation and sumoylation

Histone phosphorylation, most notably on H3 Serine 10, has an important role in chromosome condensation during mitosis (Cheung et al., 2000). It also has a role in the induction of immediate early genes, such as c-fos, in mammalian cells (Mahadevan et al., 1991). In budding yeast, ubiquitination of lysine 123 is critical for meiotic and mitotic growth (Robzyk et al., 2000). The Tafl1250 subunit of the TBP-associated TFIID possesses H1-ubiquitination activity (Pham and Sauer, 2000), thus providing a possible link to transcription. Sumoylation is a relatively new addition to the list of histone modifications, but has been shown in mammalian cells to be essential for nuclear integrity, chromosome segregation and embryonic viability (Nacerddine et al., 2005). ADP-ribosylation of histones is implicated in nuclear processes involving DNA double strand breaks, such as DNA repair, replication and recombination (Boulikas, 1989).

1.1.6 Chromatin remodelling

As discussed above, chromatin in interphase cells consists of a highly condensed nucleoprotein filament that is refractory to cellular processes such as transcription and DNA replication. Such processes can occur in this repressive environment due to the actions of highly conserved chromatin remodelling enzymes that use the energy from ATP hydrolysis to enhance or decrease the accessibility of nucleosomal DNA (Vignali
et al., 2000). These enzymes can be subdivided into three families based upon their biochemical properties and the sequence similarity of ATPase subunits: (i) the SWI–SNF group; (ii) the ISWI group; and (iii) the Mi-2/CHD group (Boyer et al., 2000). Whereas many members of the second and third groups appear to have roles in transcriptional repression (Deuring et al., 2000; Kehle et al., 1998), most SWI-SNF-like enzymes have roles in transcriptional activation (Krebs et al., 2000; Liu et al., 2001; Lomvardas and Thanos, 2001).

1.3 DNA methylation

DNA itself can be modified by the addition of methyl groups to the 5 position of cytosine residues predominantly within CpG dinucleotides (Bird, 2002). Despite being a very common modification – some 60-90% of CpG dinucleotides are methylated in typical vertebrates – it is also highly compartmentalised, with some regions such as CpG islands remaining completely undermethylated and others being fully methylated (for reviews see (Bird, 2002; Hendrich and Tweedie, 2003). This modification, generally associated with transcriptional repression, is essential; null mutations in the genes responsible for DNA methyltransferase activity produce an embryonic lethal phenotype (Li et al., 1992; Okano et al., 1999). DNA methylation has an essential role in developmental processes such as imprinting and X-inactivation (Bestor, 2000; Robertson and Wolffe, 2000), and is also important for genome stability (Walsh et al., 1998). Interestingly, not all animals use DNA methylation as a regulatory mechanism; the nematode worm C. elegans, and the yeast species S. cerevisiae and S. pombe do not contain any detectable m^5C within their genomes (Tweedie et al., 1997), and the fruitfly D. melanogaster contains very low levels of m^5C, most of which is in the CpT dinucleotide rather than the CpG (Gowher et al., 2000; Lyko et al., 2000).

1.3.1 DNA methylation and transcriptional repression

DNA methylation is strongly correlated with transcriptional repression (Razin and Cedar, 1991); methylated regions of constitutive heterochromatin are almost invariably
silenced (Jones and Takai, 2001; Martienssen and Colot, 2001), and newly integrated transgenes that are silenced are also methylated (Assaad et al., 1993; Kilby et al., 1992). The silencing of germ-cell specific genes in non-germ cells was recently shown to depend on DNA methylation of CpG islands associated with these genes (Maatouk et al., 2006). It is not clear if methylation is causative of or a result of transcriptional silencing, and could be both. It could lead to transcriptional repression in a number of ways. Firstly, as discussed below, methylated DNA, or the Dnmts themselves, can recruit other chromatin modifying activities such as histone deacetylases and methylases, which can directly affect the compaction of the chromatin and recruitment of transcriptional regulators.

Methylated DNA could also physically affect the binding of transcription factors to chromatin; for example activation of the gene encoding glial fibriallary acidic protein (GFAP) during astrocyte differentiation involves demethylation of a CpG dinucleotide that lies in its promoter region in the STAT3 binding element (Takizawa et al., 2001). However it remains to be seen if this is a commonly used mechanism in vivo. DNA methylation could also affect the ability of chromatin remodellers to move nucleosomes along the DNA to reveal or hide particular sequences of DNA. Equally, methylation of DNA could be the result of the formation of a repressive chromatin structure either preceding or following transcriptional silencing, and could serve to stabilise this repressive state (for review see (Jones and Takai, 2001)). The importance of DNA methylation in the control of transcriptional patterns is suggested by the microarray analysis of Dnmt1−/− mouse embryonic fibroblasts (MEFs). Approximately 600 genes, equivalent to approximately 10% of those on the microarray, are mis-regulated (Jackson-Grusby et al., 2001), although it is not clear if this is a direct effect of DNA demethylation resulting from a deficiency of Dnmt1.
1.3.2 Connections between DNA methylation, histone modifications and chromatin remodelling

Both histone deacetylation and DNA methylation lead to transcriptional repression, but it is not always clear how or if they interact to bring about a repressed chromatin state. In recent years several mechanistic links have been identified between them, which go some way to explain how heterochromatin is established and/or maintained. For example in vitro methyl binding domain (MBD) proteins bind selectively to methylated DNA (Hendrich and Bird, 2000; Wade, 2001) and recruit HDACs (for review see (Bird and Wolffe, 1999) and HMTases (Fuks et al., 2003b) to bring about a condensed chromatin structure. Additionally, the acetylation status of a gene stably transfected into cells appeared to depend on the methylation status of the gene prior to transfection (Eden et al., 1998). All of the Dnmts have also been shown to interact with HDACs to bring about transcriptional repression, although interestingly in every case the methyltransferase domain of the Dnmts is dispensable for transcriptional silencing (Deplus et al., 2002; Fuks et al., 2000; Fuks et al., 2001; Geiman et al., 2004b).

It is not clear whether DNA methylation occurs prior to deacetylation of histones at the target site, or if deacetylation occurs as a secondary event to ‘lock in’ a repressive chromatin state identified by methylated DNA or if both can occur. This second scenario may be the case when heterochromatin is replicated (Fuks et al., 2000; Rountree et al., 2000). However in terms of the initiation of the formation of a heterochromatic state, for example when genes are silenced in response to an extra- or intracellular signal, there is evidence that histone deacetylation and gene silencing precede DNA methylation. This is the case during X inactivation, when chromosome wide histone hypoacetylation and gene silencing are very early initiating events in the inactivation process (Heard et al., 2001), whereas DNA methylation of gene promoters occurs after gene silencing (Keohane et al., 1996; Lock et al., 1987; Wutz and Jaenisch, 2000) and is associated with maintenance of the inactive state (Csankovszki et al., 2001; Norris et al., 1991).

It was shown some time ago that a retrovirus introduced into embryonic cells was silenced within c.2 days following infection, but that DNA methylation of the gene did
not occur for c.15 days (Gautsch and Wilson, 1983; Niwa et al., 1983). The finding that
the transcriptional silencing of retroviral sequences does not require de novo methylation
activity supports the notion that methylation is a secondary, reinforcing event rather than
an initiating one (Pannell et al., 2000). Furthermore, experiments by Buschhausen et al
showed that, following microinjection into cells, transcriptional silencing of a transgenic
gene did not occur for several hours, and this timing was not affected by whether the
gene was methylated prior to injection or not. Thus DNA methylation alone is not
sufficient to cause transcriptional silencing (Buschhausen et al., 1987). More recently it
was found that following integration of a transgene into chicken erythroid cells,
transcriptional silencing, histone hypoacetylation and demethylation of H3K4 were the
primary events, followed by methylation of DNA and H3K9 (Mutskov and Felsenfeld,
2004). Most of these observations are correlative, however there is some evidence in
Neurospora Crassa and in mammalian cells that patterns of histone acetylation can
direct DNA methylation patterns (Hu et al., 2000; Selker, 1998). Moreover, as discussed
below, there is extensive evidence that histone methylation can direct DNA methylation.

Mechanistic links have been established between histone methylation and DNA
methylation, establishing histone methylation as a vital component in the pathway to
transcriptional repression. H3K9 tri-methylation (Me3H3K9), mediated in mammals by
histone methyltransferases (HMTases) such as Suv39h1 and Suv39h2, and G9a, is
selectively bound by heterochromatin protein 1 (HP1) (Bannister et al., 2001; Lachner et
al., 2001). HP1 in turn interacts with Dnmt1, 3a and 3b (Fuks et al., 2003a; Lehnertz et
al., 2003). This is particularly well exemplified by the formation of pericentric
heterochromatin in mammalian cells. Targeted mutation in the HMTases Suv39h1 and
Suv39h2 in ES cells leads to a diffuse localisation of Dnmt3b, as opposed to the
pericentric localisation normally observed in wild type cells, and a reduction in
methylation of the satellite repeats (Lehnertz et al., 2003). Dnmt3a/3b^{(-)} ES cells
however do not display mislocalisation of Me3H3K9, indicating that at pericentric
heterochromatin DNA methylation relies on prior histone methylation, and not vice-
versa (Lehnertz et al., 2003). A similar mechanism occurs in both Arabidopsis thaliana
and *Neurospora crassa*, where mutations in the H3K9 HMTases *dim-5* and *kryptonite* lead to DNA demethylation (Jackson et al., 2002; Tamaru and Selker, 2001).

Repression of the major and minor satellites, highly repetitive regions, interestingly also involves the RNAi pathway (Volpe et al., 2002). The repeat sequences are capable of being transcribed, but this can be deleterious to both yeast and mammalian cells (Bouzinba-Segard et al., 2006; Volpe et al., 2003). Transcription of repeat sequences leads to the formation of double stranded RNA, which activates the RNA interference pathway and leads to the formation of the RITS (RNA Induced Transcriptional Silencer) complex. This associates with histone methyltransferase activity, and is targeted to the repetitive DNA.

Random inactivation of one of the two X chromosomes in female ES cells occurs upon differentiation of the cells. The establishment of a heterochromatic state on the inactive X provides a further example of the links between histone methylation and DNA methylation. Unlike the centromeric repeats, which are marked by Me3H3K9, the X chromosome is hypoacetylated and decorated with Me2H3K9 during the initiation of X inactivation (Heard, 2004; Heard et al., 2001). Polycomb proteins, members of the PRC2 complex, are recruited to this mark (Silva et al., 2003), and in turn they mediate Me3H3K27. PcG proteins can also interact with Dnmts, providing a mechanism for the recruitment of DNA methylation activity to the Xi (Vire et al., 2006).

### 1.3.3 DNA methylation and chromatin remodelling

Chromatin remodelling by members of the SNF2 protein family has also been demonstrated to be linked to patterns of DNA methylation. In *Arabidopsis*, mutations in the SNF2-like ATPase family member DDM1 (for ‘Decrease in DNA Methylation’) cause a severe reduction in DNA methylation (Jeddeloh et al., 1999; Singer et al., 2001). Targeted disruption of the mammalian homologue of DDM1, *Lsh*, causes perinatal lethality along with a substantial loss of methylation throughout the genome that is attributed to a defect in *de novo* methylation by Dnmt3a and Dnmt3b during early
development (Dennis et al., 2001; Geiman et al., 2001; Zhu et al., 2006). Finally, in the human disorder X-linked \( \alpha \)-thalassemia, mutations occur in the putative SNF2-like ATPase ATRX, and this leads to decreases in methylation at the rDNA repeats (Gibbons et al., 2000).

1.3.4 Replication of heterochromatin

The replication of heterochromatin, which happens late in S phase, provides a framework for understanding how its structure is established. The maintenance methyltransferase Dnmt1 localises to the replication fork to re-instate methylation patterns (Leonhardt et al., 1992), where it interacts with both histone deacetylases and methyltransferases (Esteve et al., 2006; Fuks et al., 2000; Rountree et al., 2000). Other epigenetic modifiers have also been found to interact directly with the replication machinery, including the histone methyltransferase (HMTase) SETDB1 which interacts with MBD1, PCNA and CAF1 (Sarraf and Stancheva, 2004), and SNF2H, which is recruited by WSTF to replication sites (Poot et al., 2004), and the HMTase G9a (Esteve et al., 2006). Dnmt3a and 3b, despite their demonstrated ability to methylate DNA de novo (Hsieh, 1999; Okano et al., 1999), also have a role in the maintenance of methylation patterns, and it has been suggested that they do this in co-operation with Dnmt1 at replication foci (Chen et al., 2003b; Liang et al., 2002); Dnmt3a also interacts with the HMTase SUV39H1 and HP1, although it was not shown if this was localised to the replication fork (Fuks et al., 2003a). One model of heterochromatin establishment is that the recruitment of the above epigenetic modifying factors to replication foci allows the initiation of the formation of a heterochromatic state i.e. DNA methylation, histone hypoacetylation and methylation. However it is not clear what is the temporal order of events.

1.4 DNA methylation in development

The levels of DNA methylation change significantly during development, and this modification is involved in several important developmental processes including
imprinting and X-chromosome inactivation. It is also important for the protection of genome stability, especially during the early stages of embryogenesis. The importance of DNA methylation in the developmental process is underscored by the fact that removing either de novo or maintenance methylation activity from the embryo produces embryonic lethality (Li et al., 1992; Okano et al., 1999). Within ES cells, the presence of appropriate DNA methylation patterns appears to be dispensable for self-renewal, but not differentiation. A number of human disorders are also associated with either aberrant methylation patterns or their interpretation.

1.4.1 Imprinting

Genomic imprinting is an epigenetic phenomenon that causes some genes to be expressed according to their parental origin, thus imposing a requirement on the developmental process for the presence of both maternal and paternal genomes (McGrath and Solter, 1984; Surani et al., 1984). This developmental asymmetry between the parental genomes explains why uni-parental embryos are inviable (McGrath and Solter, 1984; Surani et al., 1984) and why uni-parental disomy is associated with severe developmental abnormalities or cancer (Brenton et al., 1995; Reik, 1989). Currently some 100 genes have been identified as being imprinted in the mouse genome (Morison et al., 2001). Many of these have roles in prenatal growth and the development of particular lineages (Bartolomei and Tilghman, 1997; Reik and Walter, 2001), and it is hypothesised that the driving force in the evolution of imprinting was genetic conflict over maternal resources (Moore and Haig, 1991).

DNA methylation is not only involved in the repression of transcription; it is also a crucial regulatory element in the control of the transcription of imprinted genes (Ferguson-Smith and Surani, 2001; Li et al., 1993). The process of imprinting frequently involves mono-allelic methylation of imprinting control regions (ICRs), cis-acting sequences that may be located many kilobases away from the genes they are involved in regulating. Imprinted genes are frequently clustered together and co-ordinately regulated by such ICRs (Paulsen and Ferguson-Smith, 2001; Reik and Walter, 2001). The
mechanisms of how this is achieved have not been elucidated for the majority of genes. However the regulation of \textit{H19} and \textit{Igf2}, some of the first imprinted genes to be described (Barlow et al., 1991; Bartolomei et al., 1991; DeChiara et al., 1991), has been shown to include, among other things, paternal methylation of a region upstream of the \textit{H19} gene (Tremblay et al., 1995). Methylation prevents binding of the insulator element CTCF, and allows the \textit{Igf2} gene preferential long range access to the enhancer downstream of \textit{H19}, leading to expression of \textit{Igf2} and repression of \textit{H19} (Ferguson-Smith, 2000; Reik and Walter, 2001; Szabo et al., 2000) (Fig.3.1.B). \textit{H19} encodes a non-coding RNA and is located 75kb away from \textit{Igf2}, which codes for a protein that is part of the insulin and insulin-like growth factor system that has demonstrated effects on prenatal growth (Zemel et al., 1992).

\textit{Gtl2} and \textit{Dlk1} are a further two linked and reciprocally imprinted genes located 80kb apart on mouse chromosome 12, which share many common regulatory features with the \textit{H19/Igf2} locus (Paulsen et al., 2001) (Fig.3.1.A). Three differentially methylated regions (DMRs) have been identified at or nearby this locus, all of which are paternally methylated; an intergenic DMR (IG-DMR) which becomes methylated in the germline, a DMR associated with the CpG island of \textit{Dlk1}, and an intronic DMR within the \textit{Gtl2} gene (Takada et al., 2002). The \textit{Gtl2} DMR is also regulated by histone modifications; chromatin immunoprecipitation analysis reveals maternal specific hyperacetylation at this locus (Carr et al., 2006). Additionally, mice lacking the Polycomb group gene \textit{Eed} lose imprinted expression of \textit{Gtl2}, but not of \textit{Dlk1} (Mager et al., 2003). The IG-DMR is a control element for imprinted genes on the maternal chromosome; deletion of the IG-DMR from the maternally inherited chromosome causes bidirectional loss of imprinting of all genes in the cluster. However imprinting is unaltered if the mutation is transmitted from the father, demonstrating that the two parental chromosomes control imprinted gene expression of this region differently (Lin et al., 2003). The observation that in \textit{Dnmt1<sup>−/−</sup>} embryos \textit{Dlk1} expression is upregulated, whilst \textit{Gtl2} expression is downregulated, demonstrates the importance of DNA methylation of this region (Schmidt et al., 2000). \textit{Dlk1} has homology to the Notch-Delta family of developmentally regulated signalling molecules that contain EGF repeats, and has a role in cellular differentiation
The *Gil2* gene produces a maternally expressed non-coding RNA transcript whose function is unknown (Schuster-Gossler et al., 1998). Several other genes have been identified in this region, most of which produce non-coding RNAs (Yevtodiyenko et al., 2002).

*Snrpn* is another example of an imprinted gene, this time on mouse chromosome 7, in a region that is syntenic to the Prader-Willi/Angelman syndromes (PWS-AS) region on human chromosome 15 (Buiting et al., 1995). Methylation of the 5’ end of the *Snrpn* gene on the maternal chromosome occurs during oogenesis and results in exclusively paternal expression of *Snrpn* (Shemer et al., 1997).

The ‘life-cycle’ of imprinted marks starts when they are established in developing gametes. They are subsequently maintained in mature gametes and embryos, despite the wave of demethylation that occurs during early embryogenesis, before being erased at an early stage of germ cell development, and then re-established at a later stage of gamete development, thus completing the cycle (Fig.1.1) (Davis et al., 2000; Reik and Walter, 2001; Warnecke et al., 1998). Paternal and maternal imprints are established at different stages of gametogenesis; the former during the gonobyte/prospermatogonia stage (E14.5 to newborn) (Davis et al., 2000; Ueda et al., 2000), and the latter later on during the oocyte growth stage (P5-20) (Lucifero et al., 2004; Obata and Kono, 2002).

### 1.4.2 X-Chromosome Inactivation

In order to achieve dosage compensation of X-linked genes, mammals inactivate one X chromosome in XX female cells (Lyon, 1961; Migeon, 1994; Riggs and Pfeifer, 1992). X inactivation in the mouse embryo proper occurs around the late blastocyst stage (McMahon and Monk, 1983), seems to coincide with cellular differentiation events (Monk and Harper, 1979), and is random (Gardner and Lyon, 1971). In those cells that undergo the earliest differentiation events to form the precursors to trophectoderm and primitive endoderm, i.e. the extraembryonic lineages, inactivation is non-random, with the paternally inherited X chromosome always undergoing silencing (Takagi et al.,
1978; West et al., 1977). This happens much earlier than random X inactivation, with Xist expression, one of the initiating events in X inactivation, being detected by the 4 cell stage (Kay et al., 1994). Interestingly, this is only the case for eutherian mammals; marsupials undergo imprinted X inactivation in every cell (Cooper, 1971). Female ES cells are an excellent model in which to study random X inactivation, because they undergo this process when they differentiate (Martin et al., 1978; Rastan and Robertson, 1985).

Initiation and propagation of X inactivation involves a large nonprotein-coding RNA, the X-inactive specific transcript (Xist), which is transcribed from the X inactivation centre, or XIC (Borsani et al., 1991; Brockdorff et al., 1991; Brockdorff et al., 1992; Brown et al., 1991; Brown et al., 1992). Xist RNA spreads over the X chromosome in cis and is thought to induce chromosome inactivation through the recruitment of silencing factors (Clemson et al., 1996; Panning et al., 1997; Sheardown et al., 1997). The XIC is also involved in ‘counting’, thus ensuring that only one X chromosome remains active per diploid autosome set (Morey et al., 2004). Silencing of X-linked genes on the Xi is a rapid process, occurring within 1-2 cell cycles following induction of ES cell differentiation and Xist activation (Wutz and Jaenisch, 2000).

Following inactivation, the entire Xi is in a heterochromatic state, replicates late (Gilbert et al., 1962), and can be visualised as a Barr body close to the nuclear periphery of interphase nuclei (Barr, 1949). As well as DNA methylation (Mohandas et al., 1981; Pfeifer et al., 1990b) and presence of the histone H3 variant macroH2A (Costanzi and Pehrson, 1998; Costanzi and Pehrson, 2001), it displays a characteristic pattern of histone tail modifications, including hypoacetylation of H3 and H4, Me2H3K9 (Boggs et al., 2002; Chaumeil et al., 2002; Heard et al., 2001), Me3H3K27 (Plath et al., 2003; Silva et al., 2003), and a lack of Me2 and Me3H3K4 (Boggs et al., 2002; Chadwick and Willard, 2003; Chaumeil et al., 2002; O'Neill et al., 2003). Many of these appear early on in the inactivation process, shortly after coating of the chromosome by Xist RNA (Chaumeil et al., 2002; O'Neill et al., 2003; Plath et al., 2003; Silva et al., 2003). Deposition of Me3H3K27 is catalysed by the polycomb group complex Eed/Enx1, and
appears to be responsible for the early maintenance of the inactive state (Kalantry et al., 2006; Silva et al., 2003), but not long-term maintenance, which is dependent on other epigenetic mechanisms including DNA methylation (Csankovszki et al., 2001; Kaslow and Migeon, 1987). Recruitment of DNA methyltransferase activity by polycomb group proteins may be responsible for this longer term maintenance (Vire et al., 2006).

1.4.3 DNA methylation is required for the differentiation but not self-renewal of ES cells

Despite being so important during development and having a clearly demonstrated role in the regulation of transcription, DNA methylation appears to be dispensable for the self-renewal of ES cells. Dnmt1<sup>(−/−)</sup>, Dnmt3a, 3b<sup>(−/−)</sup> and Dnmt1, Dnmt3a, 3b<sup>(−/−)</sup> ES cells are viable and display no obvious growth or morphological defect through up to 75 passages, despite being highly significantly demethylated (Jackson et al., 2004; Lei et al., 1996; Tsumura et al., 2006). In contrast, deletion of Dnmt1 from cultured fibroblasts, whilst leading to a progressive loss of methylation similar to that seen in Dnmt1<sup>−/−</sup> ES cells, causes a uniform p53 dependent cell death (Jackson-Grusby et al., 2001). However DNA methylation appears to be required for the differentiation of ES cells. Dnmt1-deficient embryoid bodies (EBs) aberrantly express Xist, down-regulate X-linked genes, and apoptose when induced to differentiate (Panning and Jaenisch, 1996). In addition, Dnmt3a, 3b<sup>(−/−)</sup> ES cells are blocked in their ability to terminally differentiate (Jackson et al., 2004).

1.4.4 Protection of genome stability

Approximately 96% of total DNA in the human and mouse genomes is non-coding (Lander et al., 2001; Waterston et al., 2002). 30 – 50% of this ‘junk’ DNA fraction is composed of various transposable elements including DNA transposons, retrotransposons, long interspersed nucleotide elements (LINEs) and short interspersed nucleotide elements (SINEs) (Lander et al., 2001; Waterston et al., 2002). These mobile elements have the ability to integrate into the genome at a new site within their cell of
origin and as such they have the potential to be deleterious to genome stability, as their mobilisation facilitates recombination between non-homologous loci, leading to chromosomal deletions and translocations (Kazazian, 2004). Whilst many of these elements are mutated and therefore inactive, mammalian genomes contain many copies of retroelements with intact promoters which are instead transcriptionally silenced to prevent their mobilisation (Kazazian, 2004).

Two mechanisms commonly used to achieve this are co-suppression usually mediated by small interfering RNAs (siRNAs) and DNA methylation. The former has been demonstrated for retrotransposons in *S. cerevisiae*, *D. melanogaster* and *C. elegans* (Garfinkel et al., 2003; Jensen et al., 1999; Jiang, 2002; Sijen and Plasterk, 2003). In mammals, retrotransposons tend to be heavily methylated on their DNA (Bestor et al., 1984; Jahner et al., 1982; Monk et al., 1987; Sanford et al., 1987; Yoder et al., 1997b), and this represses transcription from the promoters of at least several types of retrotransposon (Harbers et al., 1981; Kuff and Lueders, 1988; Schmid, 1996). DNA methylation of a common long terminal repeat-containing retroviral-like retrotransposon, the mouse intracisternal A particle, or IAP, has been shown to be inversely correlated with the expression of this retrotransposon (Walsh et al., 1998). A reduction in Dnmt1 levels in mouse embryos leads to demethylation and stable transcriptional activation of retrotransposable IAP elements, which persists throughout adult life and can even be passed onto the next generation (Walsh et al., 1998). The persistence of transcriptional misregulation caused by the activation of IAPs means that it is important they are protected from demethylation, and hence activation, when the embryonic genome undergoes a wave of demethylation during cleavage.

DNA methylation is also important for maintaining genome stability in ES cells; *Dnmt1* null ES cells have been shown to have an enhanced mutation rate at several endogenous loci, including *Hprt* and *Tk*, due to increased mitotic recombination leading to chromosomal rearrangements (Chen et al., 1998). The importance of protecting genomic integrity in this way is clearly demonstrated by mice carrying a *Dnmt1* hypomorphic allele leading to less than 10% expression of endogenous Dnmt1 levels; these mice are
runted at birth and within 8 months develop aggressive and lethal T-cell lymphomas (Gaudet et al., 2003). The de novo methyltransferase Dnmt3b also has a role in protecting genome stability, at least in somatic cells; inactivating Dnmt3b in mouse embryonic fibroblasts leads to DNA hypomethylation and chromosomal instabilities (Dodge et al., 2005).

1.4.5 Epigenetic reprogramming in development

In both mammals and amphibians, DNA methylation and histone modification patterns are dynamic and change substantially during certain developmental periods, in particular during pre-implantation development and gametogenesis (Fig. 1.1). Shortly after fertilisation, protamines in the paternal pronucleus are replaced with histones that are acetylated (Adenot et al., 1997; Santos et al., 2002), as well as histones that are methylated (Lepikhov and Walter, 2004; Santos et al., 2005). Subsequently, the paternal genome undergoes rapid demethylation that is completed by the one cell stage, and is therefore presumably due to an active demethylation process, although evidence for this is still lacking (Mayer et al., 2000; Oswald et al., 2000). The maternal genome also undergoes demethylation but this occurs more gradually, not reaching an equivalent level of demethylation with the paternal genome until the morula stage (Howlett and Reik, 1991; Santos et al., 2002). The level of methylation continues to drop on both genomes in a passive process caused by exclusion of the maintenance methyltransferase Dnmt1o from the nucleus (Carlson et al., 1992; Howlett and Reik, 1991; Monk et al., 1991; Rougier et al., 1998), until it reaches a minimum level by the start of the blastocyst stage (Howlett and Reik, 1991; Kafri et al., 1992; Monk et al., 1987). Certain sequences are protected from demethylation during this time, although the mechanisms behind this remain unclear. Protected sequences include all imprints, thus allowing their imprinted status to be maintained (Olek and Walter, 1997), IAP retrotransposons, preventing their transposition (Lane et al., 2003) and centromeric chromatin, possibly to maintain the stability of chromatin and the centromeric structures (Rougier et al., 1998).
During implantation, levels of DNA methylation in cells within the inner cell mass return to the levels seen in somatic cells (Monk et al., 1987; Reik et al., 2001) in a process dependent on the de novo methyltransferases Dnmt3a and 3b (Okano et al., 1999). This remethylation is not observed in trophectoderm cells, presumably due to the absence of *de novo* methyltransferase activity from these cells (Watanabe et al., 2002).

Primordial Germ Cells (PGCs) are derived from epiblast cells in the posterior primitive streak at E7.5, from where they migrate, reaching the genital ridge at E11.5 (McLaren, 2003). PGCs possess a similar epigenetic profile to epiblast cells, including random X chromosome inactivation and imprinted gene expression. These need to be erased in order to allow sex-specific marks to be imposed, and this is done at least in part by a genome wide reduction in DNA methylation during the proliferative oogonial and spermatogonial stages from E11.5 to E12.5, following arrival at the genital ridge (Hajkova et al., 2002; Reik et al., 2001; Tada et al., 1997). Following this, during gamete maturation, de novo methylation occurs, allowing the re-establishment of parentalspecific methylation marks at imprinted genes (Chaillet et al., 1991; Morgan et al., 2005; Stoger et al., 1993; Tremblay et al., 1995).
Figure 1.1 Epigenetic reprogramming during development. Schematic diagram of the major epigenetic changes occurring on the maternal and paternal genomes following fertilisation. The embryonic genome undergoes two waves of DNA demethylation; one occurring in all embryonic cells during pre-implantation development, and the second occurring in the primordial germ cells once they have arrived at the genital ridge. The methylation status of imprinted regions is indicated in red. ICM: Inner Cell Mass; TE: trophectoderm.
1.4.6 DNA methylation and disease

Changes to patterns of DNA methylation in cells is frequently associated with the neoplastic transformation of these cells. For example, loss of genomic methylation is commonly observed in tumourigenic cells and can be correlated with disease severity and the metastatic potential of these tumours (Widschwendter et al., 2004). Additionally, hypermethylation of tumour suppressor gene promoters is associated with their silencing, and is also frequently associated with neoplasia (Jones and Baylin, 2002). However it is not always clear if aberrant methylation patterns are a cause or consequence of neoplastic transformation (Prokhortchouk and Hendrich, 2002).

A number of other, less frequent, human disorders are associated with defects in methylation patterns or their interpretation; for example loss-of-imprinting disorders such as the Prader-Willi and Angelman syndromes. These are frequently associated with chromosomal deletions which result in a loss of imprinting of a cluster of genes including SNRPN and UBE3A at the human 15q11.2-13 locus (Nicholls et al., 1998). Similarly, Beckwith-Wiedemann syndrome which is associated with a lack of imprinting of a cluster of genes on human chromosome 11p15.5 (Reid et al., 1997). The autosomal recessive Immunodeficiency, Centromeric instability and Facial anomalies (ICF) syndrome is caused by mutations in the de novo methyltransferase DNMT3b (Hansen et al., 1999; Okano et al., 1999; Xu et al., 1999), and Rett syndrome is a neurodevelopmental disorder caused by mutations in the methyl binding domain protein MeCP2, leading to a failure to correctly interpret the methylation signal (Amir et al., 1999).

1.5 The DNA methyltransferases

The DNA methyltransferases, or Dnmts, are the enzymes responsible for transferring methyl groups to DNA. In mammals five members of the Dnmt family have been identified; Dnmt1, 2, 3a, 3b and 3l. All apart from Dnmt3l contain five highly conserved motifs in their C-terminus which constitute the catalytic methyltransferase domain
(Kumar et al., 1994) (see Fig.1.2). Between them they possess both de novo and maintenance methylation activities. All catalytically active members are essential genes, underlining the importance of DNA methylation in development (Li et al., 1992; Okano et al., 1999).

1.5.1 Dnmt1
The first Dnmt to be identified, Dnmt1 (Bestor et al., 1988) is considered to be responsible for ‘maintenance’ methylation of DNA, because it methylates hemimethylated DNA substrates 5-30 fold more efficiently than unmethylated DNA in whole cell lysates in vitro (Bestor, 1992; Gruenbaum et al., 1982; Yoder et al., 1997a). However Dnmt1 also has de novo methylation activity in biochemical assays (Yoder et al., 1997a), and as such a contribution towards de novo methylation events cannot be ruled out. Evidence is lacking for this in vivo, however, the Dnmt1 variant Dnmt1o accumulates to high levels in the nuclei of ooctyes during the growth phase, when genomic methylation patterns including imprints are established (Bao et al., 2000; Ueda et al., 2000). Because growing ooctyes are arrested in the dictyate stage of prophase of meiosis I, and thus lack S phases, a role for de novo methylation of these sequences during oogenesis by Dnmt1o is suggested (Mertineit et al., 1998).

The fidelity of maintenance methylation within a specific sequence of DNA was estimated to be 99.9% (Pfeifer et al., 1990a), however further studies have revealed that the maintenance methylation of some sequences of DNA has a contribution from the de novo methyltransferases Dnmt3a and 3b (Chen et al., 2003b; Liang et al., 2002). This has led to an estimate of the global average of maintenance methylation fidelity by Dnmt1 of 97.7 – 98.7% (Jackson et al., 2004).

Expressed ubiquitously in proliferating cells, Dnmt1 localises to replication foci, suggesting that maintenance methylation is coupled to DNA replication and providing a mechanism for the general heritability of methylation states in daughter cells, including the replication of heterochromatin (Leonhardt et al., 1992). Targeting to the replication
fork is via an interaction with PCNA, and can be disrupted by binding of the cell cycle regulator $p21^{WAF/CIP}$ to PCNA. PCNA is also found at sites of DNA repair and it is postulated that binding of $p21^{WAF/CIP}$, whose function it is to mediate the p53 dependent growth arrest in response to DNA damage (Li et al., 1996), prevents binding of Dnmt1 and hence inappropriate hypermethylation at sites of DNA repair (Chuang et al., 1997).

It was recently shown that low levels of Dnmt1 also localise to centromeric regions during mitosis, and that this involves a motif that is homologous to the Polybromo-1 protein, and referred to as a PBHD motif (Liu et al., 1998). It is not clear what the function of Dnmt1 at the centromeres during mitosis is. Dnmt1 also interacts with various other epigenetic modifiers including the histone methyltransferases G9a and SUV39H1, the heterochromatic protein HP1, and histone deacetylases (Esteve et al., 2006; Fuks et al., 2000; Fuks et al., 2003a). It is thought that these co-operate with Dnmt1 to re-establish a heterochromatic structure following DNA replication.
Figure 1.2 Schematic comparison of Dnmt family members; Dnmt1, 2, 3a, 3b and 3l. The five conserved motifs constituting the methyltransferase (MTase) domain are labelled as I, IV, VI, IX and X. Dnmt1: NLS; nuclear localisation signal. KG repeats; series of alternating lysine and glycine residues. Dnmt3a, 3b: PHD; PHD domain. PWWP; PWWP domain. Numbers indicate the number of amino acids. Based on Fig.2., Okano et al., 1998
1.5.1.1 Dnmt1 in vivo

Targeted mutation of the Dnmt1 gene results in embryonic lethality, with embryos failing to develop beyond E12.5, and showing a developmental delay from E9.5 (Li et al., 1992). Embryo death is ascribed to reduced proliferation and increased apoptosis of embryonic cells. Significant demethylation of embryonic cells is observed, leading to the aberrant regulation of several imprinted genes, such as H19 and Igf2 (Li et al., 1993). The Xist gene is also aberrantly activated in all cells, leading to inactivation of all X chromosomes in mutant embryos (Panning and Jaenisch, 1996). Interestingly, overexpression of Dnmt1 also leads to aberrant hypermethylation and expression of a variety of imprinted genes such as H19 and Igf2, and retroviral sequences, and also to embryonic lethality (Biniszkievicz et al., 2002). This demonstrates that the levels of methyltransferase activity and hence genomic methylation are critical for successful development.

Full-length Dnmt1 is actually not detected in embryos until E7 (Carlson et al., 1992; Mertineit et al., 1998). Prior to this, in oocytes and pre-implantation embryos, a variant of this, Dnmt1o, is expressed. This is transcribed from an alternative oocyte-specific promoter and contains a unique first exon (Mertineit et al., 1998). During oogenesis and pre-implantation development Dnmt1o undergoes complex cycles of nuclear-cytoplasmic trafficking (Cardoso and Leonhardt, 1999; Carlson et al., 1992; Mertineit et al., 1998), unlike its full-length counterpart which is ubiquitously nuclear (Leonhardt et al., 1992). Targeted ablation of Dnmt1o by deletion of the oocyte specific promoter and first exon results in a maternal effect mutation, leading to a loss of imprinting and lethality (Howell et al., 2001). Specifically, Dnmt1o<sup>-/-</sup> mice appear normal and have genomic methylation patterns comparable to wild type mice. However heterozygous offspring from homozygous females lack certain imprints and lose mono-allelic expression from these imprinted genes, and die during the last third of gestation. Because Dnmt1o displays a transient nuclear localisation at the 8 cell stage of embryonic development, it is suggested that it provides maintenance methylation activity specifically at imprinted loci during the fourth embryonic S phase (Howell et al., 2001).
1.5.2 Dnmt2

Despite containing all the conserved catalytic methyltransferase domain motifs common to all the known prokaryotic and eukaryotic Dnmts (Fig. 1.2), Dnmt2 does not appear to have a role in either de novo or maintenance methylation of DNA in ES cells; both endogenous and newly integrated retroviral DNA are fully methylated in both wild type ES cells lacking Dnmt2, and cells infected with a baculovirus producing Dnmt2 lacking the methyltransferase catalytic motif (Okano et al., 1998b). Dnmt2 is homologous to S.pombe pmt1, but contains a mutation at the putative catalytic PPC motif, which may explain why it is not catalytically active (Okano et al., 1998b). However depletion of D.melanogaster Dnmt2 by RNAi resulted in loss of the little DNA methylation found in this species, and overexpression appeared to cause hypermethylation (Kunert et al., 2003). A function has been ascribed to Dnmt2; in human cells it has been shown to methylate the small RNA aspartic acid transfer tRNA^Asp, a function that is highly conserved from flies and plants to mammals. It is suggested that an ancestral version of Dnmt2 was the founder of the Dnmt family (Goll et al., 2006).

1.5.3 The Dnmt3 family

The observation in 1996 that deletion of, at the time, the only known methyltransferase gene in mammals (now known as Dnmt1), did not impair de novo methylation activity in mutant embryos and ES cells (Lei et al., 1996), led to the conclusion that other DNA methyltransferases existed that were capable of de novo methylation. Searches of the human dbEST database using full-length bacterial type II cytosine-5 methyltransferases identified two homologous genes containing the highly conserved cytosine-5 methyltransferase motifs, and which are also conserved in mouse (Okano et al., 1998a). These were termed Dnmt3a and 3b, and shown to possess de novo methylation activity in vitro (Aoki et al., 2001; Hsieh, 1999; Okano et al., 1998a).
1.5.3.1 Dnmt3a and 3b – structural features

Both Dnmt3a and 3b share some structural and sequence similarity with Dnmt1 at the COOH-terminal catalytic region that is responsible for DNA methylation, but no sequence similarity in the N-terminal regulatory region (Chen and Li, 2004; Hermann et al., 2004). Within the N terminal region both Dnm3a and 3b contain a variable region (c.280 amino acids in Dnmt3a and c.220 amino acids in Dnmt3b) followed by two conserved regions; a PHD-like and a PWPP domain (Okano et al., 1998a; Xie et al., 1999) (Fig.1.2).

The PHD (plant-homeodomain)-like motif is contained within a cysteine rich region and is also present in the ATRX protein (Aasland et al., 1995). Within the Dnmts, it has been shown to interact with Hdac1 and to repress transcription in reporter assays (Bachman et al., 2001; Fuks et al., 2001). The PWPP domain, so-called because it contains a conserved proline-tryptophan-tryptophan-proline motif, was originally identified both in a group of proteins related to hepatoma-derived growth factor and in the protein product of the Wolf-Hirschhorn syndrome candidate gene 1 (WHSC1) (Izumoto et al., 1997; Stec et al., 1998). More than 60 proteins containing a PWPP domain have now been identified, and many of these associate with chromatin (Qiu et al., 2002; Stec et al., 2000). The general function of the domain remains unclear, however in Dnmt3a and 3b it is required for directing DNA methylation to the major satellite repeats; mutagenesis of this region disrupts the association with pericentric heterochromatin, and the ability of these enzymes to methylate this region (Chen et al., 2004). Interestingly, the Dnmt3a PWPP domain appears to have no DNA binding ability, whereas Dnmt3b appears to bind DNA non-specifically (Chen et al., 2004), suggesting functional differences between them.

1.5.3.2 In vivo function of Dnmt3a and 3b

Both Dnmt3a and Dnmt3b are highly expressed in ES cells, early embryos, and developing germ cells, but are barely detectable in somatic cells (Okano et al., 1998a; Watanabe et al., 2002). Because of their demonstrated de novo methylation activity and
their knock-out phenotypes, Dnmt3a and Dnmt3b are thought to be responsible for establishing genomic DNA methylation patterns during embryogenesis and gametogenesis (Howlett and Reik, 1991; Kafri et al., 1992; Okano et al., 1999). Both are essential genes; $Dnmt3a^{(-/-)}$ mice develop to term and appear normal at birth, however most become runted and die at about four weeks of age. $Dnmt3b^{(-/-)}$ mice display a more severe phenotype, with multiple developmental defects appearing after E9.5 and subsequent embryonic lethality (Okano et al., 1999). $Dnmt3a$, 3b$^{(-/-)}$ double knock-out mice display an even more severe phenotype, with an arrest of growth and morphogenesis shortly after gastrulation and lethality before E11.5, thus indicating that $Dnmt3a$ and $Dnmt3b$ have overlapping functions in vivo (Okano et al., 1999). An analysis of the methylation status of $Dnmt3a^{(-/-)}$ and $Dnmt3b^{(-/-)}$ embryos reveals slight demethylation of a number of sequences including endogenous C-type retroviral DNA and IAP repeats, and significant demethylation of the minor satellites. The extent of demethylation was compounded in $Dnmt3a$, 3b$^{(-/-)}$ embryos, indicating overlapping targets for Dnmt3a and 3b (Okano et al., 1999).

The difference in phenotype between the individual $Dnmt3a^{(-/-)}$ and $Dnmt3b^{(-/-)}$ mutants can perhaps be attributed both to target specificities of each Dnmt, and to their differential expression during embryonic development. $Dnmt3a$ expression is first detected at E7.5, when it is found in the embryonic ectoderm and mesoderm. At E8.5 and E9.5 expression becomes ubiquitous. $Dnmt3b$ is first detected in the inner cell mass (ICM) at E4.5, then the epiblast at E5.5, and the embryonic ectoderm at E7.5. At later stages it is detected at low levels throughout the embryo and predominantly in the forebrain and eyes (Okano et al., 1999; Watanabe et al., 2002). The co-expression of $Dnmt3a$ and 3b in the embryonic ectoderm after E7.5 could explain their partial redundancy.

1.5.3.3 Dnmt3a and imprinting

Some of the first evidence that de novo methylation activity is involved in the establishment of genomic imprints came from the observation that both Dnmt3a and 3b
co-localise in the nucleus with Dnmt3l (Hata et al., 2002), which has a role in the establishment of imprints (Arima et al., 2006; Bourc'his and Bestor, 2004; Bourc'his et al., 2001; Hata et al., 2002; Kaneda et al., 2004). Subsequently it was shown that Dnmt3a, but not Dnmt3b, is required for the establishment of both maternal and paternal imprints, although other factors are most likely involved as well (Hata et al., 2002; Kaneda et al., 2004).

1.5.3.4 Dnmt3b, ICF and cancer

ICF (Immunodeficiency, Centromeric instability and Facial anomalies) syndrome is a rare and autosomal recessive disorder causing severe immunodeficiency, making sufferers prone to infections, and usually leading to death before adulthood. Individuals with this syndrome also display centromeric instability in the form of radiated chromosomes, and facial anomalies (Hulten et al., 1978; Tiepolo et al., 1979). The chromosomal abnormality is restricted to chromosomes 1, 9 and 16 and is associated with demethylation of classical satellites 2 and 3 within the centromeric regions of these chromosomes (Jeanpierre et al., 1993). Dnmt3b has been demonstrated to be responsible for methylation of the satellite DNA in mouse ES cells (Lehnertz et al., 2003), and several studies have confirmed that ICF syndrome is caused by homozygous or compound heterozygous mutations of DNMT3B (Hansen et al., 1999; Okano et al., 1999; Xu et al., 1999). The majority of ICF causing mutations that have been characterised fall within the catalytic domain of DNMT3B, confirming that it is a defect in methylation that causes the syndrome (Hansen et al., 1999; Wijmenga et al., 2000). Additionally, ICF-causing mutations have been found which, whilst falling in the catalytic domain, do not lead to reduced enzymatic activity, but instead compromise binding of DNMT3L to DNMT3B, demonstrating that this interaction is required for correct methylation by DNMT3B (Xie et al., 2006). A mouse model for ICF has been established by introducing known ICF-causing mutations into the Dnmt3b gene, and this recapitulates to some degree the human ICF syndrome (Ueda et al., 2006). Interestingly, this has identified an essential role for Dnmt3b in the survival of T cells in the thymus of
newborn mice, possibly explaining the immunodeficiency observed in human ICF patients.

*DNMT3B* is also implicated, along with *DNMT1*, in the methylation and silencing of tumour suppressor genes in human cancer cells, contributing to the neoplastic proliferation of these cells (Rhee et al., 2002). Overexpression of *DNMT3B* is also frequently found in tumours, much more so than overexpression of *DNMT3A* or *DNMT1* (Robertson et al., 1999).

1.5.3.5 *Dnmt3a* and 3b in ES cells – a contribution to maintenance methylation

*Dnmt3a* and 3b are considered to be de novo methyltransferases because they are responsible for the methylation of newly integrated, unmethylated proviral or episomal DNA (Hsieh, 1999; Okano et al., 1999). However it is worth noting that in an in vitro assay of methylation activity, Dnm3a and 3b both displayed an equal level of de novo and maintenance methylation activity (Okano et al., 1998a). Further evidence discussed below shows that in ES cells these enzymes also have a role in the maintenance of methylation, and that they act almost redundantly in this.

Despite differences in phenotype and sequences methylated in embryos, *Dnmt3a* or 3b ES cells display demethylation of very few sequences examined, even after 5 months in culture (Chen et al., 2003b; Okano et al., 1999). Similar extended passaging of *Dnmt3a*, 3b ES cells however leads to extensive demethylation of almost every sequence examined, including endogenous C-type retroviral DNA, IAP repeats, minor and major satellite DNA, the highly methylated and tissue-specific β-globin and phosphoglycerate kinase 2 (PGK-2) autosomal gene promoters, the X-linked genes *PGK-1* and Xist, and the imprinted regions the H19 5' DMR, the DMR2 of Igf2, region 2 of Igf2r, the DMR of Peg1, and the DMR of Snrpn (Chen et al., 2003b). In agreement with this, the level of methylation in *Dnmt3a*, 3b ES cells has been shown to decrease steadily during prolonged passage, with as little as 0.6% of CpG dinucleotides methylated after 75 passages, compared to 65% in wild type cells (Jackson et al., 2004).
Thus Dnmt3a and 3b are thought to co-operate and to contribute together to the maintenance of the methylation of a variety of sequences, from repetitive sequences (Liang et al., 2002) to single copy DMRs and promoters (Chen et al., 2003b), and CpG islands (Hattori et al., 2004a).

Dnmt3a and 3b seem to be completely redundant in what sequences they are able to methylate in ES cells, but this is not the case in embryos (Okano et al., 1999). Furthermore, overexpression of Dnmt3a or 3b in human cells (which express very low levels of endogenous Dnmt3a and 3b), leads to differential, non-random methylation of both stable episomes and their corresponding chromosomal targets, suggesting the two enzymes have different target specificities (Hsieh, 1999). Also, expression of either Dnmt3a or 3b in highly demethylated Dnmt3a, 3b ES cells showed that both are capable of remethyllating the genome, however they have both common and preferred targets. For example, Dnmt3a preferentially methylated the major satellite repeats and Dnmt3b the minor satellite repeats; Dnmt3a was able to methylate the 5' region of Xist and the H19 5'DMR, but Dnmt3b was not (Chen et al., 2003b). Thus, despite the evidence suggesting that neither Dnmt3a nor 3b has any target specificity other than for the CpG dinucleotide (Dodge et al., 2002; Okano et al., 1998a), there may be some CpG dinucleotide target specificity for each Dnmt. Nonetheless there is clearly a highly significant degree of overlap between the respective specificities.

1.5.4 Regulation of Dnmt3b transcript levels

Microarray analysis comparing undifferentiated human ES cells (hESCs) and hESCs differentiating into cardiomyocytes reveals that DNMT3B is downregulated 2x fold within 1 day of differentiation, by 5 fold after 3 days of differentiation, and by 17 fold after 12 days (Beqqali et al., 2006). DNMT3B is also downregulated during hESC differentiation into embryoid bodies (EBs) (Bhattacharya et al., 2005). It is not known how this is regulated in vivo, however a few observations have been made regarding the regulation of Dnmt3b mRNA levels in several cell types. The level of Dnmt3b transcript present in the cell is regulated by multiple mechanisms. At the level of transcription,
Dnmt3b mRNA levels are regulated by both DNA methylation and histone hypoacetylation of the Dnmt3b enhancer region; treatment of ES cells with retinoic acid leads to Dnmt3b down-regulation, but subsequent treatment with either TSA or 5-aza-dC leads to reactivation of the gene (Tanaka, 2005). Additionally, the Dnmt3b promoter is bound by both Nanog and Sal14 (Wang et al., 2006), although it is not clear what this contributes to transcriptional control of Dnmt3b.

Dnmt3b mRNA levels are also regulated by affecting mRNA stability, at least in some cell types; studies in human endometrial adenocarcinoma Ishikawa cells (referred to as Ishikawa cells) have showed that expression of Dnmt3b is down-regulated upon exposure of the cells to the histone deacetylase inhibitor TSA (Xiong et al., 2005). Cloning of the Dnmt3b promoter into luciferase expression vector demonstrated that TSA treatment does not affect the rate of transcription from the promoter, instead the observed decrease in mRNA levels is shown to be due to decreased Dnmt3b mRNA stability. Treatment of these cells with the transcriptional inhibitor Actinomycin-D, followed by measurement of the Dnmt3b mRNA levels by qRT-PCR, demonstrated that treatment with TSA led to an increased rate of decay of this transcript, in a process dependent on protein synthesis. This down-regulation led to a measurable decrease in the levels of de novo methylation activity and to demethylation of the minor satellite repeats.

1.5.4.1 Dnmt3a and 3b isoforms
At least two isoforms of Dnmt3a have been reported, Dnmt3a and Dmnt3a2 (Fig.1.3). Each transcript is initiated from a separate promoter, and Dmnt3a2 is shorter, lacking the N-terminal region of full-length Dnmt3a. The N-terminal region missing in 3a2 is responsible for the localisation to heterochromatic foci of Dnmt3a, and it instead localises diffusely in the nucleus (Chen et al., 2002). Both Dnmt3a and 3a2 are capable of remethylating Dnmt3a, 3b^{(-/-)} ES cells, however Dnmt3a2 is the main isoform expressed in ES cells, germ cells, and the early embryo; Dnmt3a is instead expressed at low levels in somatic tissues (Chen et al., 2002).
Dnmt3b has been reported to have at least six isoforms (Fig.1.3), produced by alternative splicing, in both human and mouse cells (Chen et al., 2003b; Okano et al., 1998a; Xie et al., 1999). A study of the enzymatic properties of the Dnmt3b isoforms in vitro reveals that Dnmt3b(1) and (2) display similar levels of activity, whereas Dnmt3b(3) has no methyltransferase activity, due to the absence of a section of the catalytic MTase domain (Aoki et al., 2001; Okano et al., 1998a). In agreement with this, Dnmt3b(1) is able to restore the methylation of the majority of sequences demethylated in Dnmt3a, 3b− ES cells, but Dnmt3b(3) is unable to restore any of the methylation patterns (Chen et al., 2003b). Dnmt3b(1) has been shown to be most strongly expressed in ES cells and germ cells, whereas Dnmt3b(3) is expressed at low levels in almost all somatic tissues and cell lines examined (Beaulieu et al., 2002; Chen et al., 2003b; Okano et al., 1998a). Dnmt3b(4) was thought to be inactive, because it also lacks some of the catalytic domain. However overexpression of this isoform in human hepatocellular carcinoma cells may lead to hypomethylation of pericentromeric satellite regions, suggesting it may have an antagonistic role (Saito et al., 2002). Dnmt3b(5) and (6) are also presumed to be inactive because they lack some of the catalytic motifs, however their function has not been studied.
Figure 1.3 Schematic comparison of Dnmt3a and Dnmt3b isoforms. The conserved PWWP and PHD domains, the methyltransferase motifs (I, IV, VI, IX and X), and the sites of alternative splicing are indicated. The C-terminal 45 amino acids of Dnmt3b5 are out of frame and shown as an open bar. Diagram based on Fig.2, Chen et al., 2003
1.5.4.2 Dnmt interactions

Both Dnmt3a and Dnmt3b bind to a variety of other proteins in a range of cell types, including a number of epigenetic modifiers. For Dnmt3b, the list includes the chromatin remodelers hSNF2H and Lsh, HDACs 1 and 2, the HMTase Suv39h1, HP1, and the transcriptional co-repressor Sin3A (Bai et al., 2005; Geiman et al., 2004a; Geiman et al., 2004b; Zhu et al., 2006). DNMT3B also interacts with components of the mitotic chromosome condensation machinery (Geiman et al., 2004a). Dnmt3a interacts with Hdac activity (Fuks et al., 2001), HP1, and HMTase activity in the form of SUV39H1 (Fuks et al., 2003a), as well as Mbd3 and Brg1 (Datta et al., 2005). This supports the notion that DNA methylation is intimately connected to other histone modifications. Additionally, baculovirus-expressed DNMT1, DNMT3A and DNMT3B interact with each other in human cells, possibly linking DNMT3A and 3B to the replication fork (Kim et al., 2002). The only interactions confirmed in ES cells are with Lsh and HP1 (Lehnertz et al., 2003; Zhu et al., 2006), but it will be interesting to discover if the other interactions are also conserved between cell types.

1.5.4.3 Targeting of the Dnmts to DNA

None of the DNA methyltransferases have any intrinsic sequence specificity in vitro, and so they must either be actively differentially recruited, or prevented from binding (Dodge et al., 2002; Okano et al., 1998a; Tajima and Suetake, 1998; Yoder et al., 1997a). Evidence for the latter comes from the observation that Sp1 elements in the CpG island of the Aprt gene can protect the region from de novo methylation both in cells and transgenic mice (Brandeis et al., 1994; Macleod et al., 1994). The maternal factor Stella/PGC7, which is essential for early development, was recently found to protect certain imprinted sequences from demethylation during early embryogenesis, although it is not clear how this is done (Nakamura et al., 2006). The notion of differential active recruitment of the DNA methyltransferases is supported by the finding that some of the Dnmts are recruited by sequence-specific transcriptional repressors. For example, Dnmt1 interacts with the tumour suppressor protein Rb, and is targeted to Rb/E2F
responsive genes where it acts as a transcriptional co-repressor (Robertson et al., 2000). Dnmt3a binds the sequence-specific repressor RP58, and again acts to enhance transcriptional repression at RP58 responsive promoters (Fuks et al., 2001). In both these examples, the DNA methyltransferase activity of the Dnmts is dispensable for transcriptional silencing.

An example of where binding of Dnmts leads to methylation as well as silencing is the RARβ2 promoter. Both Dnmt1 and Dnmt3a are recruited by the oncogenic transcription factor PML-RAR to the RARβ2 promoter, leading to hypermethylation and silencing of the gene (Di Croce et al., 2002). Dnmt3b has also been demonstrated to be targeted by binding to HP1, but this is not sequence-specific but rather relies on underlying Me3H3K9 (Lehnertz et al., 2003); to my knowledge sequence specific recruitment of Dnmt3b has not been shown to occur. Dnmt1, and possible 3a and 3b are also recruited to the replication fork to re-establish methylation patterns following replication of heterochromatin (Chen et al., 2003b; Leonhardt et al., 1992). These examples demonstrate the principle that Dnmts can be recruited by other protein factors to help in both de novo silencing and maintenance of the silencing process.

In ES cells that lack Dnmt1, it has been shown that non-CpG methylation can occur, and that this is possibly mediated by Dnmt3a (Ramsahoye et al., 2000). In vitro, however, Dnmt3b displays a greater capability of non-CpG methylation than does Dnmt3a (Aoki et al., 2001). An involvement of all the Dnmts in non-CpG methylation is suggested by the finding that the telomeric repeats, which do not contain CpG, are heavily methylated in a manner dependent on either Dnmt1, 3a or 3b (Gonzalo et al., 2006).

1.5.4.4 Dnmt3l

Dnmt3l was originally identified on the basis of its sequence similarity to Dnmt3a and Dnmt3b, but it does not have an active methyltransferase domain, and is enzymatically inactive in vitro (Aapola et al., 2001; Chedin et al., 2002; Hata et al., 2002). Dnmt3l homozygous mutant mice are viable, but both mutant males and females are infertile.
Dnmt3l females fail to establish maternal methylation imprints in oocytes, leading to embryonic lethality of heterozygous offspring before mid-gestation and differentiation defects in the extraembryonic tissues (Arima et al., 2006; Bourc'his et al., 2001; Hata et al., 2002). Dnmt3l males showing azoospermia and loss of paternal methylation imprints at the H19 locus (Bourc'his and Bestor, 2004; Bourc'his et al., 2001; Hata et al., 2002; Kaneda et al., 2004). The establishment of maternal methylation imprints during oogenesis possibly involves an interaction with Dnmt3a. (Hata et al., 2002; Kaneda et al., 2004). A closer examination of Dnmt3l testes reveals that the germ cells are arrested and die around the early meiotic stage. Microarray analysis shows down-regulation of various gonad-specific and/or sex-chromosome-linked genes, and up-regulation of the retroviral-like IAP sequences (Hata et al., 2006), which are also demethylated (Bourc'his and Bestor, 2004; Hata et al., 2006). Thus Dnmt3l has a role in the regulation of germ-cell specific expression and the suppression of IAP retroviral sequences, which is important for the maintenance of genomic stability at this critical time (Walsh et al., 1998).

Recombinant Dnmt3l binds to both Dnmt3a and 3b directly and stimulates their DNA methylation activity in vitro, and it is thought to regulate methylation by co-operating with Dnmt3a and 3b to stimulate their de novo methylation activity in vivo (Chedin et al., 2002; Gowher et al., 2005; Suetake et al., 2004). Binding of Dnmt3l to Dnmt3a possibly involves a conformational change in Dnmt3a, which increases its affinity for DNA and the methyl group donor S-adenosylmethionine (Gowher et al., 2005). As mentioned above, the interaction between Dnmt3a and Dnmt3l is thought to be involved in the establishment of maternal methylation imprints during oogenesis (Hata et al., 2002; Kaneda et al., 2004), and mutations in Dnmt3b that are associated with ICF syndrome have been discovered that affect regulation of Dnmt3b by Dnmt3l (Xie et al., 2006), showing the functional importance of these interactions. Like Dnmt1, 3a and 3b, Dnmt3l can interact with Hdac1 and repress transcription in a Hdac-dependent manner in vitro. Both the interaction and repression capability require the conserved PHD domain (Deplus et al., 2002). Given that Dnmt3l lacks a methyltransferase domain, this substantiates the notion that the Dnmts can repress transcription in a methyltransferase-
independent manner (Deplus et al., 2002; Fuks et al., 2000; Fuks et al., 2001; Geiman et al., 2004b).

1.6 Embryonic Stem Cells

Embryonic Stem (ES) cells represent the \textit{in vitro} counterparts to cells of the epiblast, an \textit{in vivo} population of cells specific to the early embryo, which contribute to the development of the entire embryo (Gardner and Brook, 1997; Nichols, 2001). They are unique because they display the dual properties of self renewal and pluripotency; in the presence of the cytokine leukaemia inhibitory factor (LIF), mouse ES cells can be cultured indefinitely \textit{in vitro}, yet retain the ability to contribute to any tissue of the embryo even after prolonged culture (Beddington and Robertson, 1989; Keller, 2005; Smith, 2001).

1.6.1 ES cells and the Inner Cell Mass (ICM)

Cells of the early mouse embryo undergo massive changes in transcription patterns and their overall epigenetic state. During pre-implantation development a series of critical events occurs, including the first cell divisions, the establishment of cellular contacts and the first lineage differentiation which produces two cell types, the precursors of the trophectoderm lineage and the precursors of the embryo proper, the inner cell mass or ICM (also known as the epiblast) (Hamatani et al., 2004). At this stage of development, which occurs at 3.5 days post coitum (dpc), the embryo is comprised of approximately 60 cells and is referred to as a blastocyst. ES cells are derived from the inner cell mass (ICM) of a blastocyst which has been removed from the embryo and cultured \textit{in vitro} (Evans and Kaufman, 1981; Martin, 1981). They are thought to represent a developmental stage between that of the ICM/early epiblast and the late epiblast, which expresses different genes and has a different cell cycle to that of the early epiblast. It is commonly assumed that embryonic stem cells and the ICM are equivalent, but this is not actually the case – although they maintain the important ability to be able to contribute to any cell type of the embryo, there are certainly differences between them. For
example there are differences in expression of certain genes between 3.5dpc embryos, 4.5dpc embryos and ES cells (Kaji, Nichols and Hendrich, In press). Of the genes analysed so far, a number have been identified that are associated with pre-implantation development. These include Pramel4, Pramel5, Pramel6, Pramel7, Ppp2c2r, 2410076J21Rik and Sohlh2, which are down-regulated and 2 genes, HtrA1 and Pak1, which are upregulated in 4.5dpc embryos and ES cells compared to 3.5dpc embryos. Thus ES cells appear to be more like 4.5dpc embryos than 3.5dpc embryos; having undergone silencing of at least some aspects of the pre-implantation genetic programme.

Recent advances in chromatin immunoprecipitation (ChIP) technology have allowed the analysis of epigenetic marks on cells of the ICM (previously they could not be analysed in this way because far more cells are required for conventional ChIP than are available). This showed that the intensity of two epigenetic silencing marks, H3K9me2 and H3K4me1, was somewhat diminished in ES cells compared to the ICM at four genes silenced in both types of cell Cdx2, Cfc1, Hhex and Nkx2-2 (O'Neill L et al., 2006). ES cells are methylated at approximately 65 – 75 % of CpG dinucleotides, considerably more than the 3.5dpc blastocysts from which they are derived (which have c.20% of methylation), but less than post-implantation embryos which have c.60% methylation (Santos et al., 2002; Weng et al., 1995).

ES cells cannot be derived from all organisms. For example, despite the overall similarity in the genetic make-up between rats and mice, the derivation of true rat ES cells has so far not been achieved, although recently the derivation of embryonic stem-cell like cells has been described (Schulze et al., 2006). Even more remarkably there are large differences in the success rate of ES cell derivation between different strains of species from which ES cells can be successfully derived. For example the success rate of ES cell derivation in mouse varies from 0% to approximately 30%, depending on the strain and the method used (Brook and Gardner, 1997). Why this is the case is not clear but must be remembered when considering the in vivo relevance of ES cells.
1.6.2 Self-renewal of ES cells – a dependence on cytokine signalling

Mouse ES cells rely on the activation of signalling pathways by two cytokines to allow them to continue to symmetrically self-renew. These are Leukocyte Inhibitory Factor (LIF) (Smith et al., 1988; Williams et al., 1988) and Bone Morphogenetic Protein (BMP) (Ying et al., 2003). LIF, a member of the IL6 family of cytokines, binds to the cell surface LIF receptor and induces heterodimerisation of it with the glycoprotein 130 (Davis et al., 1993; Gearing et al., 1991). This leads to JAK-kinase mediated recruitment, activation and nuclear translocation of the STAT3 transcription factor (Niwa et al., 1998). In the absence of LIF, ES cells are unable to self-renew and instead differentiate into flattened epithelial-like cells. CTNF, a LIF-related cytokine, can substitute for LIF for ES cell self-renewal (Pennica et al., 1995). Interestingly, this LIF/STAT3 signalling pathway is dispensable for the self-renewal of human ES cells (Matsuda et al., 1999).

Members of the BMP family act by inducing expression of members of the Id family of transcriptional repressors, which inhibit genes promoting neural induction. BMP family members known to act in this way include BMP2, BMP4 and GDF6 (Ying et al., 2003). ES cell self-renewal is negatively impacted by ERK signalling. Activation of the mitogen activated protein kinases (MAPK) Erk1 and Erk2 can occur in response to binding of Gp130 cytokines (Burdon et al., 1999; Fukada et al., 1996), and this appears to promote differentiation (Burdon et al., 1999).

1.6.3 Pluripotency

The property of pluripotency requires that undifferentiated cells need to be able to activate the relevant cascades of transcriptional networks leading to differentiation upon receipt of the relevant stimulus. This is in contrast to more differentiated cells, which are unable to respond to stimuli to change their cell fate. Failure to correctly initiate or control differentiation is disastrous for development, as can be seen from the phenotypes of targeted deletions of a number of transcription factors involved in the transition from a pluripotent cell to a differentiated cell; all produce embryonic lethal phenotypes due to
a failure of the epiblast to expand correctly (Avilion et al., 2003; Hanna et al., 2002; Mitsui et al., 2003; Nichols et al., 1998). Importantly, the understanding of the transcriptional control of ES cells, and the capability to promote differentiation into specific cell types in vitro holds great promise for regenerative medicine in the future (Donovan and Gearhart, 2001; Loebel et al., 2003).

1.6.4 ‘Master regulators’ of pluripotency?
A number of factors have been discovered which are critical for the maintenance of pluripotency of ES cells both in vivo and in vitro. Among them are the transcription factors Oct4, Nanog and Sox2. Depletion of any of these prevents the cells renewing and causes them to differentiate; they are positive regulators of ES cell fate. As such they have been proposed to act as the master regulators of a transcriptional network controlling pluripotency (Boyer et al., 2005; Loh et al., 2006). It is likely that there are more such regulatory factors that have yet to be identified, however these four will be discussed in some detail because much work has been done to discover their target genes and mechanism of action.

1.6.4.1 Oct4
Oct4 (octamer-binding transcription factor 4), also known as Oct3, is a member of the POU (PIT/OCT/UNC) domain class of homeodomain proteins, encoded by the Pou5f1 locus. The levels of this protein are critical for specifying ES cell fate. In both humans and mice, the absence of Oct4 causes both epiblast and ES cells to take on a trophectoderm fate. However overexpression of this protein causes cells to take on an extraembryonic endodermal fate (Matin et al., 2004; Nichols et al., 1998; Niwa et al., 2000). Oct4 is a DNA binding transcription factor that binds to a classical octamer sequence, ATGCAAAAT. A number of factors have been found to regulate Oct4 expression, including DNA methylation and histone acetylation (Hattori et al., 2004b), and binding by other transcription factors (Gu et al., 2006).
1.6.4.2 Sox2

Sox2 belongs to the Sox (SRY-Related HMG box) family of transcription factors, which possess a single HMG DNA-binding domain (Kamachi et al., 2000). Like Oct4, Sox2 is essential for normal pluripotent cell development and maintenance (Avilion et al., 2003). Sox2 binds to a sox element, the sequence of which is CATTGTAA. Sox2 is positively regulated both by itself and by Oct4 (Chew et al., 2005).

1.6.4.3 Nanog

This variant homeodomain protein, closely related to the Nkx family of homeodomain proteins (Lints et al., 1993), also has a major role in directing self-renewal and maintaining pluripotency in ES cells. Nanog-overexpressing human and mouse ES cells can self-renew in a LIF- and BMP-independent fashion and are resistant to differentiation (Chambers et al., 2003; Darr et al., 2006; Mitsui et al., 2003). Positive regulators of Nanog include Oct4 and Sox2 (Rodda et al., 2005), and Foxd3 (Pan et al., 2006). p53 negatively regulates Nanog expression in response to DNA damage. Repression by p53 mediated mechanisms involves recruitment of the mSin3A transcriptional co-repressor complex, which has histone deacetylase activity (Lin et al., 2005).

1.6.4.4 Co-operativity between Oct4, Nanog and Sox2 in gene regulation

The octamer sequence bound by Oct4 is frequently found next to sox elements, which are bound by Sox2. These composite octamer-sox element binding sites allow a synergistic interaction between Oct4 and Sox2 that is known to be essential for the regulation of several target genes. Some such genes which rely on binding of both Oct4 and Sox2 for their transcription include Fgf4, Utf1, Fbx15, as well as Pou5f1, Sox2 and Nanog (Ben-Shushan et al., 1998; Botquin et al., 1998; Catena et al., 2004; Chew et al., 2005; Dailey et al., 1994; Nishimoto et al., 1999; Okumura-Nakanishi et al., 2005; Rodda et al., 2005; Tokuzawa et al., 2003; Tomioka et al., 2002). Nanog has also been shown to co-operate with Sox2 in the upregulation of the pluripotency marker Rex-1.
(Shi et al., 2006). Recently a protein known as Sall4 was found to co-occupy many Nanog binding sites in mouse ES cells, and to transactivate Nanog-binding-site-containing enhancers in somatic cells (Wu et al., 2006). It is suggested that Sall4 and Nanog form a regulatory circuit similar to that of Oct4 and Sox2.

Recent developments of chromatin immunoprecipitation (ChIP) techniques, such as ChIP cloning, CChIP and ChIP-on-chip have enabled a huge expansion in the number of genes known to be bound by either Oct4, Nanog or Sox2. These have revealed that a large number of genes in murine ES cells are bound and regulated by these transcriptional regulators, reinforcing the idea of these proteins as master regulators. Many of them are bound by more than one factor, allowing for combinatorial control. Use of the chromatin immunoprecipitation paired-end ditags method (ChIP-PET) using Oct4 and Nanog antibodies identified 1083 (Oct4) and 3006 (Nanog) binding sites for these factors in murine ES cells (Loh et al., 2006). For both Nanog and Oct4, a large proportion of binding sites were within or proximal (<10kb up or downstream) to genes. 4 miRNA genes were also found to be bound by Nanog, with 2 of them also being bound by Oct4.

RNAi-mediated knock-down of either Nanog or Oct4 resulted in either increases or decreases of a large number of these genes, showing that both factors affect levels of transcription. Although Sox2 binding was not studied in this experiment, an analysis of motifs common to the regulatory regions of genes bound by Oct4 alone showed that sox-oct composite elements were present at many of the loci, suggesting that Sox2 is also involved in the regulation of many of the genes regulated by Oct4. ChIP of some of these genes demonstrated that Sox2 does indeed bind to these loci. Of the 345 genes shown to be co-bound by Oct4 and Nanog, 30 of them encode known or putative DNA-binding regulators. Several others were shown in knock-down experiments to be involved in maintaining ES cells in an undifferentiated state. Independently, ChIP-on-Chip was used in human ES cells to identify binding sites for OCT4, SOX2 and NANOG (Boyer et al., 2005). This revealed that the three factors co-occupy the promoters of a large population of genes. Intriguingly, there is little overlap between the
putative Oct4/OCT4 and Nanog/NANOG bound genes identified in the two studies: in total, only 32 genes were found to be bound by both Oct4/OCT4 and Nanog/NANOG in both species (Loh et al., 2006). Whether this represents true differences between human and mouse ES cells or simply indicates a lack of robustness in one or both studies, needs to be addressed before the real meaning of the results can be determined.

1.6.4.5 Other regulators of ES cell fate

Increasingly, more factors are being identified in addition to the four discussed which are involved in pluripotency of ES cells. For example, c-myc and Klf4 have recently been identified as being two of only four factors (Oct4 and Sox2 being the others) able to confer pluripotency when overexpressed in mouse embryonic or adult fibroblasts (Takahashi and Yamanaka, 2006). Such somatic cells overexpressing these four factors together display the morphology and growth properties of ES cells, and express ES cell marker genes. Importantly, these 'reprogrammed' cells contributed to embryonic development when injected into blastocysts (Takahashi and Yamanaka, 2006). C-myc is a transcription factor which associates with histone acetyltransferase complexes, including the TIP60 and GCN5 complexes (McMahon et al., 1998), CBP and p300 (Vervoorts et al., 2003). There are thought to be up to 25,000 c-myc binding sites in the mouse genome, and one hypothesis is that c-myc induces global histone acetylation, promoting binding of Oct4/Sox2 to their targets (Fernandez et al., 2003). Klf4 may exert its effects by repressing p53. Because p53 is involved in repression of Nanog during ES cell differentiation, this would promote Nanog expression (Lin et al., 2005). FoxD3 is another factor identified as being necessary for the maintenance of a pluripotent state in ES cells (Hanna et al., 2002).

A number of genes, chosen for examination on the basis that their expression in ES cells is down-regulated upon differentiation, have recently been identified by shRNA loss-of-function techniques as having a role in ES cell self-renewal (Ivanova et al., 2006). As expected, these include Oct4, Nanog and Sox2, further underlining the importance of these proteins in ES cell behaviour. Esrrb, Tbx3 and Tcl1 are transcription factors also
identified as having a role here. Down-regulation of any of them results in
differentiation down particular lineages; they each repress different lineage specific
pathways in ES cells. Between them they repress multiple differentiation pathways and
promote self-renewal. It remains to be resolved whether the concept of a few master
regulators is valid, or if many more are yet to be identified. Nevertheless, the factors
already identified are clearly very important in the regulation of ES cell fate.

1.6.5 Polycomb group proteins in ES cells

ChIP-on-Chip experiments in both human and mouse ES cells has shown that
components of both the PRC1 and PRC2 polycomb group (PcG) protein complexes bind
extensively throughout the genome to hundreds or more genes, some of which code for
transcription factors known or postulated to have important roles in development (Boyer
et al., 2006; Lee et al., 2006). Consistent with the well-established function of the PRC
complexes, genes associated with PcG protein binding were associated with H3K27
trimethylation and were generally not transcribed. Mouse ES cells deficient for the
PRC2 components displayed some degree of silencing failure at several such genes
(Azuara et al., 2006; Boyer et al., 2006; Lee et al., 2006). Induction of differentiation in
ES cells led to significant derepression of PcG-bound genes (Boyer et al., 2006).
Decreases in the levels of H3K27me3 and concomitant increases in the levels of
H3K4me3 and RNA polymerase II occupancy at upregulated genes further demonstrated
the degree to which specific histone modifications continue to be associated with
transcriptional status in differentiating ES cells. This suggests that PcG proteins may
contribute to the maintenance of ES cell identity, a theory that is supported by the
observations that Eed mutant ES cells spontaneously differentiate (Boyer et al., 2006),
and that ES cells cannot be derived from blastocysts lacking the PRC2 component Ezh2
(O'Carroll et al., 2001). However, ES cells lacking the PRC2 component Eed can
contribute to most cell lineages, suggesting that PcG proteins are not necessary for
maintaining pluripotency (Morin-Kensicki et al., 2001). It remains an open question
whether PcG proteins are necessary for ES cell identity.
An interesting addition to this picture is the observation that comparisons of the data sets generated from genome-wide studies of the binding of Oct4, Sox2 and Nanog with those from PcG proteins shows a significant degree of overlap of binding of the two datasets. In human ES cells, approximately one third of SUZ12-bound genes are bound by one or more of the pluripotency-associated factors (Lee et al., 2006). It is noteworthy that many of the genes identified as being bound by both PRC1 and PRC2 components in mouse ES cells were also identified as being bound by OCT4, SOX2 and NANOG in human ES cells (Boyer et al., 2006). Thus it is suggested that combinatorial control by Oct4/Sox2/Nanog and PcG proteins is used to regulate the transcription of at least some developmentally important transcription factors in ES cells.

1.6.6 Epigenetic features of ES cells

The ES cell epigenome possesses certain features that are unique to these cell types and are likely involved in the regulation of pluripotency. Notably, ES cell chromatin contains more marks of transcriptionally active or permissive euchromatin, such as acetylated histones and increased nuclease sensitivity, than chromatin of more differentiated cells (Arney and Fisher, 2004; Francastel et al., 2000). Structural chromatin proteins are also more loosely associated with the chromatin in ES cells than in differentiated cells (Meshorer and Misteli, 2006), and the expression of several ATP-dependent chromatin remodelers is elevated in ES cells compared to other cell types (Kurisaki et al., 2005). Transmission electron-microscopy studies reveal a transition from fine granular chromatin in undifferentiated human ES cells to irregularly shaped heterochromatic nuclei in RA-induced differentiating cells (Park et al., 2004), and direct visualization of centromeric heterochromatin with a probe against the major satellite-repeat sequence reveals a more diffuse heterochromatin structure in undifferentiated ES cells versus more compact heterochromatin with well-defined foci in ES-derived neuronal progenitor cells (NPCs) (Meshorer et al., 2006). During differentiation, ES cells must establish new heritable gene expression patterns. This relies upon chromatin organisation to achieve this, and so the permissive chromatin state of ES cells presumably reflects the need for them to be able to reorganise their chromatin when
induced to differentiate. Perturbing these epigenetic processes impacts upon the ability of ES cells to differentiate correctly.

### 1.6.7 A unique permissive chromatin structure in pluripotent cells?

Recent work has shown that the promoters of a large number of developmentally important, tissue-specific genes which are silent in undifferentiated ES cells contain a unique epigenetic signature that includes both the activatory mark H3K4me3 and the repressive mark H3K27me3 (Azuara et al., 2006; Bernstein et al., 2006). These have been termed 'bivalent domains' and it is suggested that they represent a 'poised' state allowing their rapid permanent activation or inactivation upon receipt by the cell of signals to differentiate down a particular pathway. This epigenetic pattern is only found in undifferentiated ES cells, and importantly it disappears upon differentiation, frequently being replaced with histone marks reflecting the transcriptional status of the gene. Additionally, this pattern commonly coincides with the most highly conserved non-coding elements in the mammalian genome, suggesting an evolutionarily conserved role for these chromatin domains (Bernstein et al., 2006). Importantly however, this bivalent pattern was not found at the promoters of a number of lineage-control genes, such as *Myf5* and *Mash1* (Azuara et al., 2006; Bernstein et al., 2006; Williams et al., 2006), suggesting that this chromatin structure regulates only a subset of lineage-determining genes and that other regulatory mechanisms must also be used. Additionally, there was very limited overlap between the genes identified in these two studies as possessing such an epigenetic pattern, potentially calling into question the validity of the results.

As discussed above, evidence suggests that the transcription of at least some developmentally important transcription factors in ES cells is controlled in a combinatorial fashion involving binding of the pluripotency factors Oct4, Sox2 and Nanog, as well as of PcG proteins. Interestingly, many of the PcG target genes contained bivalent chromatin domains in their promoter regions (Azuara et al., 2006; Bernstein et al., 2006; Boyer et al., 2006; Lee et al., 2006). The finding that PcG target genes were
preferentially activated upon induction of differentiation implied that they were ‘poised’ for activation (Boyer et al., 2006; Lee et al., 2006). It is interesting to speculate that the binding of both positively and negatively acting transcriptional regulators, along with the presence of both activatory and repressive epigenetic marks, contributes to a fine balance between transcriptional repression and activation (Fig.1.4). This fine and flexible control of transcription patterns may be what allows ES cells to activate any one of a number of differentiation pathways, i.e. to be pluripotent.
Figure 1.4 Schematic representation of the proposed 'poised' chromatin state, and the changes that can occur during differentiation. Establishment and maintenance of this chromatin state involves a combination of both PcG proteins and pluripotency associated transcription regulators. Components of the PRC1 PcG complex also bind to target gene promoters, however only the PRC2 complex is shown here for clarity. Adapted from Boyer et al, 2006.
1.7 The NuRD complex

The Nucleosome Remodelling and histone Deacetylation complex is an abundant complex that has been isolated from mammals and amphibians (Tong et al., 1998; Wade et al., 1998; Xue et al., 1998; Zhang et al., 1998). Homologues of subunits of the complex have also been identified in *D.melanogaster, A.thaliana* and *C.elegans*, suggesting the complex is widely conserved throughout the animal kingdom (Eshed et al., 1999; Herman, 2001; Herman et al., 1999; Kehie et al., 1998; Ogas et al., 1999; Solari and Ahringer, 2000; Solari et al., 1999). The NuRD complex is also implicated embryonic viability and patterning in a number of organisms including *D.melanogaster, C.elegans* and *A.thaliana* (Eshed et al., 1999; Kehle et al., 1998; Mannervik and Levine, 1999; Ogas et al., 1999; Shi and Mello, 1998; Solari et al., 1999).

1.7.1 Function of NuRD

In mammals NuRD acts as a transcriptional repressor and is involved in several cell-fate decisions. The complex is recruited to specific genes by a variety of transcription factors such as the *Drosophila* genes *hunchback* (Kehle et al., 1998) and *Tramtrack-69* (*Tik69*) (Murawsky et al., 2001), as well as murine *Ikaros*, which has a role in the specification of the lymphocyte lineage (Kim et al., 1999). Other sequence-specific factors involved in recruiting NuRD to target sites include recruitment by BCL11B to BCL11B target genes in T lymphocytes (Cismasiu et al., 2005), recruitment by FOG-1 to FOG-1/GATA-1 target genes in erythroid cells (Hong et al., 2005; Rodriguez et al., 2005), and recruitment to BCL-6 and Snail target genes by BCL-6 in B lymphocytes and breast cancer cells respectively (Fujita et al., 2004; Fujita et al., 2003). NURD is also implicated in repression by unliganded nuclear hormone receptors (Heinzel et al., 1997; Xue et al., 1998). An interaction between NuRD and APPL1, a RAB5 effector which translocates to the nucleus upon endocytosis of the appropriate extracellular stimulation, links transcriptional regulation by NuRD to extracellular signalling (Miaczynska et al., 2004). The central NuRD component MBD3 also interacts with Aurora-A kinases, leading to its phosphorylation and localisation to the centrosomes in early M phase in HeLa cells (Sakai et al., 2002), perhaps indicating an as yet unrecognised function of
NuRD during mitosis. Mbd3 has also been reported to interact with Dnmt1 and hemi-methylated DNA, presumably at the replication fork, suggesting a link to chromatin replication (Tatematsu et al., 2000). Mbd3 also interacts with Dnmt3a in mouse lymphosarcoma cells, although it is not clear if this is in the context of the NuRD complex (Datta et al., 2005).

There is also a link to methylated DNA, because although the NuRD complex has no intrinsic methylated-DNA binding ability of its own (Zhang et al., 1999), it can interact directly with Mbd2 (Tatematsu et al., 2000), which is capable of recruiting the complex to methylated DNA (Zhang et al., 1999). When bound to Mbd2, the complex is known as MeCP1 (Ng et al., 1999). Interestingly, in *D. melanogaster*, which possesses a single Mbd2/3 homologue, the NuRD complex in association with Mbd2/3 binds CpT/A methylated DNA, rather than the more common CpG methylated DNA (Marhold et al., 2004).

1.7.2 Composition of the complex

In mammalian cells NuRD is composed of nine polypeptides: Mbd3, Hdac1, Hdac2, RbAp46, RbAp48, Gata2da, Gata2db, Mi2β and Mta2, and is approximately 2MDa in size (Feng et al., 2002). Between them they possess both histone deacetylase and chromatin remodelling capabilities. Mi-2β, encoded by *Chd4*, is the largest component of the complex and was originally identified as a nuclear auto-antigen in patients with dermatomyositis (Seelig et al., 1995; Zhang et al., 1998). It contains several motifs characteristic of chromatin-associated proteins, including two plant homeodomain (PHD) zinc fingers though which it interacts with HDAC1 (Zhang et al., 1998), two chromo domains, one SWI/SNF2 helicase/ATPase domain, and an HMG-like domain (Seelig et al., 1995; Woodage et al., 1997). *In vitro* experiments showing that recombinant Mi-2β is capable of remodelling nucleosomes by itself in an ATP-dependent fashion suggest it could be responsible for the chromatin remodelling capabilities of the complex (Wang and Zhang, 2001). In addition to a role in nucleosome remodelling, Mi-2β is also likely involved in targeting of the complex; several of the
NuRD-interacting transcription factors mentioned above interact specifically with Mi-2β within the complex (Kehle et al., 1998; Kim et al., 1999; Murawsky et al., 2001).

Expressed widely (Xia and Zhang, 2001), Mta2 is the second largest subunit of the complex and is highly similar to the metastasis associated protein Mta1, which is associated with cancer metastasis and is a potent co-repressor in estrogen-receptor mediated transcriptional repression (Mazumdar et al., 2001; Toh et al., 1999; Toh et al., 1997; Toh et al., 1994; Zhang et al., 1999). Both Mta1 and Mta2 have been reported to be NuRD components; Mta1 in adrenal cortex adenocarcinoma SW13 cells (Xue et al., 1998), and Mta2 in HeLa cells (Zhang et al., 1999). MTA3 is similar to MTA1 and 2, and is expressed in a subset of lymphocytes where it is involved in the specification of cell fate in the context of the NuRD complex, as well as being a key component of an estrogen dependent pathway regulating growth and differentiation (Fujita et al., 2004; Fujita et al., 2003). This suggests some fluidity in the NuRD complex, presumably reflecting functional specialisation of NuRD sub-complexes (Bowen et al., 2004; Yao and Yang, 2003). All members of the MTA family contain a leucine zipper domain, a GATA-type zinc finger domain, and a SANT domain found in a variety of chromatin associated proteins (Aasland et al., 1996; Solari et al., 1999). The role of the Mta proteins within the NuRD complex appears to be stimulation of the histone deacetylase activity (Zhang et al., 1999), and this is likely to be mediated by the SANT domain (Guenther et al., 2001).

Two important functional components of the complex are the histone deacetylases Hdac1 and 2. Hdac/HDAC1 and Hdac/HDAC 2 are almost 80% identical and both belong to the family of Class I HDACs, along with Hdac/HDAC3 and Hdac/HDAC8 (Cress and Seto, 2000). Hdac/HDAC1 and Hdac/HDAC2 are also found in another major transcriptional repression complex, the Sin3 complex (Knoepfler and Eisenman, 1999; Zhang et al., 1997), along with the histone binding proteins RbAp46 and 48 (Verreault et al., 1998; Zhang et al., 1997; Zhang et al., 1998). Using baculovirus expressed proteins, it has been demonstrated that HDAC1/2 together with RbAp46/48
form a deacetylase core complex that is capable of actively deacetylating histones (Zhang et al., 1999).

Gata2da/2db (formerly known as p66/p68) are two highly related, zinc-finger containing potent transcriptional repressors demonstrated to be part of the NuRD complex (Brackertz et al., 2002; Feng et al., 2002). In human cells at least these components are thought to mediate interactions between the NuRD complex, Mbd2 and histones, thus targeting the complex to methylated DNA (Brackertz et al., 2002; Brackertz et al., 2006). Interestingly, both proteins are sumoylated in vivo, and this is necessary for full transcriptional repression in vitro, possibly by affecting the interaction of the proteins with HDAC1 (Gong et al., 2006). In D.melanogaster, Gata2da represses Wg target genes and plays an essential role in development through regulation of ecdysone responsive genes (Kon et al., 2005).

The smallest component of the NuRD complex is Mbd3, a highly conserved protein belonging to the family of methyl-binding domain proteins (Hendrich and Bird, 1998). In X.laevis two xMbd3 isoforms exist which differ in the inclusion of a small exon close to the methyl binding domain region. The isoform lacking this exon is, like other MBD containing proteins, able to bind methylated DNA. However the isoform containing the extra exon is not able to do this (Wade et al., 1999). Surprisingly, mammalian Mbd3 is unable to bind methylated DNA (Hendrich and Bird, 1998), and this has been attributed to a mutation (tyrosine to phenylalanine) in the MBD (Ohki et al., 2001; Saito and Ishikawa, 2002). This suggests that mammalian and X.laevis Mbd3 (short form) may have differing roles, with xMbd3 (short form) being involved in targeting of NuRD to methylated DNA, and mammalian Mbd3 having a scaffolding role within the complex. Mbd3 interacts with most of the other NuRD components (Zhang et al., 1999), and an intact NuRD complex does not form in its absence (Kaji et al., 2006). As mentioned above, binding to methylated DNA in mammals may occur via Mbd2 (Ng et al., 1999; Zhang et al., 1999). Mbd3 exists in 3 isoforms differing in their N-termini; Mbd3a, 3b and 3c, with Mbd3a being the longest and the only isoform with an intact methyl binding domain, (Fig.1.5) (Hendrich and Bird, 1998; Kaji et al., 2006; Zhang et al.,
All isoforms have an intact C-terminus, which has been reported to interact with other NuRD components (Saito and Ishikawa, 2002), and thus is likely to be most important functionally. Mbd3b is the predominant isoform in NuRD of proliferating HeLa cells (Zhang et al., 1999), embryonic neural cells (Jung et al., 2003), as well as in ES cells, although Mbd3a and c are also detectable in ES cells (Kaji et al., 2006). Inclusion of different Mbd3 isoforms in the NuRD complex may allow for subtle differences in the composition and/or specificity of the resulting NuRD complex.

There are two MBD3-like proteins, MBD3L1 and MBD3L2, which are encoded by mammalian genomes and are similar to both Mbd2 and 3 and interact with Mbd3 and components of the NuRD complex (Jiang et al., 2002; Jin et al., 2005). MBD3L1 is expressed only in haploid germ cells (Jiang et al., 2002), whereas MBD3L2 is expressed in multiple tissues (Jin et al., 2005). Like MBD2, MBD3L2 is a transcriptional repressor in vitro. Interestingly MBD3L2 can antagonise the repressive function of MBD2, and vice-versa (Jin et al., 2005). Thus, MBD3L2 has the potential to recruit the NuRD complex away from methylated DNA and to reactivate transcription.
Figure 1.5.a Schematic representation of the NuRD complex. When the complex interacts with Mbd2 it is known as the MeCP1 complex.

Figure 1.5.b The Mbd3 isoforms. MBD: Methyl binding Domain; SD: Splice donor site; SA: Splice acceptor site; bp: base pairs; aa: amino acids; ATG: Initiating methionine codon. Mbd3a is the only isoform with an intact MBD, and Mbd3c lacks all of Exon 1.
1.7.3 Mechanism of action

The presence of both nucleosome remodelling and histone deacetylation activities within the NuRD complex suggests a function as a transcriptional repressor. This is supported by the observation that targeting the complex to reporter genes through several NuRD components, including Gata2da, Gata2db, Mi-2β, and Mbd2, resulted in transcriptional repression (Feng et al., 2002; Feng and Zhang, 2001; Kon et al., 2005; Ng et al., 1999). This repression involves histone deacetylation, because treatment with the histone deacetylase inhibitor TSA partially relieved the repression (Feng et al., 2002). When associated with Mbd2 as part of the MeCP1 complex, NuRD can preferentially bind, remodel and deacetylate methylated nucleosomes, providing a functional link between three different silencing mechanisms (Feng and Zhang, 2001; Zhang et al., 1999). Nucleosome remodelling by Mi-2β is an important part of the process, because expression of a Mi2 mutant that is defective in its ATPase activity relieves the methylation-dependent transcriptional repression (Feng and Zhang, 2001). It has not been shown definitively to have these same activities at unmethylated DNA, but the interaction of the NuRD complex with numerous transcription factors (see above) and associated repression of their target genes clearly implies a role in transcriptional silencing, presumably via both histone deacetylase and chromatin remodelling activities.

Interestingly, the association of NuRD with the zinc-finger binding protein Ikaros involves toroidal structures presumed to be associated with centromeric heterochromatin in the G1 and S phases of the T lymphocyte cell cycle (Brown et al., 1997; Kim et al., 1999), suggesting that NuRD has a function in centromeric silencing. A further link to transcriptional repression comes from the observation that NuRD can interact in vitro with H3 N-terminal tails that are either unmethylated or methylated on H3-Lys 9, and importantly this interaction is disrupted by methylation on H3-lys 4, an epigenetic mark associated with transcriptional activation (Nishioka et al., 2002; Zegerman et al., 2002).
1.7.4  *Mbd3* targeting strategy

Two ES cell lines lacking a functional *Mbd3* gene were produced by different methods (Fig.1.6). All the data presented in this thesis was done using cells produced by the second targeting strategy (B), however every experiment was confirmed with cells produced by the first targeting strategy (A). The first targeting strategy (Fig.1.6.A) involved producing *Mbd3* heterozygous ES cells by using homologous recombination to replace exons 2-7 on one wild type allele with a promoterless β-geo cassette (termed M3β6C cells). Two independently derived cell lines were used to make chimeric mice, which were used to make heterozygous mice. These were then crossed to examine the homozygous phenotype (Hendrich et al., 2001). Subsequently, a floxed hygromycin-thymidine kinase cassette was integrated adjacent to exon 1 on the wild type allele of these M3β6C ES cells to produce HyTk/- ES cells. These cells were then transfected with a floxed exon 1 construct and the (HyTk/-) ES cells selected against using Gancyclovir (III). As expected, clones were recovered in which exon 1 was flanked by LoxP sites, but surprisingly clones were also discovered in which no wild type Mbd3 allele was present (IV). The absence of *Mbd3* expression in these cells was confirmed by RT-PCR and western blotting (Kaji et al., 2006). These were most likely produced through loss of the Hygromycin/Thymidine kinase cassette containing allele, and duplication of the β-geo allele (Kaji et al., 2006).

The second targeting strategy (Fig.1.6.B) involved transfecting Cre into the HyTk/- ES cells to excise the floxed Hygromycin/Thymidine kinase cassette upstream of exon 1, then using homologous recombination to integrate a floxed Hygromycin/Thymidine kinase cassette downstream of exon 7, which was subsequently removed by transfection of Cre into the targeted cells. A further transfection of Cre into these cells resulted in the excision of the entire *Mbd3* gene, thus producing a null allele (Kaji et al., 2006).
Fig. 1.6 A) First *Mbd3* targeting strategy. Exons 2-7 of the *Mbd3* gene were replaced with a promoterless βgeo cassette in embryonic stem (ES) cells (I). Next, a floxed hygromycin-thymidine kinase cassette was integrated adjacent to exon 1 on the wild type allele of these M3β6C ES cells to produce HyTk/- ES cells (II). Subsequently these cells were transfected with a floxed exon 1 construct and the HyTk/- cells selected against using Gancyclovir (III). Clones were recovered in which exon 1 was flanked by LoxP sites. Clones were also recovered in which no wild type *Mbd3* allele was present (IV). These were most likely produced through loss of the Hygromycin/Thymidine kinase cassette containing allele, and duplication of the β-geo allele. Adapted from Kaji et al, 2006 and Hendrich et al, 2001.
Fig. 1.6.13 Second Mbd3 targeting strategy. LoxP sites were introduced sequentially at either end of the Mbd3 gene in ES cells in which one Mbd3 allele had already been replaced with a β-geo cassette. This was achieved via homologous recombination of a floxed Hygromycin/Thymidine kinase cassette, selection with gancyclovir, then transfection with Cre, to produce cells containing the floxed Mbd3 allele (Mbd3<sup>flox<sup>−</sup></sup>), (I - IV). A further Cre transfection into these cells resulted in excision of the floxed Mbd3 allele (V), producing Mbd3<sup>−/−</sup> ES cells. Adapted from Kaji et al, 2006.
1.7.5 The \(Mbd3^{(-)}\) phenotype

Targeted deletion of \(Mbd3\) reveals that it is essential for mouse embryogenesis, and its loss causes early embryonic lethality (Hendrich et al., 2001). \(Mbd3^{(-)}\) embryos are morphologically normal until 5.5dpc, when they appear runted and have a reduced number of \(Oct4\) positive pluripotent cells. Null embryos are largely resorbed by 8.5dpc. Closer examination reveals that they fail to form a mature epiblast, and do not initiate the robust proliferation, proamnionic cavity formation and distal displacement of the ICM population characteristic of wild type embryos. In addition they display a lack of extraembryonic ectoderm and organized visceral endoderm after implantation (Kaji, Nichols and Hendrich, In press). \(Mbd3\)-deficient ICMs grown in culture give rise to primitive endoderm but fail to expand their Oct4 positive cell population. Thus \(Mbd3\) is required for the transition of pluripotent cells in the peri-implantation embryo, and for the maintenance of the pluripotent cell population. The lack of phenotype until 5.5dpc may be accounted for by the presence of maternal \(Mbd3\), which is expressed until the morula stage (Kaji, Nichols and Hendrich, In press).

ES cells lacking \(Mbd3\) are viable but proliferate slowly and display profound differentiation defects; they fail to differentiate in the absence of LIF, and whilst they can initiate differentiation in embryoid bodies or chimeric embryos, they fail to commit to developmental lineages. Thus \(Mbd3\) is required for the pluripotency of ES cells (Kaji et al., 2006). Misexpression is seen of a number of genes, including some associated with pre-implantation development e.g. \(Pramel\) 6 and 7. These are normally silenced in embryonic stem cells, but are upregulated in \(Mbd3^{(+)}\) ES cells. The expression of the primordial germ cell marker \(Dppa3\) (also known as \(Pgc7\) or \(Stella\)) (Saitou et al., 2002; Sato et al., 2002) is instead inappropriately silenced (Kaji et al., 2006). Microarray analysis of \(Mbd3\) null ES cells reveals mis-expression of around 200 genes, equivalent to 1.7% of the total number of genes recorded as present on the microarray chip (K.Kaji, S.Tomlinson, personal communication). In line with NuRD having a role in repressing genes by deacetylation and/or other mechanisms, the promoters of \(Pramel\) 6 and \(Pramel\) 7 are hyperacetylated in \(Mbd3^{(-)}\) ES cells (Kaji et al., 2006). \(Mbd3^{(-)}\) ICMs also fail to
down-regulate a number of pre-implantation specific genes, possibly contributing to the failure of embryonic development beyond 5.5dp. Interestingly, the aberrations in gene expression in Mbd3(−/−) ES cells do not correlate exactly with those in Mbd3(−/−) ICMs, confirming that ES cells and cells of the ICM are different (Kaji, Nichols and Hendrich, in press). ES cells lacking Mbd3 show reduced levels of the NuRD components Mta1, Mta2 and RbAp48, showing there is some auto-regulation by these components. Importantly, Mta1, Mta2 and Hdac1 cannot be co-immunoprecipitated in Mbd3(−/−), demonstrating that Mbd3 is necessary for the formation of the NuRD complex. Thus the NuRD complex is required for the cell fate commitment of pluripotent cells.

In addition to mis-expression of genes and hyperacetylation of gene promoters, ES cells lacking Mbd3 also display demethylation of their DNA. This has been shown at the highly repetitive major and minor satellites, and the long terminal repeat regions (LTRs) of the intracisternal-A particle (IAP) DNA retrotransposon. As can be seen from Fig. 1.7, digestion of genomic DNA from either parental, Mbd3 null, or Mbd3a, 3b or 3c rescue cells with the methylation sensitive enzyme MspI results in a clearly identifiable pattern of digestion at both the satellite regions (A, B) and the IAP-LTRs (C). However digestion with the methylation sensitive isoschizomer HpaII shows that at both the satellites and IAP-LTRs, genomic DNA from the parental cell lines is methylated; it cannot be digested with this enzyme. Digestion of genomic DNA from Mbd3 null ES cells with HpaII produced a similar pattern to digestion with MspI, showing that the DNA is not methylated at these regions in these cells. Mbd3 null cells that have been rescued with either Mbd3a, 3b or 3c are not digested by HpaII at these regions, i.e. they are methylated, so the regions of heterochromatin demethylated in the absence of Mbd3 maintain sufficient marks to allow re-methylation once NuRD function is restored. Only Mbd3a possesses an intact methyl binding domain (Hendrich and Bird, 2000), so an effect on DNA methylation is not dependent on the MBD. Methylation of repetitive sequences such as these has previously shown to be affected in ES cells lacking certain epigenetic modifiers, such as Dnmt3a, Dnmt3b (Chen et al., 2003b; Okano et al., 1999), and the H3K9 histone methyltransferases Suv39h1 and Suv39h2 (Lehnertz et al., 2003).
Bisulfite sequencing was also used to examine the methylation status of a CpG island sequence associated with the germ cell specific gene Dazl (Cooke et al., 1996). Dazl is upregulated in Mbd3 null ES cells (K.Kaji, personal communication). It has been reported that CpG islands associated with some germ cell specific genes including Dazl are methylated when the genes are not expressed, and become demethylated upon activation in the appropriate stage of germ-cell development (Maatouk et al., 2006). In agreement with this, upregulation of Dazl is associated with a decrease in DNA methylation at the CpG island (Fig.1.8.A) (Powell, Costello, Back and Hendrich, manuscript in preparation). However this demethylation is not a global effect; an intron in the Pramel 7 gene, which is misexpressed in Mbd3 null ES cells (Kaji et al., 2006) shows no significant change in the levels of methylation in the absence of Mbd3 (Fig.1.8.B) (Powell, Costello, Back and Hendrich, manuscript in preparation).
Figure 1.7 Demethylation of DNA in Mbd3-/- ES cells. Genomic DNA from parental (Ex1/Flox-/-, 3Flox-/-), null (-/-) or rescue (Mbd3a, b, c) cells was digested with either Msp1(M) or HpaII(H) (A, B), or HpaII alone (C) before being hybridised with probes for the major(A) or minor satellites (B), or IAP LTRs (C). Mito: mitochondrial DNA probe used as a loading and digestion control. A: 1,2) parental. 3,4) null. 5,6,7) rescue. B and C: 1-4) parental. 5-8) null. 9-14) rescue
Figure 1.8 Demethylation of single copy sequences in Mbd3 null ES cells. Bisulfite sequencing was used to examine the methylation status of A) a CpG island 5' to the germ cell specific Dazl gene and B) an intron in the Pramel 7 gene, which is misexpressed in the absence of Mbd3. B) shows that the demethylation in Mbd3 null ES cells is not a global effect. White bars: Mbd3(flox/-) ES cells, black bars: Mbd3(-/-) ES cells. At least ten modified sequences were looked at for each CpG site, and the percentage of C-T conversion was at least 95% in all cases.
1.8 Thesis Aims

The key features of embryonic stem cells are that they can self-renew and yet retain the ability to differentiate into any cell type of the embryo. The mechanisms behind how they achieve this remarkable plasticity are not well understood, but most likely involve concerted changes in transcriptional and epigenetic patterns. A role for epigenetic processes in the transition from a pluripotent cell to a differentiated cell has been established from the finding that ES cells lacking histone deacetylase and chromatin remodelling capabilities imparted by the NuRD complex are unable to terminally differentiate in the absence of LIF. This is associated with transcriptional misregulation and hyperacetylation of the promoters of mis-regulated genes. A significant decrease in the levels of DNA methylation is also observed at the centromeric satellite sequences, at IAP retrotransposons and at a CpG island of a gene mis-regulated in \( \text{Mbd3}^{(-/-)} \) ES cells. The aims of this thesis are to investigate what, if any, further epigenetic changes are associated with a lack of Mbd3, and to attempt to establish how these epigenetic aberrations are brought about. Chapter 3 deals with an analysis of epigenetic changes at various genomic locations in \( \text{Mbd3} \) null ES cells, and associated transcriptional changes. Chapters 4 and 5 focus on understanding how these changes are caused, and how the NuRD complex may co-operate with other epigenetic modifiers to effect epigenetic changes in pluripotent cells.
Chapter Two

2 Materials and Methods

2.1 Common solutions

PBS: 140 mM NaCl, 3 mM KCl, 2 mM KH2PO4, 10 mM Na2HPO4.
PBS-T: PBS with 0.1% Tween-20
Tris-EDTA (TE) buffer: 10mM Tris-HCl, 1mM EDTA, pH adjusted to 8 with NaOH
Tris-acetate-EDTA (TAE): 40 mM Tris, 20 mM glacial acetic acid, 1 mM EDTA and pH adjusted to 8.0 with NaOH.

2.2 Bacterial techniques

2.2.1 LB Media and agar

Luria Bertani (LB) media: 10 g/l Bacto tryptone, 5 g/l Yeast extract, 10 g/l NaCl, and adjusted to pH 7.0. L.B. was then autoclaved and stored at room temperature until required. Agar: L.B. Agar – Formulated as per L.B. except that 15 g/l of bacto-agar was included. L.B. agar was autoclaved and stored at room temperature until required.

2.2.2 Production of agar plates

Agar was heated in a microwave until completely molten, allowed to cool, and ampicillin added to 50µg/ml. The molten agar was poured onto 10cm plates (approximately 20ml/plate), and allowed to cool and solidify at room temperature. Plates were stored at 4°C and used within one month.
2.2.3 Production of competent cells

A single colony of freshly streaked XL10-Blue cells (Stratagene) was inoculated into 2ml of LB, and shaken at 37°C for 16-18 hrs. The bacterial culture was diluted 1/1000 into 200ml of LB and shaken at 37°C for 10-12 hrs, until the OD$_{600}$ was 0.4 – 0.6. The cultured bacteria were split into 4 x 50ml chilled Oakridge tubes, put on ice for 30 mins, then spun at 4000rpm for 10 mins at 4°C. Pellets were resuspended in 20mls of TFB1 (10mM CaCl$_2$, 15% glycerol, 0.3M CH$_3$COOK, 8mM MnCl$_2$, 10mM RbCl$_2$, acetic acid to pH5.8), placed on ice for 5 mins, and spun at 4000rpm for 10 mins at 4°C. Pellets were resuspended in 5ml TFB2 (75mM CaCl$_2$, 15% glycerol, 1mM RbCl$_2$, 1mM PIPES, pH to 6.5 with NaOH), placed onto ice for 20 mins, and split into pre-chilled eppendorfs on ice (200μl/tube). Aliquots were snap-frozen in liquid nitrogen and stored at -80°C.

2.2.4 Glycerol storage of bacterial stocks

Single colonies from freshly streaked out plates were picked and cultured at 37°C in 5ml ampicillin-containing media for 16-18hrs. 1.4ml of culture was added to 0.6ml sterile 50% glycerol in a cryotube, vortexed briefly, snap-frozen in a dry ice/ethanol bath and stored at -80°C.

2.2.5 Transformation of bacterial competent cells

1μl of chilled ligation products were added to 100μl of chemically competent XL10-Gold (Stratagene), which had been thawed on ice. After incubation on ice for 30 mins cells were heat-shocked at 42°C for 30 secs then returned to ice for 1 min. 500μl SOC medium (2% tryptone, 0.5% yeast extract, 8.6mM NaCl, 2.5mM KCl, 20mM MgSO$_4$, 20mM glucose, 0.2μM filtered and autoclaved) was added and the cells shaken at 300rpm at 37°C for 1hr. 100μl or 250μl of the reaction volume was plated out onto pre-warmed agar plates. Where the transformed plasmid contained the lacZ gene for blue/white screening IPTG and X-Gal were added to the plates at a final concentration of 0.05 mM and 80μg/ml respectively. Plates were incubated for 16-18hrs at 37°C.
2.2.6 Miniprep procedure

White/light blue colonies were picked into 5ml LB media containing 50μg/ml ampicillin, cultured overnight at 37°C, and the plasmid DNA isolated by standard miniprep procedures. Briefly, cultures were spun down (4000rpm, 5 mins), and pellets were resuspended in Solution 1 (50mM glucose, 10mM EDTA, 25mM Tris pH8, 100μg/ml RNase) then transferred to eppendorfs. 200μl of Solution 2 (0.2M NaOH, 1% SDS) was added, the samples mixed by inverting 20 times, then 150μl of 3M sodium acetate pH 5 added and the samples mixed by inverting 10 times. Following centrifugation at 14,000rpm for 10 mins the supernatants were added to 1ml 100% ethanol, mixed by inversion and spun at 14,000rpm for 5 mins. Resulting DNA pellets were washed with 70% ethanol, dried, and resuspended in 100μl TE buffer. Samples were screened by PCR and positives purified using Qiagen spin columns prior to sequencing. Alternatively, plasmid DNA was isolated from 5 ml bacterial cultures using Qiagen Spin Miniprep kits.

2.2.7 Midi/maxi-preparation of plasmid DNA

Single colonies from freshly streaked out plates were picked and cultured at 37°C in 5ml ampicillin-containing LB media for 8hrs, Cultures were then diluted 1/1000 into 50ml (midi-prep) or 250ml (maxi-prep) ampicillin containing media and grown overnight. Plasmid DNA was isolated using the relevant Qiagen kit, and resuspended in 100μl (midi-prep) or 500μl (maxi-prep) TE buffer.

2.3 DNA techniques

2.3.1 TA cloning

Cloning of PCR products was achieved using both Invitrogen TOPO 2.1 cloning vectors and Promega pGEMT-EASY vectors. For both, PCR products were separated on a 1.2%
agarose gel, visualised and excised on a UV transilluminator, and purified into 14μl dH2O using Zymogen Gel DNA purification kits.

2.3.2 Ligations
TOPO cloning: 1μl of purified PCR product was added to 1μl (50ng) of vector, 1μl salt solution and 3μl dH2O, and incubated at room temperature for 5 mins.
pGEMT-EASY cloning: 1μl of purified PCR product was added to 1μl (50ng) of vector, 5μl of 2x ligation buffer, and 1μl T4 DNA ligase. Reactions were incubated at 4°C overnight. Positive control reactions were performed for both using the control insert DNA provided with the vectors.

2.3.3 Sequencing
4μl of sample DNA (100-300ng/μl) and 2μl of primer (1.6pM/μl) were added to a PCR tube and sequenced by the ICAPB (University of Edinburgh) sequencing service. Results were analysed using the DNASTAR seqman program.

2.3.4 UV spectrophotometry
Nucleic acid and protein concentrations were measured using a NanoDrop® ND-1000 Spectrophotometer. Nucleic acid concentrations were calculated using the A_{260}/A_{280} reading.

2.4 Bisulfite Mutagenesis of genomic DNA
1 - 5μg of genomic DNA was digested with either EcoRI or HindIII. Samples were purified by phenol-chloroform extraction then precipitated with sodium acetate pH4.8 and 100% ethanol and storage at -80°C for 1hr. Precipitated DNA was washed twice with 70% ethanol, dried and resuspended in 100μl TE buffer. Denaturation was then achieved by the addition of NaOH to a final concentration of 0.3N. Sodium Bisulfite
was added to a final concentration of 3.6M to mutagenise the DNA, and 10mM Hydroquinone added as an alkylating agent. Samples were incubated at 55°C for 18hrs in the dark then purified using Qiagen PCR purification spin-columns. Samples were denatured using 0.3N NaOH at 37°C for 15 mins then purified by phenol-chloroform extraction and precipitated according to standard procedures (Sambrooke and Russell) using ammonium acetate and ethanol. Mutagenised DNA was resuspended in 20μl TE buffer, and 1μl used in PCR reactions.

2.4.1 Nested PCR amplification of bisulfite modified DNA

Target sequences were amplified using nested PCR. The first round of amplification was achieved using Red Hot Taq (RHT) (AbGene)(0.1μl), 0.4mM dNTPs,, 0.2mM each primer and 5% DMSO, in a total reaction volume of 50μl. 40 cycles were carried out, then 2μl of reaction product used in the second round of amplification. This again used Red Hot Taq(0.05μl), 0.4mM dNTPs, 0.2mM each primer and 5% DMSO, this time in a final reaction volume of 25μl. 35 cycles were carried out. PCR products were analysed on a 1.2% agarose gel. Bands were excised, purified and cloned into T/A vectors. Clones were mini-prepped and sequenced.

2.5 Reverse-Transcription PCR

2.5.1 RNA isolation

RNase-free eppendorf tubes and pipette tips were used at all stages. 1 x T25 of cells were washed with PBS then lysed with 1ml Tri Reagent (Sigma). The suspension was scraped into a 1.5ml Eppendorf tube, 200μl chloroform added and the mixture vortexed and allowed to sit at room temperature for 10 mins. Centrifugation at 14,000rpm for 15 mins at 4°C separated the mixture into two distinct layers. The upper, aqueous phase was removed into a new tube and 0.5ml isopropanol added. The mixture was allowed to sit at room temperature for 10 mins before being spun as before. The RNA pellet was washed with 75% ethanol in DEPC dH2O, dried at room temperature for 10 mins and
resuspended in 40\(\mu\)l DEPC dH\(_2\)O. RNA was stored either in DEPC 75\% ethanol or DEPC dH\(_2\)O at -80\(^\circ\)C.

2.5.2 DNaseI treatment

2\(\mu\)g of RNA was treated with 5 units DNaseI (Promega) in a total reaction volume of 10\(\mu\)l for 30 mins at 37\(^\circ\)C. The reaction was terminated by the addition of 1\(\mu\)l Stop solution and incubation at 65\(^\circ\)C for 10 mins.

2.5.3 Reverse Transcription (oligo dT)

1\(\mu\)g of DNaseI treated RNA was incubated with 25pm oligo dT primer and 1pm dNTPs in a total reaction volume of 12\(\mu\)l for 5 mins at 95\(^\circ\)C. Following a chill on ice and a quick spin, 4\(\mu\)l of 5x first strand buffer, 2\(\mu\)l of 0.1M DTT and 1\(\mu\)l RNase-OUT (all by Invitrogen) were added and the reaction incubated at 37\(^\circ\)C for 2 mins. 1\(\mu\)l M-MLV RT was added to reaction (but not to the –RT control) and incubated at 37\(^\circ\)C for 1 hr. The reaction was terminated by incubation at 70\(^\circ\)C for 15 mins.

2.5.4 Reverse Transcription (random primers)

As above but 50ng random primers were added instead of oligo-dT.

2.5.5 PCR amplification of cDNA

0.5\(\mu\)l of a 1/10 dilution of the cDNA reaction mixture was used per PCR reaction. Amplification was achieved using RHT (0.05\(\mu\)l), 0.2mM dNTPs and 0.2mM each primer, in a total reaction volume of 25\(\mu\)l.

For expression analysis, primers were designed to span exons near the 3'end of the transcript. Minus RT samples were always used as a control.
2.6 Chromatin immunoprecipitations

2.6.1 Bead blocking

60μl protein G beads (Amersham) were washed x 3 with 1xRIPA buffer (50mM Tris-HCl pH 8, 150mM NaCl, 1mM EDTA pH 8, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate), resuspended in 1ml RIPA/salmon sperm DNA (100μg/ml) and rotated at 4°C overnight.

2.6.2 Fixation, lysis, sonication and pre-clearing

The following day, cells were trypsinised and resuspended to a concentration of 2 x 10⁶/ml. 1ml cells were fixed by rotating with 100μl 11 x fixation buffer (11.1% formaldehyde, 50mM HEPES pH 8, 100mM NaCl, 1mM EDTA pH 8, 0.5mM EGTA) for 10 mins at room temp. 100μl 1.5M glycine was added to neutralise the fixation and the cells spun down (3000rpm, 5 min, 4°C), washed (with 1ml PBS containing 1mM PMSF) and lysed with 400μl lysis buffer (50mM Tris-HCl pH 8, 10mM EDTA pH 8, 1% SDS, 1mM PMSF, 1μg/ml aprotinin, 1μg/ml leupeptin) on ice for 15 mins. Each sample was placed into a 0.5ml eppendorf and sonicated 5 times for 5 seconds each time. Samples were kept on ice during the sonication process. Following centrifugation (15,000rpm, 10 min, 4°C), the chromatin was quantified and for each sample 50μg of chromatin was diluted into dilution buffer (50mM Tris-HCl pH8, 167mM NaCl, 1.1% Triton X-100, 0.11% sodium deoxycholate, 1mM PMSF, 1μg/ml aprotinin, 1μg/ml leupeptin) to make 1ml total. Samples were rotated at 4°C for 3-4hrs. Input samples consisted of 5μg of chromatin diluted into 200μl dilution buffer, then processed and treated the same way as the samples.

2.6.3 Antibody incubation and binding to beads

Samples were spun at 14,000rpm for 10 mins at 4°C. Supernatants were transferred to new tubes, 2μg antibody was added and the samples rotated at 4°C overnight. No
2.8 Protein Techniques

2.8.1 Extraction of nuclear proteins from ES cells

1 x T75 of confluent cells were trypsinised as normal and resuspended in 10 ml ES cell media. The cell pellet was washed with 1ml PBS, then resuspended in 500μl hypotonic buffer (10mM HEPES pH7.9, 1.5mM MgCl₂, 10mM KCl, 1mM PMSF, 0.5mM DTT), transferred to an eppendorf and placed on ice for 15 mins. 31.25μl of 10% Nonidet P 40 (NP-40) was added, the sample vortexed for 30 sec, and spun for 1 min, at 10,000rpm and 4°C. The supernatant was discarded, the pellet resuspended in 60μl high salt buffer (10mM HEPES pH7.9, 12.5% glycerol, 0.75mM MgCl₂, 10mM KCl, 0.1mM EDTA, 400mM NaCl, 1mM PMSF, 0.5mM DTT), and the sample shaken on a plate shaker (300rpm) at 4°C for 1hr. Centrifugation at 14,000rpm for 10 mins, at 4°C was followed by removal of the supernatant which was either used or stored at -80°C.

2.8.2 Quantification using the Bradford Assay method

5μl of dilutions of BSA (1, 2, 3, 4, 5 mg/ml) were added to individual wells of a 96-well plate. 5μl of 2x and 5x dilutions of the protein samples were also set up in wells of the plate, All samples were diluted in the high salt buffer used for the last stage of the protein extraction. 295μl Bradford reagent was added to each well and the samples incubated at room temperature for 10 mins. A₅₉₅ readings were then taken using the UV spectrophotometer, and the concentration of the samples calculated from the standard curve.

2.8.3 Separation of proteins on acrylamide gels and detection by antibodies (SDS – PAGE and Western Blotting)

2.8.3.1 Gel preparation

7.5, 10 or 15% resolving gels were cast, depending on the size of protein to be detected, using BioRad gel casting apparatus (for a 10% gel: 0.375M Tris pH 8.8, 0.2% SDS, 10%
acrylamide, 0.1% APS, 0.05% TEMED). A 5% stacking gel (0.126M Tris pH 6.8, 0.2% SDS, 5% acrylamide, 0.1% APS, 0.1% TEMED) was cast on top. Gels were run in 1x running buffer (0.025M Tris, 0.19M glycine, 0.1% SDS) at 90V for 2hrs, or until the loading buffer had reached the bottom of the gel.

2.8.3.2 Sample preparation

2 x loading buffer (100mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.2% bromophenol blue) and DTT to a final concentration of 0.1M were added to protein samples, and the mixture incubated at 95°C for 5 mins. 10µl of prestained protein marker (NEB) was also incubated in this way. Samples were chilled on ice, spun briefly, then loaded onto the gel using fine gel-loading tips (Alpha Laboratories).

2.8.3.3 Wet transfer onto nitrocellulose membranes

Proteins were transferred onto nitrocellulose membranes (ProTran from Scheider-Schuell) for antibody detection using the BioRad transfer equipment. For every gel, 2 fibre pads, 6 pieces of Whatman filter paper and 1 piece of membrane (filter paper and membrane cut to the size of the gel) were pre-wetted in transfer buffer (0.025M Tris, 0.19M glycine, 0.1% SDS, 20% methanol (added just before use)) for 15 mins, along with the gel. A 'sandwich' of a fibre pad, 3 pieces of filter paper, the gel, the membrane, 3 pieces of filter paper and the remaining fibre pad was set up. Air bubbles were removed by gently rolling a pipette over the surface. Transfer was carried out at 90V for 1hr, or 30V overnight, in 1l of transfer buffer at 4°C. Efficiency of transfer was assessed visually by how well the prestained protein marker had transferred onto the membrane.

2.8.3.4 Blocking, antibody incubations and washing

Membranes were washed in PBS-T for 10 mins before being incubated in blocking solution for 30 mins at room temperature. Membranes were then incubated in blocking solution containing diluted primary antibody, either overnight at 4°C or for 1hr at room
antibody was added to the No Antibody controls. The next day, the samples were added to 10µl pre-washed and blocked protein G beads and rotated at 4°C for 1 – 1.5 hrs.

2.6.4 Elution from beads and reversal of crosslinks

Samples were spun at 14,000rpm for 1 min at 4°C, and the supernatant discarded. Beads were washed as follows:

i) 1 x RIPA buffer x 1  

ii) 1 x RIPA buffer/500mM NaCl x 1  

iii) LiCl wash solution (10mM Tris-HCl pH 8, 0.25M lithium chloride, 1mM EDTA pH 8, 0.5% NP-40, 0.5% sodium deoxycholate) x 1  

iv) TE buffer x 2

200µl elution buffer (10mM Tris-HCl pH8, 300mM NaCl, 5mM EDTA pH 8, 0.5% SDS) was added to beads, and incubated overnight at 65°C.

2.6.5 Purification and precipitation of chromatin fragments

1µl RNaseA (4mg/ml) was added to samples, followed by incubation at 37°C for 30 mins. 1µl Proteinase K (10mg/ml) was added to samples, followed by incubation at 55°C for 2h. 1µl tRNA (10mg/ml) was added, samples were vortexed and spun briefly, then the supernatant purified by phenol-chloroform extraction, including a back-extraction with TE buffer. DNA was precipitated by the addition of 2.5 x the sample volume of 100% ethanol and storage at -80°C for 2 hrs. Samples were resuspended in 20µl TE buffer and 1µl used for PCR.

2.6.6 Input samples

200µl of diluted, sonicated lysate from 2.10.4 was purified and precipitated as in 2.10.6. Samples were resuspended in 50µl TE and dilutions of 1, 2, 4, 8 and 16 used to produce a standard curve for real time PCR.
2.6.7 Chromatin immunoprecipitation of biotin-tagged proteins

The same protocol as for normal ChIP was used, as described above, but with the following modifications. 50μl streptavidin-coated magnetic beads (DynaBeads, Invitrogen) were blocked by 3 x 20 minute washes at room temperature in 1ml of blocking buffer (TBS + 200ng/μl purified chicken albumin, PCA). 50μg of sonicated chromatin in 1ml of dilution buffer was added to the blocked beads and incubated overnight at 4°C. The beads were collected and washed with: 1 x LiCl wash buffer, 2 x 2% SDS. Precipitated material was eluted by heating overnight at 65°C, and precipitated and analysed as described above.

2.7 Real-time PCR

Reactions were performed in triplicate on the Roche LC-480 machine. For the chromatin immunoprecipitation analysis, samples were normalised against the β-actin promoter; for the expression analyses, samples were normalised against levels of Gapdh.

2.7.1 Standard curves

For expression analyses, standard curves were produced by using serial dilutions of the relevant target cloned into a plasmid vector. The number of plasmid molecules in each standard sample was calculated so the actual transcript copy number in the samples could be accurately calculated. For the Chromatin Immunoprecipitation experiments, standard curves were produced using dilutions of the precipitated input DNA.

For all real-time experiments, a total reaction volume of 10μl was used. This incorporated 5μl of 2 x Sybr Green Master Mix (from Roche), primers at a final concentration of 0.5mM, and 3μl sample DNA.
temperature. The membranes were then washed for 2 x 10 mins in blocking solution before being incubated with the secondary antibody for 1hr at room temperature and subsequently washed with PBS-T for 3 x 15 mins.

2.8.3.5 Signal detection
Enhanced chemiluminescent reagent (Amersham) was added to the membranes, at least 0.125ml/cm² of membrane, and the membranes exposed to autoradiograph film for 1 min to 1hr depending on the strength of the signal. Films were developed using an autoradiograph developer.

2.8.3.6 Membrane stripping and re-probing
Where appropriate, membranes were stripped of the primary antibody by being washed for 10 mins in PBS-T then incubated in stripping buffer (0.5M NaCl, 0.5M acetic acid) for 30 mins at 55°C, followed by washing in PBS-T for 10 mins. Stripping efficiency was assessed by incubation with the secondary antibody and detection as above. Membranes were then washed with PBS-T and probed as above.

2.8.4 Immunoprecipitations

2.8.4.1 Preclearing
200µg undiluted nuclear extract was pre-cleared with 5µl 10% sheep serum (diluted in IP wash buffer (20mM HEPES, 0.1M NaCl, 10% glycerol, 0.2mM EDTA, 0.01% Triton-X, 1mM PMSF)) for 1hr on ice. 50µl protein A beads (Amersham) (washed x3 with IP wash buffer) were added and the mixture rotated at 4°C for 1hr, before being spun at 14,000rpm for 15 min at 4°C. The supernatant was collected.
2.8.4.2 Immunoprecipitation

The precleared nuclear extract was diluted to final volume of 500μl with IP wash buffer and 1.5μl Dnmt3b antibody added. The reaction was incubated on ice for 1hr, then 50μl protein A (washed x 3 with IP wash buffer) added and the mixture rotated overnight at 4°C.

2.8.4.3 Washing, sample preparation and gel loading

The samples were spun for 1 min at 14,000rpm at 4°C and the supernatant discarded. The beads were washed x 3 with IP wash buffer. 25μl 1 x protein sample loading buffer and DTT to a final concentration of 0.1M were added and the reaction incubated at 95°C for 10 mins. Following a quick chill on ice the samples were spun for 1 min at 14,000rpm and the supernatant loaded onto an acrylamide gel, to be analysed by Western Blotting.

2.8.4.4 Immunoprecipitation of biotin-tagged proteins

Nuclear extract was incubated with 20μl of streptavidin-coated magnetic beads (DynaBeads, Invitrogen), which had been pre-incubated for 1 hr in TBS + 200ng/μl, in a total volume of 200μl IP buffer (TBS + 0.3% NP-40 + Protease Inhibitor Cocktail) for 1 – 3 hours at 4°C. Beads were collected and washed 6 times for 5 minutes at room temperature. They were resuspended in 20μl loading buffer with 0.1M DTT, boiled at 95°C for 5 mins then run out on an acrylamide gel. Nuclear extract from the BirA expressing line was always included as a control.
2.8.5 Immunofluorescence

2.8.5.1 Fixing and permeabilisation
ES cells were plated onto 13mm diameter cover slips in 4 well dishes and grown to confluency. Following washing with PBS cells were fixed with 4%PFA in PBS for 10 mins then permeabilised with 0.1%-Triton-X in PBS for 2 mins.

2.8.5.2 Blocking, antibody incubation and washing
Fixed and permeabilised cells were washed with PBS then blocked with PBS/1% BSA/3% serum(from the species the secondary antibody was raised in) for either 1hr at room temperature or 4°C overnight. The cells were washed with PBS then incubated with the primary antibody, diluted in blocking solution, overnight at 4°C. The cells were washed for 3 x 10 mins with PBS/0.2 % Tween-20, then incubated with the secondary antibody (diluted in blocking solution) for 1hr at room temperature. Stained cells were washed 3 x 5 mins with PBS, then rinsed in dH2O.

2.8.5.3 Mounting and visualisation
Cover slips were mounted inverted onto 5μl of VectaShield mounting medium (Vector Laboratories Inc.) (containing DAPI stain) on microscope slides. Nail polish was used to fix the cover slips in place. Fluorescing cells were visualised using a Leica DM IRE2 inverted confocal microscope.

2.9 In vitro measurement of de novo DNA methyltransferase activity
A kit from Epigentek was used to measure the levels of de novo Dnmt activity in nuclear extracts. Unmethylated DNA attached to a solid substrate in wells was washed with the wash buffer, then 3-6μg nuclear extract from the cell types being examined was incubated for 1 hour with the substrate (S-adenosyl-L-methionine) and assay buffer in the pre-washed wells. Purified recombinant Dnmt enzyme was used as a positive
control. Wells were washed with wash buffer then incubated with an antibody against methylated DNA for 1 hr at room temperature. The detection antibody was added and the reactions incubated at room temperature for 30 mins. Wells were washed x 4 with the wash buffer before being incubated with the developing solution for 10 mins away from light. The reaction was stopped by the addition of stop solution, and the extent of DNA methylation in the samples assessed by absorbance at 450nm. Dnmt activity was calculated from the reading and the amount of nuclear extract used.

2.10 Tissue Culture

All spins are for 3 mins at 1300rpm, unless otherwise indicated. All reagents were stored at 4°C for no longer than one month, and warmed to 37°C before use. Plastics used for the culture of ES cells were provided by Iwaki, and were gelatinised by the addition of sufficient 0.1% gelatin in PBS to coat the base for at least 20 mins.

ES cell media: 500mls 1 x GMEM, 55mls Foetal Calf Serum, 5.5mls non-essential Amino Acids, 11mls Glutamine and Sodium Pyruvate, 555µl β-Mercaptoethanol, 600µl LIF.

2.10.1 Thawing of ES cells

Cells stored at -80°C or below were thawed quickly by placing them in a 37°C waterbath, then transferred to 10ml of pre-warmed media and spun down. They were resuspended into 10ml of warm media and transferred to a gelatinised T25. The media was changed 6-8hrs later.

2.10.2 Passaging of ES cells

Once almost confluent, cells were washed twice with PBS, then enough trypsin added to cover the base of the flask. Flasks were incubated at 37°C for 1 min, or until the cells had dissociated from the flask and each other. Media, 10 x the volume of the trypsin, was
added and the cell suspension spun down. Cells were resuspended in 10ml media and transferred to gelatinised flasks. Cells were split 1 : 3.

2.10.3 Freezing of ES cells

Confluent ES cells were washed and trypsinised as normal, and resuspended in freezing media (media/10% DMSO), to 10x the trypsin volume. The cell suspension was spun down and resuspended in an appropriate volume of freezing medium, and 1ml aliquots transferred to labelled cryotubes. Cells from a confluent T25 were split into 2 cryotubes, and from a T75 into 6 cryotubes. These were transferred into a -80°C freezer overnight then placed in the liquid nitrogen cell bank the following day.

2.10.4 Transient transfection of ES cells using Transfast

2.10.4.1 Preparation of Transfast reagent

Transfast, previously resuspended in dH₂O and stored at -20°C, was thawed and vortexed briefly. For each transfection, 6µl was added to 250µl prewarmed media in a 1.5ml Eppendorf.

2.10.4.2 Preparation of DNA to be transfected

DNA was ethanol precipitated and resuspended in TE buffer under sterile conditions. 1µg was added to the media/transfast mixture (from 2.12.4.1), and the mixture left for 15mins at room temperature.

2.10.4.3 Preparation of cells and transfection procedure

ES cells were trypsinised, counted using a haematocytometer, and resuspended in media to a density of 2 x 10⁶ cells/ml. 250µl i.e. 5 x 10⁵ cells were added to the media/DNA/Transfast. The mixture was transferred to 1 gelatinised well of a 24 well plate and incubated at 37°C for 1hr, at which stage 1ml of media was added to each
transfection. The media was changed 24hrs later. Cells were harvested 48hrs after
transfection.

2.10.5 Stable transfection of ES cells using Transfast
Cells were transfected as above, but after 48hrs they were trypsinised, resuspended in 7
ml of media and either 1ml or 6ml transferred to gelatinised 10cm plates, containing a
total of 10 mls of media. Drug selection was applied at this point and maintained
subsequently. Media was changed every few days for c.10 days until colonies were
visible by eye. These were picked into 96 well plates and expanded until a sufficient
number of transfected cells for each clone were obtained.

2.10.6 Stable transfection of ES cells by electroporation

2.10.6.1 Preparation of DNA
100μg of DNA was linearised in a total volume of 500μl then precipitated using sodium
acetate and ethanol. Precipitated DNA was resuspended in 80μl TE buffer in sterile
conditions.

2.10.6.2 Preparation of cells and electroporation
For each transfection, 1 x T75 of confluent cells were washed, trypsinised and
resuspended in 800μl PBS. The resuspended DNA was added to the cells, which were
transferred to an electroporation cuvette. The cells/DNA were pulsed with (?) kV for 1
sec (?), then allowed to ‘rest’ for 10 mins at room temperature. Then they were
resuspended in 10 ml of media, and either 1 ml or 9 ml aliquotted onto a gelatinised
150mM plate, in a total of 20ml media. The media was changed after 24 hrs, and
selection added after 48 hrs. Media (containing selection) was changed every few days
for c.10 days until colonies were visible by eye, at which point they were picked into 96
well plates and subsequently expanded.
2.10.7 Picking of Colonies

An appropriate number of wells of a flat-bottomed 96 well plate were gelatinised. 10μl PBS was added to the same number of wells of a round-bottomed 96 well plate. Plates with colonies on them were washed with PBS twice, and on the second wash enough PBS was left to just cover the colonies. Colonies were picked up using a 20μl pipette set to 10μl, and transferred to a round bottomed well containing PBS. The wells were checked for the presence of cells before 20μl of trypsin was added to each well and the plate incubated for 2-3 mins. The plate was then tapped gently before being returned to the incubator for 2 mins. 100μl of selection media was added to neutralise the tryspin, and the contents of each well transferred to the corresponding well of the gelatinised flat-bottomed 96 well plate. Media was added to 200μl. Clones were incubated until the media turned yellow (usually 2-3 days) then washed, trypsinised and transferred into a 48 well plate. Clones were expanded in this way until a sufficient number of cells were obtained.

2.11 Antibodies

2.11.1 Primary

Anti-Dnmt3b (rabbit polyclonal) was a gift from En Li, then subsequently purchased from Abcam as a mouse monoclonal, cat.no. Ab13604

Anti-Lsh2 (rabbit polyclonal) was a gift from Kathrin Muegge.

Anti-H3-K9-trimethyl (rabbit polyclonal) was a gift from Thomas Jenuwein.

Anti-Mta2: goat polyclonal from Santa Cruz Biotechnology (SCB), cat. no. sc-9447.

Anti-Oct-4: goat polyclonal from SCB, cat. no. sc-8628.

Anti-Hdac1: rabbit polyclonal from Upstate, cat.no.: 06-720.

Anti-Mbd3: Goat polyclonal from SCB, cat.no. sc-9402

Anti-HP1α: Mouse monoclonal from Upstate, cat.no. 05-698

Anti-H3: Rabbit polyclonal from Abcam, cat. no. Ab1791.

Anti-H4: Rabbit polyclonal from Abcam, cat.no. Ab10156

Anti-Acetyl-H4: Rabbit polyclonal from Upstate, cat.no. 06-598
Anti-Acetyl-H3-K9: Rabbit polyclonal from Upstate, cat.no. 06-942
Anti-Acetyl-H3-K14: Rabbit polyclonal from Upstate, cat. no. 07-353
Anti-Gata4: Goat polyclonal from SCB, cat. no. sc-1237

2.11.2 Secondary

Anti-rabbit IgG, HRP linked, from Amersham BioSciences. Cat.no. NA934V
Anti-goat/sheep IgG HRP linked, from Sigma Aldrich, clone GT 34.
Anti-rabbit IgG, Alexa-488 linked, from Molecular Probes

2.11.3 Primers

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Sequence</th>
<th>$T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major F</td>
<td>Major satellites</td>
<td>GACGACTTGAATGACGAAATC</td>
<td>64.5</td>
</tr>
<tr>
<td>Major R</td>
<td>&quot;</td>
<td>CATATTCCAGGTCTTCAATGTCATG</td>
<td>63.5</td>
</tr>
<tr>
<td>Minor F</td>
<td>Minor Satellites</td>
<td>CATGGAATGATGATGAAATC</td>
<td>61.3</td>
</tr>
<tr>
<td>Minor R</td>
<td>&quot;</td>
<td>CATCTAATATGTCTAGAGTGTGG</td>
<td>62.4</td>
</tr>
<tr>
<td>LIORF.F</td>
<td>LINEs</td>
<td>TTTGGGACACAATGAAAGCA</td>
<td>64</td>
</tr>
<tr>
<td>LIORF.R</td>
<td>&quot;</td>
<td>CGTCCCGTCTACTCCTTCTCTTCTG</td>
<td>63.8</td>
</tr>
<tr>
<td>Gapdh F</td>
<td>Gapdh</td>
<td>CCGACTAATGCTCAGAAATGGG</td>
<td>63.8</td>
</tr>
<tr>
<td>Gapdh R</td>
<td>&quot;</td>
<td>CTTTCAATGCAATGCCAAGTT</td>
<td>63.7</td>
</tr>
<tr>
<td>Oct4q.F</td>
<td>Oct4</td>
<td>GGCGTTTCTTGTGAAGGATGTTGTT</td>
<td>69.5</td>
</tr>
<tr>
<td>Oct4q.R</td>
<td>&quot;</td>
<td>CTCGAACCACATCTCTTCTCTTCTTCT</td>
<td>60.8</td>
</tr>
<tr>
<td>IAPgag.F</td>
<td>IAPs</td>
<td>CAAGCTGCCTCTCTCATAGA</td>
<td>65.6</td>
</tr>
<tr>
<td>IAPgag.R</td>
<td>&quot;</td>
<td>CGTCCCGTCTACTCCTTCTCCT</td>
<td>67.5</td>
</tr>
<tr>
<td>Dlk1.F</td>
<td>Dlk1</td>
<td>CACATCCTGAGGTGCTCCAT</td>
<td>62.5</td>
</tr>
<tr>
<td>Dlk1.R</td>
<td>&quot;</td>
<td>CGCAGCTTGTGAGGAAACAg</td>
<td>63.5</td>
</tr>
<tr>
<td>Gt12.F</td>
<td>Gt12</td>
<td>AGCGTCCATCTCAGCAAGAA</td>
<td>67.2</td>
</tr>
<tr>
<td>Gt12.R</td>
<td>&quot;</td>
<td>CGAGGATCCAGATGATGAGG</td>
<td>66.8</td>
</tr>
<tr>
<td>SnrpnFQ</td>
<td>Snrpn</td>
<td>ATTAGCGGCTCGTACAG</td>
<td>62.4</td>
</tr>
<tr>
<td>SnrpnRQ</td>
<td>&quot;</td>
<td>GAGGCGATGCTCGTACAG</td>
<td>61.1</td>
</tr>
<tr>
<td>H19FQ</td>
<td>H19</td>
<td>CACTACATACCTGCTCATCAG</td>
<td>57.4</td>
</tr>
<tr>
<td>H19RQ</td>
<td>&quot;</td>
<td>TGGGTTGGGTGTACATGAGTCT</td>
<td>64.5</td>
</tr>
<tr>
<td>B2F</td>
<td>Sohlh2</td>
<td>CCCCTTCTTCCATCATCATC</td>
<td>66.5</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
<td>Tm (°C)</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>B2R</td>
<td>GCCATCTTTTCTCCCTTGTT</td>
<td>67.2</td>
<td></td>
</tr>
<tr>
<td>DazlF</td>
<td>GTCCTTACATGTACCATTCTGTGAC</td>
<td>62.3</td>
<td></td>
</tr>
<tr>
<td>DazlR</td>
<td>GACTCCAAACAAACAGCAGACAA</td>
<td>65.2</td>
<td></td>
</tr>
<tr>
<td>Dnmt1.F</td>
<td>TACAGCAGACCCGTTGCCCT</td>
<td>72.3</td>
<td></td>
</tr>
<tr>
<td>Dnmt1.R</td>
<td>GGTGCTGAAAGACGCCATCCC</td>
<td>68.1</td>
<td></td>
</tr>
<tr>
<td>Dnmt3a.F</td>
<td>AAGAGGGACATCTCGCGATT</td>
<td>64.9</td>
<td></td>
</tr>
<tr>
<td>Dnmt3a.R</td>
<td>CTATTCTGGCCGTTGCCAGA</td>
<td>65.3</td>
<td></td>
</tr>
<tr>
<td>Dnmt3l.F</td>
<td>TTAGATGACAATGTGCTGCTG</td>
<td>63.4</td>
<td></td>
</tr>
<tr>
<td>Dnmt3l.R</td>
<td>GGTGCTGAAGAAGCCATCCC</td>
<td>68.1</td>
<td></td>
</tr>
<tr>
<td>Dnmt3b(1)</td>
<td>GTTAATAAGTGCAAGTGCG</td>
<td>58.8</td>
<td></td>
</tr>
<tr>
<td>Dnmt3b(all isoforms)</td>
<td>AACACGGGTCCCTCACAGAATT</td>
<td>62.5</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.1 RT-PCR Primer sequences**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5.F.normal</td>
<td>GT12/Dlk1 IG-DMR (M5)</td>
<td>TTTTCTATTGAACTGCAAGTG</td>
</tr>
<tr>
<td>M5.R.normal</td>
<td>AACCAGAGTGATCCCTCTTGC</td>
<td>60.7</td>
</tr>
<tr>
<td>Actin E</td>
<td>CGCCAAAAACTCTCTTCTCC</td>
<td>69.6</td>
</tr>
<tr>
<td>Actin F</td>
<td>GGGTTTTATAGGACCCACAGC</td>
<td>67.3</td>
</tr>
<tr>
<td>Oct4 A</td>
<td>ATCCGACAACTGTGTTTG</td>
<td>65.0</td>
</tr>
<tr>
<td>Oct4 B</td>
<td>CAATCCCCACCTCTAGCTT</td>
<td>64.1</td>
</tr>
<tr>
<td>IAP-LTR.F</td>
<td>IAP LTR region</td>
<td>GACAGCTTGTTCTAAGTG</td>
</tr>
<tr>
<td>IAP-LTR.R</td>
<td>CACAGACCACTATGCTTGC</td>
<td>60.9</td>
</tr>
<tr>
<td>H19.5.DMR</td>
<td>CAAGGAGACCATGCCCTAT</td>
<td>61.5</td>
</tr>
<tr>
<td>H19.5.DMR</td>
<td>CTGAGGTACTGAGTGGG</td>
<td>58.2</td>
</tr>
<tr>
<td>H19.5.DMR (region 2)</td>
<td>GAAACATTGTGGTTCTGGAG</td>
<td>57.2</td>
</tr>
<tr>
<td>Sohlh2p3</td>
<td>AAGATCTCTGCTCCTGCTCTCCC</td>
<td>72.3</td>
</tr>
<tr>
<td>Sohlh2p4</td>
<td>TAGAAACCTGTGTCGATGGC</td>
<td>69.2</td>
</tr>
<tr>
<td>Htrap11</td>
<td>Htra promoter</td>
<td></td>
</tr>
<tr>
<td>Htrap12</td>
<td>rDNA intergenic region</td>
<td>CCTGGAATTCCTCCTGAACTC</td>
</tr>
<tr>
<td>rDNA.F</td>
<td>rDNA intergenic region</td>
<td>CCTAAACTGTGACAGGTTG</td>
</tr>
</tbody>
</table>
### Table 2.2 Chromatin Immunoprecipitation primers

#### 2.1.1 Primers for Bisulfite sequencing

Primers were designed using the Meth Primer program

(https://www.urogene.org/methprimer/index.html)

<table>
<thead>
<tr>
<th>Name</th>
<th>Locus</th>
<th>Sequence</th>
<th>$T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M5.F</td>
<td>Gt12/Dlk1 IG-DMR</td>
<td>GGGATTTTTGATTTTTTTGAGGT</td>
<td>62.1</td>
</tr>
<tr>
<td>M5.R</td>
<td>&quot;</td>
<td>CCCAACCTACTATAATCAAAAATTTC</td>
<td>60.4</td>
</tr>
<tr>
<td>M5.R.BN</td>
<td>&quot;</td>
<td>GGTGGGATTTTATGGTTTTATT</td>
<td>62.7</td>
</tr>
<tr>
<td>M5.R.BN</td>
<td>&quot;</td>
<td>CTCCACCCCAATATAACTTTCTAC</td>
<td>61.6</td>
</tr>
<tr>
<td>M4.F</td>
<td>Gt12/Dlk1 IG-DMR</td>
<td>TACCATAACATATCCCATACCAA</td>
<td>60.6</td>
</tr>
<tr>
<td>M4.R</td>
<td>&quot;</td>
<td>ATGGGATTGTGTATGTAAT</td>
<td>56.9</td>
</tr>
<tr>
<td>ICR3.F</td>
<td>H19 5' DMR (region 2)</td>
<td>GGGTTTTTTTGGTTATTGAATT</td>
<td>62.6</td>
</tr>
<tr>
<td>ICR3.R</td>
<td>&quot;</td>
<td>CACCCCTATAAAATCCCTATTTAATAAC</td>
<td>61.6</td>
</tr>
<tr>
<td>ICR3.F.BN</td>
<td>&quot;</td>
<td>CCTTTTTTTTTCCCTATTTAATTTAATAAAT</td>
<td>62.2</td>
</tr>
<tr>
<td>ICR3.R.BN</td>
<td>&quot;</td>
<td>CCCAATTAAACACATTTTAAACACC</td>
<td>62.8</td>
</tr>
<tr>
<td>Snrpn.F</td>
<td>Snrpn DMR</td>
<td>TAGAGGGATAGAGATTTTGTATT</td>
<td>58.3</td>
</tr>
<tr>
<td>Snrpn.R</td>
<td>&quot;</td>
<td>AATAGGATATTTTGTGTAG</td>
<td>61.9</td>
</tr>
<tr>
<td>Snrpn.F.BN</td>
<td>&quot;</td>
<td>TAAATAGGATATTTTGTGTAG</td>
<td>53.3</td>
</tr>
<tr>
<td>Snrpn.R.BN</td>
<td>&quot;</td>
<td>TAAATAGGATATTTTGTGTAG</td>
<td>43.4</td>
</tr>
<tr>
<td>BSo2F</td>
<td>Sohlh2 promoter</td>
<td>GTTTTTTGGAGAAAAAGTTAAAATTGTG</td>
<td>60.6</td>
</tr>
<tr>
<td>BSo2R</td>
<td>&quot;</td>
<td>ACCTAATCCCCAACATCCAATCC</td>
<td>60.9</td>
</tr>
<tr>
<td>BDazP1</td>
<td>Dazl promoter</td>
<td>TTTTTTTAATGTTGGTTGAAAAGAG</td>
<td>61.0</td>
</tr>
<tr>
<td>BDazP2</td>
<td>&quot;</td>
<td>CCTATAAAAATAAATCTATAAAACAAAC</td>
<td>59.2</td>
</tr>
<tr>
<td>Pramel7P12-Bis</td>
<td>Pramel 7 intron</td>
<td>TCTCTAATCTCAATAACACC</td>
<td>53.1</td>
</tr>
<tr>
<td>Pramel7P13-Bis</td>
<td>&quot;</td>
<td>AGGTTTTATTGGAGTTGAGAG</td>
<td>52.3</td>
</tr>
</tbody>
</table>

### Table 2.3 Bisulfite Sequencing Primers
Chapter Three

3 Perturbed DNA methylation and histone modification patterns in $Mbd3^{(-/-)}$ ES cells

3.1 Introduction

Heterochromatin is characterised by histone hypoacetylation, inaccessibility to nucleases, methylation of H3K9 and H3K27, binding of HP1, DNA methylation and late replication during S phase (Grewal and Moazed, 2003; Lima de Faria and Jaworska, 1968; Schotta et al., 2004; Solage and Cedar, 1978; Wakimoto, 1998; Wallace and Orr-Weaver, 2005). Constitutive heterochromatin also tends to contain regions of highly repetitive DNA such as the major and minor satellites and transposable elements (Lohe et al., 1993; Pimpinelli et al., 1995). The transcriptionally repressive nature of the heterochromatic environment is demonstrated by the observations that regions of heterochromatin tend to be gene-poor (Richards and Elgin, 2002), and by the phenomenon of Position-Effect-Variegation (PEV). This is when translocation of an active gene that normally resides in euchromatin into heterochromatin results in unstable repression of that gene (Reuter and Spierer, 1992).

DNA methylation is a common modification of eukaryotic genomes, occurring at between 60 - 90% of CpG dinucleotides in a typical vertebrate. Patterns of DNA methylation are highly compartmentalised, with some regions such as CpG islands remaining completely undermethylated and others being fully methylated (Bird, 2002; Hendrich and Tweedie, 2003). The correct specification of DNA methylation patterns is essential for development; null mutations of the genes responsible for DNA methyltransferase activity produce an embryonic lethal phenotype (Li et al., 1992; Okano et al., 1999), and affect developmental processes such as imprinting and X-
inactivation (Bestor, 2000; Robertson and Wolffe, 2000). It is also responsible for the silencing of transposable elements (Walsh et al., 1998).

DNA methylation is considered to be transcriptionally repressive; the silencing of newly integrated transgenes is associated with methylation of their DNA (Assaad et al., 1993; Kilby et al., 1992), and constitutive heterochromatin, which is generally transcriptionally silent, tends to be methylated (Jones and Takai, 2001; Martienssen and Colot, 2001). In germ cells, demethylation of CpG islands close to the 5’ends of certain germ-cell specific genes is correlated with expression of those genes in germ cells at the appropriate developmental stage; these regions are methylated when the gene is not expressed (Maatouk et al., 2006). However it is not clear whether DNA methylation is a primary event in transcriptional silencing and heterochromatin formation, or whether it occurs as a secondary event to ‘lock in’ a repressive chromatin state.

The transcriptional regulation of imprinted genes is also frequently controlled by DNA methylation (Ferguson-Smith and Surani, 2001; Li et al., 1993). The mechanisms of how this is achieved have not been elucidated for the majority of genes, although mono-allelic methylation of ‘imprinting control regions’ (ICRs), cis-acting sequences that may be located many kilobases from the genes they regulate, is frequently observed (Reik and Walter, 2001). However the regulation of 2 genes on mouse chromosome 7, *H19* and *Igf2*, has been shown to include, among other things, paternal methylation of a region upstream of the *H19* gene (Tremblay et al., 1995). This prevents binding of the insulator protein CTCF to the paternal chromosome, and allows the *Igf2* gene preferential long range access to an enhancer element located downstream of *H19*. The result is expression of *Igf2* and repression of *H19* on the paternal chromosome, with the reverse situation on the maternal chromosome (Ferguson-Smith, 2000; Reik and Walter, 2001; Szabo et al., 2000) (Fig.3.1.13).

*Gtl2* and *Dlk1* are a further two differentially methylated, linked and reciprocally expressed genes on mouse chromosome 12 (Schmidt et al., 2000; Takada et al., 2000), which interestingly share many common regulatory features with the *H19/Igf2* locus.
Three differentially methylated regions (DMRs) have been identified at or nearby this locus; an intergenic DMR (IG-DMR), a DMR associated with the CpG island of \textit{Dlk1}, and an intronic DMR within the \textit{Gtl2} gene (Takada et al., 2002) (Fig.3.1). The 9kb IG-DMR is associated with a CpG island and inherits a paternal methylation mark from the germline (Takada et al., 2002) (Fig.3.1.A). The IG-DMR is a control element for imprinted genes on the maternal chromosome; deletion of the IG-DMR from the maternally inherited chromosome causes bidirectional loss of imprinting of all genes in the cluster. However imprinting is unaltered if the mutation is transmitted from the father, demonstrating that the two parental chromosomes control imprinted gene expression of this region differently (Lin et al., 2003). The imprinted gene \textit{Snrpn} on mouse chromosome 7 possesses a differentially methylated region at the 5' end of the gene. Methylation of the maternal chromosome occurs during oogenesis and results in exclusively paternal expression of \textit{Snrpn} (Shemer et al., 1997).

Transcriptional patterns are also affected by modifications to the N-terminal tails of histones (Strahl and Allis, 2000). Modifications include acetylation, methylation, phosphorylation, ubiquitination and ADP-ribosylation (Wolffe and Hayes, 1999). Histone acetylation was one of the first histone modifications to be identified and was first associated with transcriptional activation more than 40 years ago (Allfrey et al., 1964; Pogo et al., 1966). It is now well established that acetylation correlates with transcription (Kurdistani et al., 2004). Occurring on lysine residues, this modification neutralises the positive charge of the histone tails, thereby decreasing their affinity for DNA (Hong et al., 1993), and allowing increased access of transcriptional regulators and machinery to chromatin templates (Lee et al., 1993; Vettese-Dadey et al., 1996). Acetylation is catalysed by histone acetyltransferases (HATs), of which a variety have been identified in eukaryotes (Sterner and Berger, 2000). The reverse process, deacetylation, is carried out by a family of histone deacetylases, or HDACs. In humans, 18 have been identified and can be placed into 4 classes on the basis of their sequence and/or domain structure (Gregoretti et al., 2004). The NuRD complex contains two Class I HDACs, HDAC1 and HDAC2. These are also found in other transcriptional repression complexes such as the Sin3 complex (Ahringer, 2000).
Histone methylation is another common modification that impacts upon nuclear processes such as transcription. Catalysed by SET-domain containing proteins (Rea et al., 2000) and occurring on conserved arginine and lysine residues, it is correlated both with transcriptional activation and repression. Methylation of H3K9, H3K27 and H4K20 is associated with transcriptional repression (Lachner et al., 2003; Vaquero et al., 2003). Methylated H3K9 provides a binding site for Heterochromatin Protein 1 (HP1) (Bannister et al., 2001; Lachner et al., 2001). By its ability to dimerize, HP1 has been suggested to promote the formation of higher-order chromatin structure of heterochromatin (Nielsen et al., 2001a). Chromatin remodelling is an ATP-dependent process that allows the accessibility of nucleosomal DNA to factors involved in nuclear processes such as transcription or DNA replication to be either enhanced or decreased (Vignali et al., 2000). It is catalysed by a number of highly conserved enzymes that can be subdivided into three families based on the sequence of their ATPase subunits; the SWI/SNF group, the ISWI group and the Mi-2/CHD group (Boyer et al., 2000).

Increasingly, links are being found between the various types of epigenetic modifications described, showing that they have varying degrees of interdependence. An excellent example is the dependence of both DNA methylation and transcriptional silencing on underlying histone methylation at pericentric heterochromatin in ES cells. HP1, recruited to Me3H3K9, interacts with the DNA methyltransferase Dnmt3b, and is necessary for localization of this enzyme to, and therefore methylation of, the satellite repeat DNA (Lehnertz et al., 2003). A similar mechanism occurs in both Arabidopsis thaliana and Neurospora crassa, where mutations in the H3K9 HMTases dim-5 and kryptonite lead to DNA demethylation (Jackson et al., 2002; Tamaru and Selker, 2001).

The highly repetitive major and minor satellite repeat sequences are capable of being transcribed, but this can be deleterious to mammalian cells (Bouzinba-Segard et al., 2006), and is usually repressed by the formation of heterochromatin at these sequences. Interestingly, this involves the RNA interference (RNAi) pathway. Transcription of these repeat sequences leads to the formation of double stranded RNA, which activates
the RNA interference pathway and leads to the formation of the RITS (RNA Induced Transcriptional Silencer) complex. This associates with histone methyltransferase activity, and is targeted to the repetitive DNA, leading to silencing histone methylation, the formation of a heterochromatic structure and silencing (Volpe et al., 2002).

There is also a clear link between chromatin remodelling by members of the SNF2 protein family and DNA methylation. For example, mutations in the SNF2-like ATPase family member DDMI (for ‘Decrease in DNA Methylation’) in *A.thaliana* lead to a severe reduction in the levels of DNA methylation (Jeddeloh et al., 1999; Singer et al., 2001). Targeted deletion of the mammalian homologue of DDMI, Lsh, is associated with a substantial loss of methylation throughout the genome, due to a defect in *de novo* methylation by Dnmt3a and Dnmt3b (Dennis et al., 2001; Geiman et al., 2001; Zhu et al., 2006). Mutations in Lsh also lead to aberrant histone methylation patterns, demonstrating again the inter-dependence of epigenetic modifications (Yan et al., 2003). It is not yet clear why a failure to correctly remodel chromatin leads to DNA demethylation.

Several links have also been identified between the transcriptionally repressive modifications DNA methylation and histone hypoacetylation, although it remains unclear which is the initiating event during the formation of heterochromatin. The hypothesis that DNA methylation is the primary event in this process comes from studies showing that Methyl binding domain (MBD) containing proteins which bind selectively to methylated DNA (Hendrich and Bird, 2000; Wade, 2001) can recruit histone deacetylase and histone demethylase activity to bring about a condensed chromatin structure (Bird and Wolffè, 1999; Fuks et al., 2003b; Nan et al., 1998). Additionally, the acetylation status of a gene stably transfected into cells appeared to depend on the DNA methylation status of the gene prior to transfection (Eden et al., 1998), suggesting that acetylation patterns are specified by DNA methylation patterns. This suggests that DNA methylation may be a primary event in heterochromatin formation, with histone modifications serving to ‘lock in’ a repressive chromatin state identified by methylated DNA. This may be the case when heterochromatin is replicated.
(Fuks et al., 2000; Rountree et al., 2000). However during the formation of a heterochromatin state de novo, for example when genes are silenced in response to an extra- or intracellular signal, there is much correlative evidence that histone deacetylation and gene silencing precede DNA methylation. This is the case during X inactivation (Heard et al., 2001), and is also supported by the finding that transcriptional silencing, histone hypoacetylation and demethylation of H3K4 were the primary events occurring at a transgene following integration into chicken erythroid cells, followed by methylation of DNA and H3K9 (Mutskov and Felsenfeld, 2004). Additionally, following microinjection of a methylated TK gene into cells, transcriptional silencing relies upon the formation of chromatin to occur, and does not occur for more than 8hrs; DNA methylation alone is insufficient to cause transcriptional silencing (Buschhausen et al., 1987). Furthermore, there is some evidence in Neurospora crassa and in mammalian cells that patterns of histone acetylation can direct DNA methylation patterns (Hu et al., 2000; Selker, 1998).

The NuRD complex contains both histone deacetylase activity, via the HDAC core, and chromatin remodelling capabilities, via the SWI2/SNF2 type helicase/ATPase domain-containing protein Mi-2β (Zhang et al., 1998). Previous work has shown that ES cells lacking Mbd3, a central structural component of the NuRD complex, are highly demethylated at the major and minor satellite repeats, as well as at the long terminal repeat sequences of the Intracisternal A particle (IAP) DNA retrotransposons and a single copy sequences (Fig.1.7). However the methylation status of a different single copy sequence is unaffected by the loss of Mbd3 (Fig.1.8). Rescue cell lines expressing any of the Mbd3 isoforms, Mbd3a, Mbd3b or Mbd3c display methylation levels at these sequences comparable to those seen in wild type cells, demonstrating that hypomethylated DNA in Mbd3-/- ES cells retains the necessary heterochromatic features to allow the appropriate restoration of patterns of DNA methylation once NuRD function is restored.

In this chapter I set out to establish what, if any, further epigenetic changes are associated with a loss of Mbd3 from pluripotent cells. Evidence is presented that Mbd3
null ES cells contain reduced amounts of DNA methylation at various sites throughout the genome, and that this is associated with histone hyperacetylation and transcriptional mis-regulation.

3.2 Results

3.2.1 Reduced DNA methylation in Mbd3<sup>−/−</sup> ES cells

Because it had been established that three types of repetitive, constitutively heterochromatic DNA sequence, as well as a CpG island which is normally methylated in ES cells, undergo demethylation in Mbd3<sup>−/−</sup> ES cells, I decided to examine the methylation status of other genomic regions which are known to be methylated in wild type cells. I chose differentially methylated regions (DMRs) for the Gtl2/Dlk1 and H19/Igf2 loci because some of these genes displayed transcriptional mis-regulation in Mbd3<sup>−/−</sup> ES cells by microarray (K.Kaji, unpublished results), subsequently verified by quantitative PCR (qPCR) (Fig.3.6). Additionally, methylation of the DMRs of these genes is known to affect their transcription (Li et al., 1993). Within the 8kb Gtl2/Dlk1 intergenic-DMR (IG-DMR), two separate regions have previously been defined as being differentially methylated, these are termed ‘M4’ and ‘M5’ (Takada et al., 2002) (Fig.3.1.A). Both the M4 and M5 regions were examined here. The Snrpn differentially methylated region (DMR) was also looked at as an example of a gene whose expression does not change in Mbd3 null cells, either by microarray or qPCR (Fig.3.4).

Sequencing of bisulfite-modified DNA, following amplification by PCR and T/A cloning, was carried out in order to determine the extent of methylation at these loci in Mbd3 parental, null and Mbd3b rescue cells. At least ten sequences were examined for each CpG site, and the degree of conversion of cytosine to thymidine was more than 95% in all sequences examined.

In the null cells both the Gtl2/Dlk1 IG-DMR and H19 5’ DMR show decreases in the methylation level, with the biggest reduction at the Gtl2/Dlk1 IG-DMR M5 region. In
contrast the H19 5' DMR shows a limited but consistent decrease (Fig.3.2 A, B and C).
The demethylation occurs randomly; there is no evidence that one allele is preferentially
demethylated (Fig.3.3). Similar to what is seen for the major and minor satellites and
IAP LTRs, rescue cell lines expressing Mbd3b display levels of methylation comparable
to wild type cells. However, as demonstrated before for the Pramel 7 intron (Fig.1.7.B),
the demethylation is not global; there is no observable decrease in methylation levels at
the Snrpn DMR across 13 CpG sites (Fig.3.2.D).

As can be seen from Fig.3.2, both the Gtll/Dlk1 IG-DMR (A, B) and H19 5' DMR (C)
show complete methylation in wild type cells, suggesting full methylation at both
alleles, despite these regions having been previously shown to be imprinted and
therefore methylated only at one allele. However it has been previously shown that ES
cells frequently show epigenetic instability, including a loss of imprinting at a number of
imprinted genes (Humpherys et al., 2001). In particular the H19 5' DMR is commonly
biallelically methylated, leading to a strong reduction in expression levels (Dean et al.,
1998). Despite H19 and Igf2 being reciprocally imprinted, biallelic silencing of H19 is
not necessarily associated with biallelic expression of Igf2.

In addition to examining the DMRs of imprinted genes I looked at the promoter regions
of non-imprinted genes to see if they were also affected. It was noticed from the
microarray data that a number of up-regulated genes are germ-cell specific (B.Hendrich,
personal communication). A recent study showed that DNA methylation is a primary
mechanism for silencing germ-cell specific genes, including Dazl, both in germ and
somatic cells (Maatouk et al., 2006). Therefore we speculated that the up-regulation of
germ-cell specific factors in Mbd3<sup>−/−</sup> ES cells may be associated with demethylation of
their CpG island. This has previously shown to be the case for the germ-cell specific
gene Dazl (Fig.1.7.B), a gene coding for an RNA binding protein (Cooke et al., 1996)
which is upregulated in the absence of Mbd3 (Kaji, Nichols and Hendrich, In press). We
decided to focus on the CpG island associated with a further germ cell specific gene that
is up-regulated according to the microarray data, the bHLH transcription factor Sohlh2
(Ballow et al., 2006) (Kaji K, Nicholls J and Hendrich B; in press). As before, bisulfite-
modified DNA from the relevant cell types was amplified by PCR, and the products cloned and sequenced. At least ten sequences were examined for each CpG site, and the degree of conversion of cytosine to thymidine was more than 95% in every case.

The Sohlh2 associated CpG island displays significant demethylation in Mbd3 ES cells (Fig.3.2.E). This is again broadly rescued by the re-introduction of Mbd3b. This CpG island shows a fairly significant, but not complete, degree of methylation in wild type ES cells, in which they are silenced. This is in contrast to the imprinted gene regions and repetitive regions which are constitutively methylated in wild type ES cells. Therefore demethylation of DNA in Mbd3 ES cells is seen both at sequences that are heavily methylated in wild type cells, and those that are less heavily methylated. The lack of change in the methylation status of the Snrpn DMR and Pramel 7 intron demonstrates that there is some specificity in the sequences affected.

### 3.2.2 Demethylation is associated with histone hyperacetylation

Because the NuRD complex contains histone deacetylase activity, and links between histone acetylation and DNA methylation have previously been shown, we examined the same regions shown to be demethylated in Mbd3 ES cells for changes to levels of acetylated histone H3 lysine 9 (AcH3K9). This is considered to be a mark of 'active chromatin', and is correlated with transcriptional activation (Kurdistani et al., 2004). Mbd3 ES cells have been shown to have a 3-4 fold increase in the levels of this mark at the promoters of two up-regulated genes, Pramel 6 and 7 (Kaji et al., 2006). In contrast, the level of AcH3K9 at the promoters of several genes whose expression does not change, such as Oct4, Nanog and TCRγ, is not altered in the null cells. Here, all the regions previously shown to be demethylated in Mbd3 ES cells were examined for changes to histone acetylation.

Every locus examined displayed hyperacetylation, with the exception of the Oct4 promoter, in agreement with previous studies (Kaji et al., 2006). The largest increase, of c.10 fold increase, was seen at the Sohlh2 promoter (Fig.3.4.A). All other loci displayed
more modest increases of between 2 and 5 fold. This was rescued by Mbd3b at every locus.

In addition to the loci examined previously, the analysis was extended to include a second region of the *H19 5' DMR (H19 5' DMR (2))*, as well as the intergenic region of rDNA repeat genes (rDNA IG region) and the promoter of the *HtrAl* gene. Both the *Htral* promoter and rDNA intergenic region also displayed increased acetylation. In eukaryotes, the rDNA genes are tandemly arrayed in hundreds to thousands of copies within nucleolus organizer regions (NORs) (Grummt, 2003; Moss and Stefanovsky, 2002). However not all the genes are expressed at any one time; in metabolically active human or mouse cells about half of the c.400 repeats are silenced at any one time (Santoro and Grummt, 2005). The active and inactive genes are differentially epigenetically marked, with the promoters of active genes being hypomethylated and hyperacetylated, and the opposite pattern being observed at inactive gene promoters (Lawrence et al., 2004). Thus the rDNA repeats are an example of tandemly repeated sequences that are not constitutively heterochromatic, unlike other tandem arrays such as the satellite repeats (Lippman et al., 2004).

*HtrAl* is an example of a gene that is significantly up-regulated in *Mbd3<sup>-/-</sup>* ES cells by microarray (K.Kaji, unpublished results). The gene product is a serine protease that has been shown to inhibit signalling mediated by TGFβ family proteins (Oka et al., 2004), as well as having a role in trophoblast differentiation and/or invasion and uterine decidual regression during placental development (Nie et al., 2005).

Thus a loss of DNA methylation is associated with an increase in histone acetylation in *Mbd3<sup>-/-</sup>* ES cells. Chromatin immunoprecipitation was carried out using an antibody against AcH3K9, and precipitated DNA was analysed by real-time PCR. Values represent the average of three independent experiments and were normalised against those at the β-Actin gene promoter.
3.2.3 No significant alterations to histone methylation in null cells

I also looked to see if the levels of tri-methylated H3K9 were perturbed in $Mbd3^{+/−}$ ES cells. Along with hypoacetylation, this is a mark of constitutive heterochromatin, for example at the centromeres (Lehnertz et al., 2003; Peters et al., 2001) in mammals, and at the mating type locus in $S.pombe$ (Nakayama et al., 2001). It is also be involved in gene repression at euchromatic targets, where the Suv39h1 HMTase is recruited, along with HP1, by the tumour suppressor Retinoblastoma protein (pRb) (Nielsen et al., 2001b; Vandel et al., 2001). Evidence from Arabidopsis thaliana, Neurospora crassa, and mammals shows that Me3H3K9 can direct DNA methylation (Jackson et al., 2002; Lehnertz et al., 2003; Tamaru et al., 2003). Mechanistic links have also been established between histone acetylation and histone methylation, most clearly in $S.pombe$, where deacetylation of H3K9 by C1r3 ($Hdac1$ homologue), along with the WD-40 containing protein Rik1, is necessary for methylation of H3K9 by Chr4 ($Suv39h1$ homologue) and Swi6 ($HP1$ homologue) localisation (Nakayama et al., 2001).

Thus I was interested to see if there was any correlation between the decreased DNA methylation, increased histone acetylation, and histone methylation in the $Mbd3^{+/−}$ cells. Chromatin from parental, null and rescue cells was precipitated with an antibody against Me3H3K9, and the precipitated DNA amplified by quantitative PCR. Values were normalised against those at the β-Actin gene promoter.

Of the loci examined in this experiment, there was no clear trend and no large changes between the parental, $Mbd3$ null and $Mbd3b$ rescue cells (Fig.3.4.B. This lack of a significant change is supported by immunofluorescence data (see Fig.3.5.A), which clearly shows that similar levels of Me3H3K9 localise to centromeric DNA (marked by DAPI bright spots) in both parental and null cells. Despite the links described above between histone methylation, acetylation and DNA methylation, Me3H3K9 does not change significantly in $Mbd3^{+/−}$ ES cells, indicating that not all aspects of the heterochromatic structure are perturbed.
3.2.4 No changes to the localisation of HP1α or Lsh

HP1α (Heterochromatin Protein 1α) is a well characterised marker of heterochromatin that helps to organise and maintain the heterochromatic structure in diverse organisms from *S. pombe* to humans (for review see (Richards and Elgin, 2002). HP1α depends on the presence of Me3H3K9 for its localisation in mammalian cells (Bannister et al., 2001; Rea et al., 2000), and so I did not expect any change in its localisation in *Mbd3* null cells. Indeed this is what was observed (Fig. 3.5.B).

*Lsh* (Lymphoid Specific Helicase), a member of the SNF2 helicase family of chromatin remodelers, is required for genome-wide methylation and post-natal survival (Dennis et al., 2001; Geiman et al., 2001). To see if any changes in the localisation or levels of Lsh could be causing the demethylation observed in *Mbd3* null cells, immunofluorescence was carried out with an anti-Lsh antibody. No changes could be observed in either the localisation or the levels of this protein by this method (Fig. 3.5.C).

3.2.5 Aberrant transcription of demethylated, hyperacetylated sequences

Correlations between histone acetylation, DNA methylation and transcription have been described previously, with expressed genes tending to be found in hyperacetylated, demethylated euchromatin (e.g. (Siegfried et al., 1999)). To see if the epigenetic changes were associated with any changes in transcription patterns in the *Mbd3* null cells, the transcription of each of the sequences that are demethylated and hyperacetylated in *Mbd3*−/− ES cells, or the associated genes in the case of regulatory regions, was measured by quantitative PCR in parental, null and rescue cells. Values represent the average of three independent experiments and were normalised against those of *Gapdh*.

Transcription of both major and minor satellite repeats was increased in *Mbd3*-null ES cells, as was transcription of IAPs and LINE elements (Fig. 3.6). Surprisingly transcriptional silencing of these repeat sequences is not restored in the *Mbd3b*-rescued cells, despite the fact that these cells contain near normal levels of DNA methylation and histone acetylation at these sequences.
H19 and Gtl2 were found to be mis-regulated in Mbd3<sup>−/−</sup> ES cells by microarray previously (K.Kaji, unpublished results), and this was broadly confirmed by qPCR. As expected, H19 expression increased in the null cells in which the 5' DMR is demethylated. Igf2 also shows a slight increase in transcription, surprising given that they normally show reciprocal expression patterns (Ferguson-Smith, 2000; Szabo et al., 2000). Gtl2 and Dlk1 also do not show reciprocal changes in expression; Gtl2 transcription is reduced as would be expected, but Dlk1 transcription is also reduced. Soxlh2 and Dazl both show significant increases in transcription, which is consistent with the long-held notion that acetylation of histones and demethylation of DNA is associated with increased transcription (Kurdistani et al., 2004; Maatouk et al., 2006). Transcription in the Mbd3b rescue cells shows that expression of all single copy sequences is broadly rescued (Fig.3.6), as would be expected from the restoration of the wild type epigenetic pattern.
Figure 3.1. Schematic representation of the differentially methylated regions (DMRs) of the $Gtl2/Dlk1$ and $H19/Igf2$ imprinted loci. A) Methylation status of the $Gtl2/Dlk1$ DMRs in E12.5 embryos. IG-DMR: intergenic DMR. This 8kb region can be split into three sections termed M3, M4 and M5. White ovals: unmethylated site. Black ovals: methylated site. The $Diki$ DMR is unmethylated on the maternal allele and 50% methylated on the paternal allele. $Rtl1$ is a retrotransposon-like gene expressed from the maternal chromosome and has an antisense transcript containing 2 micro-RNAs expressed from the paternal chromosome. Adapted from Shau-Ping Lin et al, 2003 and Takada et al, 2002.

B) Methylation status of the $H19/Igf2$ intergenic DMR. Methylation on the paternal allele prevents binding of the insulating factor CTCF, and allows the $Igf2$ gene long range preferential access to an enhancer located downstream of the $H19$ gene. On the maternal allele, the region is not methylated, CTCF binds, and the $H19$ gene has preferential access to the enhancer sequences. White ovals: unmethylated site. Black ovals: methylated site. There are also at least two more DMRs in the $H19$ and $Igf2$ genes but they are omitted for clarity. Adapted from Murphy et al, 2006.
Figure 3.2. Demethylation of a variety of regions in Mbd3 null ES cells - summary of bisulfite sequencing data. A) Gt12/Dlk1 IG-DMR M4 region, B) Gt12/Dlk1 IG-DMR M5 region, C) H19 5' DMR, D) Snrpn DMR and E) Sohlh2 CpG island. White bars: Mbd3(flox/-); black bars: Mbd3(-/-); diagonally striped bars: Mbd3b rescue cells. Bisulfite-treated DNA from the indicated cell lines was amplified by PCR, T/A cloned and at least ten sequences covering all CpG sites analysed for each locus. Each locus was looked at in two independent experiments (i.e. the figures shown here represent the average methylation of each CpG site in 20 sequences). Methylation of the Sohlh2 promoter in Mbd3(-/-) cells is 0% at several CpG sites.
Figure 3.3 Schematic representation of methylation status of imprinted regions in Mbd3 parental and null cells. A) Gtl2/Dlk1 IG-DMR (M5 region); B) H19 5’DMR; C) Snrpn DMR. Each horizontal line represents an individual clone, and circles represent CpG sites. Filled circles: methylated CpG site; open circles: unmethylated CpG site.
Figure 3.4. Regions of DNA demethylated in Mbd3(-/-) ES cells contain hyperacetylated histones A) Chromatin immunoprecipitation (ChIP) with antibodies against acetylated H3K9 B) ChIP with antibodies against trimethylated H3K9. White bars: Mbd3(lox/-); Black bars: Mbd3(-/-); Diagonally striped bars: Mbd3b rescue cells. Levels of immunoprecipitated DNA were measured by real-time PCR and were normalised against those at the B-Actin promoter; PCR experiments were measured in triplicate. Values represent the average of three independent biological experiments (and PCR experiments in triplicate). Statistically significant changes are shown with an asterisk (Mann-Whitney test, n=9, p<0.05)
Figure 3.5 Immunofluorescence of both parental (Mbd3^{0/8/8}) and Mbd3
null (Mbd3^{<-<-}ES cells shows that there is no detectable change in the
levels or localisation of the heterochromatic marks Me3H3K9 (A) or HP1\(\alpha\)
(B), or the chromatin remodeller Lsh (C). All show localisation to
centromeric regions (DAPI bright spots). Images were taken on a Leica
c confocal microscope.
Figure 3.6. Sequences with epigenetic aberrations in Mbd3(-/-) ES cells also display transcriptional misregulation. Quantitative RT-PCR was used to measure levels of transcription of demethylated, hyperacetylated sequences, or genes regulated by them. White bars; Mbd3(-/-); Black bars: Mbd3(-/-); Diagonally striped bars: Mbd3b rescue cells Values were normalised against Gapdh levels, and represent the average of the levels in at least three independently derived cDNA populations. As well as three biological replicates, three technical replicates were also carried out (the PCR experiments were done in triplicate). Stastically significant differences are indicated by an asterisk (Mann-Whitney test, n = 9, p<0.05)
3.3 Discussion

I have shown that the loss of a central component of a transcriptional repression complex possessing both histone deacetylation and chromatin remodelling activities leads to alterations of both histone acetylation and another major type of epigenetic mark, DNA methylation, at numerous sites throughout the genome. This is associated with transcriptional mis-regulation of a variety of sequences. Importantly, almost all aspects of the phenotype are rescued by the reintroduction of Mbd3b. Thus despite becoming hypomethylated in the absence of Mbd3, these regions retain sufficient heterochromatic identity to have the appropriate epigenetic marks restored to be transcriptionally silenced upon the reintroduction of NuRD function.

3.3.1 How is NuRD involved in DNA methylation?

Given the demonstrated interdependence of histone methylation and DNA methylation, notably at the satellite repeat DNA, it was interesting to find that there was no change in the levels or localization of Me3H3K9 or HP1. This argues against the observed DNA demethylation being caused by a defect in their recruitment by HP1 to heterochromatic regions. Also, the observation that there is no obvious change in the localization or levels of the chromatin remodeller Lsh suggests that NuRD is not involved in Lsh-mediated DNA methylation.

DNA methylation and histone hypoacetylation have been linked both correlatively and mechanistically previously, although it remains unclear which is the primary silencing event (Bird and Wolffe, 1999; Hu et al., 2000; Mutskov and Felsenfeld, 2004; Selker, 1998). The finding that both DNA hypomethylation and histone hyperacetylation are associated with a loss of Mbd3 in ES cells reinforces this link, however it remains unclear how the NuRD complex functions to bring about these epigenetic modifications. An attractive scenario is one in which histone deacetylation is required for DNA methylation to occur, either during the replication or de novo formation of heterochromatin. Thus a failure to deacetylate histones at NuRD target sites would be
refractory to DNA methylation, swiftly leading to DNA demethylation. This is supported by the observations that histone hypoacetylation is frequently the primary event in heterochromatin formation (Mutskov and Felsenfeld, 2004), and by some limited evidence that histone acetylation patterns can direct DNA methylation patterns (Hu et al., 2000; Selker, 1998). It will be interesting to explore this hypothesis in the future.

As discussed in the Introduction, chromatin remodelling by SNF2 family members is crucial for DNA methylation in a number of organisms. Why this is so has not been established, but one plausible mechanism is that chromatin remodelling is a pre-requisite for DNA methylation, perhaps by allowing increased access to the substrate DNA. However, whether there is a clear correlation between alterations to higher-order chromatin structure and DNA methylation has not been shown. Because the NuRD complex contains chromatin remodelling activity imparted by the Mi-2β subunit, as part of this study I attempted to look for changes to the higher order chromatin structure at demethylated regions by examining the extent of digestion by micrococcal nuclease, however this was inconclusive. It will be interesting to investigate this more thoroughly in the future.

3.3.2 Do the epigenetic aberrations cause transcriptional misregulation or vice-versa?

It is noticeable that there is a correlation between the level of demethylation, hyperacetylation, and transcriptional mis-regulation in Mbd3 null cells; this is most clearly seen in the case of Sohlh2, which exhibits the highest degree of all of these. Conversely, the Snrpn DMR is not demethylated, and this gene is not mis-expressed. DNA methylation and histone hypoacetylation inhibit transcription, because they both promote the formation of a condensed heterochromatic structure, which is thought to impede access to transcription factors and the transcriptional machinery (Lee et al., 1993; Vettese-Dadey et al., 1996). DNA methylation is also known to influence the expression of imprinted genes (Li et al., 1993), and it is notable that of the c.200 genes
shown by microarray analysis to be misexpressed in Mbd3 null cells (K. Kaji, personal communication), imprinted genes are over-represented compared to their representation on the entire microarray (6% vs 0.1% of genes recorded as present on the microarray). Thus it is plausible that the observed epigenetic changes cause the transcriptional misregulation observed in Mbd3 ES cells. However it is not clear if this is the case for all mis-expressed genes; the analysis would have to be extended to determine if this is the case. It has been reported previously that transcriptional silencing of either a newly integrated transgene, or genes on the inactive X chromosome during X inactivation, occurs prior to other chromatin modifications (Gautsch and Wilson, 1983; Keohane et al., 1996; Lock et al., 1987; Niwa et al., 1983; Wutz and Jaenisch, 2000). Thus it is also possible that changes to transcription patterns lead to the epigenetic changes, not vice-versa. A close study of the temporal order of events following the loss of Mbd3 would unravel this by showing which occurs first; transcriptional or epigenetic mis-regulation, and it will be interesting to do this in the future.

There were some unexpected findings from the measurement of transcription levels of demethylated, hyperacetylated sequences in Mbd3 null ES cells (Fig.3.6). For example, H19 expression increased in the null cells in which the 5' DMR is demethylated. Igf2 also shows a slight increase in transcription, surprising given that they normally show reciprocal expression patterns (Ferguson-Smith, 2000; Szabo et al., 2000). However it has been previously shown that changes in methylation of the H19/Igf2 DMRs observed in ES cells do not always lead to the expected, reciprocal change in expression of the two genes (Dean et al., 1998). Additionally, in Dnmt1−/− MEFs which are significantly demethylated, the expression of both H19 and Igf2 is reduced, suggesting that methylation is not solely responsible for the reciprocal expression patterns (Jackson-Grusby et al., 2001). It is worth noting that the only imprinting control region (ICR) for this locus that was examined was the H19 5' DMR. Several other differentially methylated regions (DMRs) exist for this locus so it is possible that these did not undergo demethylation, and that this contributes to both H19 and Igf2 being expressed (Charalambous et al., 2004; Constancia et al., 2000; Drewell et al., 2002; Eden et al., 2001; Murrell et al., 2001).
*Gtl2* and *Dlk1* also do not show reciprocal changes in expression; *Gtl2* transcription is reduced as would be expected, but *Dlk1* transcription is also reduced. This is not what is seen in *Dnmt1*+ embryos, where *Gtl2* expression decreases, but *Dlk1* expression increases (Schmidt et al., 2000). Again, the IG-DMR was the only DMR looked at here, and it is possible that the methylation status of other control regions for the genes was affected and that this contributes towards the non-reciprocal changes in expression that are observed. *Sohlh2* and *Dazi* both show significant increases in transcription, which is consistent with the long-held notion that acetylation of histones and DNA demethylation is associated with increased transcription, particularly of germ-cell specific genes (Kurdistanti et al., 2004; Maatouk et al., 2006; Razin and Cedar, 1991).

### 3.3.3 Increased transcription of repetitive sequences

Loss of Mbd3 also leads to an increase in transcription of repetitive elements such as the major and minor satellites and IAP elements (Fig.3.6). The satellite repeats have been previously reported to be transcribed, but at a low level, in a variety of organisms and cell types from ES cells to yeast and maize, and this may contribute to centromere formation and function (Chen et al., 2003a; Gaubatz and Cutler, 1990; Lehnertz et al., 2003; Masumoto et al., 2004; Nakano et al., 2003; Rudert et al., 1995; Topp et al., 2004). Conversely, studies in murine erythroleukemic (MEL) cells show that forced accumulation of minor satellite transcripts is deleterious for the cell, leading to chromosome mis-segregation, loss of sister-chromatid cohesion, and aneuploidy (Bouzinba-Segard et al., 2006). Additionally, defects in the RNAi pathway in yeast, which results in increased levels of satellite transcripts, also lead to a mislocalisation of Me3H3K9 and a failure to build a fully developed centromere (Volpe et al., 2003), and loss of the Suv39h enzymes in ES cells leads to increased transcription of the satellite repeats, along with DNA demethylation of these regions and aneuploidy (Lehnertz et al., 2003; Martens et al., 2005; Peters et al., 2001).
In contrast the increased transcription of the satellite repeats observed in *Mbd3* cells is not associated with any observable chromosomal defect; loss of sister chromatid cohesion is not visible microscopically and the cells are not aneuploid, even after multiple passages (K.Kaji, personal communication). It is not clear why increases in satellite transcripts should have such a deleterious effect in some cell types but not others. One difference worth noting is that in both yeast, MEL and *Suv39h1, h2* ES cells an increase in satellite transcription was associated with mislocalisation of other heterochromatic marks including Me3H3K9 and HP1; this is not seen in *Mbd3* cells. Epigenetic changes such as these indicate a more profound alteration of the heterochromatic state which likely impacts on the formation of the centromeric structure. It is not known if histone methylation is a crucial component of centromere formation, but the finding that alterations to DNA methylation and histone acetylation, and increases in transcription, are not sufficient to affect centromere function suggests that histone methylation is more important than other epigenetic marks in this respect.

Increased transcription of IAP elements is also observed in *Mbd3* null cells. This is presumably deleterious to the ES cell genome; loss of DNA methylation at IAP elements in *Dnmt1* embryos leads to their stable activation, and this is associated with a high rate of tumourigenesis due to mutation-causing transposition of these elements (Walsh et al., 1998). Whether this leads to an increased mutation rate in *Mbd3* ES cells or embryos has not been examined but it is possible that this contributes to the *Mbd3* phenotype.

Surprisingly, transcriptional levels of the repetitive sequences are not rescued in the *Mbd3b* rescue cells (Fig.3.6), despite the fact that rescued cells contain near normal levels of DNA methylation and histone acetylation at these sequences. This demonstrates that the changes in epigenetic marks seen at these loci in *Mbd3* null ES cells are not simply reflecting changes in transcriptional status. The difference between repetitive and single copy sequences (whose aberrant transcription is rescued by *Mbd3b*) may reflect differences in the way in which they are silenced. For example, repression of repetitive sequences involves the RNAi pathway. It is possible that the pathway has been
compromised in Mbd3 null cells in a manner that is unable to be fully rescued by Mbd3b, so that despite reinstatement of the acetylation and DNA methylation patterns transcription is not down-regulated.

3.3.4 A contribution by transcription misregulation towards the Mbd3 phenotype?

Mbd3(-/-) ES cells grow slowly and display a failure to differentiate in the absence of LIF. Microarray analysis shows that approximately 200 genes, or 1.7% of the total number of genes recorded as present on the microarray, are mis-expressed two-fold or more in Mbd3 null cells (K.Kaji, S.Tomlinson, personal communication). Despite being considered to be a transcriptional repression complex, genes are found to be down-regulated as well as being up-regulated in the absence of Mbd3. This can be explained by considering that transcriptional regulators may be among the genes mis-regulated, thus leading to secondary effects on transcription not directly mediated by Mbd3. Given the abundance and conserved nature of the NuRD complex, it is perhaps surprising that only c.200 genes are mis-regulated. This suggests either that NuRD is targeted specifically to only a small range of genes, or that other mechanisms exist to regulate transcription in a combinatorial fashion with NuRD, and loss of NuRD is not sufficient to disturb this. In contrast, microarray analysis of Dnmt1(-/-) MEFs shows misregulation of approximately 600 genes (Jackson-Grusby et al., 2001). Some of the genes up-regulated in Mbd3(-/-) ES cells are normally expressed in pre-implantation embryos, but not in ES cells. These include the Pramel 6 and 7 genes (Kaji et al., 2006). It is suggested that the continued expression of pre-implantation genes is responsible for the Mbd3(-/-) embryonic phenotype and the failure of Mbd3(-/-) ES cells to differentiate (Kaji et al., 2006). It is not clear whether the mis-expression of other genes not necessarily associated with pre-implantation development, such as imprinted or germ-cell specific genes, contribute to the phenotype.
4 \textit{Mbd3}^{(-/-)} \text{ ES cells contain reduced levels of Dnmt3b}

4.1 Introduction

In order to investigate the cause of the methylation defect in \textit{Mbd3}^{(-/-)} \text{ ES cells} further, I next examined the levels and activities of the known DNA methyltransferases in wild type and mutant ES cells. The DNA methyltransferases, or Dnmts, are the enzymes responsible for methylating DNA. In mammals five members of the Dnmt family have been identified; \textit{Dnmt1}, 2, 3a, 3b and 3l (Fig. 1.2). All apart from \textit{Dnmt3l} contain a conserved catalytic methyltransferase domain in their C-terminus (Kumar et al., 1994). Between them they possess both de novo and maintenance methylation activities. (Li et al., 1992; Okano et al., 1999).

\textit{Dnmt1}, expressed ubiquitously in proliferating cells where it localises to replication foci (Leonhardt et al., 1992), was the first methyltransferase to be identified. Because of its localisation to replication foci, and because purified Dnmt1 protein methylates hemimethylated DNA substrates more efficiently than unmethylated DNA \textit{in vitro} (Bestor, 1992), Dnmt1 is proposed to be the primary maintenance methyltransferase that is responsible for copying the parental strand methylation pattern onto the daughter strand after the DNA is replicated. Despite significant homology to the other Dnmt family members, including the conserved methyltransferase motifs, \textit{Dnmt2} is catalytically inactive \textit{in vitro} and both endogenous and newly integrated retroviral DNA is methylated as efficiently in \textit{Dnmt2}^{(-/-)} \text{ ES cells} than in wild type cells (Okano et al., 1998b). A function has however been ascribed to Dnmt2; in human cells it has been shown to methylate the small RNA aspartic acid transfer tRNA\textsuperscript{Asp}, a function that is highly conserved from flies and plants to mammals. It is also suggested that an ancestral version of \textit{Dnmt2} was the founder of the Dnmt family (Goll et al., 2006). Dnmt3l was originally identified on the basis of its sequence similarity to Dnmt3a and 3b, but it does
not have an active methyltransferase domain. It is highly expressed in ES and germ cells (Aapola et al., 2001; Hata et al., 2002), and is thought to regulate methylation by cooperating with Dnmt3a and 3b to mediate de novo methylation (Chedin et al., 2002; Suetake et al., 2004). It is required for the establishment of both maternal and paternal imprints during gametogenesis (Arima et al., 2006; Bourc'his et al., 2001; Hata et al., 2002; Kaneda et al., 2004), as well as the regulation of germ-cell specific expression and the suppression of IAP retroviral sequences in germ cells (Bourc'his and Bestor, 2004; Hata et al., 2006).

_Dnmt3a_ and 3b, both highly expressed in ES cells, early embryos and developing germ cells (Okano et al., 1998a; Watanabe et al., 2002) possess de novo methylation activity _in vitro_ (Aoki et al., 2001), and are responsible for the methylation of newly integrated, unmethylated proviral or episomal DNA _in vivo_ (Hsieh, 1999; Okano et al., 1998a). Because of their demonstrated de novo methylation activity and their knock-out phenotypes, Dnmt3a and Dnmt3b are thought to be responsible for establishing genomic DNA methylation patterns during embryogenesis and gametogenesis (Howlett and Reik, 1991; Kafri et al., 1992; Okano et al., 1999). Both are essential genes; _Dnmt3a<sup>−/−</sup>_ mice develop to term and appear normal at birth, however most become runted and die at about four weeks of age. _Dnmt3b<sup>−/−</sup>_ mice display a more severe phenotype, with multiple developmental defects appearing after E9.5 and subsequent embryonic lethality (Okano et al., 1999). _Dnmt3a, 3b<sup>−/−</sup>_ double knock-out mice display an even more severe phenotype, with an arrest of growth and morphogenesis shortly after gastrulation and lethality before E11.5, indicating that Dnmt3a and 3b have overlapping functions _in vivo_. In humans, mutations in the DNMT3B gene are responsible for the autosomal recessive disorder Immunodeficiency, Centromeric instability and Facial anomalies (ICF) syndrome (Hansen et al., 1999; Xu et al., 1999). Along with _Dnmt1_ null ES cells, _Dnmt3a, 3b<sup>−/−</sup>_ null ES cells that have been extensively passaged lack almost all their DNA methylation and are unable to terminally differentiate, despite displaying no obvious growth or morphological phenotype (Jackson et al., 2004; Lei et al., 1996), suggesting that DNA methylation is required for the differentiation of ES cells but not their self-renewal.
An analysis of the methylation status of Dnmt3a\(^{+/−}\) and Dnmt3b\(^{+/−}\) embryos reveals slight demethylation of a number of sequences including endogenous C-type retroviral DNA and IAP repeats in Dnmt3a\(^{+/−}\) embryos, and significant demethylation of the minor satellites in Dnmt3b\(^{+/−}\) embryos. The extent of demethylation was compounded in Dnmt3a, 3b\(^{+/−}\) embryos (Okano et al., 1999). Despite differences in phenotype and sequences methylated in embryos, in Dnmt3a\(^{+/−}\) or 3b\(^{+/−}\) ES cells the only demethylation observed, even after 5 months in culture, is of the minor satellite DNA in Dnmt3b\(^{+/−}\) ES cells (Chen et al., 2003b; Okano et al., 1999). Similar extended passaging of Dnmt3a, 3b\(^{+/−}\) ES cells however leads to extensive demethylation of almost every sequence examined, including endogenous C-type retroviral DNA, IAP repeats, minor and major satellite DNA, the highly methylated and tissue-specific β-globin and phosphoglycerate kinase 2 (Pgk-2) autosomal gene promoters, the X-linked genes Pgk-1 and Xist, and the imprinted regions the H19 5’DMR, the DMR2 of Igf2, region 2 of Igf2r, the DMR of Peg1, and the DMR of Snrpn (Chen et al., 2003b). In agreement with this, the level of methylation in Dnmt3a, 3b\(^{+/−}\) ES cells has been shown to decrease steadily during prolonged passage, with only 0.6% of CpG dinucleotides methylated after 75 passages, compared to 65% in wild type cells (Jackson et al., 2004). Thus in addition to their role as de novo methyltransferases, Dnmt3a and 3b are thought to cooperate and to contribute together to the maintenance of the methylation of a variety of sequences, from repetitive sequences (Liang et al., 2002) to single copy DMRs and promoters (Chen et al., 2003b), and CpG islands in ES cells (Hattori et al., 2004a).

Apart from methylation of the minor satellite sequences by Dnmt3b, Dnmt3a and 3b seem to be completely redundant in what sequences they are able to methylate in ES cells, but this is not the case in embryos (Okano et al., 1999). Furthermore, overexpression of Dnmt3a or 3b in human cells (which express very low levels of endogenous Dnmt3a and 3b), leads to differential, non-random methylation of both stable episomes and their corresponding chromosomal targets, suggesting the two enzymes have different target specificities (Hsieh, 1999). Also, expression of either Dnmt3a or 3b in highly demethylated Dnmt3a, 3b\(^{+/−}\) ES cells showed that both are
capable of remethylating the genome, however they have both common and preferred targets. For example, Dnmt3a preferentially methylated the major satellite repeats and Dnmt3b the minor satellite repeats; Dnmt3a was able to methylate the 5' region of Xist and the H19 5’DMR, but Dnmt3b was not (Chen et al., 2003b). Thus there may be some target specificity for each Dnmt; nonetheless there is clearly a highly significant degree of overlap between the respective specificities.

Both Dnmt3a and Dnmt3b have multiple isoforms that may have different functions. At least two isoforms of Dnmt3a have been reported, Dnmt3a and Dnmt3a2 (Fig. 1.3). Each transcript is initiated from a separate promoter, and Dnmt3a2 is shorter, lacking the N-terminal region of full-length Dnmt3a. The N-terminal region missing in Dnmt3a2 is responsible for the localisation to heterochromatic foci of Dnmt3a, and instead it has a diffuse pattern in the nucleus (Chen et al., 2002). Both Dnmt3a and Dnmt3a2 are capable of remethylating Dnmt3a, 3b(+/-) ES cells, however Dnmt3a2 is the main isoform expressed in ES cells, germ cells, and the early embryo; Dnmt3a is instead expressed at low levels in somatic tissues (Chen et al., 2002).

Dnmt3b has been reported to have at least six isoforms (Fig. 1.3), transcribed from the same promoter and produced by alternative splicing in both human and mouse cells (Chen et al., 2003b; Okano et al., 1998a; Xie et al., 1999). A study of the enzymatic properties of the Dnmt3b isoforms in vitro reveals that Dnmt3b(1) and (2) display similar levels of activity, whereas Dnmt3b(3) has no methyltransferase activity, due to the absence of a section of the catalytic methyltransferase domain (Aoki et al., 2001; Okano et al., 1998a). In agreement with this, Dnmt3b(1) is able to restore the methylation of the majority of sequences demethylated in Dnmt3a, 3b(+/-) ES cells, but Dnmt3b(3) is unable to restore any of the methylation patterns (Chen et al., 2003b). Dnmt3b(1) has been shown to be most strongly expressed in ES cells and germ cells, whereas Dnmt3b(3) is expressed at low levels in almost all somatic tissues and cell lines examined (Beaulieu et al., 2002; Chen et al., 2003b; Okano et al., 1998a). Dnmt3b(4) was thought to be inactive, because it also lacks some of the catalytic domain. However overexpression of this isoform in human hepatocellular carcinoma cells may lead to
hypomethylation of pericentromeric satellite regions, suggesting it may have an antagonistic role (Saito et al., 2002). Dnmt3b(5) and (6) are also presumed to be inactive because they lack some of the catalytic motifs, however their function has not been studied.

The levels of Dnmt3b transcripts in the cell are regulated at several levels. At the level of transcription, Dnmt3b mRNA levels are regulated by both DNA methylation and histone hypoacetylation of the Dnmt3b enhancer region; treatment of ES cells with retinoic acid leads to Dnmt3b down-regulation, but subsequent treatment with either the histone deacetylase inhibitor Trichostatin A (TSA) or the DNA methyltransferase inhibitor 5-aza-dC leads to epigenetic modification of the enhancer region and reactivation of the gene (Tanaka, 2005). Additionally, the Dnmt3b promoter is bound by both Nanog and Sall4 (Wang et al., 2006), although it is not clear if this contributes to the transcriptional control of Dnmt3b. Microarray analysis comparing undifferentiated human ES cells (hESCs) and hESCs differentiating into cardiomyocytes reveals that DNMT3B is downregulated 2x fold within 1 day of differentiation, by 5 fold after 3 days of differentiation, and by 17 fold after 12 days (Beqqali et al., 2006). DNMT3B is also downregulated during hESC differentiation into embryoid bodies (EBs) (Bhattacharya et al., 2005), but it is not known how this is regulated.

Dnmt3b mRNA levels are also regulated by affecting mRNA stability, at least in some cell types; studies in human endometrial adenocarcinoma Ishikawa cells (referred to as Ishikawa cells) have showed that expression of Dnmt3b is down-regulated upon exposure of the cells to the histone deacetylase inhibitor TSA (Xiong et al., 2005). Cloning of the Dnmt3b promoter into a luciferase expression vector demonstrated that TSA treatment does not affect the rate of transcription from the promoter. Instead the observed decrease in mRNA levels is shown to be due to decreased Dnmt3b mRNA stability. Treatment of these cells with the transcriptional inhibitor Actinomycin-D, followed by measurement of the Dnmt3b mRNA levels by qRT-PCR, demonstrated that treatment with TSA led to an increased rate of decay of this transcript, in a process dependent on protein synthesis. This down-regulation led to a measurable decrease in
the levels of de novo methylation activity and to demethylation of the minor satellite repeats (Xiong et al., 2005).

In this chapter I set out to investigate the cause of the DNA demethylation observed in Mbd3 null ES cells by measuring the levels of the known DNA methyltransferases. Evidence is presented that a loss of Mbd3 is associated with a significant reduction in the transcript and protein levels of the de novo methyltransferase Dnmt3b, but no other DNA methyltransferases. However this does not cause a significant reduction in the levels of de novo methylation activity, and is unlikely to account for all of the observed demethylation. Surprisingly, ectopic expression of Dnmt3b in Mbd3 null ES cells produces growth arrest and differentiation towards the endoderm lineage.

4.2 Results

4.2.1 Mbd3<sup>−/−</sup> cells have reduced levels of Dnmt3b mRNA
As a first step to determine the cause of the demethylation observed in Mbd3 null ES cells, levels of Dnmt1, 3a, 3b and 3l transcripts were measured by quantitative RT-PCR (qRT-PCR). This revealed that the levels of Dnmt3b transcript are reduced by almost 90% in Mbd3 null cells, and that this is broadly restored in Mbd3b-rescued cells (Fig. 4.1). In contrast, there is no change in the levels of Dnmt3a or 3l. Dnmt1 also showed no change in transcription, suggesting no defect in maintenance methylation. In agreement with previous experiments, the levels of Oct4 transcripts did not change in Mbd3 null cells (Kaji et al., 2006). cDNA from five independently produced RNA samples was used, and the results averaged. Values were normalised against those of the housekeeping gene Gapdh.

4.2.2 Dnmt3b(l) is the main isoform expressed in mES cells
As discussed above, Dnmt3b(l) has been previously shown to be the most active isoform, and to be expressed in ES cells. To assess the proportion of Dnmt3b(l) transcripts out of all the Dnmt3b transcripts in wild type ES cells, the levels of
Dnmt3b(1) and Dnmt3b(all isoforms) were measured. This revealed that >95% of Dnmt3b transcript detected in wild type ES cells was Dnmt3b(1) (Fig.4.2.13). This was done by qRT-PCR using primers solely recognising Dnmt3b(1) (i.e. located within the exon missing in the other splice variants), and primers recognising a region common to all the isoforms (Fig.4.2.A). Again 5 separate cDNA populations were used and average values normalised against Gapdh. To my knowledge this has never been shown previously at the transcript level.

4.2.3 Mbd3<sup>−/−</sup> cells have reduced levels of Dnmt3b protein

The level of Dnmt3b protein was also examined in the Mbd3<sup>fl</sup>, Mbd3<sup>−/−</sup> and Mbd3b rescue cells by Western Blotting. In agreement with the decrease in Dnmt3b transcript levels, there is a significant decrease in the levels of Dnmt3b protein in Mbd3<sup>−/−</sup> cells, which is rescued by Mbd3b (Fig.4.3). It is notable that there is more than one band recognised by the polyclonal Dnmt3b antibody (Fig.4.3), despite the qRT-PCR data suggesting only one isoform is expressed in ES cells (Fig.4.2.13). All the bands disappear in Mbd3<sup>−/−</sup> cells, and also in Dnmt3b<sup>−/−</sup> ES cells (Chen et al., 2003b), showing that they are specific. The strongest band is not the highest band, making it unlikely that bands above the strongest band are different isoforms, as Dnmt3b(1), which is clearly expressed most strongly by qRT-PCR, is larger than all the other isoforms. It is possible that this higher band represents a chemically modified version of Dnmt3b.

4.2.4 The decrease in Dnmt3b(1) transcript levels is not caused by decreased stability of the transcript in Mbd3<sup>−/−</sup> cells

As discussed above, hyperacetylated human cancer cells display a reduced steady-state level of Dnmt3b transcript due to increased degradation of the transcript (Xiong et al., 2005). On the basis that both TSA treatment and Mbd3 loss lead to histone hyperacetylation, I speculated that the cause of the decrease in Dnmt3b transcription in Mbd3<sup>−/−</sup> ES cells was also due to decreased mRNA stability. In order to test this, a similar experiment was carried out in Mbd3<sup>−/−</sup> ES cells. Parental, Mbd3 null and Mbd3b rescue cell lines were cultured in the presence of either 2.5 or 5μg/ml Actinomycin-D
and RNA extracted at 0, 0.5, 2 and 8hr timepoints. The RNA was quantified, and DNaseI treated before being reverse transcribed using M-MLV-RT. Dnmt3b transcript levels were measured using quantitative PCR. Similar results were obtained for both 2.5 and 5μg/ml actinomycin-D; results from experiments using 5μg/ml are shown here.

As can be seen from Fig.4.4, the rate of decay is very similar between all the cell types. There is certainly no increase in the rate of Dnmt3b transcript decay in Mbd3−/− cells, where the transcript half-life (T[1/2]) is approximately 45 mins. In fact it is slightly decreased compared to the parental cells (T[1/2] = 35 mins). This shows that the decreased level of Dnmt3b mRNA is not due to an increased rate of degradation, and is therefore most likely due to changes in the rate of transcription.

4.2.5 An in vitro assay reveals no detectable decrease in de novo methyltransferase activity in Mbd3−/− cells

Dnmt3b has been reported to have de novo methyltransferase activity in vitro (Aoki et al., 2001) and in vivo (Hsieh, 1999), and previous reports have demonstrated that cells with reduced levels of both Dnmt3a and Dnmt3b have measurably lower levels of de novo methylation activity (Okano et al., 1999; Xiong et al., 2005). Thus I decided to examine the levels of de novo methylation activity in Mbd3−/− ES cells using an in vitro assay that measures the transfer of methyl groups from S-adenosyl-L-methionine to an unmethylated CpG rich substrate on a solid matrix (Epigentek Inc, New York, USA). The positive control is a recombinant bacterial DNA methyltransferase. The amount of methyl groups transferred during the reaction is measured after the reaction is stopped in an ELISA-based reaction. This showed that there is no change in the levels of de novo methylation activity in Mbd3−/− ES cells, compared to parental or Mbd3b rescue cells. (Fig.4.5). The experiment was carried out using either 14, 17 or 20μg of nuclear extract from the relevant cell types; similar results were obtained each time.
4.2.6 Ectopic expression of Dnmt3b(1) in Mbd3\textsuperscript{(-/-)} ES cells

Dnmt3b is exclusively responsible for the maintenance methylation of the minor satellites (Okano et al., 1999), and contributes along with Dnmt3a towards the maintenance methylation of a number of other sequences including the major satellites, IAPs, and the H19/Igf2 5’ upstream DMR (Chen et al., 2003b; Liang et al., 2002), all of which are demethylated in Mbd3\textsuperscript{(-/-)} ES cells. Given the drastic reduction in Dnmt3b transcript and protein levels in these cells I decided to test whether ectopic expression of Dnmt3b(1) would rescue any of the methylation defect.

4.2.6.1 Ectopic expression of Dnmt3b(1) leads to growth arrest and differentiation in Mbd3\textsuperscript{(-/-)} ES cells but not Mbd3\textsuperscript{(-/x)} ES cells

A pCAG-Dnmt3b(1)-IRES-Zeo vector (Appendix I) was obtained (a gift from B.Ramsahoye, University of Edinburgh) and transfected into both Mbd3\textsuperscript{(-/-)} and Mbd3\textsuperscript{(-/-)} ES cells. After 2 days the cells were plated out and selected for Zeocin resistance. Ten days later it was immediately obvious that there were significantly more colonies from the transfection into Mbd3\textsuperscript{(-/-)} cells than into Mbd3\textsuperscript{(-/-)} cells. The colonies formed from transfection into Mbd3\textsuperscript{(-/-)} cells were small, and mostly looked quite differentiated (Fig. 4.6). Ten were picked into wells of a 96 well plate but none grew at all in the wells in the plate, suggesting either a growth defect and/or differentiation. In contrast, colonies formed from Mbd3\textsuperscript{(-/-)} cells were normal looking (Fig. 4.6) and could be expanded. The transfection was repeated twice to confirm the result. Three colonies from the transfection into Mbd3\textsuperscript{(-/-)} cells were expanded until sufficient cell numbers were obtained for the level of Dnmt3b overexpression to be assayed by western blotting. This revealed considerable overexpression of Dnmt3b(1) in these cells (Fig. 4.7) compared to endogenous levels of Dnmt3b in Mbd3\textsuperscript{(-/-)} cells.

In order to confirm the phenotype observed when Dnmt3b is overexpressed in an Mbd3 null background, the experiment was done 'in reverse' i.e. Mbd3 was deleted from Mbd3\textsuperscript{(-/-)}/Dnmt3b(1) overexpressing cells by the expression of Cre in these cells (see Fig. 1.6.B for Mbd3 targetting scheme). To achieve this, a pCAG-Cre-IRES-Puro vector
(Appendix II) was transfected into both these $Mbd3^{\text{floxed}}$/Dnmt3b(1) overexpressing (OE) (transgenic) cells, as well as non-transgenic $Mbd3^{\text{floxed}}$ as a control.

Two weeks after the transfection, the number of colonies on each plate was counted revealing that fewer were present following transfection into any of the three $Mbd3^{\text{floxed}}$/Dnmt3b OE cell lines than the $Mbd3^{\text{floxed}}$ cell line. $Mbd3^{\text{floxed}}$/Dnmt3b OE clone 1 in particular showed a significant reduction in the amount of colonies; clones 2 and 3 showed a smaller decrease from the parental cell lines (Table 4.1).

The colonies from the Dnmt3b overexpressing cells also looked abnormal and differentiated. A representative picture was taken of each cell line to demonstrate this (Fig.4.8). Alkaline phosphatase staining of plates was done to reveal the extent of differentiation. As can be seen from figures 4.9.A and 4.9.B, there was a considerable increase in the number of alkaline phosphatase mixed and negative colonies following transfection of Cre into $Mbd3^{\text{floxed}}$/Dnmt3b OE cells compared to $Mbd3^{\text{floxed}}$ non-transgenic cells, indicating that cells were differentiating. There was no clear correlation between the amount of Dnmt3b(1) overexpression and the degree of growth arrest or differentiation; for example $Mbd3^{\text{floxed}}$/Dnmt3b OE clone 2 shows the greatest expression of Dnmt3b by Western (Fig.4.6), but clone 1 shows the fewest number of colonies following the Cre transfection (Table 4.1).

Colonies from each transfection were picked in an attempt to expand them. Those from the $Mbd3^{\text{floxed}}$ cells grew normally, and PCR genotyping of the cells revealed the deletion of $Mbd3$ in all of the 8 clones grown and examined, demonstrating that the transfection worked with high efficiency (Fig. 4.10). Again, colonies formed from the three $Mbd3^{\text{floxed}}$/Dnmt3b OE cell lines transfected with Cre could not be expanded beyond the 96 well plate stage, so the excision of $Mbd3$ could not be confirmed in these cells. However the high efficiency of excision of $Mbd3$ from $Mbd3^{\text{floxed}}$ cells (8 out of 8 clones genotyped were $Mbd3^{+/-}$) suggests that $Mbd3$ would have been excised from a high proportion of the $Mbd3^{\text{floxed}}$ / Dnmt3b OE cells.
4.2.6.2 Increased Gata-4, and decreased Oct4 expression in Mbd3\(^{-/-}\) ES cells overexpressing Dnmt3b(1)

In an attempt to try and characterise these cells further, immunofluorescence was carried out on the cells in 96 well plates 17 days after transfection using antibodies against the ES cell marker Oct4 and the endoderm marker Gata-4. Pictures were taken at x20 magnification and care was taken to ensure that the exposure time was the same for each primary antibody on different samples. As can be seen from Fig. 4.11.A, a decrease in the number of Oct4 highly expressing cells is discernible in the cells overexpressing Dnmt3b (Mbd3\(^{-/-}\)/Dnmt3b OE) compared to the Mbd3\(^{-/-}\) cells. Even more striking is the increase in Gata-4 expression (Fig.4.11.B) in these cells. A significant proportion of Mbd3\(^{-/-}\)/Dnmt3b OE cells show Gata-4 staining, compared to virtually none in the Mbd3\(^{-/-}\) non-OE cells. Thus combining Dnmt3b overexpression with an Mbd3\(^{-/-}\) background appears to lead to growth arrest and differentiation towards the endoderm lineage. This is in contrast to the transfection of Dnmt3b(1) into Dnmt3a, 3b\(^{+/-}\) ES cells which does not result in any obvious growth or differentiation defect (Chen et al., 2003b).
Figure 4.1 Transcription of Dnmt3b is reduced in Mbd3<sup>−/−</sup> ES cells. Transcript levels of Dnmt1, Dnmt3a, Dnmt3b and Dnmt3l were measured by qPCR following reverse transcription (qRT-PCR). Oct4 was included as a control because its expression has been previously shown to not to change in the absence of Mbd3. Values with error bars represent the average of five independently generated cDNA pools, following normalisation to the housekeeping gene Gapdh.
Figure 4.2.A Schematic diagram of Dnmt3b(1) showing location of primers used to amplify Dnmt3b(1) specifically (1F/1R), and all Dnmt3b isoforms together (AllF/1R). The dotted region is present only in the Dnmt3b(1) isoform. The black boxes represent the conserved PWWP and PHD domains, and the diagonally striped boxes the catalytic methyltransferase motifs.

![Diagram of Dnmt3b(1)](image)

Expression of Dnmt3b isoforms in wild type ES cells

![Expression graph](image)

Figure 4.2.B Measurement of Dnmt3b(1) levels shows this constitutes almost all the Dnmt3b transcribed in wild type ES cells. qRT-PCR was carried out using primers either specific for Dnmt3b(1), or recognising a region common to all 6 Dnmt3b known isoforms. Values with error bars represent the average of three independently generated cDNAs, following normalisation to the housekeeping gene Gapdh.
Figure 4.3 *Mbd3*<sup>(-/-)</sup> ES cells have a reduced amount of Dnmt3b protein compared to *Mbd3*<sup>(flox/-)</sup> cells. 5μg of nuclear extract from the indicated cell types was run out on a 10% gel and probed with anti-Dnmt3b. Membranes were then stripped and re-probed for Hdacl to confirm equal loading of the lanes. The experiment was also repeated with 3 and 7μg of protein in each lane, with a similar result.

Figure 4.4 *Mbd3*<sup>(-/-)</sup> ES cells do not have an increased rate of Dnmt3b transcript turnover compared to *Mbd3*<sup>(flox/-)</sup> cells or Mbd3b rescue cells. Cells were treated with Actinomycin-D (both 2.5μg/ml and 5 μg/ml were tested, and gave similar results. The results for 5 μg/ml are shown here) for the indicated time periods, and then RNA extracted and the levels of *Dnmt3b* measured by qRT-PCR. Values represent the average from 2 independent experiments.
De novo DNA methylation activity

Figure 4.5 The level of de novo methyltransferase (MTase) activity in Mbd3<sup>−/−</sup> (null) ES cells is no less than in Mbd3<sup>fl<sub>ox</sub>−</sup> (WT) or Mbd3b rescue (Rescue) ES cells. Either 14, 17 or 20 μg of nuclear extract from the indicated cell types was used in an in vitro assay for de novo MTase activity. The amount of methyl groups transferred from SAM to an unmethylated CpG rich substrate was measured in an ELISA-based reaction, and calculated from the A<sub>450nm</sub> reading. The positive control was a bacterial MTase provided with the kit. Values shown are the average of two independent experiments.
Figure 4.6 Overexpression of Dnmt3b(1) in Mbd3\textsuperscript{flox-/-} cells produces normal looking colonies (A), whereas overexpression in Mbd3\textsuperscript{-/-} cells produces smaller, differentiated looking colonies. 14 days following transfection of Dnmt3b(1) into the indicated cell types pictures of the colonies were taken in bright field at x10 magnification. A representative colony from each cell line is shown.

Figure 4.7 Overexpression of Dnmt3b in 3 cell lines (1, 2 and 3) derived from transfection of pCMV-Dnmt3b(1)-IRES-Zeo into Mbd3\textsuperscript{flox-/-} ES cells. 10\(\mu\)g of nuclear extract from these 3 cell lines and from the parental line (Mbd3\textsuperscript{flox-/-}) was run out on a 10% gel and probed for Dnmt3b. An antibody against Hdac1 was also used to confirm equal loading of the lanes.
Table 4.1 Fewer colonies are produced following transfection with Cre into any of the three Mbd3\textsuperscript{floxed} /Dnmt3b overexpressing cell lines (Mbd3\textsuperscript{floxed} (1, 2 or 3)), than transfection into the parental cell line expressing endogenous levels of Dnmt3b. 2 weeks after transfection the number of colonies on each plate was counted by eye.

<table>
<thead>
<tr>
<th>Dnmt3b overexpressing</th>
<th>No. of colonies following transfection with Cre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mbd3b \textsuperscript{floxed}</td>
<td>50+</td>
</tr>
<tr>
<td>Mbd3b \textsuperscript{floxed} (1)</td>
<td>&lt;10</td>
</tr>
<tr>
<td>Mbd3b \textsuperscript{floxed} (2)</td>
<td>30-40</td>
</tr>
<tr>
<td>Mbd3b \textsuperscript{floxed} (3)</td>
<td>30-40</td>
</tr>
</tbody>
</table>

Figure 4.8 Excision of Mbd3 from any of three cell lines overexpressing Dnmt3b (1, 2 and 3 from Figure 6) (B, C, D) produces colonies that are smaller and more differentiated looking than when Mbd3 is excised from cells expressing endogenous levels of Dnmt3b (A). 14 days following transfection of Cre into the indicated cell types pictures of the colonies were taken in bright field at x10 magnification. A representative colony from each cell line is shown.
Figure 4.9.A Alkaline phosphatase staining shows that cells lacking Mbd3 and overexpressing Dnmt3b are extensively differentiated. Colonies formed from the transfection of Cre into either an Mbd3^{flox-} ES cell line, or any one of three Mbd3^{flox-} /Dnmt3b overexpressing cell lines (Mbd3^{flox-}/1, 2 or 3). 2 weeks after transfection, colonies were fixed and stained for alkaline phosphatase, and viewed at x10 resolution in bright field. A representative picture from each cell line is shown.

Alkaline Phosphatase Staining

Figure 4.9.B Significantly more alkaline phosphatase mixed and negative colonies were produced when Cre was transfected into Mbd3^{flox-} cells overexpressing Dnmt3b (Mbd3^{flox-}/1, 2 or 3), than when Cre is transfected into Mbd3^{flox-} cells expressing endogenous levels of Dnmt3b. Colonies were stained for alkaline phosphatase 2 weeks post-transfection, and the number of alkaline phosphatase (AP) positive, negative or mixed colonies counted by eye (n=60)
Figure 4.10 Genotyping by PCR of 8 colonies produced from the transfection of Cre into Mbd3\(^{\text{floxed}}\) cells confirms that Mbd3 was excised from every clone of cells. p36/p43 exclusively amplify the deleted allele, p36/46 amplify both the floxed allele and the WT allele. Please see Fig.5.1.B in Chapter One for Mbd3 targeting scheme. -: dH\(_2\)O control.
Figure 4.11 ES cells that lack Mbd3 but overexpress Dnmt3b (Mbd3(-/-)/Dnmt3b OE) show differentiation towards the endoderm lineage. Colonies produced from transfection of Cre into either Mbd3(-/-) cells either over-expressing Dnmt3b, or expressing endogenous levels, were picked into wells of a 96-well plate and stained for either Oct4 (A) or Gata-4 (B). Cells were also counter-stained with the nuclear marker DAPI. Pictures were taken at x20 resolution, and care was taken that the same exposure time was used for each antibody. At least 4 colonies were stained with each antibody, the pictures shown here are representative samples.
Discussion

I have shown that \emph{Mbd3\textsuperscript{(-/-)}} ES cells contain reduced amounts of the Dnmt3b DNA methyltransferase, however this is not associated with a detectable reduction in de novo methylation. The reduction in transcript levels is not due to decreased mRNA stability, and is instead presumably regulated at the transcriptional level. Attempts to over-express Dnmt3b in \emph{Mbd3\textsuperscript{(-/-)}} cells result in extensive differentiation and growth arrest.

4.2.7 \textit{The NuRD complex has a role in the control of Dnmt3b transcript levels}

It is clear that \emph{Mbd3\textsuperscript{(-/-)}} ES cells contain significantly reduced amounts of both Dnmt3b mRNA and protein levels (Figs. 4.1 and 4.3). By measuring the rate of decay of Dnmt3b transcripts following the halt of RNA synthesis in both parental and null cells I have determined that an increased decay rate of Dnmt3b transcripts is not responsible for the reduction in steady-state levels mRNA levels (Fig.4.4). This is in contrast to the situation in Ishikawa cells (a human leukaemia cell line) where treatment with trichostatin-A (TSA), a histone deacetylase inhibitor, leads to increased degradation of and hence a decrease in Dnmt3b mRNA steady-state levels (Xiong et al., 2005). Several possible explanations for the differences from the study by Xiong et al present themselves. They used human cancer cells, which may well have different mechanisms for the degradation of transcripts than murine ES cells. Additionally, ES cells have been demonstrated to express Dnmt3b(1) (Fig.4.2). Whilst the issue of which isoform is expressed in the Ishikawa cells used by Xiong et al is not addressed in that study, previous work has shown that the major isoform expressed in human cancerous cells is Dnmt3b(3) (Beaulieu et al., 2002). Dnmt3b(3) differs from Dnmt3b(1) in that it is missing 21 amino acids in the central portion of the transcript, and 64 in the C-terminus (Okano et al., 1998a). It is possible that these differences lead to a difference in the way they are degraded. Therefore the most likely explanation for the decrease in Dnmt3b levels observed here is a decrease in the rate of transcription of Dnmt3b.

It is also worth noticing the rapid turnover of Dnmt3b transcript in mouse ES cells (\(T_{1/2} = 35\) mins), compared to what was observed in human cancer cells (\(T_{1/2} = 3\) hrs). This
rate of decay of a transcript is not unprecedented, for example the T_{1/2} of TNFα is approximately 40 mins in wild type peritoneal macrophages (Piecyk et al., 2000), and 25 mins in T cells (Raghavan et al., 2002). Microarray analysis on T cells treated with Actinomycin-D for various time periods was used to analyse degradation rates of over 6000 transcripts in these cells; approximately 4% had half-lives of 45 mins or less (Raghavan et al., 2002). It has also been reported that there are significant differences between the degradation rate of β-Actin in two human leukaemia cell lines (6.6 hrs in Nalm-6 cells vs 13.5 hrs in CCRF-CEM cells) (Leclerc et al., 2002). Therefore, the increased rate of turnover of Dnmt3b in mES cells compared to Ishikawa cells likely reflects differences in the kinetics of degradation of Dnmt3b between the two cell types, or between the two different isoforms (Dnmt3b(1) and Dnmt3b(3)). To my knowledge, the rate of decay of any Dnmt3b isoform has not been studied previously in any cell type other than those in the study by Xiong et al, or mES cells.

If the cause of the decrease in Dnmt3b mRNA levels is not increased degradation, then a likely cause of the decrease is transcriptional down-regulation. All the Dnmt3b isoforms are expressed from the same promoter, on chromosome 2 (Chen et al., 2003b; Xie et al., 1999). A number of observations have been made regarding the control of Dnmt3b transcription. Treatment of ES cells with retinoic acid leads to Dnmt3b down-regulation. Subsequent treatment with either TSA or 5-aza-dC leads to reactivation of the gene, showing that both DNA methylation and histone hypoacetylation of the Dnmt3b enhancer region are involved in this silencing process (Tanaka, 2005). Additionally, the Dnmt3b promoter is bound by both Nanog and Sall4 (Wang et al., 2006). The NuRD complex has been shown to interact with Sall4 (Ita Costello, personal communication), so it is possible that NuRD directly regulates Dnmt3b transcription. However NuRD is a transcriptional repression complex, so if it does regulate Dnmt3b, it is more likely to be indirect than direct. Nuclear run-on assays could be done to establish if a reduced rate of transcription of Dnmt3b is indeed the cause of the reduced transcript level. It remains to be shown exactly how Dnmt3b transcription is regulated both epigenetically and by transcription factors.
4.2.8 A reduction in Dnmt3b protein levels cannot solely explain the demethylation observed in Mbd3\(^{-/-}\) ES cells

The significant decrease in Dnmt3b transcript and protein levels provides an intriguing and temptingly simple explanation for the demethylation observed in Mbd3\(^{-/-}\) ES cells, namely that a decrease in the levels of a DNA methyltransferase directly leads to a decrease in the levels of DNA methylation. This is particularly plausible at the minor satellite repeats because they have been shown to be exclusively methylated by Dnmt3b (Okano et al., 1999). However, apart from the satellite repeats, Dnmt3a and 3b seem to be redundant in ES cells; loss of either one does not lead to demethylation of any of the sequences observed to be demethylated in the double mutant cells. Thus it is unlikely that a reduction in Dnmt3b levels is solely responsible for the demethylation observed in Mbd3\(^{-/-}\) ES cells, other than at the minor satellites.

Despite the significant downregulation of Dnmt3b in these cells, no defect in de novo methylation activity was found in the absence of Mbd3. Whilst it is possible that the assay is not sufficiently sensitive to pick up subtle differences between the cell types, compensation by Dnmt3a is the most likely explanation for this; ES cells only display an in vivo defect in the de novo methylation of integrated proviral DNA when they lack both Dnmt3a and 3b; no difference in activity is seen in the individual mutants (Okano et al., 1999), presumably due to redundancy in the proteins. Because ES cells overexpressing Dnmt3b(1) in an Mbd3\(^{+/+}\) background cannot be sufficiently expanded to do the relevant assays, the question of whether the DNA demethylation seen in Mbd3\(^{+/+}\) ES cells can be restored by ectopic expression of Dnmt3b(1) in Mbd3\(^{-/-}\) ES cells cannot be answered.

4.2.9 Differentiation of Mbd3\(^{+/+}/\)Dnmt3b OE cells towards the endoderm lineage?

Interestingly, overexpression of Dnmt3b in Mbd3\(^{+/+}\) ES cells leads to growth arrest and a differentiated morphology; this was shown in both early and late passage Mbd3\(^{+/+}\) ES cells. Attempts to characterise these cells showed that they display increased Gata4 expression in some cells, and decreased Oct4 expression. The zinc finger containing
transcription factor Gata4 is expressed in the extraembryonic endoderm (ExE) lineages, primitive endoderm (PrE), and the PrE derivatives: visceral endoderm (VE) and parietal endoderm (PE) (Morrisey et al., 1996). Along with Gata6 it has a role in ExE formation; Gata4\(^{(-/-)}\) mice die between E8 – E9 due to defects in heart morphogenesis (Kuo et al., 1997; Molkentin et al., 1997). Expression of Gata4 is upregulated in response to an increase in Oct4 levels or LIF withdrawal, both of which lead to ExE formation (Niwa et al., 2000). Importantly, forced expression of Gata4 in undifferentiated ES cells induces differentiation into endoderm (Fujikura et al., 2002), resulting in a morphology that resembles that of parietal endoderm (PrE).

Thus Gata4, along with Gata6, has a pivotal role in the initiation of the endoderm differentiation program. Forced expression of Gata4 in ES cells (which do not normally express Gata4 or 6) also led to a decrease in Oct4 expression (Fujikura et al., 2002), and so the upregulation of Gata4, along with the downregulation of Oct4, that is seen here, suggests that the endoderm differentiation program has been initiated. However double-staining for both Gata4 and Oct4 was not carried out, and it is noticeable that most of the cells down-regulated Oct4, but not all the cells up-regulated Gata4. Furthermore, the morphology of the cells was not consistently clearly similar to any particular endoderm lineage. Thus it cannot be said for certain that all the cells are differentiating into endoderm lineages; it remains possible that cells are differentiating down other pathways as well. It will be interesting to do a closer analysis of these cells to look for expression of other markers to either confirm differentiation into endoderm, or to see if any other differentiation pathways have been induced within the population of cells.

A possible scenario explaining why overexpression of Dnmt3b in Mbd3 null ES cells leads to Gata4 up-regulation is one in which NuRD antagonises Dnmt3b action at certain sequences, and so in the absence of NuRD and an overload of Dnmt3b these sequences are susceptible to methylation by Dnmt3b. Transcription of Gata4 is known to be activated by binding of the transcription factor USF2 to a critical E box motif less than 100bp upstream from the start site (Mazaud Guittot et al., 2006). Tissue specific transcription of Gata4 is also regulated by binding of the Forkhead protein FOXF1 and
Gata4 itself to a distal enhancer element, in a process that requires BMP4 (Rojas et al., 2005). However, the role of epigenetic modifications in the expression of Gata4 has not been examined, and it is interesting to speculate that epigenetic alterations including perhaps aberrant DNA methylation lead to the inappropriate expression of this gene. Down-regulation of Oct4 may occur in response to the up-regulation of Gata4, as forced expression of Gata4 in ES cells leads to Oct4 down-regulation and this is postulated to be due to direct repression of the Oct4 gene (Fujikura et al., 2002).

Oct4 appears to be down-regulated in all Mbd3(-/-)/Dnmt3b OE cells, whereas Gata4 is upregulated in only a subset of cells. Therefore it remains possible that more direct effects are involved in the repression of Oct4 in cells that overexpress Dnmt3b in an Mbd3(-/-) background. One plausible explanation is again inappropriate epigenetic alterations at the Oct4 promoter in this genetic background. For example the Oct4 promoter is known to become methylated during the differentiation of ES cells as Oct4 is down-regulated (Deb-Rinker et al., 2005; Gidekel and Bergman, 2002; Hattori et al., 2004b; Tsuji-Takayama et al., 2004). This involves Dnmt3a and/or 3b, as the Oct4 promoter does not become methylated after 6 days of RA differentiation in Dnmt3a, 3b(-/-) ES cells, unlike in wild type cells (Gu et al., 2006). It is possible that genomic regions such as the Oct4 promoter are normally protected from methylation in a NuRD dependent manner, but in the absence of NuRD they become methylated, leading to transcriptional changes that result in growth arrest and differentiation. In this respect it would be interesting to see whether the Oct4 promoter becomes specifically methylated in the Mbd3(-/-)/Dnmt3b OE background.

It was recently observed that repression of the Oct4 promoter during ES cell differentiation involves binding of GCNF followed by the recruitment of Mbd3, leading to transcriptional silencing. Subsequently Mbd2 binds as well and the locus becomes methylated (Gu et al., 2006). Whilst it is not clear if Mbd3 recruits DNA MTase activity, siRNA mediated depletion of Mbd3 leads to a failure to methylate and silence the gene, suggesting that Mbd3 is required for methylation of the Oct4 promoter, and making this scenario less likely.
A further possibility is that ‘free’ Dnmt3b that is not bound by NuRD triggers signalling pathways leading to the growth arrest and/or differentiation of the cell. It is plausible that in $Mbd3^{(flox/-)}$ there is enough free NuRD to ‘mop-up’ the excess Dnmt3b, thus avoiding it being recognised by the cell. As discussed in Chapter 5, it is plausible that Dnmt3b is deacetylated by NuRD, and that this is necessary for its correct function. Therefore a final possibility is that large quantities of acetylated Dnmt3b, as would be found in $Mbd3^{-/-}$ cells, are deleterious for the cell, again perhaps by inappropriately activating signalling pathways, or by affecting the expression of key genes.

It is worth noting that the levels of expression of Dnmt3b were very high compared to endogenous levels (Fig.4.7), due to expression being driven by the strong CAG promoter rather than an endogenous promoter. Thus this is not a physiological situation and as such caution should be used in interpreting the results. In order to get a more biologically relevant result Dnmt3b(1) should be ectopically expressed, but from either its endogenous promoter, or at least one weaker than the CAG promoter. Clones with levels of Dnmt3b matching endogenous levels should then be analysed. However this does not help explain why overexpression of Dnmt3b(1) should lead to endoderm differentiation in an $Mbd3^{-/-}$ background, but not an $Mbd3^{(flox/-)}$ background.
Chapter Five

5 NuRD and Dnmt3b interact and co-localise in a subset of ES cells

5.1 Introduction

The NuRD complex and the DNA methyltransferase Dnmt3b both possess epigenetic modifying activities that are involved in the formation of a transcriptionally repressive state. I have shown that loss of a component of the NuRD complex leads to histone hyperacetylation and DNA demethylation at a number of sites, including some methylated either exclusively or in part by Dnmt3b. However the mechanistic basis for how the NuRD complex affects DNA methylation patterns remains elusive. DNA methylation and histone deacytetylation are strongly correlated, and several mechanistic links have been found between them, demonstrating that the two processes are inter-dependent.

For example methyl binding domain (MBD) proteins bind selectively to methylated DNA (Hendrich and Bird, 2000; Wade, 2001) and recruit histone deacyetylases (HDACs) (Bird and Wolff, 1999) and histone methyltransferases (HMTases) (Fuks et al., 2003b) to bring about or reinforce a condensed chromatin structure. Additionally, several of the Dnmts have also been shown to interact with HDACs to bring about transcriptional repression, although interestingly in every case the methyltransferase domain of the Dnmts is dispensable for transcriptional silencing (Deplus et al., 2002; Fuks et al., 2000; Fuks et al., 2001; Geiman et al., 2004b). Dnmt3b has also previously been reported to repress transcription in a histone deacytetylation dependent manner (Bachman et al., 2001). Evidence from several organisms suggests that histone deacytetylation precedes DNA methylation (Hu et al., 2000; Mutskov and Felsenfeld, 2004; Selker, 1998), however mechanistic evidence linking the two processes remains elusive, and it is not clear exactly how the two modifications interact and if either is dependent on the other.
Recruitment of DNA methyltransferase activity to DNA cannot rely on binding of specific DNA sequences by the Dnmts, because these do not have any intrinsic sequence specificity in vitro (Tajima and Suetake, 1998; Yoder et al., 1997a). There is some evidence to suggest that certain sequences are protected from de novo methylation; for example inhibition of the Aprt CpG island by Sp1 elements located in the island (Brandeis et al., 1994; Macleod et al., 1994), and protection of some imprinted sequences from demethylation during early embryogenesis by Stella/PGC7 (Nakamura et al., 2006). However both Dnmt1 and Dnmt3a have been demonstrated to be actively recruited to chromatin by sequence-specific transcriptional repressors, where they act as transcriptional co-repressors. Examples include recruitment of Dnmt1 by the Rb protein to Rb/E2F responsive genes (Robertson et al., 2000), and recruitment of Dnmt3a by RP58 to RP58 responsive genes (Fuks et al., 2001). Additionally, both Dnmt1 and Dnmt3a are recruited by the oncogenic transcription factor PML-RAR to the RARβ2 promoter, leading to silencing of the gene (Di Croce et al., 2002). Dnmt1 is also recruited to the replication fork via an interaction with the replication machinery component PCNA (Leonhardt et al., 1992), where it interacts with other epigenetic modifiers including histone deacetylases and histone methyltransferases to re-establish methylation patterns following replication of heterochromatin (Esteve et al., 2006; Rountree et al., 2000).

Dnmt3b has also been demonstrated to be targeted by binding to HP1, but this is not sequence specific but rather relies on underlying Me3H3K9 (Lehnertz et al., 2003); to my knowledge sequence specific recruitment of Dnmt3b has not been reported. Recruitment by HP1 results in localisation to centromeric regions, which possess high levels of Me3H3K9. This has been shown to be the case in a number of cell types including ES cells (Bachman et al., 2001; Lehnertz et al., 2003). This localisation pattern is dependent on the PWWP domain of Dnmt3b (Chen et al., 2004). DNMT3B has also been shown to bind to the satellite regions in human CCL256.1 lymphoblastoid cells (Geiman et al., 2004a).
Dnmt3b interacts with a variety of other proteins in a range of cell types, including a number of epigenetic modifiers, again demonstrating the interdependence of these processes. These include the chromatin remodelers hSNF2H and Lsh, HDACs 1 and 2, the HMTase Suv39h1, HP1, and the transcriptional co-repressor Sin3A (Bai et al., 2005; Geiman et al., 2004a; Geiman et al., 2004b; Zhu et al., 2006). DNMT3B also interacts with components of the mitotic chromosome condensation machinery (Geiman et al., 2004a). Baculovirus-expressed DNMT1, DNMT3A and DNMT3B can interact with each other in insect Sf9 cells, possibly linking DNMT3A and 3B to the replication fork (Kim et al., 2002). The only interactions confirmed so far in ES cells are with Lsh and HP1 (Lehnertz et al., 2003; Zhu et al., 2006), but it will be interesting to discover if other interactions are also conserved between cell types.

Despite possessing a methyl-binding domain containing protein (Mbd3), the NuRD complex is unable to bind methylated DNA, due to a mutation in the MBD of Mbd3 (Hendrich and Bird, 1998; Ohki et al., 2001; Saito and Ishikawa, 2002). Instead, NuRD is thought to be recruited to DNA indirectly by proteins such as Mbd2, which recruits the complex to methylated promoters (Zhang et al., 1999). It has also been shown to be recruited by a variety of transcription factors such as the Drosophila gene Hunchback (Kehle et al., 1998) and Ikaros, which has a role in the specification of the lymphocyte lineage (Kim et al., 1999). Other sequence-specific factors involved in recruiting NuRD to target sites include BCL11B which recruits the complex to BCL11B target genes in T lymphocytes (Cismasiu et al., 2005); FOG-1 which recruits it to FOG-1/GATA-1 target genes in erythroid cells (Hong et al., 2005), and BCL-6 which recruits the complex to BCL-6 and Snail target genes in B lymphocytes and breast cancer cells respectively (Fujita et al., 2004; Fujita et al., 2003). The NuRD complex has also been shown to bind to CpG islands upstream of both the p14ARF and p16INK4a genes in MCF7 breast carcinoma cells (Le Guezennec et al., 2006), and the Oct4 promoter via GCNF in differentiating ES cells (Gu et al., 2006). The Mi-2β component has also been shown to localise to heterochromatic centromeric regions in activated T cells by immunofluorescence (Kim et al., 1999).
An interaction between NuRD and APPL1, a RAB5 effector which translocates to the nucleus upon endocytosis of the appropriate extracellular stimulation, links transcriptional regulation by NuRD to extracellular signalling (Miaczynska et al., 2004). MBD3 also interacts with Aurora-A kinases, leading to its phosphorylation and localisation to the centrosomes in early M phase in HeLa cells (Sakai et al., 2002), perhaps indicating an as yet unrecognised function of NuRD during mitosis. Interestingly, NuRD components have been demonstrated to interact with DNA methyltransferases previously; with Dnmt1 in 293 cells (Tatematsu et al., 2000) and Dnmt3a in mouse lymphosarcoma cells (Datta et al., 2005) although in neither case was it addressed whether this binding was necessary for Dnmt function. Interestingly, the interaction with Dnmt1 also involved binding to hemi-methylated DNA, and thus it is suggested that there is a role for the NuRD complex in the re-establishment of a heterochromatic structure at the replication fork. In support of this, in these cells the abundance of the NuRD complex components was greatest towards the end of S phase (Tatematsu et al., 2000).

In this chapter I set out to establish whether the NuRD complex functions by binding directly to its target sites, and whether it is associates with DNA methyltransferase activity in ES cells. Evidence is presented that the complex localises to at least some of its target sites, the centromeric satellite DNA, in a subset of cells and in an Mbd3 dependent manner, and that there is a direct interaction with at least one DNA methyltransferase, Dnmt3b.

5.2 Results

5.2.1 NuRD binding is not detectable at any of the demethylated, hyperacetylated sequences by chromatin immunoprecipitation

Because the sequences that are demethylated in Mbd3^{-/-} ES cells are also hyperacetylated, and the NuRD complex contains histone deacetylase activity, I asked whether this was a direct effect by looking to see if binding of the NuRD complex to any
of the demethylated regions could be detected by chromatin immunoprecipitation in wild
type ES cells.

5.2.1.1 Biotinylation strategy

Biotin is a naturally occurring cofactor for metabolic enzymes, which is active only
when covalently attached to the enzymes through the action of specific protein–biotin
ligases (Chapman-Smith and Cronan, 1999). Proteins possessing a specific 23 amino
acid peptide tag are biotinylated at this region in vivo if they are expressed in cells also
expressing the protein-biotin bacterial ligase BirA. Biotinylated proteins have a very
strong affinity for avidin/streptavidin and as such can be purified under stringent
conditions with high efficiency and purity from mammalian cells in a single step
purification (de Boer et al., 2003). Very few proteins are naturally biotinylated, so there
is very little chance of cross-reactivity with other proteins.

In order to examine interacting partners and genomic binding sites of the NuRD
complex, ES cells expressing BirA under the control of the Rosa locus were obtained
(Gift from D. Meijer). Two NuRD components, Mbd3b and Mi-2β, had the 23aa peptide
tag added to their N-termini, and the tagged sequences cloned into a pCAG-IRESPuro
vector before being separately stably transfected into the BirA expressing cell lines (Ita
Costello, unpublished results). The biotin-tagged Mbd3 (Bio-Mbd3) and Mi-2β (Bio-
Mi-2β) cell lines were used in immunoprecipitation and/or chromatin
immunoprecipitation experiments. In each experiment, the BirA expressing cells
containing no biotin tagged NuRD components were used as a control to check for
background binding to the streptavidin used to capture the tagged proteins.

In order to assess whether the NuRD complex actually bound to the regions shown to be
demethylated and hyperacetylated in Mbd3<sup>−/−</sup> ES cells, chromatin immunoprecipitation
(ChIP) of biotin-tagged Mbd3b and Mi-2β was carried out.
As discussed above, streptavidin interacts strongly and specifically with biotin tags. This allows high stringency conditions to be used for ChIP, thus reducing the high background commonly associated with this process (Viens et al., 2004). For each sample, 50μg of chromatin from either the Bio-Mbd3b, Bio-Mi2β, or BirA only cell lines was used in an immunoprecipitation experiment with streptavidin coated magnetic beads. Precipitated material was analysed by quantitative real-time PCR using primers amplifying the same regions that were shown to be demethylated and hyperacetylated. Values were normalised against the input samples and represent the averages of three independent experiments. As can be seen from Fig.5.1, none of these regions showed any enrichment in binding of either Bio-Mbd3b or Bio-Mi-2β, compared to the BirA only line, indicating either that NuRD does not directly act at these sequences, or does so but the interaction is not sufficiently stable to detect using chromatin immunoprecipitation.

Unexpectedly, Bio-Mi-2β bound reproducibly to the actin promoter, which was initially included as a control; I did not expect it to be bound by the NuRD complex because it shows no enrichment in acetylation or change in transcription in the absence of Mbd3 (K.Kaji, unpublished observations). It is not clear why Mi-2β, but not Mbd3, was found to bind to this region. Mi-2β has been reported to exist in complexes other than the NuRD complex (Shimono et al., 2005; Shimono et al., 2003), and presumably it is binding in such a capacity here. Whatever the reason for binding of Mi-2β to the actin promoter, it nonetheless provides a positive control for the detection of the binding of biotinylated proteins to chromatin. Unfortunately due to the highly tandemly repetitive nature of the satellite repeats, reliable quantification by real-time PCR is only possible with machines calibrated in a certain way (Roche representative, personal communication), and a machine able to do this was not available for use when this experiment was done. Thus binding of NuRD components to the satellite repeats could not be assayed in this way.

5.2.2 Immunofluorescence reveals localisation of the NuRD complex to centromeric
regions in a subset of wild type cells in a Mbd3 dependent manner

One possible reason for NuRD not being detected at any of the affected sequences is transient binding, perhaps in a cell-cycle dependent manner. In order to explore this possibility, and also to ask whether localisation to the satellite regions could be observed, immunofluorescence of NuRD components was carried out in Mbd3 parental and null cells.

5.2.2.1 Mta2 and Mbd3 localise to centromeric regions in a subset of wild type cells

Close inspection of Mbd3\textsuperscript{flox/\textminus} ES cells stained for the NuRD component Mta2 and the heterochromatin marker 4'6-Diamidino-2-phenylindole (DAPI) revealed that in a subset of cells Mta2 staining overlapped with the DAPI bright spots, which mark the centromeric satellite repeat regions (Fig.5.2.A). The remaining cells mostly showed diffuse staining, sometimes with small punctate foci that did not correspond to the DAPI bright spots. Visual inspection of several fields of cells revealed that approximately 30% showed overlap with the DAPI bright regions (n = 120). In Mbd3\textsuperscript{+/\textminus} ES cells however there was not a single clear case of overlap with DAPI bright regions, indicating that Mta2 does not bind to these regions outside of the context of NuRD.

Staining of wild type ES cells with Mbd3 also reveals overlap with DAPI bright spots in a subset of cells (Fig.5.2.B). Careful visual inspection shows that the proportion of cells displaying this overlap is again approximately 30% (n = 120). Mi-2\(\beta\) has also previously been shown to clearly overlap with centromeric regions in a similar subset of cells (Powell, Costello, Back and Hendrich, manuscript in preparation). In contrast to my findings, immunoﬂuorescence of the NuRD components Gata2da/2db (previously known as p66/p68) has been shown to result in a speckled nuclear distribution that did not correspond to centromeric regions (Brackertz et al., 2006), however this study looked at overexpressed proteins in NIH3T3 cells, as opposed to the localisation of endogenous proteins in ES cells.
It has not yet been investigated why the NuRD complex localises to centromeric regions in only a subset of cells, although a dependence on the cell cycle seems likely. Whatever the reason, this may explain the failure to detect binding of NuRD components to demethylated, hyperacetylated regions by chromatin immunoprecipitation; it is feasible that NuRD is bound to these regions in insufficient cells at any one time to be detected in a conventional ChIP assay.

5.2.3 The NuRD complex interacts with Dnmt3b

Based on the observations that a) the minor satellite repeats, which are exclusively methylated by Dnmt3b (Okano et al., 1999), are demethylated in Mbd3 ES cells, and b) both NuRD and Dnmt3b have been found, or previously shown, to localise to the centromeric satellite regions (Fig. 5.2) (Bachman et al., 2001; Lehnertz et al., 2003), I looked to see if the NuRD complex interacts with Dnmt3b in wild type ES cells.

5.2.3.1 Dnmt3b interacts with biotin-tagged Mbd3b

Immunoprecipitation of nuclear extract from the Bio-Mbd3 cells using streptavidin-coated magnetic beads reveals that Dnmt3b interacts with Bio-Mbd3 (Fig. 5.3.A). Nuclear extract from the cell line expressing BirA only, and no tagged Mbd3, was used as a control and reveals that the interaction is specifically with the biotin-tagged Mbd3. It is noteworthy that there are clearly two bands in the input lane, and that Mbd3b only interacts with one of them. It is possible that the higher band represents a chemically modified version of Dnmt3b, as by qRT-PCR there is only one isoform, the largest one, expressed in wild type ES cells (Fig.4.2).

5.2.3.2 Mta2 and Mbd3 interact with Dnmt3b

In order to confirm the interaction the experiment was repeated using antibodies against endogenous proteins. An antibody against Dnmt3b was used to immunoprecipitate material from parental (Mbd3^{flox/−}), null (Mbd3^{−/−}) and rescue (Mbd3 rescue) cell lines, and the material probed for the presence of Mta2; this was found to clearly interact with
Dnmt3b in both parental and rescue cell lines (Fig.5.3.B). The interaction was significantly reduced in the null cells, presumably due to the reduced amount of Dnmt3b in these cells (Figs.4.1 and 4.3). Mbd3 also interacts with Dnmt3b (Fig.5.3.C), making it likely that the interaction with Mta2 is in the context of the NuRD complex. The three bands in the input lane represent the three isoforms of Mbd3 expressed in ES cells; Mbd3a, b and c. It is evident that Dnmt3b interacts specifically with Mbd3b and c but not Mbd3a.
Figure 5.1 Neither biotin-tagged Mbd3b nor Mi-2β bind to any of the affected regions in a chromatin immunoprecipitation assay. Chromatin from cell lines expressing either biotin-tagged Mbd3b (Bio-Mbd3b) (black bars), biotin-tagged Mi-2β (Bio-Mi2B) (diagonally striped bars), or no biotin-tagged proteins (BirA) (white bars) was used in immunoprecipitation experiments with an antibody against Streptavidin. Precipitated material was analysed by real-time PCR, and normalised against the Input samples. Values represent the average of 3 independent biological experiments, and 3 technical replicates.
Figure 5.2 NuRD components localise to centromeric regions in a subset of wild type ES cells in a NuRD-dependent manner. A) Immunofluorescence of the NuRD component Mta2 in both parental (Mbd3\textsuperscript{flx/flx}) and null (Mbd3\textsuperscript{+/-}) ES cells. In some (approximately 30%, n = 120) of the parental cells Mta2 staining overlaps with the centromeres, marked by the DAPI bright spots. A representative cell is marked with an asterisk. This staining pattern is lost in the null cells. B) Immunofluorescence of the NuRD component Mbd3 in parental cells. Again overlap with DAPI bright spots can be seen in approximately 30% of cells (n = 120), and a representative cell in each field is marked with an asterisk. The same result is seen using either Alexa-488 (green) or Alexa-680 (red)-conjugated fluorophores. Images were taken using a Leica confocal microscope.
Figure 5.3 Dnmt3b interacts with the NuRD complex. A) Nuclear extract from an ES cell line expressing biotin-tagged Mbd3b (Biotin-Mbd3b) was used in an immunoprecipitation with Streptavidin, and probed for Dnmt3b. BirA is the parental cell line expressing no bio-tagged proteins. B) Nuclear extract from either parental (Mbd3\(^{\text{null}}\)), null Mbd3\(^{(-/-)}\) and rescue (Mbd3b rescue) cell lines was used in an immunoprecipitation with an antibody against Dnmt3b, and probed for the NuRD component Mta2. C) Nuclear extract from the parental cell line was used in another immunoprecipitation with the Dnmt3b antibody, and the material probed for Mbd3. Mock: serum only. In all cases Input = 10% of protein used in IP experiment. Each experiment was repeated at least twice.
5.3 Discussion

I have found that in a subset of wild type ES cells NuRD components localise to the satellite regions which are demethylated in $Mbd3^{-/-}$ ES cells, and that there is a direct interaction between NuRD and Dnmt3b. This is likely to be an important feature of DNA methylation by Dnmt3b, because without an intact NuRD complex cells become extensively demethylated.

5.3.1 NuRD binds to centromeric regions in only a subset of wild type ES cells

The finding that no enrichment of NuRD could be detected at any of the demethylated sequences examined by ChIP was surprising, because it was thought likely that the hyperacetylation of these regions was due to a lack of histone deacetylase activity that is usually imparted by the NuRD complex. However this can be explained by the observation of binding to one of the target sites, the satellites, in only a subset of wild type ES cells, in a NuRD dependent manner. It is feasible that there is an insufficient number of cells with NuRD bound at these target sites at any one time to be detected in a ChIP assay of a population of cells. Therefore it is possible that in wild type ES cells NuRD may bind to the regions that are epigenetically altered in $Mbd3$ null cells, although it is formally possible that this is restricted to the centromeric satellite regions.

This leads to the possibility that NuRD is recruited to target sites in a subset of cells to help establish a heterochromatic structure. It is not known why centromeric localisation is found only in a subset of cells, but one plausible explanation is that binding is cell cycle dependent. In support of this, the abundance of NuRD components has previously been demonstrated to reach a peak during late S phase, when heterochromatin is replicated (Tatematsu et al., 2000). However it is not known if this is also true for ES cells. ChIP and immunofluorescence with a population of ES cells that had been synchronised would be necessary to establish if the cell-cycle stage can affect binding or localisation of components of the NuRD complex to target sites.
5.3.2 *NuRD is involved in the maintenance of DNA methylation patterns by Dnmt3b*

The loss of DNA methylation from regions that are constitutively methylated in ES cells (e.g. the satellite repeats and the IAP LTRs), as well as the failure to detect a reduction in *de novo* methylation activity in *Mbd3*<sup>−/−</sup> ES cells suggests that the absence of *Mbd3* leads to a defect in the maintenance of DNA methylation of these regions. Given that i) components of the NuRD complex interact with Dnmt3b; ii) they co-localise at centromeric regions in ES cells at least some of the time, and iii) loss of NuRD function leads to demethylation of sequences including some known to have their methylation maintained exclusively, or at least partly by, Dnmt3b, it is feasible that the NuRD complex is directly involved in the maintenance of DNA methylation patterns by Dnmt3b. An interaction with Dnmt3b was initially looked for because we observed demethylation of the minor satellite DNA, which is exclusively methylated by Dnmt3b (Okano et al., 1999). A suitable Dnmt3a antibody was not available to me at the time these studies were done, however given the demethylation of other sequences known to have a contribution from Dnmt3a towards the maintenance of their methylation it is feasible that the NuRD complex also interacts with Dnmt3a; it will be interesting to pursue this in future. Indeed, an interaction of Mbd3 with Dnmt3a has been demonstrated previously, although it is not clear if other NuRD components also interacted with Dnmt3a, and it is not known if this interaction is also found in ES cells (Datta et al., 2005). Dnmt3a and Dnmt3b have also been shown to interact with each other in ES cells (Ueda et al., 2006), supporting the hypothesis that NuRD components could be found in a complex with Dnmt3a as well as Dnmt3b.

It remains unclear what the mechanistic basis is for the involvement of the NuRD complex in DNA methylation, but several possibilities present themselves. For example the NuRD complex could be required for successful recruitment of DNA methyltransferase activity to target sites. It would be interesting to see if the remaining Dnmt3b and/or Dnmt3a proteins become mis-localised in *Mbd3*<sup>−/−</sup> ES cells. The Dnmts do not appear to have any intrinsic target specificity other than for CpG dinucleotides (Yoder et al., 1997a), and so it is assumed they are recruited by other protein factors to
their target sites. One argument against a recruitment model is that the demethylation observed in Mbd3$^{(-/-)}$ ES cells is less severe than that observed in Dnmt3a, 3b$^{(-/-)}$ ES cells (Chen et al., 2003b). However it is likely that multiple mechanisms and factors are used to recruit the Dnmts to different regions of the genome, and that the NuRD complex is only found at some of these loci. Therefore loss of the NuRD complex would lead to a failure of Dnmt recruitment only at a subset of Dnmt target sites.

A further interesting possibility is that a role for the NuRD complex in specifying DNA methylation patterns may involve an underlying modification to the chromatin that is a pre-requisite for DNA methylation. Whilst convincing biochemical evidence for this is lacking, there is plenty of anecdotal evidence to suggest that histone deacetylation precedes DNA methylation (Bachman et al., 2003; Heard, 2004; Keohane et al., 1996; Mutskov and Felsenfeld, 2004). There is also some evidence to suggest that histone acetylation patterns can direct DNA methylation patterns in *N. crassa* (Selker, 1998), and perhaps in mammals (Hu et al., 2000). Thus it is possible that acetylated chromatin is refractory to DNA methylation, and so hyperacetylation due to a lack of NuRD function could prevent the DNA from being methylated. A number of other epigenetic factors are associated with DNA demethylation in their absence e.g. *Lsh*, (Zhu et al., 2006), however only in a few cases has it been shown that this is a direct effect caused by a change to the epigenetic status of the underlying chromatin structure (e.g. (Lehnertz et al., 2003)). To study this hypothesis further it would be informative to follow the temporal order of events such as histone hyperacetylation, DNA demethylation and gene mis-regulation following the loss of NuRD function. It is not clear how NuRD would be differentially recruited to different sites; presumably this would involve a combination of factors. Understanding what is the initiating event in the targeting of different epigenetic modifications to different genomic regions will be an area of great interest in the future.

A further scenario worth considering for how NuRD could affect DNA methylation by Dnmt3b is chemical modification of Dnmt3b by the NuRD complex. It is noticeable from western blots probed with anti-Dnmt3b that there is at least one band in addition to
the strongest band (Figs. 4.3 and 5.3.A), something that has been observed previously (Chen et al., 2003b). All the bands disappear in Dnmt3b−/− ES cells, demonstrating that they are specific for Dnmt3b (Chen et al., 2003b). In my hands, this extra band is above the strongest band (Figs. 4.3 and 5.4.A), and therefore represents a larger protein. This is unlikely to represent a different isoform, because I have shown that in ES cells the main isoform of Dnmt3b that is expressed, and therefore the strongest band on a protein gel, is the largest isoform Dnmb3b(1) (Fig. 4.2.B).

Additionally, when an immunoprecipitation is carried out with nuclear extract from ES cells carrying biotin-tagged Mbd3b, and the immunoprecipitated material examined for Dnmt3b, there is clearly only one band that interacts with the NuRD complex, and this is the stronger, lower band (Fig. 5.3.A). This raises the interesting possibility that the higher band is a modified version of Dnmt3b, and that NuRD interacts specifically with the smaller version. In particular, it is possible that NuRD deacetylates Dnmt3b; histone deacetylases in fact target non-histone proteins as well as histones (Glozak et al., 2005). An example of this is the deacetylation of p53 by NuRD, which modulates the activity of p53 (Luo and Su, 2000). Equally, NuRD could interact with Dnmt3b that is specifically modified in some other way. To my knowledge, the modification status of Dnmt3b has not been studied; it would be interesting to investigate whether it is modified, particularly by acetylation, and whether this has any bearing on activity.

5.3.3 A role for NuRD in re-establishing a heterochromatic structure following DNA replication?

Dnmt1 is not solely responsible for maintenance methylation in ES cells; Dnmt3a and Dnmt3b have been shown to contribute to the maintenance methylation of a number of sequences including all those demonstrated to be demethylated in Mbd3(−/−) ES cells (Chen et al., 2003b; Liang et al., 2002). It is postulated that to achieve this they act directly at the replication fork, or just after it has passed (Chen et al., 2003b), although evidence for this is currently sparse, being based on the interaction of DNMT3A and 3B with DNMT1 (Kim et al., 2002). Given that the NuRD complex and Dnmt3b interact, and both have been shown, either in this study or previously, to localise to centromeric
regions in ES cells at least some of the time, it is interesting to speculate that NuRD and Dnmt3b interact at target sites in S phase. This may be necessary for the remethylation of sequences that are incompletely methylated by Dnmt1 following DNA replication, and for the re-establishment of a repressive chromatin structure. There may even be a direct link to replication itself, because the NuRD complex has previously been linked to hemi-methylated DNA and the replication machinery (Tatematsu et al., 2000). Co-localisation of NuRD components with the replication machinery marker PCNA was looked at as part of this study, but the preliminary results were inconclusive. It will be interesting to pursue this to see if the NuRD complex is indeed ever associated with the replication machinery.
Chapter Six

6 General Discussion and Future Directions

6.1 The NuRD complex and DNA methylation

I have shown that the loss of a central component of the NuRD complex causes a significant degree of demethylation of DNA at both repetitive and single-copy sites throughout the genome. Because all of the sites examined are methylated in wild type ES cells, this suggests a defect in the maintenance of DNA methylation. The finding that the NuRD complex interacts with the DNA methyltransferase Dnmt3b, and co-localises with it to some target loci in a subset of wild type ES cells in a NuRD-dependent manner, suggests that the complex may be involved in the methylation of DNA by Dnmt3b. However it remains unclear what the mechanistic basis for this is. It will be interesting in the future to conditionally knock-out \textit{Mbd3} and to follow the temporal order of events, such as demethylation of DNA, histone hyperacetylation and gene silencing, that follows.

This would help to answer questions such as whether the observed demethylation is a passive process occurring over many cellular passages, or whether it happens more rapidly than this. Given that demethylation of \textit{Dnmt3a, 3b} ES cells occurs slowly, not reaching a minimum until after approximately 50-75 passages (Chen et al., 2003b; Jackson et al., 2004), an equally slow rate of demethylation would support the hypothesis that the NuRD complex is involved in the maintenance of methylation by Dnmt3a and/or Dnmt3b. Conversely, a faster rate of demethylation would suggest a more active role in DNA methylation, perhaps involving the primary maintenance methyltransferase Dnmt1 as well as Dnmt3a/3b. Either way, a tempting possibility for the mechanism is that the NuRD complex is directly linked to the DNA replication machinery and is involved in the re-establishment of a heterochromatic structure following the passing of the replication fork at NuRD target sites. Dnmt1 is known to act
in this way (Leonhardt et al., 1992), and Dnmt3a/Dnmt3b have been suggested to also act in this way to help maintain DNA methylation at certain types of sequence, although direct evidence for this is lacking (Chen et al., 2003b). In support of this, Mbd3 has previously been reported to interact with Dnmt1 and hemi-methylated DNA, suggesting localisation to replicating DNA (Tatematsu et al., 2000). It is possible that NuRD is required to either recruit the Dnmts to replicating/newly replicated DNA, or to modify the replicated DNA e.g. by histone deacetylation or chromatin remodelling, to provide a suitable template for the Dnmts to act on; newly replicated DNA contains acetylated histones regardless of the acetylation status prior to replication (Sobel et al., 1995). Immunofluorescence to see if the NuRD complex co-localises with components of the replication machinery at any stage would help to answer this question. Equally, chemical modification of Dnmt3b by NuRD may affect its activity.

Histone deacetylation has previously been shown to occur prior to DNA methylation during the silencing of a transgene (Mutskov and Felsenfeld, 2004), however DNA methylation has also been shown to recruit histone deacetylase activity (Bird and Wolffe, 1999) and it remains unclear whether histone deacetylation is a primary event in the de novo formation and/or maintenance of a heterochromatic state or is instead used to 'lock in' a repressive state marked by DNA methylation. It would be informative here to see if histone deacetylation precedes DNA methylation or vice-versa, and whether the process varies between different loci.

Establishing the timing of transcriptional silencing would also be informative. Although transcriptional states are responsive to histone modifications and DNA methylation states (Berger, 2002; Razin and Cedar, 1991), it has been shown in a variety of situations, including silencing of a newly integrated transgene and silencing of genes on the inactive X during X inactivation, that transcriptional silencing precedes histone modifications and/or DNA methylation (Gautsch and Wilson, 1983; Keohane et al., 1996; Lock et al., 1987; Niwa et al., 1983; Wutz and Jaenisch, 2000). Therefore, given the observed transcriptional changes in Mbd3 null ES cells, it is possible that they are causative rather than reflective of the associated epigenetic changes. However
microarray analysis of Mbd3 null ES cells (K. Kaji, personal communication) shows that a larger-than-predicted number of genes (6% rather than 0.1% of genes recorded as present on the microarray) that are misexpressed in Mbd3 null ES cells are imprinted, and thus known or likely to be regulated by DNA methylation (Ferguson-Smith and Surani, 2001; Li et al., 1993). This suggests that changes to the epigenetic status leading to changes in the transcriptional status of misexpressed genes is an unexpectedly common occurrence in Mbd3 null ES cells. Determining whether silencing occurs prior to or after the epigenetic changes would help to answer this question.

6.2 A role in de novo methylation?

The evidence presented demonstrates a defect in the maintenance of DNA methylation; sequences which are already methylated lose methylation upon loss of Mbd3. Dnmt3b has been shown to be involved in the maintenance of methylation of all the sequences demethylated in Mbd3(-/-) ES cells and so this is a feasible scenario (Chen et al., 2003b). However it is possible that there is also an effect on de novo methylation, because Dnmt3b clearly also has a role in this process (Aoki et al., 2001; Hsieh, 1999; Okano et al., 1998a). It was recently shown that the binding of Mbd3 is required for the initiation of silencing of the Oct4 promoter in differentiating ES cells, and that it is recruited by GCNF to do this (Gu et al., 2006). Subsequent to Mbd3 binding, the locus becomes methylated in a Dnmt3a and/or Dnmt3b dependent process and Mbd2 binds, leading to stable and heritable silencing. This is an example of how the NuRD complex could be involved in the initiation of de novo methylation at a gene promoter, although it does not help to define what the role of the NuRD complex is in the process. Arguing against a defect in de novo methylation is the finding that there is no measurable decrease in de novo methylation activity in Mbd3(-/-) ES cells in an in vitro assay (Fig.3.5). This is ascribed to redundancy with Dnmt3a, but it will be interesting in the future to explore whether this is truly the case, or whether a more sensitive assay such as the methylation analysis of unmethylated proviral or episomal DNA integrated into either Mbd3^flox/- or Mbd3(-/-) ES cells reveals any defects in de novo methylation in the absence of Mbd3.
6.3 DNA methylation, NuRD, and the differentiation of pluripotent cells

It has been established previously that whilst depleting ES cells of DNA methylation has no noticeable effect on the self-renewal capabilities of these cells (Lei et al., 1996; Tsumura et al., 2006), the majority of these cells cannot terminally differentiate when signalled to do so (Carlone et al., 2005; Jackson et al., 2004). Instead, highly passaged ES cells lacking both Dnmt3a and Dnmt3b, and early passage ES cells lacking Dnmt1, continue to proliferate and maintain stem cell characteristics such as expression of the pluripotency markers Oct4 and alkaline phosphatase (Jackson et al., 2004). Dnmt3a, 3b−/− ES cells are also incapable of forming teratomas in vivo, again suggesting a differentiation defect (Chen et al., 2003b). Interestingly, in contrast to the majority of Dnmt3a, 3b−/− demethylated ES cells, a small proportion of these cells can and do spontaneously differentiate into extraembryonic lineages in the presence of LIF. Embryoid bodies lacking both Dnmt3a and Dnmt3b display a spectrum of mRNA expression defects when LIF is removed, in particular displaying a tendency to differentiate towards extraembryonic rather than embryonic lineages, and failing to commit to any particular developmental lineages (Jackson et al., 2004). Restoration of DNA methylation levels via the reintroduction of the deleted Dnmt3a and Dnmt3b methyltransferases rescues these defects, demonstrating that the affected regions maintain sufficient identity to be recognised for remethylation, and that the observed phenotype is due to a lack of DNA methylation and not prolonged passage (Jackson et al., 2004).

It is not understood exactly how a lack of DNA methylation prevents differentiation from occurring correctly, but it likely involves the transcriptional misregulation of both pluripotency-associated and lineage-specific genes, and presumably involves a failure of affected cells to exclusively express the appropriate genes to allow commitment to any particular developmental pathway. Interestingly, a significant proportion (16% of those studied) of CpG islands have been shown to be differentially methylated in ES cells, embryonic germ (EG) cells, trophoblast stem (TS) cells, sperm and somatic cells,
suggesting that differential methylation of CpG islands may be involved in the specification of cell fate (Shiota et al., 2002). It has also been shown that DNA methylation of the promoters of the pluripotency determinants Oct4 and Nanog occurs during differentiation, and for Oct4 at least this is correlated with transcriptional down-regulation (Deb-Rinker et al., 2005; Feldman et al., 2006; Gidekel and Bergman, 2002; Hattori et al., 2004b; Simonsson and Gurdon, 2004; Tsuji-Takayama et al., 2004).

Mbd3−/− ES cells are significantly demethylated and whilst they have no significant defect in self-renewal, they also display differentiation defects. For example they fail to differentiate into embryoid bodies in the absence of LIF, instead continuing to proliferate and maintain expression of the pluripotency markers Oct4, Nanog, Rex1 and alkaline phosphatase, and failing to express almost all the markers of differentiated embryonic cells tested. This suggests they are unable to switch from a pluripotency-associated gene expression pattern to one associated with differentiation (Kaji et al., 2006). In neither Mbd3−/− nor Dnmt3a, 3b−/− demethylated ES cells is LIF signalling through Stat3 impaired; no increase in nuclear or phosphorylated Stat3 is observed when LIF is withdrawn (Jackson et al., 2004; Kaji et al., 2006). Therefore constitutive activation of this pathway is not responsible for the failure of these cells to differentiate.

Microarray analysis of Mbd3−/− ES cells shows that approximately 200 genes are misexpressed compared to Mbd3 parental cells (K.Kaji, personal communication). One possible cause of the differentiation failure identified in Mbd3−/− ES cells is the continued expression of genes associated with the pre-implantation stage of development such as Pramel 6 and Pramel 7, which are normally silenced in ES cells (Kaji et al., 2006). Although the functions of Pramel 6 and Pramel 7 are unknown, they are similar to the PRAME tumour antigen (Bortvin et al., 2003) which has been shown to inhibit retinoic acid induced differentiation of F9 cells (Epping et al., 2005). It is suggested that misexpression of genes such as these may interfere with the ability of Mbd3−/− ES cells to respond to differentiation signals and to initiate appropriate gene expression patterns to allow differentiation into any particular lineage (Kaji et al., 2006). It will be interesting
to analyse the function of genes misexpressed in the absence of Mbd3 to try to understand further how the block in differentiation is caused.

Exactly how these genes are regulated by the NuRD complex is not clear, but the finding that all the mis-expressed genes examined so far in Mbd3 null ES cells display epigenetic aberrations such as DNA demethylation and histone hyperacetylation suggests that NuRD may normally function by epigenetically regulating a range of genes involved in the transition from a pluripotent to a differentiated cell type. The epigenetic aberrations associated with a loss of NuRD may also be more directly involved in the failure of Mbd3−/− ES cells to either down-regulate pluripotency associated genes, or to upregulate lineage-specific genes. The ES cell epigenome contains characteristically large amounts of active euchromatic marks such as histone acetylation and low levels of repressive marks such as DNA methylation, compared to differentiated cells (Meshorer and Misteli, 2006). The process of ES cell differentiation involves a transition towards a more heterochromatic epigenome, with lower levels of histone acetylation, and increased levels of DNA methylation (Lee et al., 2004; Meshorer and Misteli, 2006). The differentiation defects displayed by Mbd3−/− ES cells are associated with decreased DNA methylation as well as histone hyperacetylation (Jackson et al., 2004; Kaji et al., 2006). It will be interesting in the future to examine by microarray which genes involved in either maintaining pluripotency or promoting differentiation are mis-expressed when Mbd3−/− ES cells are cultured in the absence of LIF. An analysis of the epigenetic changes occurring at the regulatory regions of mis-expressed genes would also be useful in delineating the function of NuRD in the regulation of gene expression patterns during ES cell differentiation.

6.4 Mbd3 and DNA methylation in non-pluripotent cells?

The functions of Mbd3 and the NuRD complex have so far been discussed exclusively with regard to embryonic stem cells. However components of the NuRD complex are widely and abundantly expressed in many different cell types in organisms as diverse as
A. thaliana, C. elegans, X. laevis and H. sapiens. (Eshed et al., 1999; Herman et al., 1999; Kehle et al., 1998; Ogas et al., 1999; Solari and Ahringer, 2000; Tong et al., 1998; Wade et al., 1998; Xue et al., 1998; Zhang et al., 1998). The complex has roles in the development of most of these organisms, and has been implicated specifically in cell-fate decisions in a number of different mammalian cell types including B and T lymphocytes (Cismasu et al., 2005; Fujita et al., 2004; Fujita et al., 2003; Kim et al., 1999) and erythroid cells (Hong et al., 2005). In these examples, NuRD is recruited by sequence specific transcriptional regulators in order to modulate the transcriptional activity of lineage specific genes. It is not known how loss of the NuRD complex affects these cell-fate decisions, nor what is the entire range of genes regulated by NuRD in these cells. How this complex comes to be recruited to and to regulate different genes in different cell types remains elusive. Recruitment by sequence specific transcriptional regulators is clearly one way in which the complex is recruited, but it seems remarkable that differential recruitment of the complex relies solely on differential expression of particular transcriptional regulators. Do the targets and role of NuRD truly differ between cell types, or are there common targets regulated by NuRD; if so what are they?

Furthermore, it would be interesting to discover if effects such as decreased DNA methylation are also found in non-pluripotent cells lacking Mbd3, or if this is a feature exclusive to pluripotent cells. Whereas DNA methylation seems to be dispensable for ES cell viability, Dnmt1−/− mouse embryonic fibroblasts (MEFs) containing a significantly reduced amount of DNA methylation apoptose in a p53 dependent manner (Jackson-Grusby et al., 2001), and MEFs in which Dnmt3b has been inactivated either senesce or spontaneously immortalise (Dodge et al., 2005). The levels of the de novo methyltransferases Dnmt3a and Dnmt3b are also highly reduced in somatic cells compared to embryonic cells (Okano et al., 1998a). Thus DNA methylation appears to have different roles in different cell types; whether regulation of DNA methylation by the NuRD complex is a feature conserved from pluripotent to non-pluripotent cells, and if so what the purpose of such regulation is, remains an open question.
Appendix I
Appendix II


domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol 203, 971-983.


structure is required to block transcription of the methylated herpes simplex virus
thymidine kinase gene. Proc Natl Acad Sci U S A 84, 1177-1181.
Cardoso, M. C., and Leonhardt, H. (1999). DNA methyltransferase is actively retained
D. G. (2005). Reduced genomic Cytosine methylation and defective cellular
differentiation in embryonic stem cells lacking CpG binding protein. Mol Cell Biol 25,
4881-4891.
DNA methyltransferase in preimplantation mouse embryos: implications for genomic
specific histone modifications regulate expression of the Dlk1-Gt12 imprinted domain.
Genomics.
Catena, R., Tiveron, C., Ronchi, A., Porta, S., Ferri, A., Tatangelo, L., Cavallaro, M.,
sites in the Sox2 upstream enhancer regulate gene expression in embryonic and neural
Chadwick, B. P., and Willard, H. F. (2003). Chromatin of the Barr body: histone and
non-histone proteins associated with or excluded from the inactive X chromosome. Hum
Mol Genet 12, 2167-2178.
Chadwick, B. P., and Willard, H. F. (2004). Multiple spatially distinct types of
facultative heterochromatin on the human inactive X chromosome. Proc Natl Acad Sci
U S A 111, 17450-17455.
methylation of an imprinted transgene is established during gametogenesis and
Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A.
(2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in
Chapman-Smith, A., and Cronan, J. E., Jr. (1999). The enzymatic biotinylation of
proteins: a post-translational modification of exceptional specificity. Trends Biochem
Charalambous, M., Menheniott, T. R., Bennett, W. R., Kelly, S. M., Dell, G., Dandolo,
expression in both imprinted and non-imprinted tissues. Dev Biol 271, 488-497.
chromosome inactivation in differentiating embryonic stem cells. Cytogenet Genome
Res 99, 75-84.


proteins homologous to the methyl-CpG-binding proteins MBD2 and MBD3: characterization of MBD3L1 as a testis-specific transcriptional repressor. Genomics 80, 621-629.


Solari, F., Bateman, A., and Ahringer, J. (1999). The Caenorhabditis elegans genes egl-27 and egr-I are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. Development 126, 2483-2494.


Website links

Harwell imprinting website: http://www.mgu.har.mrc.ac.uk/research/imprinting/

Methprimer website: http://www.uogene.org/methprimer/